THE JOURNAL OF THE LEARNING SCIENCES, *1*(3 & 4), 361–398

Interaction, Comprehension, and Instruction Usage

Richard Alterman, Roland Zito-Wolf, and Tamitha Carpenter
Computer Science Department
Center for Complex Systems
Brandeis University

This article explores the learning of common practice in a community of agents. We focus on how learning occurs "out in the world," rather than on learning as it occurs in the classroom. The emphasis of this article is that much of the learning of practice is guided by comprehension processes. We take as prima facie evidence for this the ubiquitousness of instructions. We present a concrete model of the comprehension processes that occur when interacting with an unfamiliar device. An important focus is on the usage of instructions.

The study of cognition is greatly affected by whether the research is conducted in the laboratory, the classroom, or in the context of everyday, purposeful activity. In recent years, there has been increased emphasis on studying cognition in the latter setting—"outdoors" or "in the world" (Geertz, 1983; Lave, 1988). Studies of this type have been done in the areas of planning (Suchman, 1987; Wilensky, 1981), learning (Agre & Shrager, 1990), memory (Neisser, 1978; Schank, 1982), language (Lakoff & Johnson, 1980) and mathematical reasoning (Lave, 1988). This article develops a model of how comprehension processes contribute to an individual's acquisition and internalization of common social practice.

The shared set of practices and conventions in a culture creates a sense of constancy in the world (Minsky, 1975; Schank & Abelson, 1977). Nevertheless, from the perspective of an individual agent, the shared world is

---

Requests for reprints should be sent to Richard Alterman, Computer Science Department, Ford Hall, Brandeis University, 415 South Street, Waltham, MA 02254.

slowly but constantly mutating. Economic forces, technological innovation, social trends, or local variations between communities can all be sources of variance in practice. Individuals must keep up with the changes that are occurring if they are to continue to operate within the social context. Each change represents a new learning problem for the individual.

As an example of practice and its evolution, consider the birth of video rental places. Video stores vary in their stocking and checkout procedures. One video rental store adopts aspects of the checkout routines used at the library; a second store adapts those same procedures, in a different way; and a third bases its procedures on a record store. New variations continue to appear. The trick for an individual agent, during her first encounter with a given video store, is to determine the nature of its practice by relating it to the shared background of practice.

Another example of developing practice is in the devaluation of pennies. Several years ago, containers for pennies began appearing at cash registers throughout the country. At first, the containers were a convenience for the customer: Customers deposited their extra pennies in the container after a transaction or they took out a few pennies if they were short of change. Over time, the container has come to serve other functions. Larger denomination coins now appear in the containers. Some cashiers now use the containers as an extra bin of change for cash transactions. The customer and the cashier share an understanding of the situation, which is revised in a way that is comprehensible to both actors.

Practices also drift because of technological innovation, such as the introduction of the Airfone™ which is a type of pay telephone used on airlines. The standard pay-telephone procedure of inserting coins is not appropriate for the Airfone™, because it could not be assumed that an airline passenger would have sufficient change, or access to sufficient change, to pay for a typical call. So, payment procedures involving credit cards were introduced. Other examples involving the everyday usage of mechanical and electronic devices include: microwaves, dishwashers, tape recorders, photocopiers, fax machines, automatic teller machines (ATM), video cassette recorders (VCR), self-starting coffee machines, air conditioners, video games, radio alarm clocks, CD players, and so on. In each case, as technology creates new devices and functions, the common practice for using these devices also evolves.

In this article we are broadly interested in the latter problem of how people learn and adjust to novelty in the operation of devices. The education question for this domain is: How can we develop techniques to train people to increase their learning rate for a given new device or technique (e.g., Woolf, 1988)? The flip side of the education question is: How can we present the usage of a device in a manner that makes comprehension easier (e.g., Norman, 1988)? A component of these two questions is: How is it that we are able to continuously learn and adjust to this ever-changing list of devices that demand our attention? We cannot effectively develop techniques for user training or device design without detailed models of how adjustment and adaptation occur.

In part, adaptation to changes in practice is a problem of memory and understanding. It is a problem of memory (e.g., episodic memory: Kolodner, 1983; Ross, 1989; Schank, 1982) because we are already familiar with similar sorts of devices. It is a problem of understanding (e.g., text comprehension: Kintsch & van Dijk, 1983; Schank & Abelson, 1977; Trabasso & Sperry, 1985) because our fundamental belief is that the usage of such devices is intended to be comprehensible. A third part has to do with interaction and activity (e.g., Agre, 1988; Brooks, 1985; Chapman, 1990; Suchman, 1987). The assumption here is that new practice is acquired as a result of the agent's interaction with the device.

This article develops a model of the role of comprehension processes in interactive adaptation to changes in practice, that is, the *constructive function of understanding*. Previous models of problem solving and learning have not clearly delineated the role of comprehension. The claim is that comprehension processes drive one's interaction with the world in the adjustment to changes in common practice.

This article has two parts. First, we explain, frame, and constrain the role of comprehension in interaction within a social context. In the second part, we demonstrate and extend that framework by presenting techniques for the interactive usage of instructions.

In the first part, we define the constructive function as the key to determining changes in practice for device usage. Our goal is not so much to present specific techniques as to extract a framework from an already-existing piece of work (adaptive planning; Alterman, 1988). The basic idea is that in wending her way through an interaction with a device the agent builds a description of the relevant features of the situation in which she is engaged. The view we develop is that the description that is built results in a representation, that this representation is constructed from the vocabulary of semantic memory, and that there are constraints that the constructed representation must satisfy to ensure that it "makes sense." (These latter constraints, properties, and features are largely extracted from the literature on text comprehension.) These issues are explored against the backdrop of current research in memory; we are particularly interested in the relation of adaptive planning and semantic memory (semantic memory: Lakoff, 1987; Quillian, 1968; Simmons, 1973). One of the themes of this part of the article is that because practice is constantly mutating and because the adjustment to these changes is comprehension based, skill acquisition in this domain is not just the transfer of declarative to procedural knowledge (e.g., Anderson, 1983), but also requires the constant acquisition of new semantics in the agent's domains of activity.

With that said, opinions do differ on the nature of interaction. An extreme position might argue that interaction with devices is entirely reactive and that any understanding that seemed to be associated with the activity was actually imposed post hoc. What from the vantage point of the actor is a sequence of situated reactions could be interpreted by an observer, or in retrospect, as being "planned" or "understood." We take the prevalence of instructions as one piece of evidence that in many cases understanding precedes action (and, therefore, is not merely imposed post hoc). That is, instructions require interpretation to be applied, which, in turn, requires understanding.

The second part defends and extends the comprehension-based framework for interaction by presenting techniques for the usage of instructions. Instruction usage occurs in several stages. First, text inferencing techniques are used in the initial comprehension of the instructions. Summarization and importance-measuring methods are then used to explain each instruction in relation to the agent's current activity. This understanding is converted into specific plan modifications, using a set of communication patterns rationalized by Gricean conversational postulates (Grice, 1975), which are adjusted using adaptive planning techniques as the interaction with the device proceeds. There will be several technical issues that need to be resolved. For example, the exact nature of the interface between an agent engaged in an activity and reading must be explored. The approach we describe uses semantic memory as the lingua franca in which information is exchanged.

In some ways, the approach to instruction usage developed here complements standard planning and problem-solving techniques. The SOAR model (Laird, Rosenbloom, & Newell, 1986; Newell, 1990) deals with impasses by invoking decision-making mechanisms in a higher-order problem space; in Wilensky's (1981) notion of metaplanning, impasses are dealt with in a fashion that makes explicit the rationality of the agent. The model we develop deals with impasses primarily by interpreting external events and objects of the situation. These events and objects can include the action of other agents, iconographs, maps, the affordances of the device, and instructions. The point is that this kind of external information is only available when comprehension processes are brought to bear during the engagement of the agent.

FLOABN

Throughout the article we use examples and techniques from the For Lack Or A Better Name (FLOABN) project to illustrate our points. The core of the FLOABN model is an adaptive planner. FLOABN has two other subsystems: text comprehension and spatial reasoning. It interacts with an environment implemented as a discrete event simulation. FLOABN's domain is learning to operate everyday devices such as those previously mentioned. In this article, we focus on learning about different kinds of telephones (touch-tone phone, pay telephone, Airfones) and telephone calls given an initial knowledge about local calling on a dial telephone (see Figure 1).

We extract details from the FLOABN model to make the exposition more concrete. Because the goal of this article is to present a theoretical framework for comprehension in interaction, many of the details and results of FLOABN are not relevant to the main purpose of the article and, consequently, have not been included in our discussion.

BUILDING AN INTERPRETATION OF A NEW PRACTICE

In this section, the mechanisms, properties, features, and constraints that exist during the construction of an understanding that directly leads to action are presented, that is, the constructive function of understanding. A key idea in this discussion is the necessity of treating the acquisition of practice as a phenomenon that is more than the proceduralization of

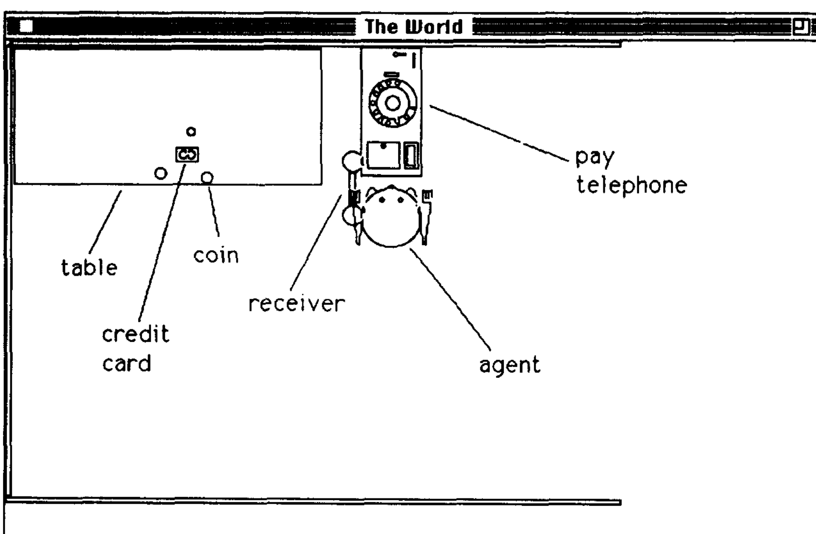

FIGURE 1 FLOABN's Telephone World: Encountering a pay telephone.

declarative knowledge. Rather, it is a dynamic process, combining the acquisition of new semantics and the adaptation and refinement of procedures. Both are necessary to engage in a changing external world defined by a community of agents, where understanding of practice plays a central role.

The Constructive Function of Understanding

In a society of agents, much of an agent's ability to successfully interact with the world is based on her ability to understand the circumstances in which she is engaged. The faculties of perception and understanding form the foundation of the interaction – the conception of the world – and allow the possibility of reasoning, problem solving, or acting.

> Consider the situation when the agent approaches an Airfone™ for the first time. It is a problem of understanding to determine the relationship of the iconograph of the credit card to the ongoing activity of the agent.

One reason for the significance of understanding during interaction arises from the fact that reality for a commonsense planner lacks "objectivity." Traditional artificial intelligence (AI) models of planning assume that objects and situations in the world can be individuated without any interpretation. In the Robby-the-Robot world (e.g., Sacerdoti, 1974), the planner is not only omniscient in the sense that it has a model of the complete layout of all the rooms, but it is also able to immediately apprehend each of the objects that it manipulates: This is a chair; those are bananas; that is a doorway, and so on. However, in the world in which people normally operate, the apprehension of situations and of objects in-the-world is an interpretive process. For commonsense planning of the sort people engage in daily, the world is underdetermined, and a central, and constructive, part of the planning process is to assign (determine the) significance to (of) the relevant features of the situation.

> The agent must interpret the iconograph of a credit card as an "iconograph of a credit card."

A second reason for the significance of understanding during interaction results from the nature of mundane day-to-day activities. From the perspective of an individual agent, the world-of-engagement is slowly but constantly mutating, as are the practices it requires. Nevertheless, there is constancy in the world because of the shared set of practices from which new practices arise (e.g., Minsky, 1975; Rumelhart, 1980; Schank &

Abelson, 1977). This "history" of practices acts as a background from which new practices are revised into being. In many cases, the agent's ability to function effectively depends on the ability to construct an interpretation, an "understanding," of this variance (novelty) in terms of shared concepts of culture, community, home, and workplace.

> The agent must interpret the "iconograph of a credit card" as indicating payment and determine a course of action related to "inserting bank cards into ATM machines."

Constructing this kind of interpretation is precisely the problem of adaptive planning.

Adaptive Planning

An adaptive planner has a memory of previous plans (routines) and retrieves from that memory a plan appropriate to the situation-at-hand. It then adapts that plan (improvises) during the period of engagement as a function of an interpretive process. A key feature of adaptive planning is that it emphasizes the constructive role of understanding in planning. A simple example of understanding is recognizing the features of the situation that allow an agent to make a cup of coffee in the morning. The constructive role of understanding accounts for variance in circumstance in a manner that directly leads to action.

The example of using the Airfone™ for the first time is an adaptive planning problem. Rather than planning from scratch, the planner retrieves from memory a related routine such as using a pay telephone. Rather than planning before engagement, the adaptive planner adjusts the routine during an interaction with the Airfone™. The agent notices the iconograph of a credit card on the device and interprets that to be an indicator of the "method of payment." This understanding of the iconograph leads directly to the action of using a credit card as the method of payment; in this example, it allows FLOABN to convert a routine for using an ATM card into a payment procedure for the Airfone™.

Previous work on adaptive planning (Alterman, 1988, 1990) was most directly influenced by the work of Wilensky (1983) on commonsense planning, Schank (1982) on the role of memory in understanding, and the early work of Simmons (1973) on semantic networks/memory. An adaptive planner works in "commonsense-planning situations"; these refer to the mundane day-to-day activities of human planners (Wilensky, 1983). A key feature of such activities is their routine nature; it is less the case that the activities of the planner vary a great deal and more the case that the circumstances under which the plans are applied vary.

The Memory Architecture

In the remainder of this section, we discuss issues in comprehension against the backdrop of FLOABN's memory architecture. The memory architecture in FLOABN consists of three levels: semantic memory, situated plans, and embodied skills. Figure 2 shows a schematic description of some commonsense knowledge about telephoning, insertions, and payment as it is distributed across FLOABN's three levels of memory.

> Semantic memory provides a vocabulary. It includes not only the meaning of words, but also the meaning aspect of plan and action notions. At the level of semantic memory are such notions as "insert card," "insert," "payment," "card," and "use telephone," related by partonomic,[1] category, role, antecedent, and consequent links.
>
> Situated plans provide a base of "practices" in the form of procedures. Situated plans are cases or routines customized to particular contexts. An example of a routine is the routine to withdraw money from the automatic teller machine at my bank in Waltham. These routines are referred to as situated to emphasize that they are sequences of steps coupled to particular contexts (environments).
>
> Embodied skills are clusters of hand/eye motor routines. Steps in situated plans can invoke embodied skill routines. The hierarchical structure of situated plans "bottoms out" in such primitive steps. For example, one of the steps involved in using a credit card is removing the credit card from the wallet. We do not discuss this level in detail in this article.

Schemata (Bartlett, 1932) cut across all three levels of memory. The schema for "inserting ATM card in bank at Waltham" is rooted at the middle level, with the notion of insert card as its semantic residue and the appropriate hand/eye motor skills as its embodiment.

Modeling Understanding as Building a Description

In symbolic AI, understanding is reduced to a problem of representation. A program is said to have understood an input to the degree to which it can represent the input and reason about the representation it builds. Examples of the sorts of reasoning that are done on the representation/understanding are summarization, question answering, and, in the case of FLOABN, the generation of action.

[1]The term *partonomic* comes out of the categorization literature and refers to the part hierarchy (Tversky & Hemenway, 1984).

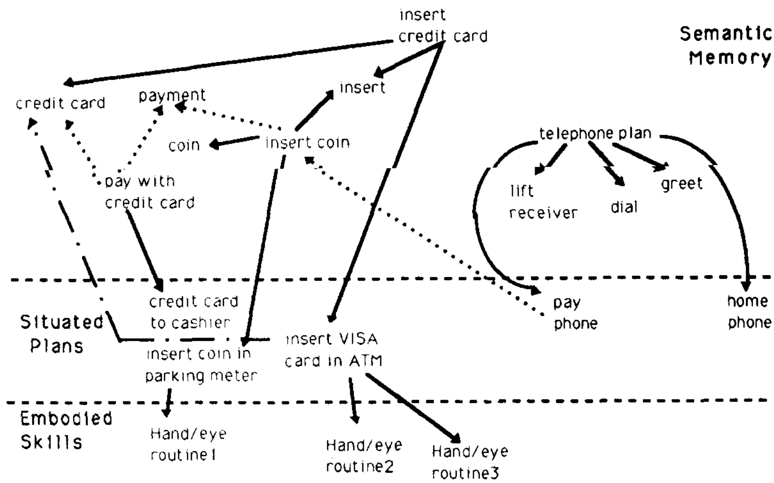

FIGURE 2 A schematic of a portion of three levels of memory.

The view here is that the constructed representation is built out of the "vocabulary" of semantic memory; there is an encoding relationship between semantic memory and the constructed representation. An important feature of this relationship is that semantic memory and the constructed representation share some of the same structure—the structure of the understanding is a "copy" of some piece of structure taken from semantic memory. This sort of copy-based notion of the understanding is inherent in semantic network-based models of text comprehension (e.g., Alterman, 1985; Charniak, 1986; Hendler, 1987; Simmons, 1973; Wilenksy et al., 1989).

For the example of the iconograph of a credit card depicted on the Airfone™, part of the constructed understanding includes the notions, copied from semantic memory:

> "Credit card" and "insert coin in telephone slot" are related through payment (dotted lines in Figure 2).
>
> One of the situated plans associated with a credit card is "insert VISA card in ATM" (dashed line in Figure 2).

Having constructed this initial understanding, the agent adapts the situated plan, "insert VISA card in ATM," to the situation of engagement.

"Making Sense": Three constraints on the constructed representation. The constructed representation captures only part of what it means to "make sense." We describe three constraints on the constructed representation that further characterize the agent's comprehension of a given situation. The actual constraints are not novel (e.g., Lakoff, 1987); our concern is to make them explicit in the computational model. The significance of these constraints is that they are used for selecting candidate interpretations of the circumstances.

The first constraint ties together the perceptual and comprehension processes:

> Correspondence. A set of points of reference between the constructed representation and the situation of engagement.

The constructed representation must be about something; this is the problem of external reference. In an interactive problem-solving situation, a large component of this is perceptual; examples of research on these problems include the procedural semantics work of Miller and Johnson-Laird (1976) and Winograd (1972) and, more recently, the work on external reference of Goodman (1986) and Chapman (1990).

The second constraint is:

> Coherence. The correspondences must be cohesive and internally consistent.

The notion of coherence as a condition of understanding has a long tradition in studies of text. Current debates center on how deeply materials are read (e.g., Black & Bower, 1980; Graesser & Clark, 1985; Kintsch, 1988; Kintsch & van Dijk, 1983; Schank & Abelson, 1977; Trabasso & Sperry, 1985; Trabasso & van den Broeck, 1985). Here, we interpret the constraint of coherence as *concept coherence* (Alterman, 1989): the constructed representation must be internally consistent within a given span of semantic memory.

Considered together, these two constraints imply that any description the agent builds on the world must have correspondents in the agent's perception of the world and that these constructed descriptions are internally consistent within semantic memory.

A third constraint on the understanding process is:

> Relation to goal. The interpretation must be related to the agent's current (sub) goal.

Any interpretation of a situation involves, or is biased by, the question of goals[2] (e.g., Carbonell, 1978; Hidi & Baird, 1986; Ram, 1990; Wilensky,

[2]The issue of intention has long been a theme in the discourse literature (e.g., Grosz & Sidner, 1986). However, these intentions are the intentions of the speaker. Here, we are interested in the intentions, or goals, of the "understander."

1983). The following example is suggested by Nicholson (1984): A passenger is on the way home on the train, and there is a certain house such that, when the passenger passes it, he knows he is within a certain distance of his destination. So, for this passenger, the house is perceived as a marker of the closeness of a destination. However, there is a boy in the backyard of that house, and he is throwing a ball against its back wall. To that boy, the house is perceived as an object to bounce a ball against. So, the same thing is interpreted differently because of the differing goals of the agents.

In other words, the current goal or subgoal serves to select among candidate interpretations. There are at least two coherent paths through the network shown in Figure 2 that can be "copied" to explain the relevance of "credit card" to the situation at hand (see Figure 3). The first path is:

> "Insert coin" is a kind of insert. Another kind of insert is "insert credit card," which has a credit card as a prop.

A second path of coherence through the network is:

> "Insert coin" has a coin as a prop, which is also a prop for "pay with coin." The latter is a subclass of "pay with cash," which is a kind of payment. "Pay with credit card" is another kind of payment, having a credit card rather than a coin as a prop.

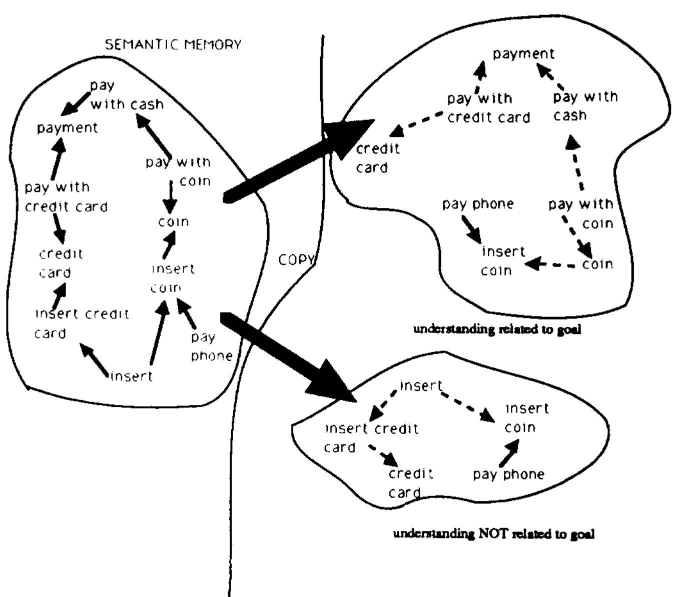

FIGURE 3 Copying from semantic memory.

The understander chooses the interpretation most relevant to the situation at hand. In this case, the second interpretation is favored because it connects the two concepts through the notion of payment, which is critical to the goal of the agent in the situation.

The Movement From Understanding to Action

When selecting action, the agent must be responsive to context. At the level of embodied skills, an agent must deal with factors like the exact position of the telephone as he reaches to pick it up. At the level of situated plans, contextual factors are also relevant. Consider three different places from which I might make a telephone call: my home, my office, and a friend's home. At home, I know I can find the numbers of friends on the bulletin board, in my little yellow book in the bookcase in my bedroom, or in Claire's address book, which is usually in her purse. At my office, numbers I consider important are taped to the wall over my telephone. I can also call upstairs to the department office to get a number, use the university directory underneath the telephone, or use the "finger" utility on the computer. At a friend's house, I have to ask to use their telephone book. In each situation described above, my ability to act is dependent on contextualizing my understanding.

There are at least two ways to contextualize a plan. One approach is to start with an abstract plan and to refine it in a piecemeal fashion (hierarchical planning: e.g., Sacerdoti, 1974). The case-based approach (CBR: Alterman, 1988; Hammond, 1990; Kolodner, 1988; Rissland & Ashley, 1986; cf. memory-based reasoning: Stanfill & Waltz, 1986) is to contextualize all at once by selecting some preexisting routine (situated plan) from memory. Because of the habitual nature of commonsense planning situations, FLOABN bases its activities on a memory of preexisting situated plans.

Thus, the movement from understanding to action has two parts:

1. Construct a representation from the language provided by semantic memory that is sufficient to select a situated plan.
2. Use the immediate circumstances as a context to step through the situated plan, selecting appropriate clusters of embodied skills by the indices provided.

This description-based notion of memory retrieval is in consonance with Kolodner (1983) and Norman and Bobrow (1979); here, we add the constraint that the constructed description is "sensible," and we focus the description-building process toward the agent's immediate external circumstances. That is, it is a copy of some internally consistent piece of semantic memory (coherence), it is about the situation of engagement (correspondence), and it is related to the agent's goal. Not every retrieved routine will fit perfectly, but the constraint that the new course of action makes sense in the current circumstances is a powerful technique for selecting action.

If FLOABN can understand the situation sufficiently—describe it in sufficient detail to select a situated plan that makes sense—then it can act on that understanding. In the case of the Airfone™, once FLOABN interprets the method of payment as involving the use of a credit card by insertion into the device, it can invoke the routine and cluster of skills involving inserting cards in an ATM.

Representing a routine or situated plan. Situated plans are sequences of steps coupled to particular contexts (environments). An interesting feature of situated plans is that they tend to drift toward the agent's habitats, as is illustrated by the different plans for looking up telephone numbers described earlier in this section.

Figure 4 shows a representative situated plan. There are two kinds of encodings used here. The semantic encoding ties the structure of a plan to semantic memory through the category structure of plans and steps, partonomic structure (how plans decompose into steps), and role knowledge (constraints on allowable fillers for roles). The second encoding involves the contingency properties of the situation. These include causal relations (e.g., reason, purpose), temporal ordering (e.g., before DIALing the number PICK-UP-RECEIVER), and context-dependent decision points (e.g., if the number is nonlocal, DIAL-1 before DIAL).

Each step in a situated plan invokes a cluster of embodied skills. For example, one of the steps involved in using a credit card is removing the credit card from the wallet. This step is embodied, invoking a cluster of hand/eye motor skills. Some contingencies are handled directly by these embodied skills: Twisting the wallet when it gets caught in the pants pocket is an embodied skill; feeling around for the credit card is another. Another cluster of skills is used to insert the credit card into the appropriate slot in the Airfone.

ADDING INSTRUCTION USAGE TO THE FRAMEWORK

There have been a few other computational approaches to instruction use, and several psychological ones, related to this work. Chapman (1990) developed a model of instruction in which the instructor and instructee are both engaged in a situation, and the instructor "kibitzes" with the instructee during the engagement. The model Chapman developed, SONJA, is a system that plays a video game, and the instructions are intended to be

Semantic Plan Elements	**Taxonomy:** telephone-plan **isa** long-dist-communication **isa** ...
	Parts of telephone-plan include:
	pick-up-receiver, listen-for-dial-tone, lookup-number, dial
	Roles include: dial-telephone **isa** device
Contextualized and Causal Elements	**Step ordering** information:
	pick-up-receiver *before* listen-for-dial-tone *before* dial ...
	Decision points:
	If non-local(call) then dial-1;
	To perform step lookup-number:
	recall-number
	else (if home) try look-on-bulletin-board
	else (if office) try look-on-wall
	Causal relations:
	The **reason** for pick-up-receiver is dial
	The **purpose** of telephone-plan is communication
	The **preconditions** of pick-up-receiver are: exist(telephone)

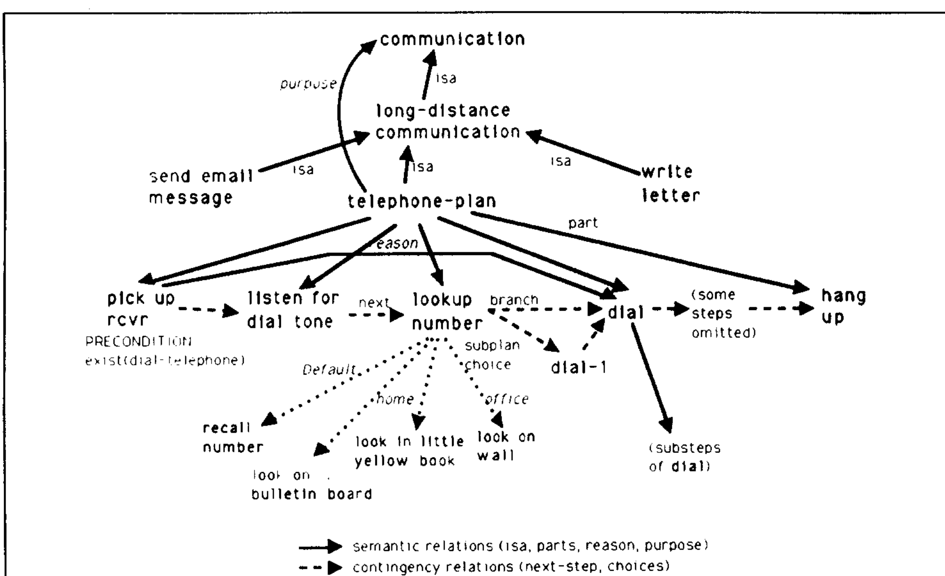

FIGURE 4(a) Components of a situated plan for telephone calling. (b) The situated telephone plan in schematic form.

real-time advice that a player might get while playing the game. Instructions are very short and are chosen from a fixed vocabulary of about 20 phrases. Mannes and Kintsch (1991) presented a model of instruction usage in which the instructions are presented before the agent begins to interact with the environment. The meaning of the instructions is assembled on-line using the construction-integration model of Kintsch (1988). Badler, Webber, Kalita, and Esakov (1990) described animated simulations that are driven by natural language instructions. Vere and Bickmore (1990) discussed a basic agent whose actions are initiated by instructions from a user. Ross (1989) presented psychological studies that show the importance of memory in the usage of instructional material; LeFevre and Dixon (1986) showed that instructions are more easily applied to future problems when presented as specific examples; Martin (1990) used instruction as a basis of knowledge acquisition; and Kieras and Bovair (1986) discussed a psychological model of instruction usage for made-up devices in which the subjects first read the instructions in their entirety and then proceduralize the understanding.

The view of instruction usage developed in this article is informed by all of the aforementioned material, but it differs in its emphasis and assumptions. FLOABN is a model of instruction usage in which memory and comprehension play a central role, and instructions are not restricted to single clauses. One critical assumption of this work is that, although events have temporal extent, in general, the agent is not particularly pressed for time as is the case when playing video games (Chapman, 1990).

FLOABN, as an adaptive planner, does not plan primarily from instructions but begins by engaging in the activity. There are three reasons for this. First, in many cases, reading instructions in their entirety would be either highly redundant or irrelevant. This is because changes in practice occur slowly, so any given situation contains a small number of relevant novel elements. Two kinds of situations can arise during activity that invoke instruction interpretation. If a problem arises that FLOABN's adaptation techniques cannot handle, FLOABN looks for instructions. FLOABN also accepts instructions addressed directly to it, such as error messages, kibitzing from an instructor, or messages appearing in the range of its attention. Although instructions are used in response to events, instruction use is nevertheless an active process: The search for, and response to, instructions is under the agent's control, and the determination of an appropriate response involves an active process of interpretation.

Second, instructions are difficult to understand when encountered outside the context of action because they tend to be abbreviated. Most instructions (written or verbal) omit an enormous number of presumably shared details, with the assumption that an agent sharing general background and an understanding of the situation can fill them in. Outside of the context of use, the agent grasps only the general sense of what the instructions mean and the operations they depict. Because the agent is already engaged in the situation when encountering the instructions, the situation provides a backdrop against which the instructions can be made concrete.

A third reason for beginning interaction before reading instructions is that instructions, although abbreviated, tend to be phrased in terms of concrete actions and advice. Where space allows, as in instruction manuals, complete examples that can be executed directly are preferred. Where brevity is important, as in interactive instructions, the instructor typically assumes the instructee possesses an almost-correct plan so that the instructor need only convey the missing details. Such instructions often refer to steps and plans the instructee is presumed to be executing, or at least, to possess. If we encounter the instruction "Do not top-off the tank" while filling up at a gas station, it is understandable precisely because we already have an adequate plan for pumping gas and because we recognize the instruction as referring to a step in a related plan for measuring fuel consumption.

A critical issue in modeling the usage of instructions is building the interface between the understanding and plan/action components of the agent. Much of the groundwork for this interface is built into the framework described in the previous sections.

Semantic memory is the common language through which the planner, the text inferencer, and other modules of the system communicate. A central idea of this interface is that the adaptive planner is building a description as the situation unfolds. In FLOABN's memory, each plan has a semantic encoding, and these encodings are used to describe the world as FLOABN interacts with a device. The constructed interpretation of the interaction results in a representation, in the language provided by semantic memory, of events that have occurred up to the point of instruction; the constructed representation obeys the constraints of coherence, correspondence, and relation to goal. The whole notion of FLOABN having an ongoing representation of the situation is important for the initial phase of instruction interpretation—because this representation is available, the first phase of instruction usage can be treated as a problem of text inferencing.

In the model presented in the next section, instruction interpretation occurs in two phases. FLOABN first establishes the coherence and correspondence of the instruction in relation to the ongoing activity. It then proceduralizes that interpretation, modifying its plan in a manner consistent with the instructions. The initial phase establishes that the instructions make sense in relation to the agent's current engagement with the situation. The second phase of interpretation must deal with the nitty gritty details of modifying some preexisting routine. Part of the proceduralization of the instructions occurs as action proceeds so that the instructions' situation-specific meaning can be excavated.

THE METHOD OF INSTRUCTION USAGE

The FLOABN model attempts to use instructions when an instruction interrupts activity or when a problem arises that cannot be handled either "reactively" at the level of embodied skills or "adaptively" at the level of situated plans.

The interpretation of instructions is a four-stage process (see Figure 5). It is assumed that the instruction text is passed initially to a parser, and instruction interpretation begins with the resulting syntactic parse. Stage 1 makes initial sense of the instructions by building a coherence representation that captures the connectivity of the text in relation to the ongoing

Stage 1: Text Inferencing. Use semantic memory to build a coherence representation that characterizes the relationship between the instructions and the ongoing activity.

Technique: Marker Passing

Output: An interpretation graph that is a copy of a piece of semantic memory.

Stage 2: Analysis of Interpretation Graph. Account for each instruction semantically.

Technique: Graph-based summarization techniques.

Output:
1. Conceptual Roots: The minimal set that covers the interpretation graph, representing the main points of the text.
2. Importance: A graded measure of importance based on the *coverage* of each node in the graph.
3. Explanation: Relationship of each instruction to the ongoing activity.

Stage 3: Proceduralization. Convert instruction elements into specific plan modifications and actions.

Input:
1. The explanation and interpretation from stages 1 and 2.
2. Current plan and execution state.

Technique:
1. Use step-ordering relationships to group instruction elements.
2. Use instruction communication patterns to account for (determine the purpose of) each instruction group.

Output:
1. Elaborated plan network.
2. List of actions to perform immediately.
3. Information about resuming action.

Stage 4: Resume engagement. Complete instruction interpretation as action proceeds.

Input: Modified plan from stage 3.

Technique: Adaptive Planning.

Output: Actions to be performed, additional plan modifications.

FIGURE 5 Reading the instructions in the context provided by the ongoing activity.

activity. The goal is to express individual instruction elements in terms of plan steps existing in memory to enable coordination of the instructions with the current plan(s). Stage 2 performs an analysis that results in an explanation of each instruction in relation to the ongoing activity. This ensures that, if an instruction has several possible interpretations, FLOABN will choose the most applicable interpretation for the situation. For further details on these techniques, see Carpenter and Alterman (1991).

Stage 3 *proceduralizes* the interpretation of the instructions, completing the understanding by making modifications to the situated plan (e.g., insertions of steps) and to the current knowledge state. Once the plan has been elaborated to handle the current situation, action can proceed (Stage 4). As a by-product of action, further plan modifications occur: overgeneral instruction steps are made specific, additional steps are inserted, and so forth. For further details on these techniques, see Zito-Wolf and Alterman (1991).

Stage 1: Text Inferencing

Semantic memory is used to build a representation, an *interpretation graph* that characterizes the relationship between the instructions and the ongoing activity. Each node in the interpretation graph is either a concept explicitly indicated by the instructions, a concept developed by the adaptive planner in the context of the ongoing activity, or a concept encountered while establishing coherence.

The interpretation graph is created by using a modification of the marker-passing techniques (see Figure 6) described in Norvig (1989; cf. Alterman, 1985; Charniak, 1986; Collins & Loftus, 1975; Hendler, 1987; Martin & Riesbeck, 1986; Quillian, 1968). The basic idea is to find paths in semantic memory between two concepts evoked by the instructions or between an instruction concept and a concept in the representation of the ongoing activity. These paths are matched by a finite-state machine to a small number of legitimate predefined patterns or *collision-types*. Two example collision types are:

Action name. This finds the action concept in semantic memory that relates the verb and the object of the instruction.

Similar goal. This indicates that an action concept from the instructions has the same goal as a step from the situated plan in use during the engagement.

If a path is recognized by the finite-state machine, it is copied to be included in the interpretation graph and is marked with its collision type.

FLOABN recognizes general classes of inference as follows:
1 FOR each node *N* in semantic memory representing the ongoing activity and each word in the text of the instructions, DO:
1a Find the nodes related to *N* via any of a relatively small number of predefined patterns of connectivity (*path shapes*).
1b Identify nodes where these paths meet (*collisions*). A collision defines the complete path between two concepts in the ongoing understanding. (Note: Path shapes are defined by the types and order of links between the nodes. Only a few collision types are defined or needed; all undefined collisions are ignored.)
2 COPY each node and link involved in the collisions. This is the interpretation graph.

FIGURE 6 Text inferencing algorithm.

The resulting representation has two critical features. First, it is a directed acyclic graph (DAG), a necessary condition for Stage 2. Second, it obeys the constraints of coherence and correspondence. Coherence is preserved because each path that is copied into the interpretation graph is constrained by the small number of legitimate collision types. Correspondence is guaranteed by definition—the inputs to the marker passing are the instructions and the ongoing activity, and, therefore, the endpoints of any path that is copied into the interpretation graph have "external reference."

Stage 2: Analysis of Interpretation Graph

Because the interpretation graph is copied directly from semantic memory—it is encoded by semantic memory—and because it is a DAG, it can be analyzed using techniques developed by Alterman and Bookman (in press). First, the semantic memory encoding provides a measure of importance that quantifies the conceptual emphasis of the understanding. Second, the *conceptual roots* can be identified. The conceptual roots are the basic framework of the understanding/representation encoded by semantic memory and can be used to explain succinctly the connection between an instruction and the ongoing activity. Third, a combination of techniques can be used to generate a description of the basic content of the instructions (i.e., a basic summary).

The techniques for finding the conceptual roots and their importance use graph-theory methods. In the terms of graph theory, the conceptual roots are the minimal set that covers the interpretation graph; that is, the smallest set of nodes in the graph from which every node in the graph can be reached. The importance of a given conceptual root is the number of nodes reachable from it.

The conceptual roots and their importance, with the collision types identified in Stage 1, are used to build a description of the instructions in the language provided by semantic memory. For each instruction, four pieces of information are determined:

- The *action* that the instruction conveys. A collision type of "action name" identifies an action concept in semantic memory that expresses the same meaning as the words of the instructions.
- The *position* of the instruction. This is simply a number that indicates where in the sequence of instructions this instruction was read or heard (e.g., the first instruction has position "1," the second has position "2,"). Knowing this ordering is a useful clue for the adaptive planner.
- The *explanation* of how the instruction is related to the ongoing activity. If the action is simply a step of the original plan, this is indicated by the explanation. Otherwise, FLOABN finds the conceptual root covering the action and some part of the representation of the ongoing activity. If two such conceptual roots exist, the root with the highest importance measure is used.
- Any *special descriptions* indicated by the instruction. These include descriptive words (e.g., "momentarily," "again") that are not used when determining either the action or the explanation, slot fillers for the action (e.g., "10 cents"), and any other information conveyed by the instructions.

FLOABN tries to learn from the novel aspects of the conceptual emphasis of the instructions by examining the most important points of the interpretation graph to see whether it can identify anything the instructions express about the device or the situation that is unfamiliar. If an unfamiliar aspect is identified, a new concept representing this novelty is then created and added to semantic memory. The actual technique used is to examine the collision types, found in Stage 1, that are associated with the most important conceptual root. Each collision type has associated with it a short procedure (e.g., see Figure 7) that works in two steps. The first step tests whether anything can be learned from a particular instance of a collision of this type. The second step details a method for adding a new concept to semantic memory as a result of a collision if it is appropriate; this includes naming the new concept and summarizing the collision.

It is important for FLOABN to perform this learning because, as social practice continues to change, the new concept can be used to comprehend

```
1   IF the collision-type is 'new step'
1a      LET N be the conceptual root covering the path (e.g., 'payment').
1b      LET N1 be the endpoint of the path that is a concept in the representation of
            the ongoing activity (e.g., 'telephone plan').
1c      LET N2 be the endpoint of the path that is a concept expressed by the instruc-
            tions (e.g., 'insert coin').
1d      CREATE a new node, X, using N and N1 (e.g., 'payment telephone plan'). X
            is a 'situated plan' of N1, and N2 is a 'step' for X.
```

FIGURE 7 An example learning procedure.

future situations. If an agent always has to reconstruct the understanding from scratch, each small change in social practice becomes increasingly more difficult to grasp, and consequently, the agent's ability to perform in the context of variance deteriorates.

Stage 3: Modifying Plan Structure (Proceduralization)

At this point, three pieces of information are available. The first is the current circumstances: the plan being executed and a partially built-up description of the situation. Second, the agent knows where in the plan the interruption occurred. The third is the analysis of the instructions.

Although the instructions have been semantically interpreted, the issue remains of how to convert these "understandings" into actions. During proceduralization, FLOABN makes specific step-oriented plan changes based on the instructions, such as step insertions, step reorderings, and decision point insertion. Second, it must determine how to proceed with the activity—whether the current plan needs to be restarted or whether it can be continued from the point of interruption.

The proceduralization algorithm is shown in Figure 8. A given instruction text can contain more than one communicative pattern, hence, the need to appropriately segment the text. FLOABN first looks for instructions indicating that the current plan will need to be restarted (Line 1). The textual analysis identifies such cases: the explicit mention of restarting, of some related term, or of a canonical terminating step for the plan (e.g., HANG-UP for TELEPHONE-PLAN). The procedure looks then for subgroups of the instruction set that represent independent communicative

To PROCEDURALIZE a set of instruction *Steps* (as returned by the text inferencer), DO:

1 LET *Restart_Steps* be the set of elements of *Steps* representing restart operations (as determined in stages 1 and 2). Remove these steps from *Steps*.

2a Chunk steps into groups by step-ordering relationships explicitly indicated by the text (e.g., **before**, **after**):

2b For each chunk, identify the communication pattern used.

2c IF a communication pattern was found, modify the plan accordingly, else treat the chunk as immediate instructions.

3 Execute any instructions designated immediate in step 2c.

4 Execute any *Restart_Steps* found.

5 IF *Restart_Steps* were found or plan structure prior to current execution point was modified, restart the *Current_Plan* at the beginning, else continue executing from the point at which the instructions were received.

FIGURE 8 Instruction proceduralization algorithm.

groups (Line 2a). It identifies such by the presence of explicit step-ordering relations,[3] such as:

You must *first* dial a one, when calling this number.

Each time such a group of steps is found, FLOABN identifies its communication pattern (Line 2b) and takes the appropriate action (Line 2c). When all such groups have been extracted, the remaining steps are interpreted as a final group.

In proceduralizing the instructions, specific communication patterns

[3] The semantic interpretation performed by the text inferencer includes the translation of lexical step-ordering information (e.g., "before," "then") into semantic primitives (e.g., before, after).

account for (portions of) the instruction text. Communication patterns focus on groups of steps because instructions are often not simply of the form "do <new-step>," but also mention existing plan steps to establish context and coherence with the current plan. These patterns generalize forms found in actual instructions. These forms, though schematic, can be justified using Grice's conversational postulates (Grice, 1975). One such pattern is shown in Figure 9, where mentioning an existing step in a plan implicitly informs the agent of where a new plan step is to be inserted. Three such patterns used by FLOABN are:

Insert step in indicated place. If the instructions refer to existing step(s) in the current plan as predecessors or successors to a new step, this indicates that the new step is to be added to the plan in the specified relation to the existing step(s). The relations may be explicit, as in the "dial a 1" instruction previously mentioned, or implicit, as in the message given by a pay telephone to encourage one to insert a dime (see the section An Example of Instruction Use).

Implicit assertion of a fact. Where the previous pattern deals with the insertion of steps, this pattern acquires their correct conditions of use. If the instructions refer to some step that is present in the situated plan being executed but was not executed in the current situation, it draws the agent's attention to some missing precondition of a decision point controlling execution of that step. Stating that this instruction should have been executed implicitly informs the agent that the said precondition is true in this situation. An example would be a situation where the agent's telephone plan knew about dialing prefix 1 but where the agent lacked the information that the particular number it was dialing was nonlocal.

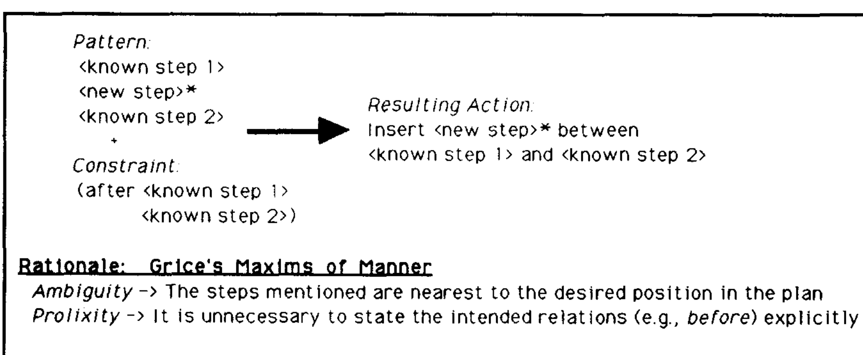

FIGURE 9 Communication pattern: "Insert step into indicated place."

Immediate execution. If no specific connection to the current plan can be established, assume the instructions are to be executed immediately (i.e., at the agent's current position in the plan) and in the order given. This case handles single instructions, for example, "Please deposit $.80" or the residue of compound instructions. This is the default case.

The common element in these instruction patterns is the search for partial matches to a current plan, to establish a context for proceduralization. The existing plan subsequence most like the instruction sequence is found and used to determine what modifications to the plan or world knowledge are required to reconcile the two. To find the best match of an instruction sequence to a plan, all possible step sequences generated by that plan (allowing for varying circumstances and by excluding loops) are examined, and the sequence maximizing the size of intersection(Instruction_Sequence, Sequence) while preserving the order of the steps in the instruction sequence is selected.

Once all the instructions have been processed, action can resume (Figure 8, Steps 4 and 5). If any Immediate_Steps were found, they are performed now. Plan execution then resumes. If explicit restart steps were found, or if a plan modification was made that changes our previous actions (i.e., a modification to the portion of the Current_Plan that has already been executed), then that plan must be executed from the beginning. Otherwise, FLOABN proceeds from the point of interruption.

Each plan modification incorporated into the situated plan represents an extension of the range of situations to which the plan applies and, hence, an extension of the agent's competence—the agent has learned something. The agent also learns indirectly from instructions through his episodic memory of the instruction interpretation episode. If in the future a similar instruction is encountered, it will know what to expect. It will not need to completely reinterpret the situation and instructions because it already knows what they are telling the agent to do.

Stage 4: Adaptation During Engagement

Instruction proceduralization is completed while interacting with the device. In many cases, there are unanticipated interactions that need to be sorted through during the actual engagement. These sorts of on-line interpretative adaptations are the bread and butter of an adaptive planner. For example, suppose that the system is instructed to insert a certain amount of money for a long-distance call. This causes an INSERT-MONEY step to be added to the plan. The details of finding and inserting the specific amount of money requested need not be addressed until the step is (about to be) executed.

As engagement proceeds, several kinds of learning occur. Adaptations made during this stage elaborate the base plan (with appropriate conditionals) and, thereby, extend the base plan to new situations. Second, the details of each episode are stored in episodic memory, which in later encounters can be used to generate expectations (cf. Kolodner & Simpson, 1989; Rissland & Ashley, 1986; Schank, 1982). Also, information about role fillers is learned, for example, (the overgeneralization) that all PAYPHONEs will require a $.10 deposit. (This generalization will be refined in later encounters.)

AN EXAMPLE OF INSTRUCTION USAGE

This section illustrates the methods described in the previous section with an example of FLOABN making a local call on a pay telephone for the first time. We assume that FLOABN has a basic set of routines for using a home telephone. When placing a call, sometimes the normal course of events is interrupted by a message from the telephone. If this occurs, the system automatically goes into instruction interpretation mode. (We assume the ability to distinguish instructions from other percepts, such as a greeting from the desired party.) The example in this section shows FLOABN applying its home-telephone plan to a pay-telephone situation (shown in Figure 1), and assimilating the instructions received from the telephone to extend its plan for placing calls. (Detailed traces are given in the Appendix.)

The agent approaches the pay telephone, lifts the receiver, hears the dial tone, and dials the number. The telephone rings twice, there is a beep, and the instructions shown in Figure 10 are heard. The text is parsed, and FLOABN interprets the result.

A Semantic Interpretation of the Instructions (Stages 1 and 2)

FLOABN begins this stage with a semantic memory containing relevant background knowledge, including knowledge about the current situation of

1. Please hang up momentarily,
2. listen for dial tone,
3. deposit the required coin,
4. and dial your call again.

FIGURE 10 The instructions received from the pay telephone.

engagement as constructed by the adaptive planner, but with no representation of a pay telephone or its practice (because this is the first time FLOABN has encountered a pay phone). FLOABN passes markers (see Figure 6, lines 1 and 2) from each significant word in the instructions (i.e., words other than "please," "the," and "and") and from the current plan being adapted: "telephone plan."

When the marker passing is completed, all the paths that were involved in defined collisions are copied from semantic memory (Figure 6, line 3), and summarization is performed. The result is a list of the conceptual roots and their relative importance, shown in Figure 11. The conceptual roots provide the information necessary to explain the connection between each instruction and the ongoing activity.

Using the results of both the marker passing and the summarization, FLOABN constructs the description of the pay-telephone instructions shown in Figure 12. The first pay-telephone instruction (see Figure 10) matches the step "hang-up." The purpose of executing this step is to "restart call," the conceptual root covering hang up in the interpretation graph. The descriptive word "momentarily" is also included for completeness.

The second instruction indicates the step "listen for dial tone." This step is explained as a reference to the original plan and has no further significance. The fourth step has the same explanation, but the step is "dial" and it requires that the thing dialed be a "number." This is somewhat ambiguous because number can indicate either the type of the object of dial or an already-determined telephone number. However, because the dial step in the current situation of engagement already has the value of a telephone number, this interpretation is used by default. Dial also has the descriptive modifier of "again," indicating that dial was executed before instruction interpretation began.

The third instruction illustrates three aspects of this phase of instruction interpretation. First, FLOABN finds that the instruction "deposit the required coin" conveys the same action as the concept insert coin. Second, it finds that the value of the coin to be inserted is $.10 (from a related instruction found on the front of the pay telephone). And third, it explains

conceptual root	importance
PAYMENT	10
INSERT	5
RESTART-CALL	1
CALL	1

FIGURE 11 The conceptual roots and their relative importance for the pay telephone example.

```
((HANG-UP :POS 1 :ISA RESTART-CALL :MODIFIER ((MOMENTARILY)))

 (LISTEN-FOR-DIAL-TONE :POS 2 :STEP-OF TELEPHONE-PLAN)

 (INSERT-COIN :POS 3 :ISA PAYMENT :WITH ((:VALUE * 10 :TYPE CENTS)))

 (DIAL :POS 4 :MODIFIER ((AGAIN)) :STEP-OF TELEPHONE-PLAN :WITH ((:VALUE * NUMBER))))
```

FIGURE 12 The results produced from the pay telephone instructions.

that this instruction indicates the method of payment for a pay telephone. In the original telephone plan, payment was satisfied by the step of paying a monthly bill.

The interpretation graph and conceptual roots are then used to add to semantic memory a new concept that represents what the instructions say about the pay telephone. This is done by observing that the most important conceptual root is payment, and the collision at that node is as shown in Figure 13. In this case, the collision is of type "new step"—the intersection of a "goal of action" path with a "goal of plan" path—which indicates that a new concept, "payment telephone plan" is to be created. This concept will become a situated plan associated with the original telephone plan, and will have a step of insert coin.

Instruction Proceduralization (Stage 3)

At this point, FLOABN has three pieces of information. The first is the current circumstances: the current plan is TELEPHONE-PLAN, the telephone is in the hand, and an instructional message has been received. The second piece of information is where in the plan the interruption occurred: while waiting for the call to be answered. The third is the analysis of the instructions.

FLOABN assumes that instructions received during engagement provide two kinds of information: plan modifications and recovery information. FLOABN must determine how the plan modifications are to be incorporated—where in the plan they fit and in what circumstances they apply.

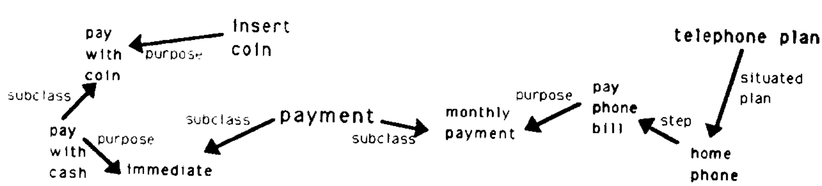

FIGURE 13 The interpretation copied from semantic memory.

Then, it must use the recovery information to determine what action (if any) is needed to resume plan execution.

First, the instructions are segmented (Figure 8, Lines 1 and 2a). The HANG-UP instruction is separated out as a restart instruction. The remaining instructions contain no explicit grouping, so they are treated as one group by default.

Second, the communication pattern for this group is identified (Line 2b). This group matches the pattern shown in Figure 9, in that it positions a new plan step (INSERT-COIN) by referencing two steps already in the plan. Mentioning existing steps in a plan implicitly informs the hearer of where in the plan the new plan step is to be inserted. This instruction group, therefore, indicates a new *deposit* step is to be inserted between the existing instructions *listen for dial tone* and dial.

The insertion of INSERT $.10 before DIAl in the context of a pay telephone (note the overgeneralization) and the idea of getting the dime from the agent's pocket (see next section) are forms of learning because they extend the agent's range of competence. These steps will now be available without explicit cogitation in similar situations in the future. The resulting telephone plan is shown in Figure 14.

Interaction Resumes (Stage 4)

Having accounted for all the instruction elements, FLOABN can now execute the HANG-UP instruction (Figure 8, Line 4). FLOABN resumes plan execution at the start of TELEPHONE-PLAN because restart instructions were given (Line 5).

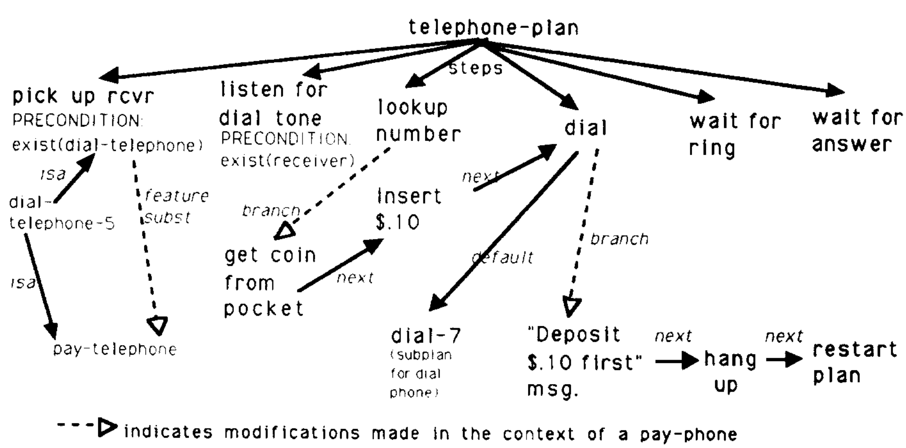

FIGURE 14 Telephone plan after pay telephone encounter.

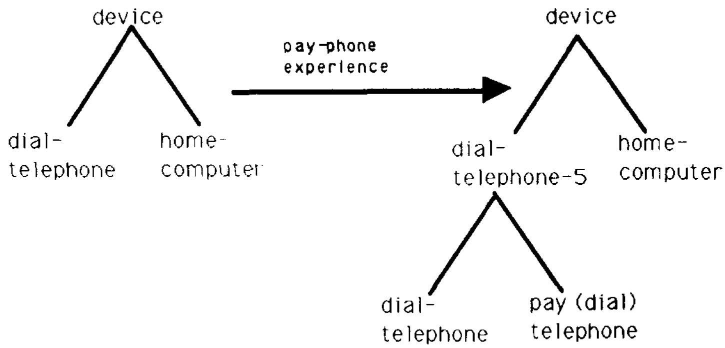

FIGURE 15 Ad hoc category creation.

After restarting, execution of telephone plan proceeds without further problems until the newly inserted step is encountered. FLOABN finds that further decisions need to be made because INSERT-COIN has a failing precondition that a coin be in hand. The context of execution helps FLOABN find an appropriate adaptation. FLOABN combines two indices—the failing precondition and the environmental feature POCKET—to retrieve the step GET-COIN-FROM-POCKET. This step is inserted, and the plan is completed.

This pay telephone episode, including the experience of receiving the instructions, is stored in episodic memory and incorporated into the plan. If FLOABN encounters this message in the future, and the appropriate reminding occurs, it will not need to reinterpret the message because it already knows what the message is telling it to do.

During interaction, an ad hoc category representing the functional commonality between the dial and pay telephones is created. This is due to an adaptation occurring at the start of the pay-telephone episode. FLOABN's TELEPHONE-PLAN initially knows only about desk-style telephones (DIAL-TELEPHONE). Because there is no dial telephone handy, it must adapt some other telephone-like object to satisfy the precondition of TELEPHONE-PLAN that a DIAL-TELEPHONE exist; it finds PAY-TELEPHONE-004.[4] It binds PAY-TELEPHONE-004 to DIAL-TELEPHONE for the remainder of execution and incorporates this as a change to the plan. This binding prompts the learning component to construct a new telephone category with DIAL-TELEPHONE as one subcategory plus a second, new subcategory representing pay telephones, the latter based on exemplar PAY-TELEPHONE-004 (see Figure 15). This

[4]We assume a perceptual apparatus adequate to identify PAY-TELEPHONE-004 as an object; however, the connection of PAY-TELEPHONE-004 to the DIAL-TELEPHONE concept does not occur until the need for a DIAL-TELEPHONE substitute is discovered.

"ad-hoc category" facilitates transfer of telephone knowledge in future episodes because it suggests that for the purposes of telephone plans, dial and pay (dial) telephones are substitutable.

CONCLUSION

This article is concerned with the problem that within a community of agents practice is in constant flux. From the perspective of an individual agent, to continue to operate in an effective manner, continuous learning must occur. The approach detailed here focuses on the role of comprehension as a blend of memory and interaction.

An example of a domain of activity where this problem is present is in the usage of mechanical and electronic devices. Current technology moves at such a rapid pace that it can easily outdistance the community's ability to absorb new device functions into common practice. As an example, current photocopier technology delivers an impressive array of functions, but a problem exists in encouraging the average office worker to learn the various capabilities of that device (e.g., Suchman, 1987).

A fundamental question to ask is: What is the nature of the agent's practice of such devices? There are several reasons to ask this question. One reason is that if we know how an agent approaches and uses new devices then we are one step closer to building devices that present themselves in a revealing fashion (e.g., Norman, 1988). A second reason is that such a model also brings us a step closer to the technology that will increase the rate of transfer of practice. A third motivation is that the usage of new technology is a miniature of the larger question of accounting for the cognitive aspects of mutating social practice.

The first part of this article is a somewhat self-conscious attempt to be precise and concrete about what is meant by such inherently vague notions like making sense or constructing an interpretation. Some have argued that the concretion function of computational models is their primary role in the cognitive sciences, for example, Hayes (1975) and Dennett (1988). Much of the analysis of this part was culled from contemporary usage in the text comprehension community; our task was to convert these sorts of assumptions into a specific computational framework for interactive comprehension-based problem solving.

The constructive function of understanding during engagement results in a representation, in the language provided by semantic memory. The constructed representation makes sense because it relates to the immediate circumstances (correspondence), is internally consistent (coherence), and is partially determined by the agent's current goal. A key idea is that the construction of an interpretation leads to the selection of a known routine.

The second part of the article expands this framework to account for the usage of instructions. The goal of this part is not so much to present a complete theory of instruction usage in its many forms, but to build a solid computational foundation. The model we describe used instructions only after the interaction had begun. There were three reasons for this design/model decision: (a) Interaction with new device technology is largely made up of familiar actions, and for the average user, much of the instruction is redundant and repetitive; (b) by beginning to engage in the situation, the agent creates a context for deciphering the meaning of the instructions; and (c) instructions frequently refer to plans or steps in plans that the agent is presumed to be using.

Because the problem-solving task of determining new practice was focused largely as a comprehension task, we were half way toward tying the reading of the instructions to the ongoing activity.

There are two phases of instruction usage. The first phase establishes the relevance of the instructions, and makes sense of them. The second phase proceduralizes the initial understanding and deals with details of, for example, when to precisely take a given action or how to resume action from an impasse.

One of the lessons to draw from this article is the nature of skill as social practice. Because comprehension is central to the determination of new practice and because practice is constantly in flux, skill and its acquisition are not just the transfer of declarative knowledge to procedural form; rather, they also include the constant acquisition, maintenance, and usage of semantic knowledge. In other words, from the perspective of FLOABN, to acquire a skill means not only to acquire a set of procedures, but also to acquire the semantic knowledge that supports the usage, deployment, and adaptation of these procedures during the engagement. The importance of the acquisition of this semantic knowledge is especially apparent in the use of instructions; instruction usage is predicated on the assumption that the agent has acquired a sufficient set of understandings to serve as a basis for extracting the procedural content from the instructions. However, even without any explicit instructions, in a community of agents that shares a common set of practices, it is the shared set of concepts, stored by each individual in his semantic memory, that provide crucial support to an individual agent in the identification of a course of action while negotiating his way through a slowly, but constantly, mutating world.

ACKNOWLEDGMENTS

Research for this article was supported in part by the Defense Advanced Research Projects Agency, administered by the U.S. Air Force Office of Scientific Research under contract F49620-88-C-0058.

We thank Jim Hendler, Alan Lesgold, Suzanne Mannes, and Ashwin Ram for their comments on earlier versions of this article. We also thank Janet Kolodner for her editorial advice. We take full responsibility for the contents.

REFERENCES

Agre, P. E. (1988). *The dynamic structure of everyday life* (Tech. Rep. No. AI-TR 1085). Cambridge, MA: MIT Artificial Intelligence Laboratory.

Agre, P., & Shrager, J. (1990). Routine evolution as the microgenetic basis of skill acquisition. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 694-701). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Alterman, R. (1985). A dictionary based on concept coherence. *Artificial Intelligence, 25,* 153-186.

Alterman, R. (1988). Adaptive planning. *Cognitive Science, 12,* 393-421.

Alterman, R. (1989). Event concept coherence. In D. Waltz (Ed.), *Semantic structures: Advances in natural language processing* (pp. 57-87). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Alterman, R. (1990). An adaptive planner. In J. Allen, J. Hendler, & A. Tate (Eds.), *Readings in planning* (pp. 660-664). Los Altos, CA: Kaufmann.

Alterman, R., & Bookman, L. A. (in press). Reasoning about a semantic memory encoding of the connectivity of events. *Cognitive Science Journal, 16*(2).

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Badler, N., Webber, B. Kalita, J., & Esakov, J. (1990). Animation from instructions. In N. Badler, B. Barsky, & D. Zeltzer (Eds.), *Making them move: Mechanics, control and animation of articulated figures* (pp. 51-93). Los Altos, CA: Kaufmann.

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge University Press.

Black, J., & Bower, G. (1980). Story understanding as problem solving. *Poetics, 9,* 223-250.

Brooks, R. A. (1985). *A robust layered control system for a mobile robot* (Tech. Rep. AI Memo 864). Cambridge, MA: MIT Artificial Intelligence Laboratory.

Carbonell, J. (1978). Politics: Automated ideological reasoning. *Cognitive Science, 1,* 27-51.

Carpenter, T., & Alterman, R. (1991). Reading instructions. In *The Thirteenth Annual Meeting of the Cognitive Science Society* (pp. 653-657). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Chapman, D. (1990). *Vision, instruction, and action* (Tech. Rep. No. AI-TR 1024). Cambridge, MA: MIT Artificial Intelligence Laboratory.

Charniak, E. (1986). A neat theory of marker passing. In *Proceedings of the Fifth National Conference on Artificial Intelligence* (pp. 584-588). Los Altos, CA: Kaufmann.

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407-428.

Dennett, D. C. (1988). When philosophers encounter artificial intelligence. *Daedalus—Journal of the American Academy of Arts and Sciences, 117,* 283-296.

Geertz, C. (1983). *Local knowledge: Further essays in interpretive anthropology.* New York: Basic.

Goodman, B. A. (1986). Reference identification and reference identification failures. *Computational Linguistics, 12*(4), 273-305.

Graesser, A., & Clark, L. (1985). *Structure and procedures of implicit knowledge.* Norwood, NJ: Ablex.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and* *semantics 3: Speech acts* (pp. 41-58). New York: Academic.

Grosz, B., & Sidner, C. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics, 12*(3), 175-204.

Hammond, K. J. (1990). Case-based planning: A framework for planning from experience. *Cognitive Science, 14,* 385-443.

Hayes, P. J. (1975). Computer programming as a cognitive paradigm. *Nature, 254,* 563-566.

Hendler, J. (1987). *Integrating marker-passing and problem-solving: A spreading activation approach to improved choice in planning.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hidi, S., & Baird, W. (1986). Interestingness—A neglected variable in discourse processing. *Cognitive Science, 10*(2), 179-194.

Kieras, D. E., & Bovair, S. (1986). The acquisition of procedures from text: A production-system analysis of transfer of training. *Journal of Memory and Language, 25,* 507-524.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182.

Kintsch, W., & van Dijk, T. (1983). *Strategies of discourse comprehension.* New York: Academic.

Kolodner, J. L. (1983). Maintaining organization in a dynamic long-term memory. *Cognitive Science, 7,* 243-280.

Kolodner, J. (1988). Retrieving events from a case memory: A parallel implementation. In J. Kolodner (Ed.), *Case-based Reasoning Workshop* (pp. 233-249). San Mateo, CA: Kaufmann.

Kolodner, J. L., & Simpson, R. L. (1989). The MEDIATOR: Analysis of an early case-based problem solver. *Cognitive Science, 13,* 507-549.

Laird, J. E., Rosenbloom, P., & Newell, A. (1986). Chunking in SOAR: The anatomy of a general learning mechanism. *Machine Learning, 1,* 11-46.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the world.* Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lave, J. (1988). *Cognition and practice.* Cambridge, England: Cambridge University Press.

LeFevre, J. J., & Dixon, P. (1986). Do written instructions need examples? *Cognition and Instruction, 3,* 1-30.

Mannes, S. M., & Kintsch, W. (1991). Routine computing tasks: Planning as understanding. *Cognitive Science, 15*(4), 305-342.

Martin, C. E., & Riesbeck, C. K. (1986). Uniform parsing and inferencing for learning. In *Proceedings of the Fifth National Conference on Artificial Intelligence* (pp. 257-261). Los Altos, CA: Kaufmann.

Martin, J. H. (1990). *A computational model of metaphor interpretation.* San Diego, CA: Academic.

Miller, G., & Johnson-Laird, P. (1976). *Language and perception.* Cambridge, MA: Harvard University Press.

Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The Psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.

Neisser, U. (1978). Memory: What are the important questions. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory.* London: Academic.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Nicholson, G. (1984). *Seeing and reading.* Atlantic Heights, NJ: Humanities Press International.

Norman, D. A. (1988). *The psychology of everyday things.* New York: Basic.

Norman, D., & Bobrow, D. (1979). Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology, 11,* 107-123.

Norvig, P. (1989). Marker passing as a weak method for text inferencing. *Cognitive Science, 13*(4), 569–620.

Quillian, R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 227–270). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ram, A. (1990). Knowledge goals: A theory of interestingness. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 206–214). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Rissland, E., & Ashley, K. (1986). Hypotheticals as heuristic device. In *Proceedings of the Fifth National Conference on Artificial Intelligence.* Los Altos, CA: Kaufmann.

Ross, B. H. (1989). Remindings in learning and instruction. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 438–469). Cambridge: Cambridge University Press.

Rumelhart, D. (1980). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sacerdoti, E. (1974). Planning in a hierarchy of abstraction spaces. *Artificial Intelligence, 5,* 115–135.

Schank, R. (1982). *Dynamic memory: A theory of reminding and learning in computers and people.* Cambridge: Cambridge University Press.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Simmons, R. (1973). Semantic networks: Their computation and use for understanding English sentences. In R. Schank & K. Colby (Eds.), *Computer models of thought and language* (pp. 63–113). San Francisco, CA: Freeman.

Stanfill, C., & Waltz, D. (1986). Toward memory-based reasoning. *Communications of the ACM, 29*(12), 1213–1239.

Suchman, L. A. (1987). *Plans and situated actions.* Cambridge: Cambridge University Press.

Trabasso, T., & Sperry, L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language, 24,* 595–611.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and representation of narrative events. *Journal of Memory and Language, 24,* 612–630.

Tversky, B., & Hemenway, D. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113,* 169–193.

Vere, S., & Bickmore, T. (1990). A basic agent. *Computational Intelligence, 6,* 41–60.

Wilensky, R. (1981). Meta-planning: Representing and using knowledge about planning in problem solving and natural language understanding. *Cognitive Science Journal, 5,* 197–233.

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

Wilenksy, R., Chin, D. N., Luria, M., Martin, J., Mayfield, J., & Wu, D. (1989). *The Berkeley UNIX consultant project* (Tech. Rep. No. UCB/CSD 89/520). Berkeley: University of California.

Winograd, T. (1972). *Understanding natural language.* New York: Academic.

Woolf, B. (1988). Intelligent tutoring systems: A survey. In H. E. Shrobe (Ed.), *Exploring artificial intelligence* (pp. 1–43). San Mateo, CA: Kaufmann.

Zito-Wolf, R., & Alterman, R. (1991). *The representation and use of a body of related cases: Continuous acquisition and evolution.* Unpublished manuscript.

APPENDIX

A Trace of FLOABN for the Telephone-Call Example

The following trace shows the details of FLOABN's operation on the task of making a local call in a situation offering only a pay phone. FLOABN begins by attempting to apply its existing telephone-call plan. The first significant event is the adaptation that results in the ad hoc "telephone" category [discussed in section Interaction Resumes (Stage 4)]:

```
FLOABN: starting new run situation = PAYPHONE
   plan = TELEPHONE = PLAN
Working on step TELEPHONE-PLAN [node =#<PNODE P90>]
  Precondition (EXIST DIAL-TELEPHONE) is failing for
     TELEPHONE-PLAN
FIND-PLAN called to fix failing precs ((EXIST
   DIAL-TELEPHONE))
FIND-PLAN found plans NIL
     Can neither delete TELEPHONE-PLAN nor reorder steps
     Will try abstraction and specialization
       (Trying to substitute for feature DIAL-TELEPHONE)
Generated candidate items
  ((PAY-DIAL-TELEPHONE-004 200) (POCKET 0) (SLOT-004 0)
    (SWITCH-HOOK-LEVER-004 0) (PHONE-DIAL-004 0)
    (TELEPHONE-RECEIVER-004 0)) for DIAL-TELEPHONE at
       detail level 1
Substituted feature PAY-DIAL-TELEPHONE-004 for
   DIAL-TELEPHONE based on similarities
```

FLOABN then executes the steps PICK-UP-RECEIVER, LISTEN-FOR-DIAL-TONE, and DIAL without difficulty. Then, while waiting for the ringing tone, a message is heard:

```
Working on step WAIT-FOR-RING [Node =#<PNODE P128>]
Expecting EVENT WAIT-FOR-RING to take   0.0(+ - 0.0) seconds
Firing rule WAIT-FOR-RING

* Message heard: "The call you have made requires an initial deposit.
Please hang up momentarily. Listen for dial tone, deposit the required
coin, and dial your call again."
```

FLOABN passes the message to the text interpreter and then proceduralizes the result, identifying a new step to be inserted in the plan.

Details of this process can be found in the section Instruction Proceduralization (Stage 3.) After interpretation is complete, a new step is inserted:

```
trying to interpret LOCAL-RESP
IIMP returning ((HANG-UP:TYPE INSTRUCTION:ISA
      RESTART-CALL)
                  (LISTEN-FOR-DIAL-TONE:TYPE INSTRUC-
                    TION:STEP-OF TELEPHONE-PLAN)
                  (INSERT-COIN:TYPE INSTRUCTION:WITH
                    ((VALUE * 10)))
                  (DIAL:TYPE INSTRUCTION:STEP-OF
                    TELEPHONE-PLAN:MODIFIERS ((AGAIN)))
as interpretation of (PERCEPT
                    HEAR
                    LOCAL-RESP
                    "The call you have made requires an initial
                    deposit. Please hang up momentarily. Listen for
                    dial tone, deposit the required coin, and dial
                    your call again.")

Applying instruction pattern REF-TO-PRIOR-PLAN:
Interpreted instructions as refs to in-use plan TELEPHONE-PLAN:
  refs=(1 3) new instrs=(2) agains=(3)
Steps 1 [(LISTEN-FOR-DIAL-TONE :TYPE INSTRUCTION
   :STEP-OF TELEPHONE-PLAN)]
     and after are reference marks into prior plan
        TELEPHONE-PLAN
This interpretation is supported by "again" refs in steps (3)

Step HANG-UP is part of restarting TELEPHONE-PLAN and will
   be executed now
Step INSERT-COIN is new and will be inserted into
   TELEPHONE-PLAN.
Inserting step INSERT-COIN before step DIAL of plan
   TELEPHONE-PLAN.
```

FLOABN then resumes the interaction, by first executing the immediate instruction HANG-UP and then restarting TELEPHONE-PLAN:

```
*** Will run step HANG-UP immediately ***

Working on step HANG-UP [Node =#<PNODE P130>]
```

"I HANG-UP"
Expecting STEP HANG-UP to take 4.5(+ – 0.0) seconds
Firing rule HANG-UP02 ; embodied skill RAPs
Firing rule HANG-UP02
Firing rule HANG-UP01
EXECUTE-1 (HANG-UP) returned T

* TELEPHONE-RECEIVER-004 has been replaced on phone
STEP HANG-UP actually used 3.3 seconds
(No failing outcome)

*** Will run step (RESTART-NODE WITH ((NODE #<PNODE P90>))) immediately ***
Working on step RESTART-NODE (Restart a plan) [Node =#<PNODE P131>]
"I RESTART-NODE #<PNODE P90>"
Preparing to restart Step TELEPHONE-PLAN from node #<PNODE P90>
(No failing outcome)

Finished adapting steps of RESTART-NODE
Restarting Step TELEPHONE-PLAN from node #<PNODE P90> as node #<PNODE P132>
Working on step TELEPHONE-PLAN [Node =#<PNODE P132>]

The TELEPHONE-PLAN proceeds with one further adaptation, the insertion of a step to procure a dime to be inserted into the phone:

Working on step LOOKUP-NUMBER [Node =#<PNODE P135>]
Step LOOKUP-NUMBER can be skipped because outcomes ((EXIST-VALUE DESIRED NUMBER)) already hold.
Before step (DIAL) I will run step INSERT-COIN

Working on step INSERT-COIN [Node =#<PNODE P136>]
Precondition (HOLDING COIN) is failing for INSERT-COIN
FIND-PLAN called to fix failing precs ((HOLDING COIN))
FIND-PLAN found plans ((GET-COIN-FROM-POCKET 0 NIL))
Step GET-GOIN-FROM-POCKET being inserted before INSERT-COIN
Next step is an inserted step GET-COIN-FROM-POCKET.

Working on step GET-COIN-FROM-POCKET [Node =#<PNODE P137>]
"I GET-COIN-FROM-POCKET the POCKET"
Firing rule GET-COIN-FROM-POCKET
EXECUTE-1 (GET-COIN-FROM-POCKET) returned T

* Moved hand RIGHT-HAND to POCKET
* Getting coin COIN-001 from POCKET
PLAN GET-COIN-FROM-POCKET actually used 2.9 seconds
    (No failing outcome)

Working on step INSERT-COIN [Node =#<PNODE P138>]
"I INSERT COIN
Assuming value of AMOUNT is 10 as suggested by instructions.
Firing rule INSERT-COIN
EXECUTE-1 (INSERT-COIN) returned T

*Inserting coin COIN-001 into slot SLOT-004 of phone
  PAY-DIAL-TELEPHONE-004.
PLAN-INSERT-COIN actually used 2.5 seconds
    (No failing outcome)
Working on step DIAL [Node =#<PNODE P139>]
"I DIAL"

Execution completes without further incident.