Submitted to *The Journal of the Learning Sciences.*

Creative Conceptual Change

Ashwin Ram, Kenneth Moorman, Juan Carlos Santamaría

College of Computing

Georgia Institute of Technology

Atlanta, Georgia 30332-0280

E-mail: ashwin.ram@cc.gatech.edu

Abstract

Creative conceptual change involves (a) the construction of new concepts and of coherent belief systems, or theories, relating these concepts, and (b) the modification and extrapolation of existing concepts and theories in novel situations. We discuss these and other types of conceptual change, and present computational models of constructive and extrapolative processes in creative conceptual change. The models have been implemented as computer programs in two very different task domains, autonomous robotic navigation and fictional story understanding.

Contents

1 Introduction

Much research in conceptual change has focused on either developmental conceptual change in children or scientific conceptual change in expert adults. Keil (1989), for example, is concerned with the nature of children's concepts, their differences from concepts that adults have, and how children's concepts change through cognitive development. Such conceptual change is qualitative; not only do children learn new concepts, the nature of the concepts themselves changes through development. The study of scientific conceptual change is concerned with how new conceptual structures in a scientific community come to replace existing ones (see, for example, the description of the geological revolution taking place in the 1950s and 1960s (Giere, 1988)) either through scientific revolutions (Kuhn, 1962) or through longer-term enterprise (Gruber, 1989).

In contrast, the conceptual change that we are concerned with here is the everyday kind. It involves everyday reasoning by intelligent systems, human or machine, in situations that allow (or require) creativity and learning. Nersessian (1992) argues that "the problem-solving strategies scientists have invented and the representational practices they have developed over the course of the history of science are very sophisticated and refined outgrowths of ordinary reasoning and representational processes." However, no comprehensive models of these "ordinary" forms of reasoning exist; the focus of our research is on that topic. Furthermore, as opposed to the mechanisms of cognitive development or the construction of shared concepts through social interaction (as in, for example, Roschelle's (1992) account of conceptual change through social collaboration), our focus is on conceptual change through the cognitive processes of learning and creativity. To emphasize this focus, we will refer to this kind of conceptual change as *creative conceptual change*. In this article, we will argue that conceptual change requires two kinds of creative processes: the construction of new concepts from input information, and the extrapolation of existing concepts in novel and unfamiliar situations. We will present computational models of these processes and discuss the properties of conceptual change as carried out by computer implementations of these models.

The first kind of process involves reformulating low-level information, such as perceptual or sensorimotor data, into higher-level abstractions. Consider a reasoner in a strange environment. It may improve its ability to act in that environment by learning about the effects of its actions in that environment (for example, learning to control a car on the highway). The actions themselves may be new and unfamiliar; a reasoner may need to learn about its own actions and the interactions of these actions with the environment (for example, learning to drive a car in the first place). The reasoner may also need to learn about the structure of the environment itself (for instance, learning the layout of the roads in a city). All of these scenarios require creative conceptual change of a particular kind: the construction of conceptual representations to represent causal and predictive relationships between sensory inputs, motor actions, and the environment. We will call this *constructive conceptual change* since it involves the construction of new concepts from sensorimotor experience. Although this process is not usually thought of as “creative,” we will argue that the process is in fact so because it results in representations that are novel, useful, and qualitatively different from those that the reasoner initially starts out with.

Another kind of process involved in creative conceptual change is that commonly associated with fictional and imaginative scenarios. Consider reading a science fiction story. This requires a temporary suspension of disbelief and the extension or adaptation of existing concepts to create a conceptual model of the described situation (which may be very different from the reasoner’s real-world experience). We will call this *extrapolative conceptual change* since it involves extrapolation from existing concepts to create new ones. Consider, for example, reading a *Star Trek* story (or watching a *Star Trek* episode on television). In order to understand the concept of time travel, one has to extend one’s everyday understanding of travel, a physical concept, and create a new concept which is the temporal analog of the physical concept. In some cases, extrapolation may not require a large conceptual leap. For example, one can initially model a “phasor” as a kind of gun and gradually refine that model while reasoning with it—a process not unlike the gradual model revision process discussed by Clement (1989) in a more real-life domain. The outcome of such a process is a new concept (and an associated model) that is initially created by adapting an existing concept and refined through further experience.

In addition to guiding the reasoner in the current situation, the new concepts (or systems of concepts) may be useful in other contexts as well. As we will argue, the mechanisms and knowledge involved in such reasoning are not unique to understanding fiction; they are really no different from the mechanisms and knowledge involved in reasoning in non-fictional or real-world situations. Although models of creativity and conceptual change have traditionally been developed separately from models from everyday reasoning, the constructive and extrapolative processes discussed here are not viewed as being extraordinary or special; they, and the creative conceptual change that they result in, are an integral part of everyday reasoning.

We believe that "conceptual change" does not necessarily imply the radical restructuring of pre-conceptual knowledge that is often prerequisite to the use of this term in the literature. While the construction and extrapolation of concepts through the creative processes presented in this article is often a gradual and incremental process, the properties of our model are similar enough to the properties of the more radical type of conceptual change to merit the use of the term "conceptual change" in the broader context. Specifically, both constructive and extrapolative conceptual change have much in common with each other, as well as with developmental and scientific conceptual change. Keil (1989) argues that systematic belief systems, or "theories," are important in developmental conceptual change, and that causal relations are essential and more useful in such theories than other sorts of relations (see also Neisser, 1987). Causal belief systems are critical in extrapolative conceptual change as well since they guide and constrain the creative adaptations performed by the reasoner. Keil views concepts as partial theories in that they embody explanations or mental models of the relations between their constituents, of their origins, and of their relations to other clusters of features (see also Johnson-Laird, 1983; Murphy & Medin, 1985; Thagard, 1989). Similarly, the representations constructed through extrapolative and constructive conceptual change also embody such explanations (albeit not always "correct" ones). Analogy and mental modeling play a crucial role in theories of scientific conceptual change (e.g., Nersessian, 1992), and in extrapolative conceptual change as well. All these types of conceptual change rely both on inductive and analytical reasoning processes, though sometimes to different extents. Typically, analytical processes are used when appropriate theories are available to support analysis (such as in experts), and inductive processes are used when such theories are not available (such as in novices). In addition to the creation of individual concepts and their gradual evolution through experience, conceptual change may also involve the reorganization of an entire system of concepts.

Our decomposition of the processes of conceptual change into constructive and extrapolative forms is a functional one. Rather than discuss conceptual change in children and adults, in laypersons and scientists, or in physics and mathematics, we will focus on the underlying *functions* of conceptual change (the construction and evolution of concepts), on the *mechanisms* or processes that achieve these functions, and on the *knowledge* that these mechanisms rely on. Such a decomposition is methodologically useful because it allows us to study the types of knowledge and processes that underlie conceptual change and their commonalities across different performance tasks, domains, and levels of expertise of the reasoners. Furthermore, our theory of the functions, mechanisms, and knowledge involved in conceptual change will be a computational one (see Boden, 1988, 1991, and Simon & Halford, 1995, for a discussion of the importance and role of computational modeling in psychological inquiry). Such models provide a computational or information-processing account of conceptual change and suggest an explanation for the observed behaviors and properties of this phenomenon. The primary contribution of this article, therefore, is a computational model of creative conceptual change that explains the epistemological basis for and mechanisms underlying the properties of conceptual change that have been noted in the literature. Furthermore, we suggest that the processes of creativity that bring about conceptual change in our theory are an intrinsic part of both incremental and radical conceptual change; thus, our theory accounts for a range of phenomena in conceptual change in everyday situations that have traditionally not been included in theories of conceptual change.

In this article, we will discuss computational models of constructive and extrapolative conceptual change, focusing in particular on two computer programs that instantiate the models in two very different "everyday" task domains. The computer programs aid in the development and evaluation of the models, and provide an experimental framework for further exploration of theoretical ideas. We will conclude with a discussion of a framework for the integration of these (and other) methods of conceptual change into a single "multistrategy" system, and discuss implications from the results of these case studies for the nature of conceptual change in general.

2 Case studies in creative conceptual change

The computer programs presented here serve as case studies of constructive and extrapolative processes in conceptual change. The first program, called SINS (Self-Improving Navigation System), is an autonomous robotic navigation system that learns to navigate in an obstacle-ridden world (Ram & Santamaría, 1993). Autonomous robotic navigation is the task of finding a path along which a robot can physically move through a given environment and then executing the actions to carry out the movement in a real or simulated world. The ability to adapt to changes in the environment, and to learn from experiences, is crucial to adequate performance and survivability in the real world. SINS uses fast robotic control augmented with multiple learning methods that allow the system to adapt to novel environments and to learn from its experiences. The core of the system is a constructive conceptual change mechanism that autonomously and progressively constructs representational structures that encapsulate the system's experiences. These structures comprise a higher-level representation of the system's perceptual and sensorimotor interactions with its environment, and are used to aid the navigation task in two ways: they allow the system to dynamically select the appropriate robotic control behaviors in different situations, and they also allow the system to adapt selected behaviors to the immediate demands of the environment.

The second case study is based on a computer program called ISAAC (Integrated Story Analysis And Creativity), which is a natural language understanding system that reads short stories from the science fiction genre (Moorman & Ram, 1994a, 1994b). Such stories require creative understanding, in which the reader must learn enough about an alien world in a short text in order to accept it as the background for the story, and simultaneously must understand the story itself. ISAAC implements a process of extrapolative conceptual change which is based on the creative extrapolation, modification, or extension of existing concepts and theories to invent new ones. The extrapolation is constrained by the content of the story, by the system's existing concepts and theories, and by the requirements of the reading and understanding task.

There is much debate on when representational construction is truly "conceptual change" as opposed to merely "learning". In this article, we will avoid this debate and take as our criterion for creative conceptual change the construction of new representations that are novel, useful, and qualitatively different from the initial representations. Thus, a robot's creation of high-level tactical or strategic concepts (such as retreat) from low-level sensorimotor observations constitutes creative conceptual change, as does a science fiction reader's invention of a fictional theory of physics for time travel. As the case studies will reveal, there is much in common between these two kinds of conceptual change, as exemplified by the computer systems, despite their superficial differences. Both systems use multiple types of knowledge, and multiple types of reasoning processes. Both rely on multiple sources of constraints on these processes, including theories, knowledge and knowledge organization, and both are situated in actual experience on a complex, real-world task.

Both systems learn autonomously through experience. Creative conceptual change in both systems is a process of gradual evolution of concepts to create better approximations of the observed world. The new concepts contribute significantly to the systems' abilities to carry out their respective tasks, and may be very different from those that the systems initially started out with.

The differences between the systems are also of interest. SINS relies directly on its experiences in the real world as it is known to us[1], whereas ISAAC's real world is that of natural language texts which vicariously describe fictional world experiences of fictional characters. ISAAC integrates its processes using explicit arbitration and control; thus, conceptual change in ISAAC is guided by the particular needs and goals of the program. SINS, in contrast, learns "automatically" through its task performance, and thus is better characterized as having an implicit orientation or goal to learn (Barsalou, discussed in Leake & Ram, 1993). The two systems are discussed in more detail below.

3 Constructive conceptual change

Constructive conceptual change is the process of constructing conceptual representations that designate causal relationships between sensory inputs and motor outputs. Such conceptual representations are novel and useful: novel because new concepts are created as a result of the agent's experiences to represent causal relationships which were not known beforehand, and useful because they enable the agent to form expectations about what might result after a specific action is executed under a particular situation. Thus, the ability to perform constructive conceptual change results directly in the ability to enhance task performance autonomously, since the newly learned concepts can be used to predict, with some amount of certainty, the outcomes of possible actions and therefore the agent can decide which action to take to best accomplish its goals. Without such predictive capability, an agent's behavior can only rely on feedback from sensory information alone, resulting in actions that are not anticipatory or sufficiently adaptive in complex, dynamic environments.

Concepts that designate casual relationships contain two types of knowledge: knowledge that represents relevant situations and knowledge that represents the casual mappings from actions to

[1]The system has been implemented on an actual Denning MRV-III robot as well as on a simulate platform which allows controlled experimentation over a wide range of system configurations and environmental situations.

situations. Representing relevant situations is helpful because it is often the case that sensory inputs do not convey all the required information to assess the true situation or state of the environment, and without this information the agent may decide to perform actions that will result in poor performance. For example, a mobile robot that decides where to move next based on only the current "snapshot" of its sonar sensors may fail to detect it is entering into an area of higher density of obstacles and thereby move along a collision direction. Additionally, representing the causal mappings between actions and situations is useful because the agent can decide which action to select in a given situation if it knows what situation will result after performing that action. For example, if the robot knows that by slowing down when entering a higher density area will avoid collisions, then it may decide to do so in order to avoid damage.

While simple stimulus-response rules might capture some of the necessary predictive relationships, it is useful to construct concepts that designate generalized causal relationships that not only are useful in making inferences about expectations but also represent causal information at a higher level of abstraction than the basic perceptual-motor information that the agents starts with. This level of representation is generalizable to novel and unexpected situations. As such, creation of these concepts involves a true conceptual change in terms of the fundamental nature and level of abstraction of the representations, and not simply clustering of information at the level of existing perceptual-motor representations.

Initially, the agent starts with elemental concepts about perception and action. Without loss of generality, we may regard these as sensory inputs and motor control outputs. Such concepts are typically represented in continuous, numerical, or analog terms rather than as discrete, symbolic structures as traditionally used in artificial intelligence systems. With experience, the process of constructive conceptual change creates new, more complex, and more abstract concepts that embody the elemental concepts in a way that causal relationships are preserved. These new concepts designate patterns of elemental concepts that capture some invariant aspect or regularity of the environment with respect to the actions the agent is able to perform. The higher level of abstraction enables the agent to understand and reason about its environment (and about its interaction with the environment) in terms of those regularities, which are unique to the particular ecological niche in which the agent is performing its task and the sensory and motor capabilities of the agent. These regularities provide a reliable source of inferential knowledge which can be used to create expectations and act accordingly. Furthermore, with experience, the new concepts might become reliable enough to serve, in turn, as elemental concepts for constructive conceptual change at higher levels of abstraction.

As an illustration of what we mean by constructive conceptual change, consider the problem of spatial representation and exploration in a real-world environment. An agent learning about its physical environment through exploration might learn about relevant situations (e.g., streets and intersections) and build a cognitive map (i.e., a causal mapping between situations and actions) representing topological and metrical information about the space around it (e.g., making a right in this intersection takes me from 7th street to 2nd avenue). Several studies have suggested that cognitive maps are organized into layers (e.g., Lynch, 1960; Piaget & Inhelder, 1967; Siegel & White, 1975). The cognitive map contains information about space, locations, connectivity, and distance, learned gradually through interaction with and exploration of the environment. These studies have motivated computational models of robot map-learning as well. For example, Kuipers & Byun (1991) describe a simulated robot, NX, that learns a hierarchy of types of spatial knowledge organized into sensorimotor, control, procedural, topological, and metrical knowledge. At the lowest level, the robot has access to raw sensory data from the environment. The robot's representation of the space surrounding it undergoes a series of conceptual changes as sensorimotor data (which is continuous and numerical) is reformulated and abstracted into successively higher-level descriptions (which are discrete and symbolic).

The SINS system discussed here also learns from continuous sensorimotor information, but addresses a somewhat different problem in constructive conceptual change: that of learning the appropriate concepts for dynamic and adaptive control of action. The control of action or decision-making process of an agent is implemented by executing a *policy*. Adaptive control of action occurs when the agent adjusts or adapts its policy according to the situation it is currently facing, resulting in improved performance. This problem is different from the map-learning problem which involves constructing representations of the environment; here, the agent must not only learn about the environment but also create representations of its interaction with the environment and learn appropriate policies for different situations.

In SINS, the perception-action task and the adaptation-learning task are integrated in a tightly knit cycle, similar to the "anytime learning" approach of Grefenstette & Ramsey (1992). Together, the mechanisms that carry out these tasks result in the construction of conceptual structures that encapsulate continuous sensorimotor experience. These structures are modified continuously even as they are used to guide action. The problem solving and learning process must operate continuously; there is no time to "stop and think," nor a logical point in the process at which to do so. Each conceptual modification represents a step in a sequence of successive approximations in which the concepts evolve into stable perception-action models representing the agent's "understanding" of the world and of its interactions with the world. These models in turn result in improved performance on a wide range of task environments.

4 Technical details: The SINS system

4.1 Task: Autonomous robotic navigation

Autonomous robotic navigation is defined as the task of finding a path along which a robotic agent can move safely from a source point to a destination point in a obstacle-ridden terrain (path planning) and executing the actions to carry out the movement in a real or simulated world (plan execution). Several methods have been proposed for this task, ranging from high-level planning methods to reactive methods.

High-level planning methods use extensive world knowledge and inferences about the environment they interact with (Fikes, Hart & Nilsson, 1972; Georgeff, 1987; Maes, 1990; Sacerdoti, 1975). Knowledge about available actions and their consequences is used to formulate a detailed plan before the actions are actually executed in the world. These methods can successfully perform the path-finding required by the navigation task, but only if an accurate and complete representation of the world, and of available actions and their effects, is available to the agent. Situated or reactive control methods have been proposed as an alternative to high-level planning methods (Arkin, 1989; Brooks, 1986; Kaelbling, 1986; Payton, 1986). In these methods, no planning is performed; instead, a simple sensory representation of the environment is used to select the next action that should be performed. Actions are represented as simple behaviors, which can be selected and executed rapidly, often in real-time. These methods can cope with unknown and dynamic environmental configurations, but only those that lie within the scope of predetermined behaviors.

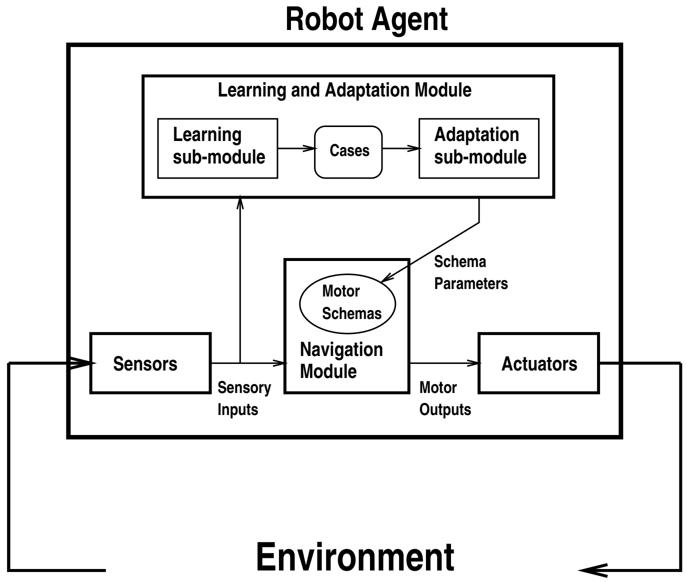

Figure 1: Functional architecture of the SINS system.

In a complex and dynamic environment, an agent needs to develop a combination of the above abilities: the ability to respond accurately and in real-time when facing unfamiliar situations, and the ability to use high-level knowledge to improve its performance when facing familiar situations. Furthermore, every experience should contribute to improve the agent's performance in the future: experiences with unfamiliar situations should contribute towards making such situations more familiar, and experiences with familiar situations should contribute towards improving the accuracy and reliability of the agent's knowledge about such situations. In the SINS system, we have focused on developing a system with the above capabilities.

The SINS architecture is shown in Figure 1. The two main modules execute concurrently to perform the task of autonomous navigation. The navigation module implements a reactive policy that uses sensory inputs to determine motor control outputs. A reactive policy consists of a function that decides appropriate motor control outputs considering only the current set of sensory inputs. For example, a robot that senses its goal directly ahead and an obstacle immediately to its right may decide to move perpendicularly towards its left to avoid a possible collision (a conservative policy that prioritizes collision avoidance above all else). In SINS, the policy can be adapted at any time by modifying a set of control parameters, thereby modifying the decision-making process. For example, the above reactive policy can be modified to be less conservative by changing its control parameters of the function that implements it. A robot operating under the modified policy in the above situation may decide to move left (away from the obstacle) and forward (towards the goal) at the same time, thereby avoiding the obstacle but also making some progress towards the goal. The learning and adaptation module in SINS is responsible for characterizing the situation faced by the robot at a given time and adapting the robot's reactive policy so that the robot may navigate efficiently through that situation. Furthermore, this module is also responsible for learning about new situations and about appropriate control parameters the reactive policy should use under those situations.

The navigation module implements its reactive policy using schema-based reactive control (Arkin, 1989). The policy results from the combination of basic behaviors or *motor schemas* such as obstacle avoidance and movement towards a goal. Each motor schema individually recommend specific motor actions which are then combined to produce the final action of the agent. Potential field or vectors are used to represent the recommendations of the motor schemas; thus, a final motor action can be computed as the sum of the vectors of each motor schema and be delivered to the robot's effectors. Each motor schema uses current sensory information from the environment and control parameters to compute its potential field which recommends the direction and speed at which the robot is to move given current environmental conditions. For example, the motor schema AVOID-STATIC-OBSTACLE directs the system to move itself away from detected obstacles, and the motor schema control parameter **Obstacle-Gain** determines the magnitude of the repulsive potential field generated by the obstacles perceived by the system. The vectors produced by all the schemas are then combined to produce a potential field that directs the actual movement of the robot. Simple behaviors, such as wandering, obstacle avoidance, and goal following, can combine to produce complex emergent behaviors in a particular environment. Different emergent behaviors can be obtained by modifying the simple behaviors through their control parameters. A detailed description of schema-based reactive control methods can be found in Arkin (1989).

Different policies resulting from different combinations of motor schema control parameters values would cause different behaviors to be exhibited by the system (see Figure 2.) Since environments with different characteristics require different behavioral configurations for successful performance, on-line selection and modification of the appropriate parameters based on the current environment can enhance navigational performance. This has been demonstrated in the ACBARR system (Ram, Arkin, Moorman & Clark, 1992). In this system, hand-coded strategies are used

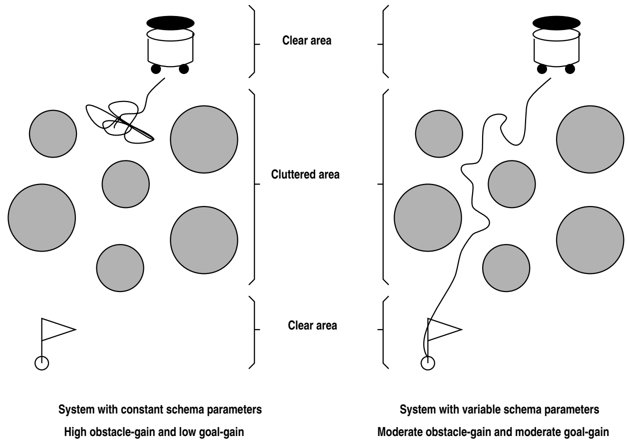

Figure 2: Typical navigational behaviors of the autonomous robotic system. The figure on the left shows the non-learning system with high obstacle avoidance and low goal attraction. On the right, the learning system has lowered obstacle avoidance and increased goal attraction, allowing it to "squeeze" through the obstacles and then take a relatively direct path to the goal.

to modify motor schema control parameter values according to perceived information. These strategies are represented as *cases* and used by a case-based reasoning method to dynamically select and modify control parameter values.

The issue we are concerned with here is how an agent might construct its own conceptual representations of its environment and what such representations might look like. Since these representations are based on (and capture) the agent's experiences, and since case-like representations have shown to be useful in guiding sensorimotor tasks, we hypothesize that constructive conceptual change is fundamentally a case learning activity, where the term "case" is used in a broader and somewhat different sense from traditional case-based reasoning as will be discussed in more detail below. The approach used in SINS is similar to ACBARR's in that on-line adaptation of motor schema control parameter values is used to enhance task performance at the sensorimotor level. However, the concepts that are used to guide such adaptations are constructed autonomously by the system through its own experiences. Casual relationships between sensory inputs and motor schema control parameters are learned, exploited for task performance, and further modified through experience during performance on those very tasks.

SINS uses four sensory inputs to characterize the environment and discriminate between different environmental situations: **Obstacle-Distance-Ahead**, **Obstacle-Distance-Behind**, **Obstacle-**

Distance-Right, and **Obstacle-Distance-Left**. These sensory inputs or variables provide a measure of the occupied areas that impede navigation in the direction towards, contrary to, right of, and left of the goal respectively. The values of these variables are constantly updated with the information received from the robot's sensors. Additionally, SINS uses four motor schema control parameters associated with three motor schemas: **Obstacle-Gain**, which determines the magnitude of an obstacle avoidance vector generated by the AVOID-STATIC-OBSTACLE schema; **Goal-Gain**, which determines the magnitude of a goal attraction vector generated by the MOVE-TO-GOAL schema; **Noise-Gain**, which determines the magnitude of a noise (random exploration) vector generated by the NOISE schema; and **Noise-Persistence**, which determines the duration for which the noise value generated by the NOISE schema is allowed to persist. The gain parameters define in what proportions the outputs of the individual motor schemas should be combined; higher values indicate greater influence of a particular motor schema in the final action of the robot. Parameter values are set periodically according to the recommendations of the case that best matches the current environmental situation. The new values remain constant over a "control interval" until the next setting period.

4.2 Representation: Continuous prototypical cases

SINS constructs representations of its interaction with the environment by observing and detecting sequences of sensorimotor associations that occur regularly. These sequences represent an internal model of the interaction between the robot and its environment and provide information about actions and their outcomes. SINS uses this information to enhance the performance of the navigation task since it is able to anticipate events and act in accordance with the representations rather merely react to the current environmental stimuli. Figure 3 illustrates a stylized example of the process.

As the agent approaches the first cluster of obstacles (towards the left of the figure), the **Obstacle-Distance-Ahead** input variable decreases. If **Obstacle-Gain** is too low, the agent will be trapped in a local minima inside the trap. Assume that by some mechanism (random exploration will suffice), SINS first increases the value of this control parameter and then the value of **Goal-Gain**, causing the agent to move away from the obstacles, lean towards its left, and overcome the first cluster of obstacles. This sequential pattern between input and output variables is detected

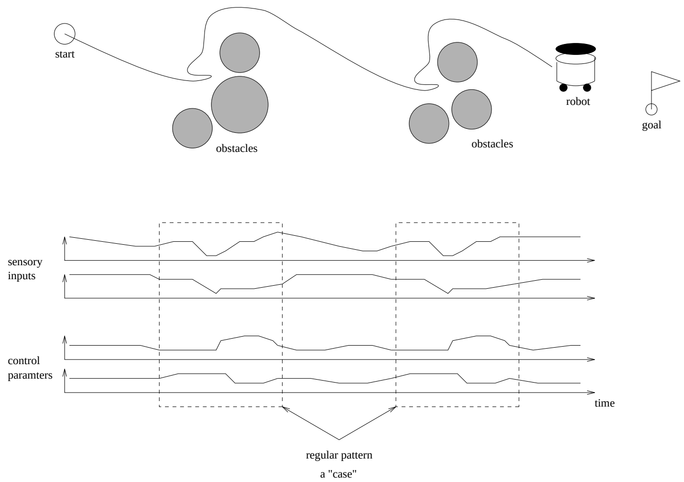

Figure 3: Example of a robot navigation experience. The top figure shows the world and the trajectory of the agent. The bottom figure shows the values of the sensory inputs and control parameters during the execution. The regular sensorimotor pattern around the obstacles represents the concept of a trap.

again as the agent approaches the second cluster of obstacles (towards the right of the figure), and is used again to guide the adaptation of the control parameters in a similar manner to overcome a similar perceived situation. From the point of view of the robot navigation task, the regularities in this sequence represents the concept of a trap known as a “box canyon” or “V”. Since the concept is a generalized representation of actual experiences, we refer to it as a “case”.

While this example illustrates the basic idea underlying SINS, it is simplified in many respects. First, an agent will in general not be able to detect and learn regularities with only two experiences. There are several solutions to this and similar problems, and several experiences may be necessary to distinguish between them. For example, increasing **Noise-Gain** might help in escaping a trap if at the same time the NOISE motor schema happens to direct the robot away from the obstacles, but this is not a generally useful strategy. Second, the robot’s perceptual input in general may not completely determine the environment. For example, there may be environmental situations in which the **Obstacle-Distance-Ahead** variable is decreasing even though the environment does not consist of obstacles arranged in a manner that produces a trap. For example, the cluster of obstacles may be arranged in an inverted “V” shape, which does not require an increased **Obstacle-Gain** to navigate.

Finally, the agent must be able to distinguish between the perceived situation inside a cluster of obstacles when it is entering the cluster and the same perceptual state when it is exiting the cluster. Representing this distinction requires not only an instantaneous “snapshot” description of sensory inputs, but also the recent history of changes of the sensory inputs. Distinguishing between “V” and “inverted V” environments requires an additional ability. These environments appear similar when first approached (similar **Obstacle-Distance-Ahead**) but differ in **Obstacle-Distance-Left** and **Obstacle-Distance-Right** during actual navigation. In SINS, situations are represented as patterns over consecutive samples of the sensory input variables. These patterns are represented as the sequences of associations that make up SINS’s cases. SINS can compare these sequences and remember differences between apparently similar sequences; these differences, over time, enable it to distinguish between sensorimotor concepts that require different actions, such as “V” and “inverted V”.

More precisely, learned situation-action parameterization concepts in SINS are represented by a set of cases, which are sets of associations between samples of sensory inputs and samples of motor schema control parameters over a suitable time interval. Sensory inputs provide information about the situation of the environment, and schema control parameters specify how to adapt the motor schemas of the navigation module in the environments to which the case is applicable. Each type of information is represented by a sequence of samples of a quantitative variable (a sensory input or a schema control parameter) at regular time intervals. An association consists of a set of sensory inputs values that characterize a particular environment situation grouped with a set of control parameter values that the agent should use in that environmental situation (see Figure 4.[2]) Thus, a case contains knowledge about environmental situations (i.e., combinations of samples of input variables) and knowledge about causal mappings between motor schema control parameters and situations (i.e. what situation commonly results when using a specific combination of control parameter values is used under the current situation).

This representation has three properties essential for the constructive conceptual change process.

[2] Although this representation does not capture truly analog action models in a real continuous sense, it does capture a very good approximation of them. The values of continuous variables are quantized in magnitude and over time. As long as the quantization process is relatively fine-grained with respect to the variables being quantized, the variables can be regarded as continuous (see, for example, Shanmugam, 1979, and the sampling theorem).

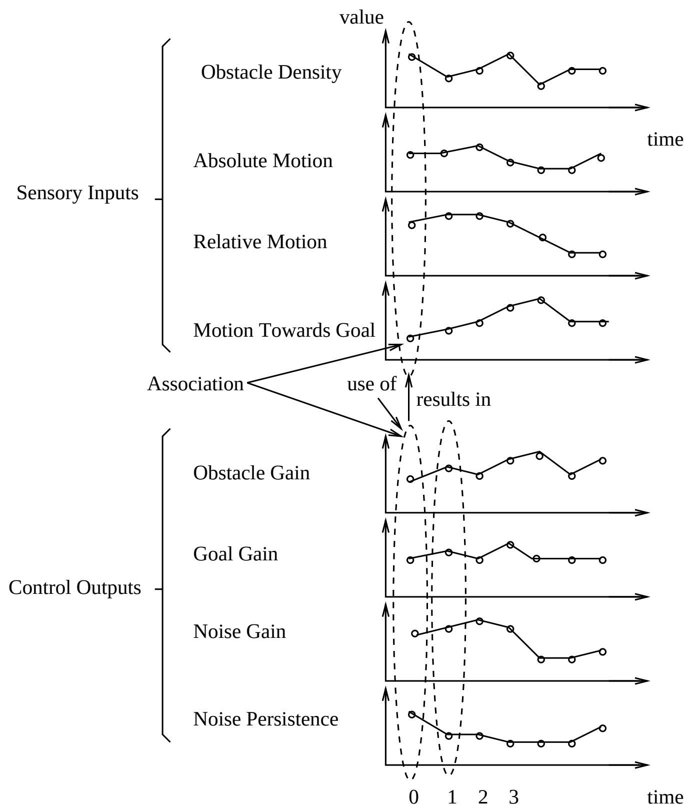

Figure 4: Sample representations showing the time history of analog values representing sensory inputs and control outputs. Associations between sensory inputs and control outputs are arranged vertically, and the sequence of associations over time is arranged horizontally. Each case in the system is represented in this manner, as is the current on-going navigational experience of the system.

First, the representation is capable of capturing a wide range of possible associations between of sensory inputs and motor schema control parameters. The continuous nature of the representation allows it to capture every possible value for each sensory input and control parameter. Second, it permits continuous progressive refinement of the associations. Mathematical formulas, such as gradient descent, can be used to update the variables continuously and progressively. Finally, the representation captures trends or patterns of input and output values over time. This allows the system to detect patterns over larger time windows rather than having to make a decision based only on instantaneous values of perceptual inputs.

There are also other advantages to using continuous values to represent associations. First, the fine-grained nature of the representations support the use of metric functions to assess the similarity among different environment situations; the learning and adaptation module can then use similarity measures to cluster the experiences. Second, continuous representations allow incremental incorporation of new information. The patterns are updated by progressively integrating new similar experiences together. Finally, continuous representations exploits the spatio-temporal continuity embedded in the world of physical objects. This characteristic allows the learning and navigation module to extrapolate the values of new environment situations from the values of previous similar situations.

4.3 Process: Concept construction and modification

There are four fundamental activities involved in the process of learning and using sensorimotor concepts represented as regular patterns of associations: discrimination of environmental situations, extrapolation of control parameters from previous situations, incremental refinement of associations, and anytime learning. Discrimination of environmental situations refers to the ability to assess the similarity between the current and previous experiences. This allows the system to cluster its experiences and organize its memory. Extrapolation of control parameters refers to the ability to infer appropriate control parameter values using previous experiences and through exploration. Incremental refinement of associations refers to the ability to incrementally improve the learned associations based on feedback from the environments in which they are deployed. Finally, anytime learning refers to the ability to use every experience for learning additional information in a continuous, on-line manner (Grefenstette & Ramsey, 1992).

The system's cases are automatically constructed using a hybrid case-based and reinforcement learning method without extensive high-level reasoning. The learning and navigation modules function in an integrated manner. The learning module is always trying to find a better model of the interaction of the system with its environment so that it can tune the navigation module to perform its function better. The navigation module provides feedback to the learning component so it can build a better model of this interaction. The behavior of the system is the result of an equilibrium point established by the learning module, which is trying to refine the model, and the environment, which is complex and dynamic in nature. This equilibrium may shift and need to be re-established if the environment changes drastically; however, the model is generic enough at any point to be able to deal with a very wide range of environments. Thus, adaptation is performed by retrieving the case that is most similar to the current situation and using the control parameter values suggested by the case. Learning is performed both by updating the case situation to be more similar to the current situation and by remembering, forgetting, or modifying the associated control parameters based on the observed outcome in the current situation.

The learning methods are based on a combination of ideas from case-based reasoning and learning, which deals with the issue of using past experiences to deal with and learn from novel situations (e.g., Kolodner, 1993; Hammond, 1989), and from reinforcement learning, which deals with the issue of strengthening the tendency to produce actions leading to satisfactory states of affairs (e.g., Sutton, 1992). Each case in SINS represents an observed regularity between a particular environmental configuration and the effects of different actions, and prescribes the values of the control parameters that are most appropriate (as far as the system can determine based on its previous experience) for that environment.

The learning and adaptation module performs the following tasks in a cyclic manner: (1) **perceive** and represent the current environment; (2) **retrieve** a case which represents an environment most similar to the current environment; (3) **adapt** the motor control parameters in use by the navigation module based on the recommendations of the case; and (4) **learn** new associations and/or adapt existing associations represented in the case to reflect any new information gained through the use of the case in the new situation to enhance the reliability of their predictions.

Learning is performed in a progressive manner by performing successive approximations to a conceptual space representing an efficient mapping between situations and control parameters. Each perceive/retrieve/adapt/learn cycle is an opportunity for creation and refinement of cases. Since learning is not supervised by an outside expert, an important question is how the system can determine whether the current experience should be used to modify and improve an existing case, or whether a new case should be created. In SINS, this is done through two means: an inductive procedure that uses information about prior applications of the case, and a reward signal which provides feedback based on the observed outcome of using a case. When a case is retrieved and applied to the current situation, a "relative similarity measure" is used to quantify how similar the current environment configuration is to the environment configuration encoded by the case, relative to how similar the environment has been in previous utilizations of the case. Intuitively, if a case matches the current situation better than previous situations it was used in, it is likely that the situation involves the very regularities that the case is beginning to capture; thus, it is worthwhile modifying the case in the direction of the current situation. Alternatively, if the match is not quite as good, the case should not be modified because that will take it away from the regularity it has been converging towards. Finally, if the current situation is a very bad fit to the case, it makes more sense to create a new case to represent what is probably a new class of situations. The modifications are mediated by the reward feedback, which is used to remember control parameters which produce favorable outcomes and discard those that do not. Over time, the system learns a set of cases representing prototypical situations that are most relevant to the navigation task.

4.4 Discussion: Constructive conceptual change

SINS is fully implemented in C++ on a simulated and an actual robotic platform. Detailed performance results from the SINS system are presented elsewhere (Ram & Santamaría, 1993) and show that SINS can improve the performance of the robot navigation system significantly along several performance metrics of interest. Here, we will discuss the system from the point of view of conceptual change. Constructive conceptual change is the process of creating concepts that are original and useful, and at a higher level of abstraction than the level of description of the original experiences. The main characteristics of this process is that it is automatic and on-going. The process requires several experiences from which concepts are induced; these concepts are then used to improve the performance on the task. Concepts are used even as they are created; however, they are modified incrementally with experience and are likely useful only after they have been proved to be consistent over time.

SINS carries out a constructive conceptual change process in which new conceptual representations of regularities in system-environment sensorimotor interactions are created through experience. The process results in a qualitative shift in the system's internal "theory" of perception and action, and results in new concepts that are creative by virtue of being both original and useful (Koestler, 1964; Turner, 1991). The function of the new concepts is to provide a predictive perception-action model that can be used to guide action and predict outcomes. This function is similar to that observed in early developmental research. For example, Canfield and Haith (1991) examined the formation of expectations for visual stimulus sequences in 2- and 3-month-old infants. In their studies, infants' visual fixations were monitored as they were exposed to predictable and unpredictable sequences of stimuli. Their results showed that young infants developed representations of the spatial and temporal parameters of stimulus sequences to anticipate future events. As discussed earlier, the concepts of expectation and anticipation are important in everyday adaptive action because they guide behavior to accommodate and adapt to unusual environments. Without expectations, behavior can only rely on feedback from sensory information alone, resulting in action that is too slow to be adaptive in dynamic environments since it must be appropriately synchronized (Schmidt, 1968).

The process of progressive modification of concepts guided by consistency and feedback has a similar parallel in developmental research. Vosniadou and Brewer (1992) investigated the change conceptual knowledge about the earth in elementary school children. When 1st-, 3rd- and 5th-grade children were asked a series of questions about the earth, their responses revealed considerable apparent inconsistency. For example, many children said that the earth is round but also stated that it has an end or edge from which people could fall. Vosniadou and Brewer state:

> A great deal of this apparent inconsistency could be explained by assuming that children used, in a consistent fashion, a mental model of the earth other than the spherical earth model... Some of these ... seem to be initial models children construct before they are exposed to the culturally accepted information that the earth is a sphere. In the process of knowledge acquisition, children appear to modify their initial

> models to make them more consistent with the culturally accepted model by gradually reinterpreting their presuppositions. [p. 535]

Note that the term "consistent" used in Vosniadou and Brewer's research implies consistency in expectations, rather than consistency over time as implemented in SINS. However, in the context of sensorimotor representations and the robotic navigation task, consistency over time is the same as consistency in expectations. Also note that, as suggested by this experiment, children are able to use incorrect models to solve problems. As inconsistencies are detected during problem solving activity, the model is successively refined until convergence to a model that produces good performance and no inconsistencies.

In SINS, sets of sensory inputs and control parameters are associated through the case representations that the system constructs. This grouping induces (albeit implicitly) a set of concepts that can be used to describe a control strategy or an environmental regularity. For example, if SINS is getting deeper into a crowded area, the values of the sensory inputs responsible for object detection will increase over time. A useful strategy in such a situation might be to back out and go around the obstacles. However, such a strategy cannot be expressed in purely perceptual terms; it requires the concepts of "crowdedness", "retreat", and so on, which are qualitatively different from the sensorimotor information that is initially available to the system.

Since learning and adaptation are based on a relative similarity measure, the overall effect of this process is to cause the cases to converge on stable associations between environment configurations and control parameters. Stable associations represent regularities in the world that have been identified by the system through its experience, and provide the predictive power necessary to navigate in future situations. The assumption behind this method is that the interaction between the system and the environment can be characterized by a finite set of causal patterns or associations between the sensory inputs and the actions performed by the system. The method allows the system to learn these causal patterns and to use them to modify its actions by updating its motor control parameters as appropriate.

As one might expect, the creation of new concepts in SINS (and in other systems such as Kuipers & Byun's (1991) NX) is an incremental process and involves, in addition to the abstraction of low-level inputs into higher-level representations, the modification of such representations in response to future experiences. In this sense, constructive conceptual change involves some degree of extrapolation as well. In addition, there is some amount of extrapolation at the time of adapting the suggested schema parameter values for a given situation. For a given sequence of perceived environmental situations, the agent selects the most similar sequence of associations in memory and extrapolates the schema parameter values suggested by the association to the perceived environment situation. In this way, the agent is using existing concepts to understand a new situation and to respond to it. However, since this extrapolation does not require the kinds of creative leaps as those needed in the ISAAC system, the latter provides a better case study of extrapolative conceptual change and is discussed next.

5 Extrapolative conceptual change

In developing the SINS system, we were interested in the problem of constructing new conceptual representations from continuous sensorimotor experience. Another type of conceptual change exists is that which occurs when existing conceptual representations are used to understand concepts from a new and unfamiliar domain. In order for the novel concepts to be understood, some change may need to take place in the existing conceptual framework of the reasoner; the more different the domain, the more radical the change.

In the ISAAC system, we have focused on the construction of new concepts (and associated theories) through creative theory-guided transfer of existing concepts to a new domain. Unlike the SINS process which is largely inductive, the process implemented in ISAAC is largely analytical and involves analogical and metaphorical reasoning. There are two central issues here: what are the processes by which existing theories are extrapolated, and what is the nature of the constraints on these processes? ISAAC explores these ideas in the domain of reading short stories from the science fiction literature. Science fiction is a particularly good choice for our research domain since stories can introduce concepts which are entirely novel to a reasoner. Consider the following short story, *Men Are Different* by Alan Bloch (1963).

> I'm an archaeologist, and Men are my business. Just the same, I wonder if we'll ever find out about Men—I mean *really* find out what made Man different from us

Robots—by digging around on the dead planets. You see, I lived with a Man once, and I know it isn't as simple as they told us back in school.

We have a few records, of course, and Robots like me are filling in some of the gaps, but I think now that we aren't really getting anywhere. We know, or at least the historians say we know, that Men came from a planet called Earth. We know, too, that they rode out bravely from star to star; and wherever they stopped, they left colonies—Men, Robots, and sometimes both—against their return. But they never came back.

Those were the shining days of the world. But are we so old now? Men had a bright flame—the old word is "divine," I think—that flung them far across the night skies, and we have lost the strands of the web they wove.

Our scientists tell us that Men were very much like us—and the skeleton of a Man is, to be sure, almost the same as the skeleton of a Robot, except that it's made of some calcium compound instead of titanium. Just the same, there are other differences.

It was on my last field trip, to one of the inner planets, that I met the Man. He must have been the last Man in this system, and he'd forgotten how to talk—he'd been alone so long. I planned to bring him back with me. Something happened to him, though.

One day, for no reason at all, he complained of the heat. I checked his temperature and decided that his thermostat circuits were shot. I had a kit of field spares with me, and he was obviously out of order, so I went to work. I pushed the needle into his neck to operate the cut-off switch, and he stopped moving, just like a Robot. But when I opened him up he wasn't the same inside. And when I put him back together I couldn't get him running again. Then he sort of weathered away—and by the time I was ready to come home, about a year later, there was nothing left of him but bones. Yes, Men are indeed different.

In order to understand this story, the reader must infer that the narrator is a robot, that robots are the dominant lifeform in the future, that humans have practically died out, that robots are capable of making logical errors such as the ones that the narrator made, and so on. The reader must construct an appropriate model of this world, and interpret the story with respect to this model even as the model evolves. The reader must also be willing to suspend disbelief to understand concepts which do not fit into a standard world view.

In ISAAC, new concepts are constructed through extrapolation and modification of existing theories and concepts. The extrapolation is constrained by the actual content of the story, by the system's existing theories and concepts, and by the cognitive constraints on the reading and understanding mechanisms that are responsible for processing the story. No reader, machine or human, could have the time, memory, and other resources to read every single word in a story in-depth and to consider all the ramifications of each word. The reader's environment (the story), knowledge (existing concepts), goals and tasks (e.g., Ram & Hunter, 1992), and cognitive resources available to the processing machinery (e.g., Just & Carpenter, 1992) interact to constrain the possible extrapolation to a more manageable level.

The story understanding processes in ISAAC are not unique to science fiction stories, of course. Understanding any fictional story requires similar kinds of processing. The same is true of non-fictional stories as well as unfamiliar real-world scenarios, although the types and degree of conceptual modifications required may be different.

6 Technical details: The ISAAC system

6.1 Task: Reading natural language texts

The ISAAC reading system consists of six "supertasks," each of which is made up of several tasks that interact with each other. The tasks are based on research in psycholinguistics (e.g., Holbrook, Eiselt & Mahesh, 1992; van Dijk & Kintsch, 1983), reading comprehension (e.g., Black & Seifert, 1981; Graesser, Golding, & Long, 1991), story understanding (e.g., Birnbaum, 1986; Ram, 1991; Rumelhart, 1977), episodic memory (e.g., Kolodner, 1984; Schank, 1982), analogy (e.g., Falkenhainer, 1987; Gentner, 1989), creativity (e.g., Gruber, 1989; Schank & Leake, 1990), and metacognition (e.g., Gavelek & Ram & Cox, 1994; Raphael, 1985; Schneider, 1985; Weinert, 1987; Wellman, 1985). The supertasks and their functions are summarized below; a high-level system architecture is shown in Figure 5.

aids the reader in understanding the story.

Scenario comprehension: The tasks making up the scenario comprehender are the *event parser*, which identifies various components such as agents, actions, states, objects, and locations; the *agent modeler*, which maintains descriptions of the agents, including their goals, knowledge, and beliefs; and the *action modeler*, which maintains descriptions of the acts with which the agents are involved. This supertask would be used in any experience in which the reasoner had to understand the actions of agents around it. As such, it is equally useful in day-to-day encounters with other reasoners as it is in the understanding of text.

Explanation and reasoning: This supertask performs high-level reasoning and learning. *Creative understanding* attempts to understand concepts which do not fit the reader's world view, including words which may be novel to the reader. *Interest management* controls the reader's level of interest in the story. An avid science fiction fan, for instance, would be more interested in the example story than a fan of Westerns. *Belief management* reasons about the beliefs of the agents involved in the scenario. Did the robot actually believe that the Man was capable of being turned off, and if so, why? The *explanation* task builds the inferences needed to connect the events of the story, enabling the reader to learn from the material. Finally, the *metareasoning* task reflects on the reader's own actions during the reading process; this information is also used for learning.

Memory management: This supertask handles general memory storage and retrieval, including spontaneous reminding. It is made up of *case building*, which constructs the various cases which result from a reading experience; *memory retrieval*, which returns information from memory which the system needs; and *memory storage*, which places new information and cases into memory. In addition to the conscious requests to store or retrieve elements in memory, the memory supertask should also handle the unconscious storing of material and spontaneous remindings. During the reading of *Men...*, the reader will constantly be retrieving and storing concepts related to the material in the text. Some of these may be spontaneous; for example, the story may trigger memories of other stories which concern the extinction of humankind, or other stories written by Alan Bloch. Other memory processes will be more deliberate; the reader may decide, for example, to try to remember the last science fiction story they read about robots.

Metacontrol: Metacontrol integrates the other supertasks. It includes the tasks of *focus control*, which manages the depth of reading based on interest and understanding; *time management*, which allows the reader to make decisions based on time resources; and *suspension of disbelief*, which enables a reader to accept, at least temporarily, a text which violates their world view. This last function is particularly important in the case of reading a story containing unfamiliar concepts. A rational reader of the *Men...* example knows that the story cannot be true. Mankind still exists, space travel doesn't, and today's robots are merely mechanical devices used on assembly lines. In order to read, understand, and enjoy the story, however, the reader must be willing to accept these unfamiliar ideas for the duration of the reading experience.

6.2 Definition: Creative understanding

As noted by teachers and education researchers for decades, in order to read at the most comprehensive level, a person needs to read creatively (e.g., Popov, 1993). This means that the reader can engage the text fully, use their own experiences to bridge comprehension failures, and come to an understanding of a story which contains concepts with which they were previously unfamiliar. In order to successfully read a story which contains novel ideas, a reader must have some method for understanding and interpreting those ideas. We call this process *creative understanding*. A central requirement for this process is the willingness of the reader to suspend his or her disbelief of the material being presented or the assumptions being made about the fictional world (a process noted by Coleridge as described in Corrigan, 1979). Consider the ambiguous title of a Larry Niven (1973) story, *Flight of the Horse*. This phrase could refer to a fleeing horse, a horse on an airplane flight, or to a flying horse. If a story understanding system relied on a belief in the validity of world knowledge, it would disambiguate the phrase to eliminate the latter meaning since it "knows" horses cannot fly. This may be incorrect if the story was about a flying horse (or a pegasus), which is perfectly reasonable in a science fiction or mythological story. As argued earlier, these considerations are not unique to science fiction stories; even factual stories (such as newspaper stories) in domains that are not completely understood may require the system to consider the possibility that its current understanding of the domain is incomplete or incorrect (e.g., Ram, 1993).

To understand concepts which do not fit into a standard world view, the ISAAC system attempts to modify existing concepts (Schank, 1986). This usually involves extending or adapting not just a single concept, but systems of concepts—that is, theories. This modification can occur in several ways. Definitional constraints may be relaxed to produce concepts with alternative constraints. For example, relaxing the definitional constraint that a horse's primary mode of locomotion is its legs may result in a "horse" with wings—a pegasus. Another option is to add new constraints or features to existing concepts, or to combine two concepts together. Suitcases, for example, do not normally have a mode of locomotion; adding one may result in an independently mobile suitcase, much like the one depicted in Terry Pratchett's (1983) story, *The Colour of Magic*. Creativity may also result from relaxed constraints on memory search processes, such as in the "imaginative memory" of Turner's (1991) MINSTREL system. The question, then, is exactly how do these modifications get performed, and what knowledge is used to perform them?

Before discussing issues of process and knowledge, however, we first need to formally define what we mean by creative understanding (Moorman & Ram, 1994b):

> **Creative understanding**: A directed, internal process by which a novel artifact is understood by the reasoner. In the case of reading, this understanding is inherently useful to the overall task of understanding the text.

Definitions of creativity abound in the literature; we present the above as a working definition sufficient to describe the process in which we are interested. We require a creative process to be internal, to ensure that the reasoner is not simply repeating a piece of knowledge just received from another source, and directed, to ensure that the reasoner is not simply a random generator of solutions, where one may eventually be "creative" through sheer chance. In addition, we require the successful understanding of a novel concept to be inherently useful to the reader, in that it allows reading to continue and to ultimately result in a coherent understanding of the story. The system cannot simply devise a random, bizarre "understanding" of a novel concept and thereby claim that true understanding (or creativity) has taken place.

The artifact—the novel concept being understood—may be a physical object in a story (such as an intelligent robot) or a story concept without a physical instantiation. Any artifact can be described by a set of attributes which define its characteristics. One of these, *function*, represents the best-known uses of the artifact. The remaining attributes are divided into *primary attributes*, which contribute to an explanation for why the artifact can perform its function; and *secondary attributes*, the rest of the attributes which are incidental to its primary function.

The term “novel” has numerous connotations, especially in creativity research (e.g., Boden, 1991; Stewart, 1950; Thurstone, 1952). For our purposes, there are four ways in which an artifact may be novel. These will determine what sorts of creative understanding is required to “make sense” of a given artifact in order to come to a coherent interpretation of a story. With respect to the function or goal (G) of an artifact (M), the four types of novelty are:

- Absolute Novelty (A-Novel): M is defined to be A-Novel iff M is unknown to the reasoner whose point of view is being considered. A longsword is A-Novel to a reasoner who has never seen one before.

- Instantiation Novelty (I-Novel): M is defined to be I-Novel iff M is unknown to the reasoner whose point view is being considered, and if M accomplished G in a way which is similar or identical to other artifacts which also accomplish G. This is generally the case if a secondary attribute is altered without changing M’s function. If a reasoner, for example, only knows about black longswords and sees a red one, the red longsword would be I-Novel with respect to the reasoner.

- Evolutionary Novelty (E-Novel): M is defined to be E-Novel iff M is A-Novel and M accomplishes G in a better way than other examples of artifacts which accomplish G. This is generally the result of altering one of the primary attributes of M. For the reasoner from above, a shortsword and a two-handed sword would be E-Novel if all they knew about were longswords.

- Revolutionary Novelty (R-Novel): M is defined to be R-Novel iff M is A-Novel and M accomplished G in a new way than other examples of artifacts which accomplish G. This form of novelty requires more extensive changes than to simply alter the primary attributes of M. Secondary attributes may be altered in ways which cause them to now participate in the function of M, attributes may be added to either set, attributes may be removed, and so on. The function of the artifact remains the same; the underlying explanation of how it

Input:	Initial state (I) Goal state desired (G) Set of constraints (C) (optional) Current solution (S) (optional) Critique why S is not good enough solution (K) (optional)
Output:	Solution path (S′) which achieves G given I and does not violate C

Figure 6: FUNCTION **Problem Solver**

accomplishes this function is what has altered. The light saber from the Star Wars series would be an example of an R-Novel artifact; it accomplishes its function in a revolutionary way compared to the known longsword.

6.3 Process: Algorithms for creative understanding

With a task definition in place, we can now develop an algorithmic model of the understanding process. Since understanding can be described as the complementary operation to problem solving (e.g., Wilensky, 1983), we start with a formal view of problem solving. Problem solving begins with the reasoner in an initial state. A reasoner knows of operations that it can perform which will move it through a search space. The search stops when a desired goal state is achieved. The output from the process is a solution path which takes the reasoner from the initial state to the goal state (Newell & Simon, 1972). Also important in the formalism are constraints: conditions which cannot be violated in the final solution (Sacerdoti, 1975; Sussman, 1973). Finally, in situations when the reasoner may already possess a solution, problem solving may discover a better solution if the reasoner possesses a critique of why the current solution is not a viable one (Hammond, 1989). The complete formulation is shown in Figure 6.

Based on Figure 6, we propose a formal specification of understanding (Figure 7). There are three behaviors which can result from this understanding formulation. Using the example of *Men...*: if a reader sees a robotic character “turn off” a man and then open him, they may understand the episode by reasoning that the robot had the goal of repairing the man (*abduction*, reasoning from

Input	Output	Function
Solution (S)	Goal (G)	Abduction
Goal (G)	Solution (S)	Prediction
Solution (S) and Goal (G)	Critique (K) of why S is a good solution	Explanation

Figure 7: FUNCTION **Understander**

a conclusion to an antecedent, in this case, from an action back to a goal); a reader who learns that Mankind has become extinct and that the remaining robots are curious as to the fate of their creators may understand this by reasoning about upcoming story actions (*prediction*, reasoning from an antecedent forward to a conclusion, in this case, forward in story time to hypothesize about upcoming events based on what has happened so far); finally, the reader may attempt to understand why the robot felt that field repairs was a good solution to the man's discomfort (*explanation*, reasoning with both an antecedent and conclusion and trying to determine how one may move from one to the other, in this case, trying to determine why a certain belief in the story was motivated by the events of the story).

Mundane understanding is primarily a recognition task. If a reasoner knows explanations relating goals and solutions, and the reasoner is then presented with both a goal and a solution to explain, the problem is simply recognizing the configuration and retrieving the proper items from memory. Similarly, a reasoner may know what goals and solutions are logically linked; when presented with one, it is fairly straightforward to retrieve the corresponding other. On the other hand, creative understanding is the process by which new concepts are understood by a reasoner. This is a more difficult task than mundane understanding in the sense that there are no in-memory items which directly answer the understanding questions. As a result, the reasoner must resort to creative techniques in order to attempt to understand the novel concepts.

Given the input and output requirements for creative understanding, the problem is to identify a set of mechanisms which accomplish the function as a whole. In order to achieve this, both the relevant processes and the required knowledge needs to be identified. We suggest that a central component of creativity is the ability to perform extrapolative conceptual change and, furthermore, that it is possible to identify a core set of cognitive mechanisms which carry out this function.

Since the reasoner is dealing with new concepts not in memory, there are three basic ways which it may attempt to understand them. First, it may be the case that a new concept is sufficiently like a known concept that a process of analogical reasoning can be employed. The reasoner's understanding of the known concept is used to understand the new concept. The second approach is to reason from general knowledge which the reasoner possesses. This is the application of abstract knowledge to the problem of understanding the new concept. Lastly, the reasoner may attempt to creatively reverse engineer the new concept to something they are familiar with.

In ISAAC, the three basic ways to achieve creative reasoning are combined into one iterative algorithm, the *creative understanding process*. The cycle of creative understanding is shown in Figure 8. (Note that, while the description in the figure considers the situation in which S is an input and G is the desired output, the same approach is used for the other examples of understanding discussed previously.) If the reasoner only attempts memory retrieval and straightforward analogical mapping, the process results in mundane understanding. Each cycle through the complete function increases the potential for creative understanding. Although there is no theoretical limit to the "amount" of creativity which may be generated, at some point the new concept may be so far removed from the original concept that further iterations will be useless. The point at which creativity degenerates into bizarreness is, of course, dependent on the context and the background knowledge of the reasoner; this issue will be discussed below. Regardless of the extent of the creative extrapolation, however, the reasoner must make a decision to suspend its disbelief in potentially bizarre concepts in order to explore the creative possibilities they represent.

The first step of the algorithm involves a memory retrieval attempt. If concepts are retrieved which can directly be used to understand the novel concept and incorporate it into memory, the cycle ends successfully. If nothing is returned which is immediately useful, processing continues. This may occur if nothing relevant is available in memory or if the relevant items are simply not returned (e.g., due to an indexing problem). In *Men...*, for example, the reasoner is confronted with the idea of a sentient robot. If the reasoner's memory contains only concepts describing an industrial robot, understanding will fail since the usual concept of an industrial robot is insufficient to explain the robot's actions.

The second step of the creative understanding algorithm deals with the case in which the retrieved concept is not an exact match. Since the memory system had some matching criteria that

Input:	Solution (S)
Output:	Goal state desired (G) Constraint set (K)
Process:	REPEAT 1. Perform *memory retrieval* 2. If (1) fails ⇒ attempt *analogical mapping* 3. If (2) fails ⇒ attempt *base-constructive analogy* 4. If (3) fails ⇒ *reformulate the problem* UNTIL (successful OR concept is too bizarre)

Figure 8: FUNCTION **Creative Understander**

suggested that the retrieved concept is a potentially good candidate with which to understand the new concept, the reasoner must now attempt to determine why this is the case. To do this, the reasoner tries to use analogical mapping, a process by which two concepts are seen to be related in some way (e.g., Falkenhainer, 1987; Gentner, 1989). Analogical mapping is the process of transferring known facts or inferences from the base domain, suitably modified, to the corresponding concepts in the target domain. Past research has shown that this framework can be the springboard on which creative ideas are based (Schank, 1986).

The next two steps of the creative understanding process deal with the situation when analogical mapping fails to result in an understanding of the new situation or concept. For step three, we propose a process that we call *base-constructive analogy* to creatively construct an understanding of a target even in a situation for which no base exists in memory.[3] It would be highly unlikely that a reasoner would possess just the right knowledge of a base domain, represented using just the right kind of mental models, for use in analogical mapping for all the novel situations it might encounter. In base-constructive analogy, the reasoner dynamically builds a base concept in an attempt to understand a concept, and then uses that base to understand the target. For ISAAC, this allows new concepts to be constructed dynamically and then used in analogical reasoning.

[3]We are not the first researchers to notice this form of analogy (see, e.g., Clement, 1989, and Nersessian, 1992, for descriptions of a similar idea), but we do not know of any previous computational models of this process.

The target in question is described within the context of a specific situation or domain (in the case of ISAAC, the story being read). If the reasoner possesses sufficient knowledge about some other domain (albeit not necessarily formulated as a ready-to-map mental model), and at least a shallow mount of knowledge about the target domain, it is possible to map the target problem from its original domain into the more familiar one. In doing this, the reasoner creates a base by relying on its knowledge of the new domain. This form of analogy is most apparent in human discourse when a concept is being explained to someone. A teacher, for example, may explain the concept by appealing to the background information possessed by a student.

Finally, if all of the above steps have failed to produce a satisfactory level of understanding, the reasoner must resort to *problem reformulation*. There are some cases in which the initial statement of a problem or task description is not the one which will lead to an optimal solution. By recasting the problem in a new way, a reasoner may gain a previously unseen insight into the resolution of the situation. Recent work on this topic has been done by Jones (1992). His thesis is that the reformulation should be from an actual problem to a more abstract description; in particular, a primary source of culturally shared abstract knowledge exists in the form of proverbs which reasoners can use in solving novel problems. We view problem reformulation as being the transformation of one instantiation of a problem into another one, rather than the transformation of a problem into a more abstract form, but there are similarities in the computational mechanisms that underlie both processes.

Problem reformulation is also the process by which creative reverse engineering occurs in the creative understanding algorithm. By successively reformulating the given problem description, it is possible to reverse engineer a chain of creative reasoning which would lead from a mundane object to the creative artifact in question. Consider a reader attempting to understand a unicorn who was totally unfamiliar with the concept. If the problem becomes *understand an animal which is like a unicorn but has no horn* (the removal of an attribute), the next cycle will retrieve the concept of *horse* from memory. The reasoner has reverse engineered *unicorn* to *horse* by removing the attribute of *horn*. To understand the unicorn in this way, then, the reasoner need only transform $horse + horn \Rightarrow unicorn$.

In ISAAC, the tasks of problem reformulation/creative reverse engineering and base-constructive analogy are carried out by a mechanism known as *function-driven morphological synthesis* (Moorman & Ram, 1994b). This mechanism builds on the work of two researchers from the 1950s whose goal was to increase the creative potential in individuals. *Attribute listing*, proposed by Crawford (cited in Finke, Ward, & Smith, 1992), begins with a reasoner considering an object which it thinks can be the springboard to a new, creative artifact. Each of the object's primary attributes are examined, with the reasoner thinking about ways in which each could be manipulated. Finally, the reasoner filters out the manipulations which lead nowhere; those that remain are considered to be creative. *Morphological synthesis*, proposed by Allen (cited in Finke, Ward, & Smith, 1992), begins in the same fashion as attribute listing. The reasoner considers both attributes and commonly accepted values for each. Then, new combinations of attributes and values are examined. The reasoner looks for unknown combinations which may prove to be creative. While this is similar to attribute listing, in morphological synthesis the reasoner does not concentrate on changing attributes one-at-a-time but instead explores interesting combinations. This leads to faster generation of potentially creative artifacts.

The obvious drawback to these approaches is that a reasoner will never develop an R-novel concepts, just relatively close variations of the original concept (i.e., only E-novel concepts). This is due to the fact that the reasoner considers only primary attributes for modification. The main advantage of morphological synthesis is the capability of attribute combinations; this decreases the time needed to produce potentially creative artifacts, but neither technique allows for the introduction of new attributes. Function-directed morphological synthesis, discussed in Moorman and Ram (1994a), starts with the concept which needs to be understood (either a goal (G), a solution (S), or a critique (K)). The reasoner applies a set of manipulator functions to the artifact, altering its attributes and producing new artifacts that still possess the original functionality but achieve that functionality in novel ways.

The modifications themselves can be performed in two ways: by examining other known concepts with the same functionality to see how they achieve that functionality, and by a search through the reasoner's knowledge of a given attribute and its possible values. An intelligent robot, for example, might achieve its intelligence in the same manner as humans do; knowledge of humans can allow the reasoner to postulate a robotic brain to be added to the concept of an industrial robot, resulting in an a new concept describing an intelligent robot. This modification reformulates the original problem into one of discovering how such a brain might be designed, implemented, and integrated into an industrial artifact.

6.4 Representation: Constraints on creativity

The above techniques for creative concept manipulation are very powerful; they can effectively alter an artifact's attributes to produce new artifacts, thereby producing conceptual change. The problem with such concept manipulation is that it is difficult to specify principled constraints on this process. For example, while wings might be added to a horse to provide new locomotion functionality (resulting in a pegasus), could a toaster be a good mode of horse locomotion? Up to a certain limit, constraint manipulation will result in concepts which could be called creative, after which the resulting concepts may be too bizarre to be useful. However, utility and interestingness are not inherent in particular concepts, but can only be evaluated with respect to the reasoner's knowledge, the organization of this knowledge, the reasoner's goals, the task at hand, the environment in which the reasoner is carrying out its tasks (in the case of ISAAC, the story), and general processing heuristics (Pinto, Shrager & Berthenthal, 1992; Ram, 1990).

There are three ways in which the alterations performed by ISAAC are constrained. Firstly, the use of known concepts and real-life artifacts in analogy and in problem reformulation provides a springboard that is at least based on reality. While aspects of these artifacts may be bizarre in the current context, the use of them ensures that those aspects *were* relevant at some point. Second, as mentioned above, the story itself acts as a powerful constraint. ISAAC is conservative in its conceptual manipulations; when an artifact is created which is sufficient to allow story processing to continue, manipulation stops. In other words, ISAAC's creative processes are "situated" in its task and task context. Finally, ISAAC's knowledge organization scheme provides a structure for the conceptual change process. ISAAC's knowledge base is organized into a semantic network, which is indexed through a multidimensional grid (see Figure 9). The rows of the grid represent "thematic roles" for adaptation; for concepts representing events, these include action, agent, state, and object. The columns of the grid represent "conceptual domains," such as physical, mental, social, emotional, and temporal. For example, a transfer is a generic action. Different types of transfers can be represented as physical (e.g., the PTRANS primitive of Schank, 1972), mental (e.g., MTRANS), and social (e.g., ATRANS). The grid also allows the system to represent emotional and

	Physical	Mental	Social	Emotional	Temporal
Agents	person	consciousness	boss	Ares	entropy
Actions	walking	thinking	selling	loving	getting closer to March
Objects	rock	idea	teacher-student relationship	hatred	second
States	young	lack of knowledge	public dishoner	being angry	early

Figure 9: Knowledge representation grid.

temporal transfers (see also Domeshek, 1992). Although it was developed independently, the grid bears a strong resemblance to the work of Chi (1993) and Carey (1992), both of whom attempt to discover the ontological relations inherent in human reasoning.

Each step of the creative understanding process has the potential to allow concepts to move around the grid. Each type of movement incurs a cost to the system, depending on the degree to which the concept has been altered. Movement within a single cell is the easiest type to perform, movement along a single row or a single column is more difficult, and adaptations requiring movement across both rows and columns are the most difficult. The system tries to perform the least amount of adaptation necessary, guided by the knowledge representation grid, such that the resulting concepts can explain and provide a structure for the input.

For example, many temporal metaphors can be represented as analogies between the physical and temporal columns of the grid (Lakoff & Johnson, 1980). In a sentence such as "Time has passed her by," for example, a temporal event is described in physical terms, and an abstract object (time) is described as the agent of the physical action. Similarly, in the second paragraph of *Men Are Different*, "we aren't really getting anywhere" is a metaphorical use of knowledge of physical actions to describe a mental action. Such a metaphor requires a larger creative leap than an adaptation within the physical column alone, such as in Schank's (1986) example in which an analogy is drawn between a jogger and a racehorse. Continuing with the earlier horse locomotion examples, a horse with wings involves an adaptation in which a known mode of locomotion (wings) is substituted for another one (legs), and is less bizarre than an independently mobile suitcase with wings in which an inanimate object is viewed as an animate agent with an invented (but plausible) mode of locomotion where none existed previously. As before, however, utility and interestingness are not absolute; a suitcase with wings (perhaps airplane wings rather than bird wings) might make sense in the right context.

In *Men Are Different*, robots, which in the real world are physical objects used as tools in manufacturing, are conceptualized as independent volitional agents. The reader must adopt this view to build an appropriate story model. Interestingly, the irony in this story derives from the fact that the robot in the story performs what one might view as the reverse inference, conceptualizing the man as a physical object to be repaired in a manner that one might use to repair a physical robotic device. It is important to note that the invented concepts are "real" within the context of the story, in contrast to the "bright flame of Men" which is metaphorical even within the fictional world. Similarly, a sentence such as "Winter is rapidly approaching" uses a spatial metaphor to describe a temporal event, whereas time travel may in fact be a "real" concept in a story. Understanding this concept involves adapting knowledge about actions, states, and causality from the physical column of the grid to the temporal. Such adaptation is the heart of the extrapolative conceptual change process. Once the new concepts and theories are built, they can be used to understand the story within the framework of these concepts; in turn, this may result in further modification of the concepts.

In addition to aiding in the reading process, the new concepts and theories can also provide a basis for future problem solving in the real world (e.g., Koestler, 1964). For example, reading about a fictitious device may prompt the reader to develop a similar device in the real world, or may help the reader understand a similar device when it is actually encountered at some later point. Motorola's MicroTAC hand-held personal cellular phone, for instance, has a strong resemblance to the hand-held personal communicators used in the *Star Trek* television series. Goodman, Waterman & Alterman's (1991) SPATR system uses a similar case-based reasoning process to understand novel devices (such as an Airphone) and natural language instructions for using these devices based on hierarchical spatial models of known devices (such as an ATM). Finally, reading about a creative problem solving episode may also allow the reader to replay the observed solution process on a real-world problem in a manner similar to Carbonell's (1986) derivational analogy.

Stories that are not creative can also be understood and used in such ways, of course. The mechanisms of conceptual change discussed here are an integral part of ordinary reasoning. Creative understanding in ISAAC is not implemented through a separate "creativity" process, but rather through normal processes of reasoning and learning (Gruber, 1989). Similarly, conceptual change in SINS also occurs through the normal processes of perception and control of action. Everyday reasoning is robust, adaptive, and creative; no special process need be postulated to model or explain these capabilities.

6.5 Implementation: The computer model

The current implementation of ISAAC can read several stories from the science fiction literature, including *Men Are Different*. ISAAC is built in Common Lisp and uses the KR frame package (Giuse, 1990) for basic knowledge representation. ISAAC currently uses the COMPERE parsing system (Mahesh, 1993) as a drop-in module for its sentence processing supertask. The input to the system is the story text, unaltered from its original published form. At present, ISAAC is capable of building a detailed story structure model, a scenario model of the events depicted, and a metareasoning model of its own activities during the reading process. On the knowledge side, there are currently several hundred concepts stored within the system. This knowledge is organized according to our knowledge organization grid. A partial view the comprehension built for *Men Are Different* is shown in Figure 10.

7 Discussion

While the individual case studies presented above are of interest in themselves, our concern here is with the nature of creative conceptual change. To this end, let us discuss the case studies with the goal of drawing implications about the functions of conceptual change, the mechanisms that carry out these functions, and the knowledge and knowledge representations that are required to support these functions. The function of conceptual change is to create new concepts through constructive and extrapolative processes. The mechanisms that implement these processes are multiple and complex; they are described using computational models and implemented in computer programs that are the testbeds for our theories.

Table 1: Differences in constructive and extrapolative conceptual change

Constructive conceptual change	Extrapolative conceptual change
Automatic (occurs always)	Strategic, goal-oriented (occurs when needed)
Occurs primarily across several task experiences	Occurs primarily within a task experience (but continues to occur across experiences)
Conceptual changes are persistent and ubiquitous	Conceptual changes are persistent but defined within a specific situation context
Concepts are modified incrementally (small changes)	Concepts are modified drastically (big changes)
Change is primarily inductive	Change is primarily analogical

It is an open question how these models which are, in some sense, at opposite ends of the spectrum of creative conceptual change might be unified into a single framework. Quine (1977) suggests that early concepts may be more perceptual, being defined inductively using an "innate similarity notion or spacing of qualities," and later concepts may become more "scientifically sophisticated," conceptual, and theory-embedded (see also Keil, 1989). Quine was interested in the issue of development of natural kinds, but perhaps a similar idea could be used to integrate perceptual and conceptual change in an "adult" reasoning system.

To facilitate integration, it is useful to look at commonalities between the models (see Table 2). Although the SINS model is closer to actual perceptual features in real world and the ISAAC model is closer to theories and mental models, both are based in real experience (whether personal or vicarious), and are constrained by the interaction between the system and the environment. Both are creative processes, and result not just in learning but in conceptual change as well. In SINS, raw sensorimotor information is encapsulated into predictive perception-action models, and in ISAAC, existing theories are modified to provide a belief structure for new and unfamiliar concepts. Both require inductive and analytical processes (although to different degrees), and both combine multiple methods of learning, concept formation, and conceptual change. Both are based on multiple types of knowledge. In both, existing knowledge provides constraints on reasoning and learning processes. Both types of creative conceptual change model a gradual evolution of concepts to better approximate the observed world and in both, evolving concepts are used in the performance task even as they are modified.

These points also highlight many of similarities between the models of constructive and extrapolative conceptual change presented here and other theories of conceptual change, including theories of developmental and scientific conceptual change discussed earlier in this article. The items in Table 2, therefore, also summarize the conclusions from this research; while in our case they were derived from computational modeling we believe that these observations are consistent with other kinds of studies (such as psychological or historical studies) in the conceptual change literature. We hypothesize that the incremental, everyday type of conceptual change discussed in this article is accomplished using similar knowledge and reasoning mechanisms as the more radical type of conceptual change in, for example, scientific discovery, but further research is necessary to develop computational accounts of the latter.

One framework for integration of these (and other) methods of conceptual change is through a multistrategy learning model, in which various learning methods are combined into a unified framework. Recent attention to such models is evident in machine learning (e.g., Carbonell, Knoblock & Minton, 1991; Michalski & Tecuci, 1993) and cognitive psychology (e.g., Anderson, 1983; Wisniewski & Medin, 1991). Multistrategy approaches provide the flexibility and power required in practical, real-world domains.

There are several methods of integrating multiple learning algorithms into a single system (see Michalski & Tecuci, 1993). One such framework is that used in the Meta-AQUA and Meta-TS systems (Ram & Cox, 1994; Ram, Narayanan, & Cox, in press). In this model, the reasoning system actively selects and combines learning methods based on an analysis of its learning goals which are represented explicitly in the system. Some learning goals may be low-level and always active, such as in SINS and NX. These systems can be described as performing “goal relevant” learning, in that learning is relevant to the overall goals of the system (Thagard, discussed in Leake & Ram, 1993) but the system only has an implicit goal to learn (Barsalou, discussed in Leake & Ram, 1993). Other learning goals may be selected based on a higher-level analysis of utility

Table 2: Similarities between constructive and extrapolative conceptual change. The methods presented here provide a computational account of these properties.

- Concepts are theories about the world: coherent belief systems containing relations of explanatory coherence
- Conceptual change relies on multiple types of knowledge
- Conceptual change is carried out by multiple types of reasoning processes
- Conceptual change utilizes multiple sources of constraints (theories, knowledge, knowledge organization, real-world experiences)
- Conceptual change occurs through actual experience
- Conceptual change is gradual
- Concepts are used even as they are learned and modified
- Conceptual change is situated in the real world
- Conceptual change exploits regularities in the environment
- There is no separate "creativity" module; creativity emerges when everyday reasoning processes are challenged by extremely novel problems
- Constructive conceptual change requires extrapolation, and extrapolative conceptual change requires construction

of knowledge and relevance to the system's tasks, such as in ISAAC, Meta-AQUA, IVY (Hunter, 1990), and PAGODA (desJardins, 1992). These systems are better described as "goal directed" since goals are explicitly represented and used to drive the selection and execution of reasoning and learning strategies (Leake & Ram, 1993; Ram & Cox, 1994; Ram & Hunter, 1992). Although we do not want to suggest that humans have perfect or conscious metacognitive knowledge of, and control over, their learning processes, such a model could be used to take an intentional stance (Dennett, 1987) towards a computational theory of multistrategy reasoning, both as a description of human reasoning processes and as a basis for the design of creative AI systems.

In conclusion, creative conceptual change is an everyday process involving multiple integrated mechanisms that are constrained by existing knowledge and by the task at hand. This process is situated in, and therefore also constrained by, the real world, and results in original, useful and qualitatively different representations of systems of beliefs. The process involves the ongoing construction and extrapolation of concepts and theories in the context of, in service of, and in response to a real-world performance task. The constructive and extrapolative processes are modeled computationally through specification of functions (tasks), mechanisms and knowledge; these models are then instantiated as computer programs and evaluated empirically. In this article, we have used robotic navigation in dynamic environments and comprehension of actual science fiction short stories as the task domains in which to present two case studies of creative conceptual change. These case studies highlight the issues involved in conceptual change and provide a basis for the development and evaluation of models that address these issues.

8 Acknowledgements

We thank Mike Cox, Eric Domeshek, and Janet Kolodner for their comments on earlier drafts of this article, Ron Arkin for his help with the robotics project, and Anthony Francis for his help with the story understanding project. The research was supported in part by the Advanced Research Projects Agency, monitored by ONR under contract N00014-91-J-4092, in part by a Fannie and John Hertz Foundation fellowship, and in part by the Georgia Institute of Technology.

9 References

Anderson, J.R. (1983). *The Architecture of Cognition*. Harvard University Press, Cambridge, MA.

Arkin, R.C. (1989). Motor Schema-Based Mobile Robot Navigation. *The International Journal of Robotics Research*, *8*(4):92–112.

Birnbaum, L. (1986). *Integrated Processing in Planning and Understanding*. Ph.D. thesis, Research Report #489, Yale University, Department of Computer Science, New Haven, CT.

Black, J.B. & Seifert, C.M. (1981). The Psychological Study of Story Understanding. Technical Report #18, Yale University, New Haven, CT.

Bloch, A. (1963). Men Are Different. In I. Asimov & G. Conklin, editors, *50 Short Science Fiction Tales*, Macmillan Publishing Company, New York.

Boden, M.A. (1988). *Computer Models of Mind*. Cambridge University Press, Cambridge.

Boden, M.A. (1991). *The Creative Mind: Myths and Mechanisms*. BasicBooks, New York.

Brooks, R. (1986). A Robust Layered Control System for a Mobile Robot. *IEEE Journal of Robotics and Automation*, RA-*2*(1):14–23.

Canfield, R.L. & Haith, M.M. (1991). Young Infants' Visual Expectations for Symmetric and Asymmetric Stimulus Sequences. *Developmental Psychology*, *27*(2):198–208.

Carbonell, J.G. (1986). Derivational Analogy: A Theory of Reconstructive Problem Solving and Expertise Acquisition. In R.S. Michalski, J.G. Carbonell, & T.M. Mitchell, editors, *Machine Learning II: An Artificial Intelligence Approach*, Morgan Kaufmann Publishers, San Mateo, CA.

Carbonell, J.G., Knoblock, C.A., & Minton, S. (1991). PRODIGY: An Integrated Architecture for Planning and Learning. In K. VanLehn, editor, *Architectures for Intelligence*, pages 241–278, Lawrence Erlbaum Associates, Hillsdale, NJ.

Carey, S. (1992). The Origin and Evolution of Everyday Concepts. In R.N. Giere, editor, *Cognitive Models of Science*, University of Minnesota Press, Minneapolis.

Chi, M.T.H. (1993). Barriers to Conceptual Change in Learning Science Concepts: A Theoretical Conjecture. *Proceedings of the Fifteenth Annual Cognitive Science Conference*, pages 312-317, Lawrence Erlbaum Associates, Hillsdale, NJ.

Clement, J. (1989). Learning via Model Construction and Criticism: Protocol Evidence on Sources of Creativity in Science. In J.A. Glover, R.R. Ronning, & C.R. Reynolds, editors, *Handbook of Creativity*, Plenum Press, New York.

Corrigan, R.W. (1979). *The World of the Theatre*. Scott, Foresman and Company, Glenview, IL.

Dennett, D. (1987). *The Intentional Stance*. Bradford Books/MIT Press, Boston, MA.

desJardins, M. (1992). Goal-Directed Learning: A Decision-Theoretic Model for Deciding What to Learn Next. In *Proceedings of the ML-92 Workshop on Machine Discovery*, pages 147–151, Ninth International Machine Learning Conference, University of Aberdeen, Scotland.

Domeshek, E.A. (1992). *Do the Right Thing: A Component Theory for Indexing Stories as Social Advice*. Ph.D. thesis, Yale University, Department of Computer Science, New Haven, CT.

Falkenhainer, B. (1987). Scientific Theory Formation through Analogical Inference. In *Proceedings of the Fourth International Workshop on Machine Learning*, Morgan Kaufmann Publishers, Los Altos, CA.

Fikes, R.E., Hart, P.E., & Nilsson, N.J. (1972). Learning and Executing Generalized Robot Plans. *Artificial Intelligence*, *3*:251–288.

Finke, R.A., Ward, T.B., & Smith, S.M. (1992). *Creative Cognition: Theory, Research, and Applications*. MIT Press, Cambridge.

Gavelek, J.R. & Raphael, R.E. (1985). Metacognition, Instruction and the Role of Questioning Activities. In D.L. Forrest-Pressley, G.E. MacKinnon, & T.G. Waller, editors, *Metacognition, Cognition, and Human Performance, Volume 2 (Instructional Practices)*, pages 103–136, Academic Press, New York.

Gentner, D. (1989). Mechanisms of Analogical Learning. In S. Vosniadou & A. Ortony, editors, *Similarity and Analogical Reasoning*, Cambridge University Press, London.

Georgeff, M., (1987). *Planning*. Annual Review of Computer Science, 2, 359–400.

Giere, R. N. (1988). Explaining the Revolution in Geology. Chapter 8 in *Explaining Science: A Cognitive Approach*, The University of Chicago Press, Chicago.

Giuse, D. (1990). Efficient Knowledge Representation Systems. *The Knowledge Engineering Review*, *5*(1):35-50.

Goodman, M., Waterman, S., & Alterman, R. (1991). Interactive Reasoning about Spatial Concepts. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, pages 734–738, Lawrence Erlbaum Associates, Hillsdale, NJ.

Graesser, A., Golding, J.M., & Long, D.L. (1991). Narrative Representation and Comprehension. In R. Barr, M.L. Kamil, P. Mosenthal & P.D. Pearson, editors, *Handbook of Reading Research*, volume 2, chapter 8, Longman Publishing Group, White Plains, NY.

Grefenstette, J.J. & Ramsey, C.L. (1992). An Approach to Anytime Learning. In *Proceedings of the Ninth International Conference on Machine Learning*, pages 189–195, Aberdeen, Scotland.

Gruber, H.E. (1989). The Evolving Systems Approach to Creative Work. In D.B. Wallace & H.E. Gruber, editors, *Creative People at Work*, pages 3–24, Oxford University Press, New York.

Hammond, K.J. (1989). *Case-Based Planning: Viewing Planning as a Memory Task.* Academic Press, Boston, MA.

Holbrook, J.K., Eiselt, K.P., & Mahesh, K. (1992). A Unified Process Model of Syntactic and Semantic Error Recovery in Sentence Understanding. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, pages 195–200, Lawrence Erlbaum Publishers, Hillsdale, NJ.

Hunter, L.E. (1990). Planning to Learn. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, pages 261–276, Boston, MA.

Johnson-Laird, P.N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Harvard University Press, Cambridge, MA.

Jones, E.K. (1992). *The Flexible Use of Abstract Knowledge in Planning.* Northwestern University, The Institute for the Learning Sciences Tech Report # 28.

Just, M.A. & Carpenter, P.A. (1992). A Capacity Theory of Comprehension: Individual Differences in Working Memory. *Psychological Review*, *99*(1):122–149.

Kaelbling, L. (1986). *An Architecture for Intelligent Reactive Systems*. Technical Note #400, SRI International.

Keil, F.C. (1989). *Concepts, Kinds, and Cognitive Development.* MIT Press, Cambridge, MA.

Koestler, A. (1964). *The Act of Creation*. MacMillan Publishers, New York.

Kolodner, J.L. (1984). *Retrieval and Organizational Strategies in Conceptual Memory: A Computer Model*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Kolodner, J.L. (1993). *Case-Based Reasoning*. Morgan-Kaufman Publishers, Inc., San Mateo, CA.

Kuhn, T.S. (1962). The Structure of Scientific Revolutions, University of Chicago Press, Chicago, IL.

Kuipers, B.J. & Byun, Y.-T. (1991). A Robot Exploration and Mapping Strategy Based on a Semantic Hierarchy of Spatial Representations. *Robotics and Autonomous Systems*, *8*(1-2):47–63.

Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press, Chicago, IL.

Leake, D. & Ram, A. (1993). *Goal-Driven Learning: Fundamental Issues and Symposium Report*, Technical Report #85, Indiana University, Cognitive Science Program, Bloomington, IN.

Lynch, K. (1960). *The Image of the City*. MIT Press, Cambridge, MA.

Maes, P. (1990). Situated Agents Can Have Goals. *Robotics and Autonomous Systems*, 6, 49–70.

Mahesh, K. (1993). *A Theory of Interaction and Independence in Sentence Understanding*. Technical Report GIT-CC-93/34, College of Computing, Georgia Institute of Technology, Atlanta, Georgia.

Michalski, R.S. & Tecuci, G. (1994), editors. *Machine Learning: A Multistrategy Approach, Volume IV*, Morgan Kaufman Publishers, San Mateo, CA.

Moorman, K. & Ram, A. (1994a). *A Functional Theory of Creative Reading*. Technical Report GIT-CC-94/01, College of Computing, Georgia Institute of Technology, Atlanta, Georgia.

Moorman, K. & Ram, A. (1994b). *A Model of Creative Understanding*. In *Proceedings of the 12th National AAAI Conference*, Seattle, WA.

Murphy, G.L. & Medin, D.L. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review*, *92*:289–316.

Neisser, U. (1987), editor. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization.* Cambridge University Press, 1987.

Nersessian, N.J. (1992). How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science. In R.N. Giere, editor, *Minnesota Studies in the Philosophy of Science, Volume XV (Cognitive Models of Science)*, University of Minnesota Press, Minneapolis, MN.

Newell, A. & Simon, H.A. (1972). *Human Problem Solving*, Prentice-Hall, Englewood Cliffs, NJ.

Niven, L. (1973). *The Flight of the Horse*. Ballantine Books, New York.

Payton, D. (1986). An Architecture for Reflexive Autonomous Vehicle Control. In *Proceedings of the 1986 IEEE Conference on Robotics and Automation*, pages 1838–1845, IEEE.

Piaget, J. & Inhelder, B. (1967). *The Child's Conception of Space*. Norton Publishers, New York.

Pinto, J., Shrager, J. & Berthenthal, B.I. (1992). Developmental Changes in Infants' Perceptual Processing of Biomechanical Motions. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, pages 60–65, Lawrence Erlbaum Associates, Hillsdale, NJ.

Popov, A. (1993). Creativity and Reading Comprehension. *Journal of Creative Behavior*, *26*(3): 206-212.

Pratchett, T. (1983). *The Colour of Magic*. Penguin Books, New York.

Quine, W.V.O. (1977). Natural Kinds. In S.P. Schwarts, editor, *Naming, Necessity, and Natural Kinds*, Cornell University Press, Ithaca, NY.

Ram, A. (1990). Knowledge Goals: A Theory of Interestingness. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, pages 206–214, Lawrence Erlbaum Associates, Hillsdale, NJ.

Ram, A. (1991). A Theory of Questions and Question Asking. *The Journal of the Learning Sciences*, *1*(3&4):273–318.

Ram, A. (1993). Indexing, Elaboration and Refinement: Incremental Learning of Explanatory Cases. *Machine Learning*, *10*:201-248.

Ram, A., Arkin, R.C., Moorman, K., & Clark, R.J. (1992) *Case-Based Reactive Navigation: A Case-Based Method for On-Line Selection and Adaptation or Reactive Control Parameters in Autonomous Robotic Systems*. Technical Report GIT-CC-92/57, College of Computing, Georgia Institute of Technology, Atlanta, Georgia.

Ram, A. & Cox, M.T. (1994). Introspective Reasoning using Meta-Explanations for Multistrategy Learning. In R.S. Michalski & G. Tecuci, editors, *Machine Learning: A Multistrategy Approach, Volume IV*, Morgan Kaufman Publishers, San Mateo, CA.

Ram, A. & Hunter, L. (1992). The Use of Explicit Goals for Knowledge to Guide Inference and Learning. *Applied Intelligence, 2*(1):47–73.

Ram, A., Narayanan, S., & Cox, M.T. (in press). Learning to Troubleshoot: Multistrategy Learning of Diagnostic Knowledge for a Real-World Problem-Solving Task. *Cognitive Science*.

Ram, A. & Santamaría, J.C. (1993). Multistrategy Learning in Reactive Control Systems for Autonomous Robotic Navigation. *Informatica*, *17*(4):347-369.

Roschelle, J. (1992). Learning by Collaborating: Convergent Conceptual Change. *The Journal of the Learning Sciences*, *2*(3):235–276.

Rumelhart, D.E. (1977). Understanding and Summarizing Brief Stories. In D.L. Berge and J. Samuels, editors, *Basic Processes in Reading and Comprehension*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Sacerdoti, E.D. (1975). *A Structure for Plans and Behavior*. Technical Note #109, SRI International. Summarized in P.R. Cohen & E.A. Feigenbaum, *Handbook of AI, Volume III*, pages 541–550.

Schank, R.C. (1972). Conceptual Dependency: A Theory of Natural Language Understanding. *Cognitive Psychology*, *3*(4):552–631.

Schank, R.C. (1982). *Dynamic Memory: A Theory of Learning in Computers and People*. Cambridge University Press, New York.

Schank, R.C. (1986). *Explanation Patterns: Understanding Mechanically and Creatively*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Schank, R.C. & Leake, D.B. (1990). Creativity and Learning in a Case-Based Explainer. In J.G. Carbonell, editor, *Machine Learning: Paradigms and Methods*, MIT Press, Cambridge, MA.

Schmidt, R.A. (1968). Anticipation and Timing in Human Motor Performance. *Psychological Bulletin*, *70*:631-646.

Schneider, W. (1985). Developmental Trends in the Metamemory-Memory Behavior Relationship: An Integrative Review. In D.L. Forrest-Pressley, G.E. MacKinnon, & T.G. Waller, editors, *Metacognition, Cognition, and Human Performance, Volume 1*, Academic Press, New York.

Shanmugam, K.S., (1979). *Digital and Analog Communication Systems*. John Wiley and Sons, New York.

Siegel, A.W. & White, S.H. (1975). The Development of Spatial Representations of Large-Scale Environments. In H.W. Reese, editor, *Advances in Child Development and Behavior*, Academic Press, New York.

Simon, T. & Halford, G.S. (1995). Computational Models and Cognitive Change. In T. Simon & G.S. Halford, editors, *Developing Cognitive Competence: New Approaches to Process Modeling*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Smith, N.B. (1986). *American Reading Instruction*. International Reading Association, Newark, Delaware.

Stewart, G.W. (1950). Can Productive Thinking Be Taught. *Journal of Higher Education*, *21*:411-414.

Sutton, R.S. (1992), editor. *Machine Learning*, 8(3/4), special issue on Reinforcement Learning.

Sussman, G.J. (1973). *A Computational Model of Skill Acquisition*, MIT Tech Report # AI TR-297.

Thagard, P. (1989). Explanatory Coherence. *Behavioral and Brain Sciences*, 12(3):435–502.

Thurstone, L.L. (1952). Creative Talent. *Applications of Psychology*, Harper & Row, New York.

Turner, S.R. (1991). A Case-Based Model of Creativity. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, pages 933-937, Lawrence Erlbaum Associates, Hillsdale, NJ.

van Dijk, T.A. & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Academic Press, New York.

Vosniadou, S. & Brewer, W.F. (1992). Mental Models of the Earth: A Study of Conceptual Change in Childhood. *Cognitive Psychology*, *24*:535–585.

Weinert, F.E. (1987). Introduction and Overview: Metacognition and Motivation as Determinants of Effective Learning and Understanding. In F.E. Weinert & R.H. Kluwe, editors, *Metacognition, Motivation, and Understanding*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Wellman, H.M. (1985). The Origins of Metacognition. In D.L. Forrest-Pressley, G.E. MacKinnon, & T.G. Waller, editors, *Metacognition, Cognition, and Human Performance, Volume 1,* Academic Press, New York.

Wilensky, R. (1983). *Planning and Understanding*. Addison-Wesley, Reading, MA.

Wisniewski, E.J. & Medin, D.L. (1991). Harpoons and Long Sticks: The Interaction of Theory and Similarity in Rule Induction. In D. Fisher & M.J. Pazzani, editors, *Concept Formation: Knowledge and Experience in Unsupervised Learning*, Morgan Kaufman Publishers, San Mateo, CA.