Toward a Soar Theory of Taking Instructions for Immediate Reasoning Tasks

Richard L. Lewis, Allen Newell, and Thad A. Polk
School of Computer Science
Carnegie Mellon University

Abstract

Soar is a theory of the human cognitive architecture. We present here the Soar theory of taking instructions for *immediate reasoning* tasks, which involve extracting implicit information from simple situations in a few tens of seconds. This theory is realized in a computer system that comprehends simple English instructions and organizes itself to perform a required task. Comprehending instructions produces a *model of future behavior* that is interpretively executed to yield task behavior. Soar thereby acquires task-specific problem spaces that, together with basic reasoning capabilities, model human performance in multiple immediate reasoning tasks. By providing an account of taking instructions, we reduce the *degrees of freedom* available to our theory of immediate reasoning, and also give more support for Soar as a unified theory of cognition.

Soar is a theory of human cognition (Newell, 1989), embodied in a computer system. Soar specifies the cognitive architecture, which is the relatively fixed set of mechanisms that permit goals and knowledge of the task environment to be encoded in memory and brought to bear to produce behavior. Soar is proposed as a unified theory of cognition, and it has been applied to human behavior in a broad spectrum of domains. This paper reports progress in getting Soar to take instructions and organize itself to perform a required task.

There are three important reasons for wanting a theory of instructions. First, taking instructions is a domain of cognitive activity, with interesting phenomena and practical importance. Any unified theory of cognition must ultimately provide such a theory. Second, a major issue for psychology has always been the radical underdetermination of theory by data. Though an issue for all sciences, it is particularly irksome for psychology (and the social sciences) because humans bring massive knowledge to a task and dynamically organize themselves accordingly. Taking instructions to perform tasks is an important instance of such self-organization (e.g., as it takes place in psychological experiments). By specifying how task specific organization arises from instructions, a theory of instruction comprehension would go some way toward removing what can be called *theory degrees of freedom*. Instructions are only one source of knowledge determining behavior, but understanding them could pave the way for dealing with other sources. Third, including both instruction taking and task performance in the same theoretical account provides mutual constraint. This constraint is an instance of the gains to be made from a unified theory of cognition.

In this paper we take some steps toward a theory of instruction taking.[1] The system we present here, NL-Soar, comprehends instructions given in elementary English for *immediate reasoning* tasks, such as the *relational reasoning* task (Johnson-Laird, 1988) shown in Figure 1. This comprehension is part of a system, IR-Soar, that models the way humans perform immediate reasoning (reported in a companion paper (Polk, Newell & Lewis, 1989)). Here we focus on the internal representation of instructions, how

[1]Building on earlier work in (Newell, 1989, Chap. 7; Yost & Newell, 1988).

Instructions

Read four premises.
Then read a statement.
If the statement is true say "true".
Then stop.

Task input

Premises	A plate is left of a knife. A fork is left of the plate. A jug is above the knife. The fork is below a cup.
Statement	The cup is left of the jug.

Figure 1: **Relational reasoning task.**

Soar organizes itself to do the task, and the associated psychological claims. The process of comprehending the language of instructions to create these representations is also part of the total theory, and involves both linguistic and psycholinguistic issues. Although we do not deal with these issues here, NL-Soar does embody a theory of language comprehension (Newell, 1989, Chap. 8). We also do not present direct behavioral evidence for instruction taking. For the moment, the psychological relevance of the instruction taking is that it leads to an organization of IR-Soar that explains how people do immediate reasoning tasks.

There has been relatively little work on the psychology of instruction taking. The most notable was the UNDERSTAND program (Simon & Hayes, 1979), which took instructions for the Tower of Hanoi and constructed a problem space in which to do the task. Our account is consonant with the broad thrust of that work, the main advances being in the plausibility of the processes and representations used, and in the embedding of this in a unified theory. Our account is also consonant with the implications from ACT* (Anderson, 1983): that conversion from declarative to procedural form occurs by an interpretive process that leads to creating chunks of conditional behavior (productions).

We first present the psychological claims of the Soar theory of taking instructions, and in passing review the basic assumptions of the Soar architecture. We then illustrate the theory by tracing the behavior of the system in detail on the relational reasoning task in Figure 1. Finally, we briefly describe how the theory has been applied to two other immediate reasoning tasks.

THE SOAR THEORY OF TAKING INSTRUCTIONS

Soar as a cognitive architecture has been described in several places (Laird, Newell & Rosenbloom, 1987) and we will take its major outlines to be familiar. All tasks are formulated in *problem spaces*; all long term knowledge is held in a *recognition memory* (realized as productions); processing proceeds by a sequence of *decision cycles* that accumulate knowledge about what spaces, states and operators to select; *subgoals* are generated in an attempt to resolve *impasses* that occur when the decision-making knowledge is insufficient or conflicting; and the experience gained in resolving impasses is learned in the form of *chunks* (new productions in recognition memory).

One additional assumption is that states in problem spaces are *annotated models*.[2] Models consist of

[2]This is not yet an architectural assumption for Soar, which only assumes a representation consisting of attributes and values.

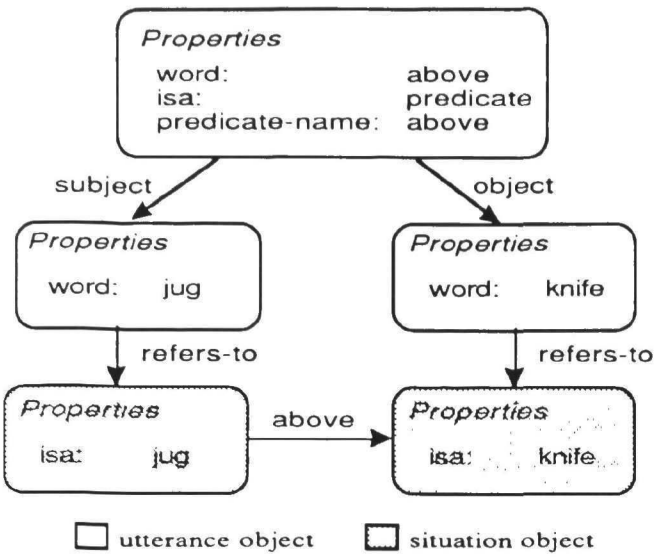

Figure 2: **Utterance and situation models for "A jug is above the knife."**

objects, properties, and relations (model elements) and satisfy the semantic assumption that each element in the representation corresponds to an element in the referent. This assumption may be violated in principled ways by attaching annotations to model elements. An annotation specifies a non-standard interpretation for the single element to which it is attached (e.g., an element annotated **many** corresponds to multiple elements in the referent). There are computational advantages to processing annotated models and there is also evidence that humans use them (Johnson-Laird, 1983; Polk & Newell, 1988). We do not review these considerations, but simply assume annotated models. Beyond these assumptions, the Soar theory of taking instructions embodies the following psychological claims:

1. **Situation model.** The objective of comprehending an utterance is to represent the situation that the utterance is about. To do so, comprehension builds a *model of the situation.*
2. **Utterance model.** As a processing side effect, comprehension produces a *model of the utterance*—a model that reflects the logical form of the utterance, and that can be interpreted as the abstract proposition or description asserted by the utterance.
3. **Behavior model.** If instructions are comprehended, the situation model that comprehension produces is a *model of the subject's future behavior.*
4. **Performance by interpretation.** Task performance proceeds initially by *interpretively executing* the behavior model. During this interpretation process, chunks are learned that directly perform the task.

The theory of comprehension

Comprehension (construction of the situation model) occurs in the **comprehend** problem space by applying a *comprehension operator* to each incoming word (Newell, 1989). These operators iteratively

---

Nevertheless, the current work in Soar on human cognition assumes models (Newell, 1989, Chap. 7; Polk & Newell, 1988).

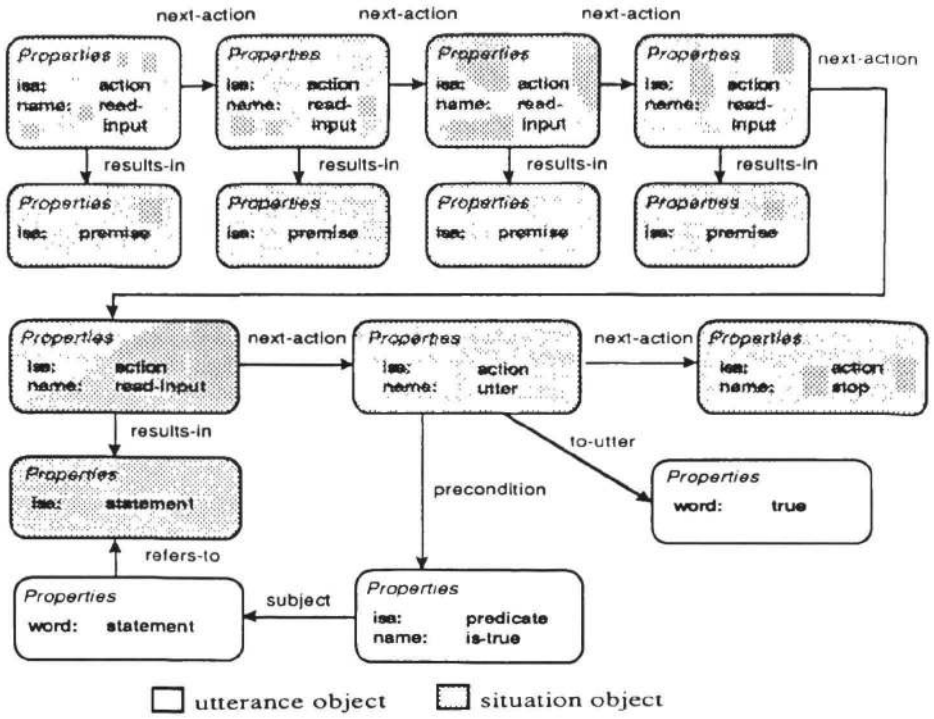

Figure 3: **Behavior model for the relational reasoning task of Figure 1.**

augment and refine the situation model as the utterance is comprehended. Figure 2, bottom, shows a simple situation model produced by **comprehend** for the third premise of Figure 1.

Since all the knowledge that a word contributes to an utterance may not be available at the moment the word is read, the comprehension process must have some means of holding partial comprehension knowledge. In the **comprehend** space, this knowledge is held by *expectation* data structures. Expectations can be syntactic, semantic, or pragmatic. In particular, part of what is delivered by a comprehension operator for a word is knowledge about what is expected to come in the rest of the utterance. Thus, there must be some way of modeling the utterance itself. The utterance model that serves this processing requirement is a structured linguistic form. As comprehension proceeds, it evolves into a model that reflects the underlying logical form of the utterance. In contrast to the situation model, the utterance model is closer to a predicate calculus-like language.

As an example, Figure 2, top, gives the final utterance model for "A jug is above the knife." [3] It is useful to compare the two models in this figure. The objects in the situation model correspond to the jug and knife in the situation, and the relation "above" corresponds to the spatial relation in the situation. In contrast, the objects in the utterance model correspond to linguistic objects that can be interpreted as predicates and arguments. Thus, "above" in the utterance model corresponds to the predicate "above", and the relations correspond to relations between the predicate and its arguments. Since objects in the utterance model originate as words, the language of predicates is as rich as the

[3]The utterance model in this figure is actually somewhat simplified for expository purposes.

natural language expressing the utterance.

The utterance model is not a deliberate product of comprehension; it is the means by which **comprehend** deals with language. However, since some knowledge cannot be encoded in the situation model (e.g., universal quantification), the total knowledge provided by an utterance may reside jointly in both models.

The theory of task performance

Comprehending instructions produces a model that represents future actions to be taken. An element in a behavior model corresponds to an action or an object related to an action (such as the expected input or output). For example, the model in Figure 3, produced by NL-Soar for the task of Figure 1, specifies that the task begins with four acts of reading input, and that each act should yield a premise.

After reading the instructions, the system attempts to perform the task. It is initially unable to proceed, because it lacks operational knowledge of the task in recognition memory (the knowledge is in the static data structures of the behavior model). This leads to interpreting the behavior model. Earlier work with Soar has shown how such processes can yield chunks that directly implement the task and bypass interpretation (Yost & Newell, 1988; Newell, 1989). Currently, this capability is embodied in BI-Soar (behavior-model interpretation), a set of problem spaces that are independent of NL-Soar and IR-Soar. For immediate reasoning tasks, the interpretation of the behavior model gives rise to problem spaces whose operators are implemented in the three basic IR-Soar spaces (**comprehend**, **test-proposition**, and **build-proposition**). Polk, Newell & Lewis (1989) show that the task spaces so acquired can indeed be used to model human performance.

AN EXAMPLE: RELATIONAL REASONING

We illustrate the theory by tracing through the task of Figure 1. Figure 4 shows the behavior of the system as it comprehends the instructions and attempts to perform the task. The system begins in the **read** problem space (see (1) in Figure 4), and applies a series of comprehend-input operators to read the instructions. Each operator application comprehends one statement, and builds up the behavior model accordingly. (2) The comprehend-input operator is implemented in the **comprehend** space, where a series of comprehension operators fire for each incoming word. These operators actually fire multiple times as expectations build up and are satisfied.

(3) Processing continues until comprehension of the word "begin," which the system takes to mean it should start the task. (4) It deliberately sets the goal of doing the task by selecting the do-new-task operator. (5) Since the knowledge required to perform the task is not directly available in recognition memory, Soar impasses and creates a new problem space (**relation**) to implement the operator.

(6) Once in the **relation** space, Soar impasses again because it has no operators to propose for this new space. Resolving the impasse requires consulting the behavior model for what to do next. (7) This is the function of the **fetch-operator** problem space (a space of BI-Soar), which contains the knowledge required to locate the next action in the behavior model and interpret it as an operator in the task space. (8) In this case, the impasse is resolved by selecting read-input, an instantiation of the comprehend-input operator that will yield a premise (Figure 3). (9) The premise object from the behavior model is set up

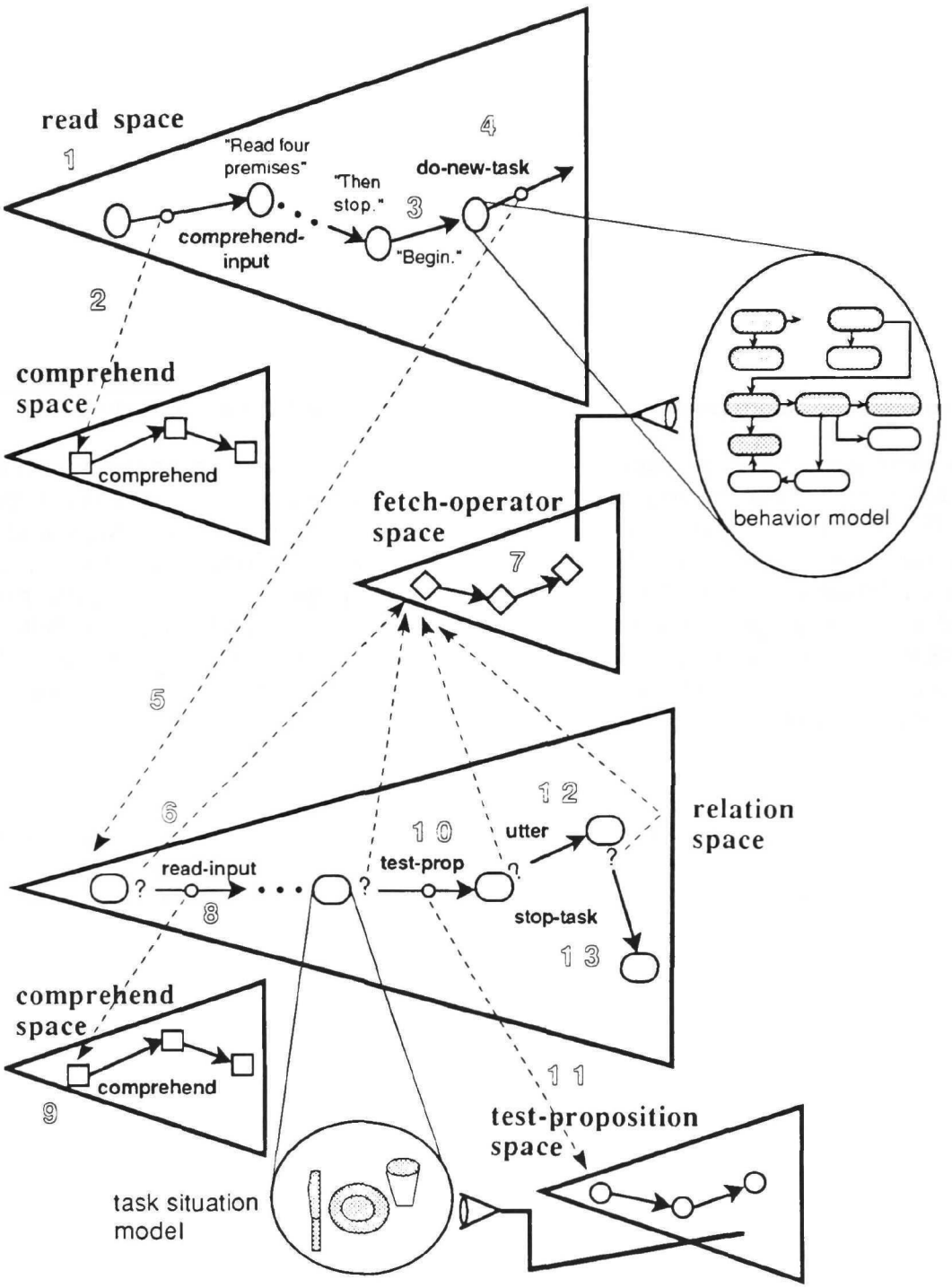

Figure 4: **Acquiring and performing the relational reasoning task.**

Categorical Syllogisms

Instructions	*Syllogism space operators*
1. Read two premises that share a term.	1. Read-input [comprehend]
2. Then produce a statement that follows from the premises.	2. Make-conclusion [build-proposition]
3. The statement relates the unique terms of the premises.	3. Make-conclusion (goal-test)
4. Then stop.	4. Stop-task

Elementary Sentence Verification

Instructions	*ESV space operators*
1. Examine the picture.	1. Read-input [comprehend]
2. Then read a statement.	2. Read-input [comprehend]
3. If the statement is true press the t-button.	3. Test-prop [test-proposition]
4. If the statement is false press the f-button.	4. Test-prop [test-proposition]
5. Then stop.	5. Stop-task

Figure 5: **Other immediate reasoning tasks.**

as an expectation in the **comprehend** space, and thus provides the goal test for read-input.

This *impasse-fetch-apply* cycle continues until Soar arrives at the utter action in the behavior model. This action has a precondition (namely, that the statement just read is true), interpreted as a proposition (Figure 3). This proposition originated as part of the utterance model for one of the comprehended instructions. (10) Determining if this action should be taken requires verifying the proposition, so the impasse in the **relation** space is resolved by selecting the test-prop operator. (11) This operator is implemented in the **test-proposition** space, one of the three basic IR-Soar problem spaces (Polk, Newell & Lewis, 1989). (12) Once the proposition has been verified by consulting the situation model, the **fetch-operator** space selects and instantiates the utter operator in the next fetch cycle. (13) Finally, the stop-task operator is selected and terminates the task.

OTHER TASKS

NL-BI-Soar has acquired two other immediate reasoning tasks. Figure 5 shows the instructions and problem-spaces for the elementary sentence verification task (Clark & Chase, 1972), and the categorical syllogisms task. As in the relational reasoning task, task-specific behavior arises by interpreting the behavior model, and applying operators in the new task space that are implemented in IR-Soar's **comprehend**, **test-proposition**, or **build-proposition** problem spaces[4].

The elementary sentence verification task differs from relational reasoning in the simplicity of the initial situation, and the form used to present the situation (a picture). The latter difference shows up in the behavior model as a comprehend-input action that expects a picture rather than a linguistic utterance. The difference in simplicity is a function of the task input, not the task instructions.

The syllogisms task (Polk, Newell & Lewis, 1989) is interesting because the subject must utter a conclusion that conforms to a particular specification given in the instructions— namely, that the conclusion relate the *unique terms of the premises*. Soar realizes this as an application of the build-proposition operator instantiated to relate the correct terms. Knowledge in the comprehension operators for the words "unique" and "relate" leads to construction of the appropriate behavior model

[4]NL-Soar will deal with the conditional reasoning and Wason tasks, which are the additional examples in (Polk, Newell & Lewis, 1989); as of submission of this paper, the runs are not completed, but no difficulties are expected.

that captures this constraint.

CONCLUSION

We have presented a Soar theory of taking instructions for immediate reasoning tasks. This theory is implemented in two collections of Soar problem spaces, NL-Soar and BI-Soar. NL-Soar uses the **comprehend** problem space to read simple English statements and produce an annotated model of the situation being described. As a side effect of comprehending these statements, **comprehend** produces a model that reflects the logical form of the utterance. When reading task instructions, **comprehend** creates a model of the behavior described by the instructions. By repeatedly consulting this behavior model, BI-Soar can acquire the problem spaces necessary to perform the task.

Besides being interesting in its own right, this theory opens up some interesting possibilities. For one, it begins to significantly alleviate the problem of the underdetermination of theories by data. In a companion paper (Polk, Newell & Lewis, 1989), we have presented a theory of immediate reasoning that depends on the task-specific problem spaces that arise from the instructions given to NL-Soar. The degrees of freedom available to that theory are significantly reduced as a result. Further, the two subtheories of taking instructions and immediate reasoning mutually constrain each other, making both significantly stronger. For instance, it is not an independent assumption of the immediate reasoning theory that both the utterance model and the situation model are available as sources of knowledge to do the task. Similarly, the problem spaces acquired through task instructions must be used to model behavior in immediate reasoning tasks, significantly constraining the theory presented here. Finally, this theory represents another step toward making Soar a unified theory of cognition.

Acknowledgements

Many thanks to Erik Altmann, Norma Pribadi, Kathryn Swedlow, and Shirley Tessler for useful comments and invaluable assistance in preparing the draft and the figures, as well as to Chris Tuck for helping bridge the Pittsburgh-Ann Arbor gap. This work was supported by the Information Sciences Division, Office of Naval Research, under contract N00014-86-K-0678, and by the National Science Foundation and Kodak fellowship programs in which Richard Lewis and Thad Polk, respectively, participate. The views expressed in this paper are those of the authors and do not necessarily reflect those of the supporting agencies. Reproduction in whole or in part is permitted for any purpose of the United States government. Approved for public release; distribution unlimited.

References

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

Clark, H. and Chase, W. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3,* 472–517.

Johnson-Laird, P. (1988). Reasoning by rule or model? In *Proceedings of the Annual Conference of the Cognitive Science Society,* pages 765–771.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness.* Harvard University Press, Cambridge, Massachusetts.

Laird, J. E., Newell, A., and Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33(1),* 1–64.

Newell, A. (1989). *Unified Theories of Cognition.* Harvard University Press, Cambridge, Massachusetts. In press.

Polk, T., Newell, A., and Lewis, R. (1989). Toward a Unified Theory of Immediate Reasoning in Soar. To appear in the Proceedings of the Annual Conference of the Cognitive Science Society, August, 1989.

Polk, T. A. and Newell, A. (1988). Modeling human syllogistic reasoning in Soar. In *Proceedings of the Annual Conference of the Cognitive Science Society,* pages 181–187.

Simon, H. and Hayes, J. (1979). Understanding written problem instructions. In Simon, H., editor, *Models of Thought,* pages 451–476, Yale University Press, New Haven, Conneticut.

Yost, G. R. and Newell, A. (1988). Learning New Tasks in Soar. Unpublished.