INFORMATION THEORETIC CONSIDERATIONS IN THE DESIGN OF AN INTELLIGENT MACHINE

Don J. Orser

Robot Systems Division
Center for Manufacturing Engineering
National Institute of Standards and Technology
Gaithersburg, Md 20899

ABSTRACT

A fundamental problem any design for an intelligent machine must address is the following: *How will the machine, immersed in a physical environment whose number of states is orders of magnitude larger than its own number of internal (information theoretic, or "mental") states, select which external states will cause modification of its own internal state?*

This paper elaborates on this problem, as it applies to man-made intelligent machines, in terms of an informal information theoretic model, one primarily concerned with what is to be done, rather than how. We argue that current models of automated control for a robot do not address this problem, primarily because so much emphasis is placed on model-based logical inference.

A proposed solution framework is offered to the sensory selection problem based on an information theoretic definition of *event*. The resulting architectural paradigm is then applied to what we call *event-driven hierarchical control*.

"Substances are not the units of things and events are not their motion, but events are the units of things and what is described as a material object is just a feature of events." ALFRED N. WHITEHEAD

1. INTRODUCTION

Within the last several years a number of research projects concerned with the design and implementation of computer controlled vehicles, with various levels of "autonomy", have appeared. This "autonomous vehicle" would have the ability to decompose high level descriptions of a "mission" into an appropriate plan and subsequently execute it, all with relatively remote human involvement. Projects involving land vehicles [19,23], underwater vehicles [2,16,29] and space vehicles [3,18] have been funded.

It is the intent of these efforts to integrate results from artificial intelligence (AI), sensory (vision) understanding research, modern control theory, etc., all driven by the economics of low cost, high performance microprocessors. Initially, these projects will address issues of low level control etc., but the expectation is that this will evolve upward to the point where real-time planning and execution of multi-vehicle coordinated activity, in support of complex missions, becomes feasi-

This work was supported by the Defense Advanced Projects Agency (DARPA), Naval Technology Office, Arlington, Va. 22209-2308.

ble. While much of the funding has come from the defense community, giving a distinct military flavor to these projects, it is also clear that an autonomous vehicle is not only an attractive context within which to integrate these technologies, but would provide an invaluable capability in many non-military situations.

There is clearly a host of engineering issues and design decisions related to sensory data acquisition and the representation of the workspace, all in support of task execution. In this paper we will not be addressing these issues in any detail, but rather wish to place in perspective what is *happening from a more abstract point of* view. In particular, we wish to address a fundamental theoretical issue whose recognition, articulation and "solution", we see as having a number of practical implications for the functional interaction among the subsystem components making up an "intelligent machine".

From the point of view of design, this issue concerns the decision process which selects that subset of the information potentially available about the environment which will be allowed to modify the machines "mental behavior". This decision process involves numerous consecutive decision steps of which we single out three major levels for the purpose of discussion. These are

(1) *Hardware Selection:* By choosing a particular suite of sensors, the designer bounds the information available to the machine, and hence the potential effect of the environment, to just those sensor modalities, frequencies, resolutions, etc., characterized by the sensors chosen.

(2) *Algorithmic Feature Selection:* Given the suite of sensors, the effect of the environment is further reduced to just those features of the environment for which the designer supplies an algorithmic procedure for their extraction from the sensory data.

(3) *Information Theoretic Model Selection:* The provision by which a "world/task model" is incorporated into the machine, the use of which provides the basis for the machine to "select" its sensory input from the much larger amount available to it.

All three levels of design are concerned with narrowing the effects of the environment to just those which will provide sufficient information for the machine to accomplish its task. The first two levels are explicitly made by the designer, while the third is made somewhat less explicitly through whatever inferencing mechanisms are built into it. It is the method whereby this third level, the information theoretic, selects information from the environment that we are concerned with here.

We view any definition of what constitutes "intelligent behavior" by a machine as picking an arbitrary point on a continuum. More important for the discussion here is the distinction between deterministic closed form algorithmic methods based on a tractable mathematical analysis of the desired behavior, and heuristic methods. We will argue that "classical robotics" has emphasized methods of the former, but must, if it is to address the new class of problems to be encountered by these "autonomous vehicles", adopt an extended architectural paradigm.

Another point of view motivating our ideas is the apparent impasse which AI seems to have encountered, and suggest that AI has limited itself by addressing a form of intelligence based solely on *model-based logical inference,* one devoid of sensory input, and hence suffering from "sensory depravation". Within AI research, one such problem associated with a priori model-based reasoning has become known as the "frame problem" [6,22,26]. In the frame problem, an algorithm for planning is sought which operates entirely by model-based logical inference. The assumption is that by providing for a sufficiently rich internal modeling of the external environment and the task, the problem may be solved a-priori (at least to within some "resolution", so that the rest is only a matter of "detail"), all without recourse to sensing the physical world. That is, success in solving an instance of the planning problem in information space is tantamount to its successful physical execution. For "sufficiently simple" environments this will work. However, as documented in the above references, this paradigm appears to break down at some level of complexity, if not in principle, then in practice.

We argue here that the frame problem is an artificial problem precisely because it requires a solution to take place entirely within an internally held information theoretic model. In observing "natural intelligence", one observes that its goals are achieved through the use of *primitive thought processes* acting *on* and *sensing* the physical environment[4].

This is to say that at a fundamental level, at least for natural intelligence, the external world, to a great extent, is its own model. In this way, behavior becomes as rich and open ended as the environment. This is not to say that an internal model of the environment does not occur, but rather that it is more in the form of a repository of learned past successful behavior in terms of generic "mental landmarks". These latter include, for example, rote fragments of formal methods, e.g., logic and mathematics, utilizing sensed symbols either in the environment or in the "minds eye", and hence provide the basis for what appears to an observer, the ability to perform complex inference based problem solving.

We argue here that this "natural organism" paradigm of side stepping the frame problem appears at the moment to be the only one available for use by an artificially intelligent machine operating in real time, and suggest that the integration of AI techniques and sensory processing provides a vehicle for accomplishing this.

In what follows we will use the term *autonomous agent* to refer to a member of this (as yet) hypothetical class of machines. We use the term *machine* deliberately so as to include, not just isolated information theoretic algorithmic processes, but the physical mechanisms by which these abstract processes interact with the environment, i.e., its ability to sense and act on the environment. More specifically, we are interested in how a machine can take action in a world whose events it can only partially predict, and argue that the solution lies in treating both the information theoretic and physical aspects simultaneously.

2. THE PROBLEM OF SELECTIVE SENSORY PROCESSING

Sensory processing for an intelligent machine is concerned with "making sense" out of the wealth of information potentially available to it about the immediate environment. Naively, it would seem that "natural intelligence", e.g., human intelligence, somehow uses a "bottom up" process by which it successfully accomplishes this. However, a simple information theoretic argument demonstrates that this bottom up model must be an illusion. This argument makes the observation that the number of potential states, and hence the information content, of the physical environment within sensor range is much larger than the number of potential "internal" logical states encoding some model of the environment. Hence any attempt to let this information modify its own internal states will cause it to eventually run out of "new" states, thus in effect negating any effect the information up to that point may have had on it.

The above problem may be stated:

> *How does an autonomous agent, immersed in a physical environment whose number of potential physical states is orders of magnitude larger than the agents own number of internal logical states, select which physical states will cause modification of its own internal state?*

The solution, clearly, must lie within the ability of the machine to convert the task given it, not only into just actuator, or as we shall refer to them, effector controls, e.g., motion control etc., but also into criteria whose purpose is to match with and hence select predefined changes of external state space relevant to carrying out the task.

More pointedly, this means that for a machine to exhibit adaptive behavior with respect to its sensing of the environment, *it must be prepared to select its own input amongst the much larger set of data available to it.*

For a ''computer'', the selection criteria are in the mind of the designer and incorporated implicitly in the ''input'' hardware and associated input decoding algorithms for selecting a set of predefined signals. A more physically adaptive behavior results from selection criteria which respond to generic environment features, e.g., ''tropisms'' which result from being attracted or repelled by predefined sensory features such as light, temperature or specialized geometric features [5,11]. A third ''level'' of selection criteria is that based on an *information theoretic model,* resident in the ''mind'' of the autonomous agent, which serves to translate the needs of the task/goal at hand into specific selection criteria. It is this idea which we are interested in developing here.

The word ''event'', through common usage, has come to mean ''any change'', or ''thing which happens'' in external physical state space. However, from an information theoretic point of view concerned with how physical state space modifies an agent's behavior, we argue that events exist only in the context of an individual agent's information theoretic state space. They are precisely identifiable with matches made between selection criteria and sensory features coming from physical state space. In general, an agent's behavior must be explainable in terms of the selection criteria and the subsequent acquisition of external physical state space changes, i.e., the agents personal sequence of events. An autonomous agent's behavior both determines its events and is determined by its events.

This brings us to a second related issue, namely the need to distinguish between an agents world model space and an ''observer's'' (e.g., the ''designer's'') world model state space. This distinction is important as an agent has access *only to its own world model space in performing a task, and not to the observer's.*

For example, a list of desirable features for an autonomous agent often includes the ability to deal with ''unexpected events''. But this ''unexpectedness'' exists only within the mind of the designer, for if an autonomous agent's design and subsequent behavior take that information into account, its behavior, from its point of view, can hardly be said to be dealing with the ''unexpected''. Hence the idea that an autonomous agent must be able to deal with the unexpected is a naive and ill-defined one.

For example, consider a mobile robot which has been designed to move around in an environment containing obstacles made known to it through the use of a ''vision sensor''. If now an obstacle in the form of a sheet of glass is placed in the robots path and is hit by it, (e.g., a bee against a window), the robots physical state behavior as observed and modeled by an observer's world model will have been modified to be different than what it otherwise would have been had the glass not been there. But in fact the robot's internal, or ''mental'' model of what has happened has not changed, but only gotten ''out of sync'' with its environment due to the inadequacy of its sensors. It is the observer's events which have been modified by the introduction of the obstacle, not the robot's.

Hence we must be on guard to always distinguish between changes in a first agent's physical state space as observed by a second agent, and changes in the information theoretic model of the first agent. The observed physical state can not always be used as an indication of the information theoretic.

We will return to this point later to show that what is intended to be captured by ''unexpected event'' is captured by a particular type of information theoretic expectation on the part of the autonomous agent.

3. THE LIMITATION OF CONTROL THEORY FOR "CLASSICAL" ROBOTICS

If a robot can answer a question about a hypothetical experiment involving its own effectors without actually performing the experiment, then it has exhibited one of the components of intelligent behavior.

Classical robotic control theory consists of establishing ''closed'' form answers to questions of (1), kinematics, and (2), dynamics [10]. In both cases, what is meant by closed form answers is a method of establishing the *functional homomorphism,* as we shall informally refer to it, between certain internal logical control states of the robot and that subset of physical state space constituting the possible states the robot's effectors may be in.

For example, from knowing the values of certain joint state variables, one wants to deterministically calculate the position and orientation of a gripper. Conversely, from an initial position and orientation, one would like to calculate the sequence of logical state variable values associated with the joints in question so as to move the gripper in a particular manner to a desired location and orientation.

A fundamental observation of classical robotics is the (ideal) requirement that there be a one to one mapping (via the kinematic and dynamic model) between the set of (legal) values for the logical state control variables and the set of physical state variable values used to characterize the physical state the robot is in. This insures that the successful generation of a "plan" for effector control is logically tantamount to the plan's successful execution. Elaboration on this static model to allow the "sensed" entry and exit of objects in the workspace only requires that we include all the physical environment as constituting the "robot". The fundamental observation remains: the robot must maintain an internal "simulation" or "world-model" of the workspace which will maintain the robot's ability to compute elements of the homomorphism modeling a subset of physical state space.

In short, all "external events" associated with certain physical states have an analogous internal logical state with which the external states of the workspace may be identified, and conversely, all "internal events" associated with certain logical states may be identified with certain corresponding external world model states. *It is by the matching of these internal and external events, i.e., the ability to compute the homomorphism, that synchrony is maintained between internal logical state space and external physical state space.*

Classical robotics is based on this fundamental relationship: the ability to predict (evaluate the homomorphism), detect (sense the difference between the current external state and the desired), and correct (change the external state through the predicted use of effectors). This classic paradigm is depicted in figure 1.

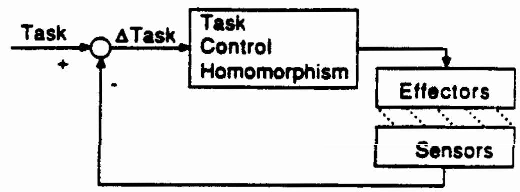

Figure 1. A conceptual schematic of a closed servo loop. Its simplicity results from the simple (closed form) relationship between the two sets of equivalence classes induced by the effectors and sensors respectively.

The ability for a world model to always compute the homomorphism (for a given environment/task), will be referred to as a *complete world model.* In general, the ability to compute the homomorphism will come from both *a priori* information, and from the ability to sense external physical state space. If only a priori knowledge is required to compute the homomorphism, the world model is said to be an *a priori complete world model* ("open loop"), otherwise if sensing the environment is required to maintain the homomorphism, it is said to be a *sensory extended complete world model* (servoed or "closed loop"). In the extreme in this latter, behavior becomes predominantly determined by the environment through what is sensed. For example, see [5,11].

The entire thrust of control theory for classical robotics is concerned with the articulation of complete workspace-world model homomorphisms in the form of algorithms for computing this homomorphism for specific effectors and sensors operating in a given workspace performing specific tasks. The precise characterization of the robots effectors, sensors and workspace serve the purpose of providing an information theoretic complete world model which in turn provides the basis for computing the homomorphism.

We have attempted to capture, by the notion of the world model homomorphism, all control processes, from the simplest thermostat through the most complex multiprocessor multiprocess control system. The algorithms controlling these processes are by design intended to deal with some set of expected contingencies by maintaining synchrony between their world model state space and the physical state space within which they find themselves. For a complete world model no selective process is required to select among the contingencies as all possibilities have been accounted for a priori within the three levels of design mentioned in the introduction. Only their particular sequence and time of arrival is unknown.

If in fact the world model for a machine does not provide the basis for computing the homomorphism between the machine's information state space and its physical state space, then the world model is inadequate, and we will refer to it as an *incomplete world model.* We will (arbitrarily) reserve the term *autonomous agent* to refer to an intelligent machine whose world model is incomplete, and argue that the ability to modify ones own world model homomorphism, in the case of incompleteness, is a necessary component of autonomous behavior. (We also note, that as a practical matter, a world model will consist of complete and incomplete submodels, but wish only to discuss the fundamental issues here.) To the extent that a world model is inadequate in its ability to distinguish (i.e., predict) between task-induced differentiated states, it must selectively acquire the information to make the needed differentiation through its sensors.

Given a particular task or class of related tasks, issues which a theory of robotics should address with respect to completeness include:

(1) The basis by which a given robot, i.e., a world model together with a suite of sensors and effectors including associated algorithms, may be proved complete or incomplete with respect to carrying out a task or class of tasks.

(2) The basis by which, for a given robot, the existence or nonexistence of a complete world model, sufficient to perform a given task or class of tasks, may be proven.

For example, if sensors do not exist for detecting changes of physical state space produced by causes other than the robots own effectors and the task requires the ability to distinguish between these states, then a complete world model, *with respect to that task,* clearly cannot exist.

The conceptual limitation of control theory for classical robotics is that it presupposes the existence in principle of a complete world model: given enough sensors, effectors, computational power and the proper world model and task representation, the homomorphism may be computed to a sufficient "resolution", so that the machine may accomplish the task. It presumes on the part of the designer the ability to a priori incorporate the sufficient sensory selection criteria by the choice of sensor hardware and sensory feature extraction algorithms. We argue here that this conceptual model is inadequate for an autonomous agent, primarily because it does not address the problem of selective sensory processing at the information theoretic modeling level required for an incomplete world model.

In this paper we are concerned primarily with intelligent machines whose world model is incomplete. From the point of view of the machine, not all external states are representable nor of interest to the machine. In fact we will argue that a major component of "intelligence" is bound up in this selection process required of machines whose world model is incomplete.

4. THE INFORMATION THEORETIC MODEL

The designer of an intelligent machine will naturally divide physical state space into "machine" and "non-machine''. However, this dichotomy exists only in the world model, i.e., ''mind'', of the designer, and is not even then very well defined. For example, is the road the wheeled mobile robot uses part of the robot? And what about the ''landmarks'' it uses for navigation? We contend that any separation of physical state space into ''machine'' and ''non-machine'' serves only the needs of the designer in reasoning about the machine. From the point of view of the machine, its ability to make distinctions about physical state space, including physical aspects of itself, is that determined by its own functional homomorphism. This homomorphism may or may not include an elaborate representation of self depending upon the class of tasks the machine is to perform.

As a consequence of this, the point of view taken here is that based on a ''functional dualism'' [14] between the following two components: (a) a *physical state space*, and (b), an *information theoretic state space*. Physical state space is characterized by *physical*, or, ''external'' *state variables*, which constitute a description of the physical component of the machine together with the physical environment. Information theoretic state space consists of a finite set of distinguishable symbols, encoded as a special set of physical state variables, i.e., memory, whose distinct configurations over time constitute the succession of ''internal'' states the machine is in, and among other things, serves to ''implement'' the machine's functional homomorphism.

Two classes of transducers provide for the respective modification of these two disjoint state spaces: (1) *Sensors* by which physical state space is selectively allowed to modify information state space, and (2), *effectors* by which information state space selectively ''operates'' on and hence modifies a certain subset of physical state space. See figure 2.

Processes, encoded in information state space, control the interaction between information state space and physical state space. These process ''maps'' are:

(1) *External State Acquisition*(Ω): The process by which certain physical state variables are ''measured'' by various sensor modalities with the intent of (a), maintaining the ability of the world model to compute the homomorphism, and (b), selecting information relevant to the carrying out of some task.

(2) *World Model homomorphism*(Ψ): A process implementing the homomorphism: on demand knowledge of the current state of selected physical state variables, and as an adjunct the ability to ''predict'' future states of them under hypothesized conditions.

(3) *Planning and Execution*(Γ): The process by which certain information state variables, derived from task/goal definitions by ''planning'', are ''executed'' via the effectors so as to cause a change in physical state space.

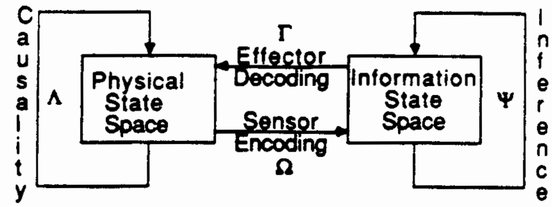

Figure 2. An intelligent machine may be thought to consist of all physical state space, together with an encoding of information state space, in which all interactions between the two are limited to the sensor and effector functions.

Collectively, we identify the information contained within these processes the logical or information theoretic state space of the machine. This information state space is ''closed'', i.e., does not ''interact'' with any other state variables of physical state space, except through the use of its effectors and sensors.

In figure 2 the four maps Λ, Ω, Ψ and Γ, respectively, depicting ''physical causality'', sensor encoding, world modeling and effector decoding have, for a complete world model, the following relationship,

$$\Lambda|_{task} = \Gamma|_{effectors} \bullet \Psi|_{task\ model} \bullet \Omega|_{sensors},$$

where • is function composition and the subscripted bar means that the function domain is "restricted to".

As mentioned earlier, physical state space must include not only the physical workspace of the machine but the physical machine as well, as there is no way for the machine to disambiguate the two beyond what is modeled by its world model homomorphism. We will collectively refer to physical state space as the machine's *workspace*. It is the information theoretic nature of the interaction between physical state space and the world model homomorphism which we are concerned with here.

As an intelligent machine performs some task, its world model will acquire new information in support of maintaining its ability to compute the homomorphism. This sequence of homomorphism transitions, taking place entirely within information state space, will be called its *objective behavior*. This is in contrast to its overt, or *observed behavior*, created in the information theoretic state space of an observer. This latter includes projected purposive behavior, e.g., the projection of "intent" into the observed behavior of the agent by the observer, and is the source of much potential confusion. The contrast between these two is depicted in figure 3.

The school of psychology known as behaviorism is based on the tenet that since only observed behavior is known to an observer, any theory of that organism must be in terms of just these observations. Purposive behavior is then seen to exist only in the observer's mind. However, for designed machines, predictive, as opposed to merely descriptive understanding on the part of an observer, must be entirely in terms of objective behavior[7].

For a machine with a complete world model, the sequence of world model states (objective behavior) is homomorphic, modulo the task, to an appropriate subset of physical state space transitions (observed behavior), and hence its behavior could be identified with either. However, if synchrony is not maintained, these state transitions are not equivalent, but the knowledge of where this discrepancy lies, potentially exists only in

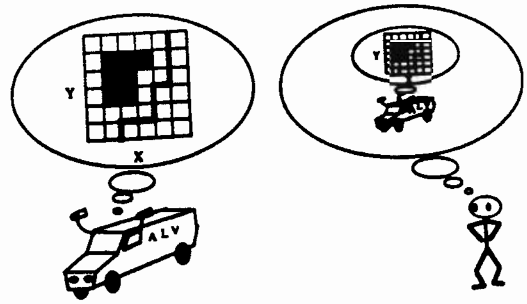

Figure 3. The *objective* behavior of an autonomous agent is identified with its world model homomorphism transitions. Its observed, or *subjective* behavior, is identified with that portion of an observers world model homomorphism transitions modeling some subset of the first agents physical and information theoretic state space variables.

the world model of some observer agent. When such a discrepancy between objective and observed behavior exists, it will be called *undefined behavior*. The detection of self-undefined behavior on the part of an agent is of course of considerable interest to that agent and corresponds to self knowledge about an incorrect model of its own physical state space.

The definition of an *event* is then identified with transitions in the world model homomorphism. An event is a source of new information which results in an (updated) homomorphism.

The problem of selective sensory processing then becomes:

> *How does an autonomous agent choose its own events?*

We argue that the distinction between the classical model for a robot and that for an autonomous agent lies not in the physical relationship between logical state space and physical state space, but rather in the ability of an autonomous agent to generate from its task, selection criteria, or as we shall refer to them, *expectations*, which will serve the purpose of selecting its input from physical state space. A successful match between an expectation and its physical state space instance results in the modification of the current world homomorphism and is identified with an event. Generality of design, on the part of an autonomous agent's designer, determines whether modification of the homomorphism is, at one extreme, a parametric modification, or at the other extreme, one of functional form modification, and embodies what is meant by "learning".

In the classical model for a robot, the selection of its "input" is preordained by relatively fixed hardware and program expectations associated with a particular sensing device, order of occurrence, discretization, predefined codes etc. Examples of such sensors and their "measurement " include a clock's "ticks", navigation sensors and the resulting coordinates, communication channels and the characters they transmit (but not necessarily the interpretation of these characters), etc. It is able to do this only because associated with each input is an (implicit) expectation to match it. We will refer to such sensor data as *signals,* and note that a signal's interpretation, i.e., its effect on subsequent behavior by the robot in a given state, is predefined.

When the input available becomes mixed with irrelevant information or when its interpretation must be extracted from a larger set of potential interpretations, e.g., the control of a camera and the processing of its imagery for the purpose of extracting knowledge concerning obstacle avoidance by a vehicle, the process by which the input is selected becomes central to both practical and theoretic considerations in the design of an intelligent machine.

5. KNOWLEDGE BY HYPOTHESIZED QUERY

In order to make more specific this process of acquiring information, and hence beliefs about the environment, we perform the following mental experiment.

Imagine for a moment that you, the reader, were placed in a steel tank, i.e., a "submarine", and placed underwater. Imagine that this submarine were equipped with controls for moving and orienting the submarine, and also, as part of the submarine, a number of sensing devices, more or less equivalent to your own senses are available, but in a peculiar way: two teletype like machines are available to you in the submarine, one labeled *control,* the other labeled *sensing*.

The control teletype accepts English like commands for operating the motion control mechanism of the submarine, while the other teletype will accept and answer queries about the immediate environment at the moment the query is made.

Since the sensors are to be "general purpose", they make no interpretation of what they "sense", but rather respond in terms of a set of "features" which are extracted by the application of processing algorithms to raw data obtained by the sensing devices. These sensing devices and associated algorithms for the extraction of features may be of your own choosing.

Our claim is that this is the situation which an algorithm implementing the "mind" of an intelligent agent would find itself, with you the imaginary submarine pilot simulating such an algorithm.

If you have some model of the underwater terrain with you, either in the form of a map or in your head, then you may perform certain tasks by commanding that the submarine move to a particular location by coordinate or landmark etc. If you know a priori that there are no obstacles, or you know their location, you can still move without hitting them by breaking the motion up into sequences around the obstacle(s). This is the case of having an a priori complete world model with respect to a simple "move to location x" class of tasks.

If your model of the environment you have been placed in is not an a priori complete one, e.g., you know that all obstacles are lying on the bottom and are less than five feet in diameter, but you do not know their location, then it is still possible to perform the task of moving about without bumping into an obstacle by using the sensor query teletype. Assuming you have a feature called "obstacle ahead" you may avoid bumping into these objects as you proceed to another location by continuously utilizing this query. The knowledge you have of the objects together with the ability to sense their location provides you with the ability to perform the same simple class of tasks as before, and exemplifies a sensory extended complete world model.

If the nature of the objects is unbounded, either with respect to your ability to sense them or spatially so that they possibly isolate you from your destination, then success, assuming the existence of a solution, depends on how "clever" you are in forming task relevant hypotheses about the environment, and follows from the following claim: the only method for obtaining information about the outside is to ask questions of the form *"Is there an instance of primitive query feature X in the environment?"*, or questions which are compounded of such queries.

Your "plan" to achieve the task then, is, together with the actions necessary to achieve it, a sequence of hypothesized queries sufficient to find the additional information needed to make the actions sufficient. Your mental activity consists of (incrementally) first "imagining", i.e., hypothesizing, the situation, in terms of the available features, prior to finding whether each is the case or not. Your timeliness and effectiveness in achieving the task is, for a given primitive query feature set, dependent on the order in which the compounded queries are asked.

It might be argued that a query of the form *"What's out there?"*, resulting in an enumeration of all instances of the primitive query feature set, will contain as a subset all relevant information. But not only is this computationally expensive, but one is still left with the essence of the original problem: the algorithmic determination of the criteria by which the task relevant subset is found.

A positive response to a compound query of the form *"Is there an instance of either one of the features A or B?"*, contains intrinsic uncertainty as to which of either feature A or B has occurred. Most queries will contain some degree of generality compounded of such disjoint features, and hence will have an *intrinsic uncertainty* associated with positive responses. (For example, see [27] for an application of this to the problem of motion (uncertainty) determination from a sequence of images.)

The point we wish to make is that in attempting to find out about the environment you are in, you are intrinsically limited to making queries which can always be answered by a simple "yes" or "no" as to whether a (predefined) feature or complex of features does or does not exist within the sensed region you are in. Stated another way, your paradigm is limited to making "matches" between a fixed set of internally held or generated patterns and their occurrence as feature instances in the environment.

There is no algorithmic procedure by which the environment is forced to reveal itself, except that by which an hypothesis is first generated and then matched against the environment.

We will refer to this method of acquiring information about the physical environment as the *hypothesized query paradigm*.

A good argument can be made for supposing that human sensor processing is based on the hypothesized query paradigm to acquire information about the external physical world. This is especially well documented for human vision [17]. At a higher level, it is the "mind set" (expectations) whose characterization is attempted in the psychologist's Rorschach test. This paradigm of matching predefined features, patterns, hypotheses etc., is a fundamental one and is ubiquitous within human endeavors. The *scientific Method*, mathematics, diagnostic medicine and even a court trial provide examples whereby hypotheses, propositions, statements etc., are first generated, and then solicitation made as to their truth or falsity.

An exception is the logician's explicit purpose in creating a methodology for establishing complete world models for some area of discourse (usually mathematical in nature). By starting from some set of axioms, he attempts to provide a method by which the body of knowledge is generated systematically. As a practical matter, Godel's incompleteness theorem established that the logician's objective is not possible (at least for any model complex enough to include arithmetic), as there will always be statements which are correct but not capable of being generated by any algorithmic process. This in fact supports the argument that an information theoretic model adds to its knowledge about its physical environment only by hypothesized query.

For an autonomous agent, expectations, generated as a by product of planning, serve as the basis for hypothesized queries in search of the events out of which the world model homomorphism is maintained.

We suggest that an *information theoretic theory of questions and answers,* based not on model directed inference, but rather on task directed sensing of the environment, is needed in support of the design of such an agent. Such a theory would provide the basis for an algorithmic articulation of the *scientific method* within closed or "engineered" environments, and would provide the basis for *artificial epistemology*.

6. ELEMENTARY ARTIFICIAL EPISTEMOLOGY

By epistemology is meant the study of the grounds for belief and the nature of knowledge, and is operationally identifiable with the scientific method. Artificial epistemology deals with similar issues, but within closed artificial (i.e., designed by a human) subsystems, such as a robot's information state space model of it's workspace.

Artificial epistemology is concerned with characterizing the necessary and sufficient information contained in information state space in order that the machine be able to perform a given *task*. A task is defined as a pair of initial and final states of physical state space, in which the machine is to transform the initial state into the final state through the use of its effectors. Operationally, artificial epistemology is just the algorithmic articulation of (a task induced subset of) the scientific method utilizing a suite of artificial sensors.

From the point of view of the design of a machine to accomplish a given task, physical state space has three independent partitions superimposed on it. The first partition, induced by (1), the machine's task definition, induces in turn the necessary discriminating ability of (2), the sensor, and (3), effector, partitions.

More specifically, a particular sensor differentiates between discretized measurements performed on some set of physical state variables. This results in the partitioning of all physical state space into equivalence classes, in which all elements of an equivalence class have the same value with respect to that sensor measurement. Two *orthogonal* sensors measure two independent physical state variables and generate the cross product of their respective equivalence classes. (See [24] for the application of this idea to navigation.)

Similarly, a given set of effectors serve to partition physical state space into equivalence classes in which members of each such class all correspond to the same control variable state setting.

A task is a given partition together with two identified states: the current state and the desired task completion state. A successful execution of the task corresponds to going from the current state to the task completion state through a sequence of intermediate states, each of which is obtainable and distinguishable by effectors and sensors, respectively, from the previous state.

A given task and workspace in which the task is to be carried out also determine a set of predicates. These predicates in turn determine the sensor and effector partitioning needed to carry out the task: sensors must provide sufficient information about a sensor equivalence class that it be distinguishable from other classes of the partition. Analogously, effector predicates, corresponding to effector states, must be sufficient to provide the class of physical state transitions between effector equivalence classes needed for carrying out the task.

Artificial epistemology is concerned with the mathematical characterization of task definitions in terms of workspace equivalence class partitioning, together with methods of proving that given effector and sensor partitions provide a "belief" structure sufficient for carrying out that task.

Kripke structures [15,21] are related to the above in that they attempt to provide a language and associated semantics within the framework of modal logic for world model homomorphism partitions, called *possible worlds* by Kripke. We have suggested here, that in general, three such structures are needed i.e., task, sensor (world model), and effector structures. However, we also suggest the possibility that two or more of these partitionings be identical. For example, simple servoing results when the sensor and effector partitions are identical: associated with each effector control variable is an identical sensor variable whose interpretation is the same or computationally simply related. Hence a simple predict, detect and correct algorithm exists for their control, as depicted in figure 1.

7. PLANNING EVENTS IN AN INCOMPLETE WORLD MODEL

A collection of partially ordered (with respect to time, cost and other priorities) tasks constitutes a *mission*. Planning is the compilation of task(s), stated in terms of an information theoretic task model, into *actuation sequences* stated in terms of the planning and execution state space subset of information theoretic state space, and is used for generating physical state space changes through the use of its effectors. An actuation sequence is made up of *(atomic) actions* which reflect the elementary changes of physical state space by a particular suite of effectors.

An action is not only determined by task intermediate physical states, but also by *goals* implicit in the definition of a task and the cost of using the resources available to the machine. These goals are used to choose among alternative action sequences for achieving the task in ways which tradeoff the "value" of intermediate task states against their "cost". An action also contains the instantiation of information descriptive of physical state space where the task is to be performed. This latter information may be known a priori, or only become known incrementally through the use of the machine's sensors.

For example, the task of going from location A to B is compiled into an action entailing the control of effectors for moving from A to B. The particular coordinates used are determined by A and B and obstacles between them. Among the many routes, the one best meeting the goals of least time, fuel etc., is chosen.

The compilation of a task into an actuation sequence for controlling effectors is still not sufficient for intelligent behavior. As an action is "executed", changes in physical state space must be kept in synchrony with the plan. For a machine whose world model is complete, this is performed via the homomorphism, but for incomplete world models, these physical state changes must be selected from all those available to it through the use of its sensors. We argue here that this selection process is critical, and its elaboration provides the constructive basis for designing an autonomous agent.

The concept of an event is central to this basis, and from an external point of view is identified with precisely those changes of external physical state space which result in a modification of the autonomous agent's world model homomorphism, i.e., an event is defined operationally in terms of whether the subsequent *objective* behavior of the agent is different from what would have been the case had the event not occurred. The definition is made in the context of a hierarchy of set inclusions with event being the most exclusive. This hierarchy and the corresponding selection criteria is depicted in figure 4.

The most inclusive set of information sources for an intelligent machine are the changes of external physical state space *(world space potential)*. In order for these to have any effect on an intelligent machine however, they must be capable of being detected by a sensor. This set of potentially detectable changes of physical state *(sensory space potential)* is further restricted by those actually detected, since most sensors will not be "omniscient"

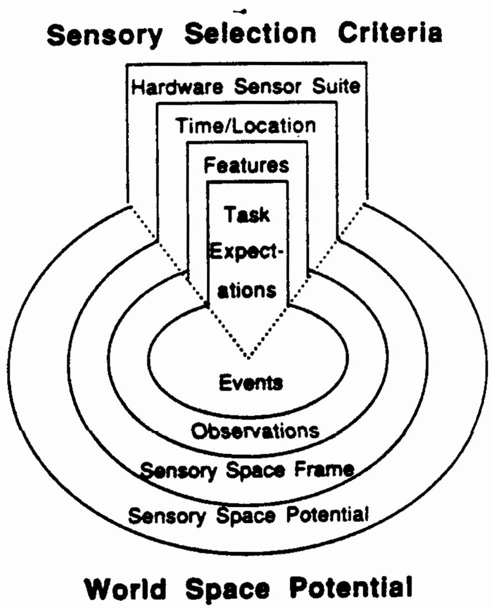

Figure 4. An autonomous agent's events are jointly selected by a hierarchy of sensor selection criteria, starting with those determined by the designer through those which are task specific, together with the corresponding hierarchy of induced instances contained in the world space potential.

with respect to the modality they measure, but will be dependent upon, for example, their location and the time *(sensory space frame)*. Further, these detected changes of state must be transformed into changes in a subset of information state space which we will call *observations*, which in general are in terms of a priori defined generic "patterns" of atomic features. The machine at this point may record these observations for latter perusal etc., but one additional restrictive process must be brought to bear in order for the external change of state to cause an information state space change and hence be an event. This restriction is that the observation must be "matched" with an expectation associated with the achievement of an action resulting from the "decomposition" of higher level action sequences.

For an autonomous agent r then, an *information theoretic* r*-event*, is a change of external physical state space whose occurrence is matched to a world model state space expectation stemming from the "decomposition" of r's task. In other words, all r-events, or r's "events", must be anticipated by a "plan", which in effect is the "organized" representation for achieving these events within some predefined set of "eventualities" significant to the success of the task. It is in this way that exclusively relevant external changes of state are (1), extracted from the environment, (2) kept in synchrony with the execution of the plan, and (3), made "meaningful" by being identified with a particular action of the plan. Observations not matched represent changes of external state which are "uneventful" (irrelevant) with respect to the current plan.

A plan then consists of *action sequences* and associated *expectation sequences*. As presented above, the events resulting from a match between the expectations and the corresponding features from sensory space are the *enabling events*, and are identifiable with the achievement of corresponding actions. Enabling events are based on the a priori information available to the planner when the *feasible plan*, consisting of action and expectation sequences, was generated. The execution of this plan entails the plan's *resolution* against reality, and must include, due to the incomplete world model, the possibility of new information being acquired which negates the possibility of achieving what would otherwise be the next enabling event. Hence the planner must also incorporate *disabling expectations* which potentially match and become *disabling events*. The significance of a disabling event is that it serves as a source of new information which must now be used to modify the currently *disabled plan* and turn it into an updated feasible plan.

Planning has two aspects. *Feasibility planning* is concerned with the evaluation of alternative strategies for accomplishing a given task by resolving consequences of these alternatives until one alternative is, to some degree of confidence, superior to the remaining. Feasibility planning requires of the world model the ability to hypothesize actions and derive their task relevant consequences. We have downplayed this process, not because it is not important, but because its very importance has generated a large literature [8,25,28].

Once the feasible plan is chosen, *plan decomposition* proceeds by resolving to successively higher precision the instantiation of that single plan. For a complete world model, feasibility planning is simply the ability to compute the homomorphism. If only an incomplete world model is available, then disabling expectations must be generated so as to acquire the appropriate additional information as it becomes available.

From a logical point of view, the current feasible plan is a representation of the known necessary actions needed to generate the events required to carry out the task. However, because the information about the world is incomplete, these actions and resulting events are not necessarily sufficient. Hence there must also be a method of acquiring these unknown events in a manner that allows them to determine additional enabling events to be incorporated into the plan. The disabling expectations serve this purpose.

For example, a plan for going from point A to point B with no a priori known objects in the way might also want to incorporate the ongoing (disabling) expectation of encountering an obstacle. The expectation may be a generic one, but when matched against a specific instance results in the information needed to generate new actions and enabling events associated with going around the object.

Events take place in four dimensional space-time, so that expectations may be viewed as variously dimensioned elements in this space. Expectations may be thought of as having *windows* which open and close as these elements are entered and exited. The failure of a disabling event to occur before the associated window closes is of no concern. For enabling events however, some may be *unconditional enabling events*, so that their failure to occur within their window is a disabling event in itself. Other enabling events may be *conditional enabling events* whose occurrence is optional. For example, given the task of detecting an object "presumed" to be in some designated area determines a spatial window within which a conditional enabling event corresponding to the objects detection is expected. Its failure to occur before its associated window closes signals only that it was not detected, an *internal event* which must be one of the possible outcomes incorporated in the plan. The various expectation categories and their relationships provide for an "event calculus" whose interrelationships potentially provide the basis for the generation of, reasoning about, and subsequent execution of plans.

So far all events have been in terms of atomic features taken from observations. These are *atomic events*, and represent the most primitive level of state transitions. *Compound events* are defined in terms of other lower level events and provide the basis by which matching of higher level *compound features* are matched with higher level expectations. This provides the basis by which event acquisition followed by *event assimilation* is performed. Event assimilation is what occurs after an expectation is matched, hence generating the event, and relates this event instance back to the task via the plan from which the original expectation was associated, hence giving the event "meaning".

The importance of event assimilation is demonstrated by its absence in the following:

"There were a total of six of them, three red, one black and two blue."

Atomic events consisting of letter, word and semantic identification etc., took place in the readers mind, but the result was not assimilated due to the lack of a proper higher level expectation associated with the readers task.

For a given task there are in general an *infinite number* of potentially disabling events. Since not all of these eventualities can be looked for as execution of the plan is carried out, some must be discarded, with the result that the autonomous agent will be oblivious to (most) changes in physical state space. A large component of intelligent behavior consists of the process by which a finite number of them are selected and prioritized for inclusion in what might go "wrong" with the current plan at its various junctures.

8. EVENT-DRIVEN HIERARCHICAL CONTROL

We have argued that for any autonomous agent whose world model is incomplete with respect to carrying out some class of tasks, a mechanism for matching task specific expectations with predefined features of sensory space provides the basis for selecting just that information necessary and sufficient (if this is possible) to the carrying out of the task. This space of potential expectations is anticipated by the designer in selecting sensors, e.g., modalities and frequencies etc., and by the algorithms chosen to extract predefined features. Even further, we have argued that potential "soft" events must be "programmable" from expectations generated from planning, i.e., information theoretic level expectations. The result is computational efficiency as well as keeping an autonomous agent's behavior as task specific as possible.

This presents the designer of an autonomous agent with the central design issue: *At what level, hardware, algorithmic or information theoretic modeling, should a given expectation be incorporated?*

In this section we apply these ideas to the *hierarchical control* architecture for real time control [1,20,30].

The general form, depicted in figure 5, may be thought of as imposing three "legs" on information state space. These legs are hierarchical in that they resolve space and time most finely at the bottom. The legs perform the role of (1), feasibility planning, plan decomposition and execution, (2), multi-sensor data acquisition, fusion, interpretation and assimilation, and (3), a world model homomorphism acting as an intermediary between (1) and (2).

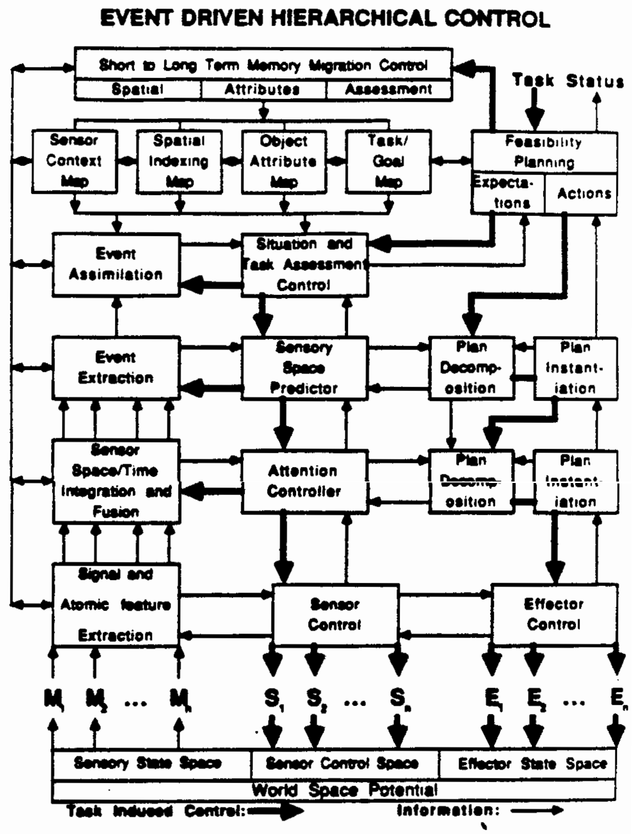

Figure 5. The dynamics of event-driven hierarchical control include, on the right, a descending hierarchy of feasibility planning, plan decomposition and instantiation, terminating in execution, and on the left, an ascending hierarchy of multi-sensor data extraction, fusion, interpretation and assimilation, with the world model homomorphism acting to match their respective expectations and features. The resulting matches constitute the agents events.

Figure 5 is not to be thought of as an actual architecture, but rather the graphical depiction of the relationship between the concepts described here. For example, in an actual implementation, the actual number of levels, e.g., of plan decomposition and instantiation, will reflect the particular requirements of the task(s) to be performed.

An additional component, in the form of common memory, supports several types of "two way mapping" (bijective) functions. As these are central to the ability to compute the homomorphism between information state space and physical state space we list them explicitly:

(1) *Spatial indexing map:* Given a location in space, e.g., a point, region etc., determine what object(s) are there, or inversely, given these objects, determine their location(s).

(2) *Object attribute map:* Given a set of attributes and/or attribute values, determine the objects satisfying them, and inversely.

(3) *Sensor context map:* The dynamic context within which all sensory interpretation, i.e., the transcription of time/space dependent features of a sensory frame into time/space invariant features, takes place.

(4) *Task/Goal Map:* A map relating the task, and goals constraining its achievement, to expectations and events.

Information state space interacts with physical state space, or as it will be referred to here, the world space potential, at the bottom. On the left, sensors provide data to an ascending hierarchy of increasing integration. Roughly these levels are, starting at the bottom:

(1) *Signal Acquisition and Feature Extraction:* Signals, whose interpretation is predefined, e.g., clock ticks, fuel left, velocity, location etc. and are generated by sensors specific to the signal, and predefined sensory features are continuously looked for and extracted in the current sensor frame. We call these extracted features *observations*.

(2) *Space/Time Integration and Fusion:* Signals and atomic features are joined in patterns coming from an *attention controller,* forming *compound features*.

(3) *Event Extraction:* The matching of an observed (instance of) compound feature and a (generic) expectation, e.g., a man is coming toward me.

(4) *Event Assimilation:* The process whereby an event is integrated within the context of the current task directed behavior, e.g., the man coming toward me is my brother, whom I am meeting for lunch.

The first two levels constitute what might be called a *preattentive phase* while the latter two levels constitute a *postattentive phase*. The first makes up the pool of information aquired by the sensors in a sensor-frame, while the latter represent that select subset extracted from that pool by virtue of their having been matched via the outgrowth of an expectation.

Clearly, what we have in mind here for *sensory processing* is a kind of generalized *model-based vision* paradigm [30]. The central idea is that at each level, other than at the lowest, the only information to become assimilated into information state space is that which has been anticipated by expectations which are in a form to be matched from a set of predefined features. At the lowest level, the features to be extracted are determined during the design process by the class of tasks which the machine is expected to carry out. The so called "blackboard" model, used in the Hearsay-II architecture [13] seems appropriate for performing this hierarchical integration [9,12]. "Knowledge sources" coming from below in the form of feature observations are matched hierarchically to predictions derived from hypotheses in turn stemming from expectations.

In general, feasibility planning cannot a priori place a strict chronological order on when the expected events will actually occur, but can only place a partial order on them [28]. However, the actual occurrence of some event during execution of the plan may cause the order, time or time range of other expected events to be determined. In addition, some expected events will only conditionally occur, or will be mutually exclusive with others etc. Since it is not desirable that all possible expected events be looked for simultaneously, a technique borrowed from discrete simulation can be used to order expectations and hence control the opening and closing of the windows associated with them, namely the *event queue*.

All expectations are associated with either a predicted time range or location range, i.e., their window in space-time. Expectations whose time range to occur can be predicted are placed on a partially ordered event queue at the proper location.

The top(s) of this queue, when real time catches up to the internal simulated time, will contain the expectation(s) which can be looked for. The power of this technique lies in the fact that expectations can be "chained" by causing an event to spawn later expectation(s). In this way, the exact time and order of events hypothesized by the planner, in the form of the expectation sequence, need not be determined until execution of the plan.

For example, the event associated with the time at which a course trajectory is completed and a new heading started, can cause the insertion of an analogous expectation corresponding to the completion of the second trajectory, etc. In this way each trajectory of a multi-trajectory route spawns expectations for instantiating the specifics for the following trajectory, so that the events associated with the execution of become self-sustaining for the entire route. The same technique may be applied to other expected events which are partially ordered with respect to one another.

In analogy with the partially ordered *temporal event queue*, we define a *spatial event queue*, which orders with respect to location and orientation. Given the current location, this queue contains the anticipated events for which predictions can be made about sensory space features at that location. This is then used to predict and focus attention in the processing of sensory data as a function of location and orientation.

On the right hand of figure 5 is the descending hierarchy of feasibility planning, plan decomposition and instantiation, and execution, i.e., the generation of sequences of actions via the effectors to accomplish a task given at the top of the hierarchy.

For an a priori complete world model, in which the successful generation of a feasible plan is tantamount to its successful execution, the hierarchy is as stated. In the case where a priori information must be augmented by sensory data for world model completeness, planning must also, in conjunction with effector commands, generate expectations to be used by sensory processing to acquire information not known a priori.

In an incomplete world model, the linear hierarchy becomes much less a depiction of what must occur. In the case of a complete world model, the feasible plan need only be instantiated with information either known a priori or known a priori that it can be acquired through sensory processing. In an incomplete world model, any or all levels of the decomposition and instantiation process must be prepared to reject as *infeasible* the (sub)task given it. Hence the rather static hierarchy of the complete world model must be replaced with the much more dynamic process by which each level of the hierarchy must be prepared to perform feasibility planning for the (sub)task at that level. Knowledge of when this feasibility planning is required to take place is critically dependent on the occurrence of disabling events, which event also (potentially) provides the additional information required to generate the new feasible (sub)plan.

Status information ascends from each level to the level above it, and contains information as to the completion of (sub)tasks. For example, if in response to a disabling event, a particular level is unable to generate a new feasible plan, then this information together with the disabling event instantiation information is passed up to the next higher level, where a broader set of alternatives for feasibility planning can be evaluated within the (new) altered context.

Another consequence of the inability to compute the state of the environment as a function of effector actions is the need to continually assess the task relevant events extracted from physical state space. This activity is controlled by an *assessment sequence* which has been generated in parallel with the actuation control sequence. A major component of this assessment sequence is the hierarchically descending expectation sequence described earlier by which the enabling and potential disabling events were acquired. After event acquisition, the hierarchically ascending assessment sequence is also used to identify these events with various stages of task accomplishment, and hence provides at the top of the hierarchy the basis for *situation assessment*, with respect to the task.

The third leg of information state space is the world model. At each level of the hierarchy it acts as an intermediary by providing appropriately resolved enabling and disabling events acquired by sensory processing to planning and execution. In addition, world modeling performs three other tasks of a hierarchical nature:

(1) Provides the means by which hypothesized actions generated as part of feasibility planning are evaluated for their consequences. Evaluation may occur in the form of closed form solutions, model-based simulation or incomplete heuristic methods.

(2) Provides the basis for *task assessment* by resolving assimilated events (obtained from task level sensor processing) against the assessment sequence to generate an *assessment history*. This history is used for, among other things, feasibility (re)planning in which planning must start from an intermediate state of affairs.

(3) Provides for the hierarchical decomposition of expectations for the purpose of predicting features and the controlling of focus of attention in sensor processing by prioritizing the efforts at matching predicted features and observations.

Sensor processing is controlled from below by the observations it makes and from above by the patterns looked for as determined by predictions generated by world modeling based on expectations generated in conjunction with the decomposition and instantiation of a feasible plan.

9. SUMMARY

We have argued in this paper that the classical model of control, within the context of what we have called a complete world model, is inadequate for a control architecture for an autonomous agent. This is in large part because the classical model does not address the problem of selective sensory processing: *In order that a change in physical state space make a change in information theoretic state space, the change in physical state must have been anticipated in the form of an appropriate expectation.* This expectation serves as a template against physical state space in order that an instance be found, and is mandatory in an agent which has only a partial model of its environment.

The concept of an event, identified with the acquisition of sensory information which results in behavior different than what otherwise would have occurred, is central to an information theoretic understanding of an autonomous agent. We have argued that these events, unlike in the classical control model, must be actively selected from the much larger available set, by matching expectations generated by task specific planning.

Events may be further broken down into enabling and disabling events. The enabling events are just those which, based on a current incomplete world model, are sufficient for achieving the task. The disabling ones are not necessary, but if they occur, provide the needed additional information to generate the new set of sufficient events. Since not every potentially disabling event can be looked for, a large part of intelligence is constituted in the ability to select just those which have a high probability of occurring.

We have argued that the basis for understanding these issues lies in treating physical state space and information state space simultaneously.

ACKNOWLEDGEMENTS

The author wishes to thank members of the Robot Systems Division for many stimulating discussions concerning the needs of the Multiple Autonomous Undersea Vehicle project (MAUV), and particularly Hoosh Abrishamian, Jim Albus, John Fiala, Martin Herman, Russell Kirsch and Christoph Witzgall for providing draft comments.

REFERENCES

[1] Albus, J. S., Barbara, A. J. and Nagel, R. N., "Theory and Practice of Hierarchical Control," *Proceedings of the 23rd IEEE Computer Society International Conference*, September

1981, pp. 18-39.

[2] Albus, James S. and Blidberg, D. Richard, "A Control System Architecture for Multiple Autonomous Undersea Vehicles," *Proceedings of the Fifth International Symposium on Unmanned Untethered Submersible Technology,* Merrimack, New Hampshire, June 22-24, 1987, pp. 444-466.

[3] Albus, J. S., Lumia, R. and McCain, H., "Hierarchical Control of Intelligent Machines Applied to Space Station Telerobots," in *Proceedings of the Workshop on Space Telerobotics,* July, 1987, Jet Propulsion Laboratory Publication 87-13, ed. by G. Rodriquez, pp. 155-165.

[4] Arbib, Michael A., "Perceptual Structures and Distributed Control," in *Handbook of Physiology: The Nervous System, II, Motor Control,* ed. by V. R. Brooks, 1981, pp. 1449-1480.

[5] Brooks, Rodney A., "A Robust Programming Scheme for a Mobile Robot," in *Languages for Sensor-Based Control in Robotics,* edited by U. Rembold and K. Hörmann, NATO ASI Series, Vol. F29, Springer-Verlag, 1987.

[6] Brown, Frank M., *The Frame Problem in Artificial Intelligence,* Proceedings of the 1987 Workshop, April 12-15, University of Kansas, Lawrence, Kansas, pp. 5-28.

[7] Casti, John L., "Behaviorism to Cognition: a System-Theoretic Inquiry into Brains, Minds, and Mechanisms," in *Real Brains, Artificial Minds,* ed. by John L. Casti and Anders Karlqvist, Elsevier Science Publishing Co., Inc., 1987, pp. 47-75.

[8] Christiansen, Alan D., "The History of Planning Methodology: An Annotated Bibliography," *SIGART Newsletter,* published by the Association of Computing Machinery, New York, No. 94, October 1985, pp. 44-46, 43, 18, 21, 25, 14, 23.

[9] Corkill, Daniel D., Gallagher, K. Q. and Murray, K. E., "GBB: A Generic Blackboard Development System," notes made available at the *Blackboard Workshop* held at Carnegie-Mellon University and organized by S. Shafer and A. Stentz, June 12-13, 1986.

[10] Craig, John J., *Introduction to Robotics: Mechanics and Control,* Addison-Wesley Publishing Co., Reading, Massachusetts, 1986.

[11] Dewdney, A. K., "Computer Recreations," *Scientific American,* Vol. 256, No. 3, March 1987, pp. 16-24.

[12] Durfee, Edmund H., Lesser, Victor R. and Corkill, Daniel D., "Coherent Cooperation Among Communicating Problem Solvers," *IEEE Trans. Comput.,* Vol. C-36, No. 11, November 1987, pp. 1275-1291.

[13] Erman, L. D., Hayes-Roth, F., Lessor, V. R. and Reddy, D. R., "The Hearsay-II Speech Understanding System: Integrating Knowledge to resolve Uncertainty," *Computing Surveys,* Vol. 12, June, 1980, pp. 213-253.

[14] Fodor, Jerry A., "The Mind-Body Problem," *Scientific American,* Vol. 244, No. 1, January, 1981, pp. 114-123.

[15] Halpern, Joseph Y., "Reasoning about Knowledge: An Overview," in *Theoretical Aspects of Reasoning About Knowledge,* Proceedings of the 1986 Conference, March 19-22, Monterey, Calif., edited by Joseph Halpern, Morgan Kaufmann, 1986, pp. 2-17.

[16] Herman, Martin and Albus, James S., "Real-Time Hierarchical Planning for Multiple Mobile Robots," Proceedings for *DARPA Workshop on Knowledge-Based Planning,* Austin, Texas, December 8-10,1987.

[17] Iarbus, A. L., *Eye Movements in Vision,* Plenum Press, New York, 1967.

[18] Jenkins, L., "Space Telerobotic Systems: Applications and Concepts," in *Proceedings of the Workshop on Space Telerobotics,* July, 1987, Jet Propulsion Laboratory Publication

87-13, ed. by G. Rodriquez, pp. 29-34.

[19] Kanade, T., Thorpe, C. and Whittaker, W., "Autonomous Land Vehicle Project at CMU," in *Proceedings 1986 ACM Computer Conference,* Cincinnati, OH, February, 1986.

[20] Kent, Ernest W. and Albus, James S., "Servoed World Models as Interfaces Between Robot Control Systems and Sensory Data," *Robotica,* Vol. 2, 1984, pp. 17-25.

[21] Kripke, Saul, "Semantical Analysis of Modal Logic," *Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik,* 9, 1963, pp. 67-96.

[22] McCarthy, J. M. and Hayes, P. J., "Some Philosophical Problems From the Standpoint of Artificial Intelligence," *Readings in Artificial Intelligence,* Tioga Publishing Co., Palo Alto, Calif., 1981, pp. 431-450.

[23] Meystel, A., *Intelligent Mobile Autonomous System,* progress report, Drexel University, November, 1985.

[24] Moravec, Hans P., "Certainty Grids for Mobile Robots," in *Proceedings of the Workshop on Space Telerobotics,* July, 1987, Jet Propulsion Laboratory Publication 87-13, ed. by G. Rodriquez, pp. 307-312.

[25] Morgenstern, Leora, "A First Order Theory of Planning, Knowledge, and Action,", in *Theoretical Aspects of Reasoning About Knowledge,* Proceedings of the 1986 Conference, March 19-22, Monterey, Calif., edited by Joseph Halpern, Morgan Kaufmann, 1986, pp. 99-114.

[26] Pylyshyn, Zenon W. editor, *The Robot's Dilemma: The Frame Problem in Artificial Intelligence,* Ablex Publishing Corp., Norwood, New Jersey.

[27] Jasinschi, Radu, *Space-Time Filtering, Sampling and Motion Uncertainty,* Computer Science Department, Carnegie-Mellon University, 1988.

[28] Sacerdoti, Earl D., *A Structure for Plans and Behavior,* Elsevier North-Holland, Inc., New York, 1977.

[29] Schweizer, P. F. and Oravec, J. J., "Intelligent Control and Signal processing for an Autonomous Underseas Vehicle," Westinghouse R&D paper 86-1c5-AVERS-p1, August 19, 1986.

[30] Shneier, M. O., Lumia, R. and Kent, E. W., "Model-Based Strategies for High-Level Robot Vision," *Computer Vision, Graphics, and Image Processing,* Vol. 33, No. 3, 1986, pp. 293-306.