Understanding actions in relation to goals

Carolyn L. Foss and Gordon H. Bower

INTRODUCTION

We assume that people understand the actions of others by viewing those actions as purposive, as goal-directed. People use their knowledge of human intentionality, of the types of goals people have and of the types of plans they devise in service of those goals, to understand action sequences that are described in narratives or observed directly. Many recent approaches to comprehension emphasize the role of goal-planning knowledge when understanding narratives (e.g. Cullingford, 1978; Schank & Abelson, 1977; Wilensky, 1978, 1983) and conversations (e.g. Perault, Allen, & Cohen, 1978; Schank et al., 1982), and when remembering observed sequences and goal-directed actions (Lichtenstein & Brewer, 1980; Brewer & Dupree, 1983).

According to these approaches, understanding involves inferring the intentions (i.e. the plans and goals) of the characters, speakers, or actors. Such inferences are ubiquitous because narratives frequently provide only sketchy descriptions of the character's actions and goals; speakers rarely state their intentions directly; and observers rarely see all the events preceding or following the action to be explained. Therefore, people are forced to use their general knowledge of human intentionality to fill in the missing information; they do this by generating expectations and drawing inferences in order to come up with a plan that explains an actor's behavior. Although the importance of this type of knowledge for understanding natural discourse and action sequences has been recognized, only recently have cognitive scientists begun examining the psychological processes involved in drawing inferences about human intentionality. In this paper we examine what people know about goal-action relationships, and how they retrieve this knowledge to make sense of goal-directed actions.

We will assume that information about plans, goals, and actions is retrieved in much the same way as is other information from memory. The classic experiments on retrieval of information from semantic memory assume that information is organized in a semantic network of concepts and relations, with certain concepts organized hierarchically (e.g. Collins & Loftus, 1975; Holyoak & Glass, 1975). For example, properties are stored at the most general level possible. Thus, in an animal hierarchy where a canary is a bird which is an animal, a property like "has skin" would be stored only at the most superordinate level (attached to "animal" only), whereas a unique, distinguishing property like "canaries are yellow" would be directly attached to the subconcept of canary. The model assumes that to answer such queries as "Do canaries have skin?" a person derives the relationship between the subject and predicate by retrieving and combining individual facts. Thus, because a canary is a bird, and a bird is an animal, and animals have skin, it follows that a canary has skin. Allegedly, the longer the derivation, the more elementary facts that have to be retrieved and combined to connect the subject and predicate, and the longer the time people need to answer such questions.

This paper concerns the connecting-up processes involved in understanding actions described in narratives. Although we will deal exclusively with subjects understanding actions stated in text, we assume that similar processes occur when subjects comprehend actions they observe directly (see Lichtenstein & Brewer, 1980). A plausible proposal is that similar linking-up processes will be involved in interpreting the goal-directed actions of real-life actors as well as in filling in missing information while trying to understand events described in narratives. In both cases understanding involves generating explanations and predictions related to the actor's plans.

When understanding narratives, an important step in generating predictions is to know or guess a character's goals, because these can predict actions. For instance, if a reader knows that a character's goal is to go on an overseas trip, then actions such as packing or going to the airport are expected and readily understood. Similarly, actions can imply their underlying motives. We try to explain an action by linking it to one of the actor's known goals or to a goal-generating theme or role for that actor (see Schank & Abelson, 1977, for a detailed discussion of this approach to comprehension).

Let us briefly consider what people might know about goal relationships and how that information may be organized. Goals are typically arranged in a hierarchy of subgoals. Goals and subgoals can be represented in a goal reduction tree, which is constructed by decomposing a top-level goal into a conjunction or disjunction of subgoals which may in turn be decomposed further. The decomposition ends when only immediately attainable goals (or executable actions) are left. The goals, subgoals, and actions so generated may be represented as nodes in a tree linked by branches. The relationships among goals are revealed by their placement along the branches. For any given node, we can designate all its ancestors (items
above and on the same branch) as “goals”, and all its descendants (items below the node and on the same branch) as “actions”. Thus, any node in the tree, except the top node and the terminal nodes, can be a “goal” or an “action”, depending upon whether it is viewed from “below” or “above”, respectively.

![Diagram of a goal reduction tree](image)

**Fig. 1** — An example of goal reduction tree, also referred to as a goal hierarchy.

To illustrate, the top goal of the goal reduction tree schematized in Fig. 1 is to steal money. For the person whose limited knowledge is represented by this tree, this top goal can be achieved by either of three means or subgoals. To steal money, he believes only that he can either embezzle or rob or burglarize. Nodes that stem from such or branches are not preconditions or subgoals for one another. For instance, performing a robbery is not a precondition for embezzlement. The vertical links in the tree represent the “is a subgoal of” relationship. This relationship is transitive for nodes that lie along a single vertical branch; this relationship thus imposes a partial ordering on to the nodes. For example, “fixing the books” is a subgoal of the top goal “stealing money” because it is directly linked to “embezzle” which in turn is directly linked to “stealing money”.

We began our investigations with the intuition that goals and actions would vary in their “psychological distance” from one another. We will coordinate this notion of “psychological distance” to the distance in the goal-reduction tree (the number of links) separating the action from the goal state. Notice in Fig. 1 that the same action (e.g. getting a gun) can be near to one goal (e.g. finding a weapon), but far from another goal (e.g. stealing money) that lies along one branch of a goal-reduction tree.

We hypothesize further that people will often understand the relationship between a goal and an action by finding, retrieving, or computing a link between the intervening subgoals. For instance, understanding a sentence like, “John wanted to go overseas; so he caught a bus” involves guessing that John is probably taking a bus to the airport so he can catch a flight overseas.

Wilensky (1978, 1983) created a computer simulation program that understands simple plan-based episodes by applying knowledge about goals and actions and how they fit in with plans. Although no experimental evidence was offered for the program as a psychological model, it does produce explanations (answers to why questions about human-action episodes) somewhat similar to those of human understanders. Wilensky’s story understanding program, called PAM for Plan Applier Mechanism, follows a specific algorithm for understanding each action as it occurs in a story. The program begins by using the first few lines to guess at a plan or goal for the actor. If a goal is stated, then an action which will lead to that goal is expected to follow in the next few story statements. If no goal is stated, then it is assumed that the statement will suggest a plan, so that the plan’s goal will be inferred. If actions or goals mentioned in the next few story statements match the inferred plan, then the plan is taken to be the explanation for those actions. Consider these statements:

1. The company accountant wanted to steal some money.
2. He fixed the company’s books.

The first step in understanding (2) would be to infer a plan from the action (fix company’s books). A guess is that the plan is to embezzle from the company. The plan is instrumental to a known goal, of stealing money. Thus, the explanation of the action is assumed to be that the accountant fixed the books because he wanted to embezzle from the company.

Some episodes are more difficult to understand than others because they may contain actions or goals that are difficult to incorporate into any plan. One way this difficulty arises is when the goals and actions in an episode are psychologically distant and so require many inferences to associate them. If understanding an action involves linking it with a known goal, then we would expect comprehension of an action statement to take longer the greater the psychological distance between the goal and action. In terms of our goal reduction tree in Fig. 1, we may conceive of the listener linking up the action (e.g. “fix books”) to the stated goal (e.g. “steal money”) by a memory search process. The greater the link distance, the longer the search for an intersection point in the tree, and hence the longer it will take the reader to understand the action in the light of the goal. The same reasoning would apply if the action is stated first and we measure how long the reader requires to understand the goal statement.

In the experiments below, we set out to find such “distance effects” in people’s comprehension of actions in the light of the actor’s goals. We are not the first ones to search for a “distance” effect in comprehension of
narratives. Bower, Black, and Turner (1979, Exp. 6) used stereotypic, script-based narratives, and measured how quickly a critical script-action description was read as a function of its hypothetical "distance" from the script-action mentioned in the prior sentence. The distance was either one, two, or three steps back in the underlying script. For example, a restaurant story might have the target action sentence: John ordered lamb chops. Its reading time would be measured immediately following the reading of a context sentence, either (a), (b), or (c):

(a) He decided what he wanted to eat.
(b) He took up the menu.
(c) He sat down at a table.

In the underlying script, the target action is closest to event (a), intermediate to event (b), and farthest from event (c), both in terms of temporal distance and in terms of chains of causes or enablements. Although they expected a monotonic distance effect on the basis of spreading activation from the context event to the target event, Bower et al. observed a non-monotonic trend: target actions at the intermediate distance were understood more slowly than the closest actions, but the farthest actions were as fast as the closest actions. Thus, the "distance" hypothesis received only equivocal support.

Abbott, Black, and Smith (1985, Exp. 2) hypothesized that the non-monotonic distance effect observed by Bower et al. could have arisen from the way they constructed their script narratives. Specifically, scripts are hierarchically structured, so that certain actions name larger, higher chunks of the script ("Script headers" such as "the customer ordered"), whereas other actions specify lower, more detailed aspects of the scene ("actions" like "the customer read the menu"). Some of the script narratives written by Bower et al. used target and context actions that came from different levels of the script hierarchy. Such texts are stylistically somewhat "choppy" and require the reader to shift levels, and this may have obscured a simple distance effect. When Abbot et al. repeated the Bower et al. experiment with script actions in a narrative at the same level (all scene headers, or all detailed actions), a monotonic distance effect was observed: a given target action took longer to comprehend the greater the scripted distance ("gap") between it and the prior action mentioned in the text.

Recall that we are interested in how rapidly an action is understood in the light of its "psychological distance" from its goal. Thus, the aforementioned script distance studies are not directly relevant, since early events in a scriptal sequence only rarely name the goal of a later event (action). For example, "He sat down" is not a goal satisfied by "The waitress took his order". Rather, the earlier context statement may refer to a temporally preceding event and/or an enabling precondition of the target event. Thus, an experiment is needed in which explicit goal-action distances are investigated.

A directly relevant study was reported at the 1981 Cognitive Science Conference by Smith and Collins (1981). They studied the time subjects took to understand an action that seemed near or far to an earlier stated (or implied) goal. An example of their material is:

John needed money to pay for his wife's operation.
(a) He decided to borrow money from his Uncle Harry.
(b) He would give his Uncle Harry a quick call.
(c) He had to find out Uncle Harry's phone number.
John reached for the most recent suburban directory.

Subject's reading times were measured for the first sentence, then for either (a), (b), or (c), and then for the final sentence. Based on a goal-subgoal analysis, the authors note that sentence (a) is close to the top goal (need money) but far from the last sentence (get directory); that sentence (b) is intermediate; and that sentence (c) is far from the top goal but near to the final sentence. Smith and Collins predicted that reading time for the second sentence would be faster the "nearer" its action was to the top goal sentence, i.e. (a) fastest, (c) slowest, but that the reading time for the final common sentence (get directory) would show the reverse trend, i.e. slowest after (a), fastest after (c). They reported that their results generally confirmed these predictions, whether the story statements emphasized the character's intentions (like "decided" or "would give") or emphasized his actions (e.g. "he called" or "he asked for a loan"). Thus, these results support the goal-action distance hypothesis that we stated in the Introduction.

Our experiments, which were completed before the Smith and Collins report, avoid one inelegancy of the Smith and Collins design, and attempt to extend the distance effect to new domains. The inelegancy of their design is that the critical "gap size" variable was manipulated by varying which of three target sentences (a), (b), or (c) subjects read, and so the actual target material was confounded with the gap size. This "materials confound" may have created problems, since Smith and Collins noted that the magnitude of the gap-size effect varied considerably across different lines and episodes. In our experiments, the crucial target sentence was fixed within a given narrative and the prior context-sentences were varied, so that the distance variable was not confounded with the actual target sentence.

The first experiment examined how distance between goals and actions in natural episodes about familiar situations affects comprehension. We worried about a potential problem with the use of naturalistic materials, however. "Distance" is not always easy to estimate for familiar plans because of people's differing experiences; therefore, the second experiment drew on subject's knowledge about a newly learned artificial goal hierarchy, for which we could calculate the distance between any goal and action. Experiment 3 returned to a more naturalistic reading procedure: subjects first casually read over an actual folktales, and then performed a goal-action verification task similar to that in Experiment 2.
EXPERIMENT 1

Wilensky's (1983) story understanding algorithm is a useful hypothesis about the steps people follow when filling in missing information in order to understand narratives. If the relationships between story events are not stated, the reader has to infer the necessary connections. Readers search for explanations of story events by successively matching actors' goals to their actions. If an action does not fit an inferred goal, then it is necessary either to establish a new goal for the character, to redefine the character's old goal, or to decompose an old goal into one or more subgoals that may match the action better. Different goal-action pairs may require differing numbers of steps to link them together. Consider these goal-action sequences:

(3) A far pair
   Goal: Marge wanted to stop construction of a nuclear power plant.
   Action: She made a sign.

(4) A near pair
   Goal: Marge wanted to participate in an anti-nuke protest rally.
   Action: She made a sign.

Compared to the far goal-action sequence, the near pair would probably require fewer inferential steps to integrate. Some sequences will be easier to understand than others depending on the transparency of the goals and subgoals. Sequences with clearly stated or easily inferred goals that lead to close ('obvious') actions should be relatively easy to comprehend.

Our hypothesis is that comprehension will be more difficult the greater the number of intervening subgoals between the stated goal and its actions. This will be called the distance hypothesis, where 'distance' refers to the number of subgoals in a chain of subordinates separating a goal and an action. Distance will be used here in a comparative manner, with comparisons restricted to two nested goals and an action, all of which lie along the same branch of a goal reduction tree. Hence, the second sequence in the example is near (closer) compared to the first (far) sequence.

If the two goal-action sequences in the example were charted in a goal reduction tree, the goal (stop construction) would be at the top; below that would be several subgoals, including goal 2 (participate in rally). The action (make a sign) would lie further along the same branch; hence, the action is a descendant of both goals. Since goal 1 is a greater distance in the tree from the action, the distance hypothesis predicts that this sequence would take longer to understand than the goal 2-action sequence. The distance hypothesis will be tested by constructing a collection of short episodes that contain both a far event sequence and a near subgoal sequence nested within it. As in the example, we tried to select sequences of top goal-subgoal-action that would be nested along one branch of a goal reduction tree, so that the near event sequences (subgoal-action) were embedded within the far sequences (top goal-action).

So far we have discussed the case where the goal is stated first, followed by the action. But a distance effect should also occur in situations where the action is mentioned before the goal is stated, as in sequences such as "John caught a bus; he wanted to go overseas." The same linking up processes should be taking place, so we expect that far action-goal sequences will take less time to understand than near action-goal sequences. However, the goal-then-action sequences may be understood quicker than its action-then-goal mate, since the former follows the canonical order of antecedent-then-consequent in most causal schemata.

The distance hypothesis was tested in Experiment 1 with a set of episodes written to exemplify near and far event sequences. We measured reading time for a critical goal or action statement when it was embedded in a near or far episode.

Method
We constructed a base of 48 episodes similar to this example:†

   Opening (O): Marge was concerned about the environment.
   Goal (G): She wanted to stop construction of a power plant.
   Subgoal (S): She participated in an anti-nuclear rally.
   Action (A): She made a sign.
   Closing (C): Marge believed atomic power plants are harmful.

Every episode presented began with an opening, ended with a closing, and presented two of the three G, S, or A statements in some order. The critical statement whose reading time was measured was always located in the third position. Four experimental conditions were created by varying the second (context) statement in the episode and its relation to this critical third statement. For 24 of the 48 base episodes, the critical third sentence was always the action, A (e.g. "She made a sign"), which was preceded by its near goal (S) half the time and by its far goal (G) the other half of the time. For the other 24 base episodes, the critical third sentence was always the goal, G (e.g. "She wanted to stop construction of a power plant"), which was preceded by its far action (A) half the time and by its near action (S) otherwise. With respect to the critical third sentence, these materials permitted us to record twelve reading times for each of four episode types—near goal-action, far goal-action, near action-goal, far action-goal.

Subjects read the episodes one at a time on a computer screen by pressing a button that cleared the screen and caused the next sentence in the episode to appear. They were told to read each line in the episode, and to push the button when they were sure they understood the sentence; they were not told that their reading time was being measured. They were also told to read each episode carefully, in preparation for a comprehension test to follow after reading the entire set of episodes. The subjects were 14 Stanford University students who participated in the experiment for extra credit in their introductory psychology class. Each subject was run individually; the session lasted 45 minutes.

† The materials used in these experiments were obtained by the first author.
Results and discussion

The critical measure is the reading time on the third sentence of each four-line episode. To check for possible performance differences between the two sets of episodes, an overall subject by episode set analysis of variance (ANOVA) was performed on the average reading times. Since no reliable differences were found, the data from the two sets of episodes were combined for the following analyses.

Our first question is whether subjects take longer to comprehend action statements when they are preceded by "distant" goals. The results, shown in Fig. 2, reveal that target sentences were indeed easier to comprehend when preceded by near rather than far context sentences. A three-way ANOVA, with factors of subjects, distance (near or far), and episode-type (goal-action or action-goal), was performed on the target comprehension times. This yielded a significant main effect for distance ($F(1, 12) = 13.34, p < 0.01$). As illustrated in Fig. 2, the differences in reading time were in the predicted direction for all conditions: when preceded by a near versus far goal, reading time for an action target sentence was 1.80 sec. versus 2.10 sec., respectively. Similarly, a goal target sentence was read faster when it followed a near action (2.19 sec.) rather than a far action (2.61 sec.). Each

mean cited is based on approximately 168 observations.

The distance hypothesis was supported by these data. Subjects needed more time to integrate a target sentence into an episode as its inferential distance from the context sentence increased. This was true for both goal–action and action–goal episodes. That is, it took more time to integrate an action target statement when it was preceded by a far goal rather than a near goal. Similarly, more time was needed to understand a goal target statement when it was preceded by a far rather than near action. This suggests that subjects understood the episodes by inferring plausible connections between the context and target statements. Our hypothesis says that this reading time was quicker for near episodes because fewer inferential steps were needed to relate the goals and actions to each other.

As can be seen in Fig. 2, target sentences were read faster in goal–action episodes (mean = 1.95 sec.) than in action–goal episodes (mean = 2.40 sec.), $F(1, 12) = 16.99, p < 0.001$. Furthermore, examining reading time for the second or context sentences revealed that they were read faster in goal–action episodes (mean = 2.15 sec.) than in action–goal episodes (mean = 2.39 sec.), $F(1, 12) = 6.79, p < 0.02$. We may relate this order effect to the fact that goals are typically followed by actions in episodes. For example: "John wanted to go downtown, he caught a bus" is more conventional in text than the reverse order. This order also follows the rule of "tell the cause first, then the effect." This may explain why episodes containing action-goal sequences are read slower. This effect was probably heightened by the "one-line-at-a-time" presentation, since that forced subjects to comprehend each sentence as it appeared on the screen. Normally, unusually ordered statements in an episode can be re-read until their relationship is uncovered. As expected, there were no significant effects for distance in the context sentences. In other words, time to read the goal statement, for goal–action episodes and time to read the action statement, for action–goal episodes, was not reliably affected by being near to or far from the target statement which followed. Similarly, the reading times for opening and closing statements were not affected by distance (near or far) or episode-type (goal–action of action–goal) nor was there a distance by episode-type interaction.

The comprehension and recall test was only given to motivate the subjects to read the episodes carefully. As expected, no differences between conditions were found in number of target sentences recalled on the 24-item comprehension test. The mean number of correct items based on goal–action episodes was 22.36 and 22.57 for far and near episodes, respectively. Similarly, scores for items taken from far action–goal episodes (21.78) did not differ from scores for near action–goal episodes (21.93).

This experiment provided evidence for the distance hypothesis. However, as alluded to earlier, several problems arise from the use of short naturalistic episodes. One problem stems from assumptions about subjects' pre-experimental knowledge. There is no way of assuring that the episodes activate the same goal structure for every subject. We do not know how accurate we were in guessing at culturally prototypical goal trees. Related to
this problem is the “functional inequality” of different subgoals, which surely affects the ease of linking up actions with them. Some actions such as “look into the yellow pages” are appropriate for a vast number of long-range goals, while others may have narrower ranges of application. Additionally, different goals may have varying numbers of subgoals and actions beneath them. Thus, necessary subgoals may enjoy priority over other subgoals.

Another problem stems from generation and selection of the test episodes. In this first experiment, there was no formal way of determining near versus far event sequences beyond our judgement that they appeared to lie nested along one branch of a goal reduction tree. The problem of selection can be compounded by pre-experimental associations that subjects may have between various goals and actions. Experiment 2 used an artificial goal structure to overcome some of these problems.

EXPERIMENT 2

Experiment 1 found that difficulty of comprehension increased with psychological distance separating goals and actions in a goal hierarchy. However, “distance” was relative since it was determined only within a triad of goal–subgoal–action elements judged to lie along the same branch of a goal hierarchy. Goal-to-action distance or number of links could be manipulated with more precision if the goal hierarchy in question was created by the experimenter and taught to the subjects prior to the comprehension study.

Experiment 2 used an artificial goal hierarchy to circumvent the difficulties associated with the test episodes in Experiment 1. An artificial goal hierarchy minimizes the likelihood that pre-experimental knowledge will introduce unwanted variability between subjects and materials. An advantage of using an artificial goal hierarchy is that the experimenter determines exactly the number of subgoals or distance between goals and actions. Thus, since we can select goal–action pairs of varying distances with certainty, we can study the effects of distance between goals and actions on comprehension in a carefully controlled way.

Experiment 2 used the time to verify a true–false assertion as a dependent measure rather than reading time. Although reading time was used successfully as a measure on comprehension in Experiment 1 and has been used extensively in many classic experiments (e.g., Haviland & Clark, 1974), the measure is not without its problems. From the subjects’ point of view, the task of simply presenting themselves with sentences on a computer screen has uncertain goals and criteria. Surely, subjects’ criteria for reporting their “comprehension” varies in uncontrolled ways. Further, reading time varies with task demands. A subject instructed to skim the sentences for gist approaches a reading task differently than will one instructed to learn the sentences for recall. Thus, reading time itself can be a complex measure.

The time to verify a statement (decide whether it is true or false) was used in Experiment 2 to avoid some of these problems associated with reading times. Since comprehension is a precondition for making a correct verification, we assume verification time is composed of comprehension time plus the time to verify or disconfirm a statement.

The statements to be verified in Experiment 2 were phrased similarly to test questions used in a verification task studied by Farley and McCarty (1981). They presented subjects with some single-episode short stories and asked them to rate the acceptability of statements in the following form:

“The protagonist did Action 1 in order that the protagonist could do Action 2.”

They found that subjects’ acceptability judgments decreased with greater distances in the goal-tree analysis of the narrative episodes. That is, the closer Action 1 was to Action along a branch of a goal reduction tree, the higher was the acceptability rating for the pair. Following the lead of Farley and McCarty, the statements in our experiment consisted of two-part questions containing a goal phrase and an action phrase. An example test item is, “In order to do Goal X, the protagonist did Action Y.” The goal phrase was selected to be either near or far from the action in the artificial goal hierarchy the subject had just learned.

We expected that subjects would verify or disconfirm the questions by inferring the subgoals between the goal and action phrase. Assuming that inferences take time, we predict that verification time should increase as “distance” between the goal phrase and action phrase increases in the artificial goal hierarchy. In order to build up the sample size, the test questions were repeated a second time. We expected the verification times to shorten from both a general practice effect and due to storage of answers to earlier questions that originally required slow derivation of the answer.

Method

An artificial goal hierarchy was constructed that depicted a lengthy procedure for becoming a member of the “Top Secret Club”. This goal hierarchy is illustrated in Fig. 3. The goal hierarchy consisted of 18 subgoals nested along six branches below the top goal. In actuality, the subjects never saw a representation of the goal hierarchy such as that in Fig. 3. Rather, they read a 600-word description of the procedure for joining the Top Secret Club, which embodied the goal–subgoal relationships represented in the hierarchy.

A set of 32 true and 24 false “in order to” statements were constructed as possible plans for individuals to become members of the Top Secret Club. The subjects decided whether each plan was correct (true) or incorrect (false). False statements were used as foils in the verification task. An “In order to do X, John did Y” statement is false either if X is a descendant of Y, or X is on a branch to the side of Y. An example of a false “plan” is: “In order to spy on the Zero Club, John did the initiation rites”.

The questions were constructed from a base set of triplets sampled from the tree. A triplet consisted of three randomly spaced goals or actions nested
along one branch of the tree. None of the triplets contained the top goal. Referring to Fig. 3, an example of a “nested” triplet is: (1) spy on the Zero Club, (2) find the location of the hidden barn, (3) buy daffodils. Four types of questions were created from such triplets: true-near, true-far, false-near, and false-far. True-near and true-far questions were constructed by designating the lowest element in the triplet as the action and the two higher members (up the branch) as near or far goals. The question sets contained 32 true (16 near, 16 far) and 24 false (8 near, 8 far, and 8 lateral).

To illustrate, a true-near question from the triplet described above (see Fig. 3) is: “In order to find the location of the hidden barn, John bought some daffodils” and the true-far version is: “In order to spy on the Zero Club, John bought some daffodils.” False-near and false-far items were constructed by reversing the two clauses in the question, thereby altering the truth of the assertion. An example of a false-near question is: “In order to buy some daffodils, John located the hidden barn”; its false-far counterpart is: “In order to buy daffodils, John spied on the Zero Club”. Lateral false,

the third type of false question, were constructed by using a goal and action

selected from completely different branches of the tree. An example is: “In order to buy daffodils, John drank the magic fluid”. Lateral false

were identical in both sets of questions. Note that the three types of false questions differ in their second clause. The questions were presented in a

new random order for each subject.

Distance, defined as the number of links between the goal and action phrase in a question, was computed by subtracting the height of the action phrase from the height of the goal phrase (see Fig. 3). Setting the top goal in the tree to a height of 1, any subgoal or action has a height equal to the height of its immediate goal plus 1. For example, the test item “In order to do the initiation rites, John drank the magic fluid” has a distance of 4 − 2 = 2 (see Fig. 3). The questions tested distances ranging from 1 to 5, since no questions contained the highest goal (“Join the Top Secret Club”). The average distance for nears was 1.2 and for fars was 2.77.

Subjects learned the goal hierarchy by working through some training materials. The entire goal hierarchy had to be thoroughly memorized to assure the subjects’ low error rates on the verification task. To help them memorize all the goal-action relationships described in the manual, a questionnaire was prepared containing short-answer “why” and “how” questions that could be answered by moving one step up or down the goal hierarchy. The questionnaire was designed so that the subjects could work through it at their own pace and receive immediate feedback.

The experiment consisted of a training phase followed by a test phase. After reading introductory instructions, subjects began the training phase, which involved reading the “Top Secret” manual, recalling its contents, re-reading the manual, filling out the questionnaire and, finally, re-reading the manual a third time. The testing phase began by giving subjects instructions to read that familiarized them with the “in order to” phrasing of the questions, gave examples of true and false items, and that asked them to respond as fast and accurately as possible to the test questions. Next, subjects were seated before a computer screen where a “warm-up” instruction paragraph appeared that they read one line at a time by pressing a button after they read each sentence. There followed four practice questions and then the 56 test questions for which verification time was measured. After this, the 56 questions were re-randomized and presented a second time, making a total of 112 test questions per subject.

Subjects presented themselves with the two-phrase questions, one phrase at a time on the computer screen. One second after the prior trial, a new trial began with the goal phrase (“In order to do Goal X”) of the next question appearing automatically. When subjects finished reading this phrase, they pressed a button labeled “Next” to clear the screen, start the clock, and present the action phrase (“John did Action Y”). Subjects decided whether or not this second phrase named an action that was part of those needed to achieve the stated goal, i.e. whether the action was a “direct descendant” of the goal named in the first phrase. If subjects agreed with the plan they pressed a button labeled “True” with their dominant hand; if they
judged the plan to be incorrect, they pressed the "False" button with the other hand. Either button press cleared the screen, stopped the clock, and presented feedback (four stars for correct, or the word "error" for incorrect responses), which remained on the screen during the inter-trial interval of 1 second. Verification time was recorded for the target (action) phrase, and reading time was recorded for the context (goal) phrase. Twelve Stanford University students participated for credit in their Introductory Psychology class. Each subject was tested individually; the session lasted 1.5 hours.

Results and discussion
Subjects understood the phrasing of the "in order to" questions as evidenced by a low overall error rate of 7%, with no reliable differences in error rate across types of question. To check for possible performance differences between the two matched sets of questions, an initial five-way ANOVA (subjects by run by question set by distance by question type) was performed on the verification times and no differences involving the question sets were found. Thus, data from the two sets were combined in subsequent analyses. Very long reaction times greater than 10 seconds were removed (about 1% of the observations). Only error-free data were used in the following analyses.

The distance effect
Our first question is whether subjects used the organization of the goal hierarchy to answer the questions. If so, then verification times should be positively related to distance between the goal phrase and action phrase in the artificial goal hierarchy. Fig. 4 shows that the predicted relation is strongly upheld in the true data. The data for the false questions will be discussed shortly.

For the first run through the 56 questions, mean verification time was 1.89 seconds for true–near questions, and 2.96 seconds for true–far questions, a 58% increase. (Each mean is based on approximately 192 observations.) Decision times for the second run were shorter and differences were again in the predicted direction. Mean verification times were 1.32 seconds for true–nears, and 1.94 seconds for true–fars, an increase of 47%. In a four-way ANOVA, with factors of subjects, run, distance, and answer (true versus false, excluding lateral falses), the distance effect was highly reliable ($F(1, 11) = 16.18, p < 0.002$).

Turning to the differences between Run 1 and Run 2, we see two types of practice effects. Recall that Run 2 consisted of the same questions (re-randomized) that appeared in Run 1. First, Run 2 verification times were shorter than Run 1. Mean verification time for Run 2 was 1.74 seconds compared to 2.38 seconds for Run 1 ($F(1, 11) = 30.38, p < 0.001$). Second, the distance effect was reduced slightly in Run 2. This run by distance interaction ($F(1, 11) = 5.92, p < 0.032$) could result either as a "scaling effect" of lowered times or if subjects remembered some of the answers to the inference questions they saw in Run 1, making it unnecessary to recompute the answer in Run 2. Although the distance effect was weakened slightly by practice, the general finding is that for the trues, questions involving more distant goal–action relationships take more time to confirm.

We examined this relationship between distance and verification time in more detail and found it to be quite orderly. Recall that distance between the goal and action phrases ranged from 1 to 5. Mean verification times grouped according to distance are displayed in Fig. 5 (combining data for distances 4 and 5 due to the small samples). Since the pattern of results was the same for Run 1 and Run 2, the data from both runs were combined for the following analyses. For the trues, verification time increased linearly with distance between the goal and action phrases in the questions ($F(1, 44) = 37.73, p < 0.001$).

These results suggest that subjects used the goal hierarchy in a systematic way to verify the questions in Experiment 2. Decision times were largely determined by the positions of the goal and the action in the hierarchy. As distance between the goal and action statements increased, more inferential steps were needed to relate them. So far, then, the results for the true questions strongly suggest a simple, linear relationship between verification time and distance. However, data for the false questions indicate otherwise.
Fig. 5 — Mean verification time for the action phrase as a function of distance separating the goal and action phrases in Experiment 2, collapsed over both runs.

Although a reliable distance effect occurred for the true questions, Figs. 4 and 5 show its complete absence for the false. The interaction between question type (true versus false) and distance (near versus far) was highly significant ($F(1, 11) = 12.06, p < 0.005$). A three-way ANOVA (subjects by run by distance) on the false data alone revealed no reliable effects of distance on verification times. Each mean for the false questions is based on approximately 96 observations. As expected, the time to reject false plans shortened with practice, from 2.32 seconds for Run 1 to 1.82 seconds for Run 2, ($F(1, 11) = 16.21, p < 0.002$).

One may ask how we could observe such a robust distance effect for trues but no effect at all for false. Rather than an embarrassment, this difference in outcome strongly implies a consistent strategy by which our subjects verified the conjectured plans. We call this the downward search model, to which we now turn.

**A downward search model**

We hoped that the strategies subjects used to link the goal and action statements in the verification task would be similar to those they use to make sense of events in narratives. In both tasks, relationships between the goal and the action have to be figured out by calling to mind the connecting subgoals. However, people are rarely faced with the problem of detecting absurd goal–action relationships. For this reason, we were primarily interested in the way people verified the true questions and did not plan to focus on how they disconfirmed the false. Those questions were mainly included as foils for the trues.

But the false data can help us develop a more precise description of the inference strategy used by our subjects. First, the strategy used has to fulfill the dual requirement of verifying true goal–action relationships as well as disconfirming false relationships. Second, the strategy must not only produce a robust distance effect for the trues, but also the absence of an effect for the false. These constraints narrow the possibilities substantially. The downward search model meets these requirements.

In the downward search model, a goal–action plan would be verified by entering memory at the node stipulated by the first ("goal") phrase, and successively retrieving the chain of subgoals beneath the goal statement, looking for the action mentioned in the plan. The more subgoals separating the goal and action of a "true" plan, the more memory retrieval required, and so the longer it should take to connect the goal to the action statement in the goal hierarchy. On the other hand, plans can be rejected when the action is not among the descendants of a goal phrase. To understand the expected outcomes for trues and false in our experiment, let us first review how true and false questions were constructed.

In the actual hierarchical tree, (Fig. 3), goal nodes are ancestors of action nodes (higher up), so this was the case for the true questions in the experiment; the first (goal) phrase was logically above the second (action) phrase. For the false questions, though, this relationship was violated. The first (alleged goal) phrase (e.g. "In order to drink the magic fluid") was either below the second (action) phrase ("John carried out a courageous mission") or on a side ("lateral") branch.

A downward search model would decide about a statement such as, "In order to locate the hidden barn, John bought some daffodils" by entering the tree at the first goal phrase node ("Locate hidden barn") and then searching down the branch for the second, action phrase. If the action phrase is found, then the conjectured plan is verified; if the action is not uncovered in the downward search, then the plan is disconfirmed.

This model predicts that, for the true questions, verification time increases with distance. Recall that near and far true questions were constructed by fixing the action phrase and moving the goal phrase up or down one branch of the tree. The model assumes that (1) the tree is always entered at the node corresponding to the first (goal) phrase, and (2) search time is proportional to the distance searched downward in the tree. Thus we expect true–far questions to take longer because the tree is entered at a point farther from the action than is the case for a true–near question.

However, for false statements, this model predicts that verification time will not be affected by distance separating the first and second phrases. To explain the absence of differences among false items, recall that the false
questions were constructed differently from the trues. The first (goal) phrase (rather than the action phrase) was held constant and the second (action) phrase was selected to be near or far from it in the tree. The action phrase was always above the goal phrase. This arrangement has consequences for the downward search model, in which a plan is evaluated by entering the tree at the first-named node and searching down among its descendants until either the second-named node of the plan is found or the terminal node is hit. Since the second phrase named a node above that named by the first phrase, a downward search from the first-named node would terminate when it reached the bottom of the tree. Moreover, the time it takes to reach the bottom of the tree is independent of how far the second-named node is above the first-named node. Therefore, this model predicts no difference between the near and far false, because in both cases the tree is entered at the same point, so the search distance to the bottom of the tree will average the same for both false-near and false-far questions.

In general, the downward search model accounts for the interaction between distance and question type, but since the experiment was not originally designed to test this model per se, certain factors (e.g., sentence length) that are important for testing its implications were not counterbalanced in a controlled way. It is proposed here only as one possible avenue of exploration and future experiments could refine this model better.

The fan effect
Earlier we mentioned that retrieval of information about goals probably proceeds in the same way as retrieval of other information from memory. One well-known effect produced in many sentence verification experiments is an associative interference effect (also called a “fan” effect by Anderson & Bower, 1973; Anderson, 1976). As more facts are learned about a concept, the more time it usually takes to retrieve any particular fact.

In this experiment we observed similar interference effects. A node in the goal hierarchy can have one or two immediate descendants (subgoals). For example (see Fig. 3), the node “Do initiation rites” branches to two immediate descendants, “Read blue manual” and “Do courageous mission” (i.e. one branch point). On the other hand, the node “Locate hidden barn” has only a single immediate descendant (i.e. no branching). If subjects verify the “in order to” questions by searching down the tree for a link between the action and the goal, then verification time should increase if there are branching subgoals between the goal and action phrases.

The questions in this experiment could span either 0, 1, or 2 branching points. Referring again to Fig. 3, a question with distance between the goal and action phrase equal to, say, two links can span either two branching points (e.g. “In order to do the spy on the Zero Club, John went to the meetings”); one branch (e.g. “In order to do the initiation rites, John drank the magic fluid”); or zero branchings (e.g. “In order to locate the hidden barn, John bought some daffodils”). For a given distance, reaction time increased with the number of branch points in 83% of the cases for the trues and 80% of the cases for the falses. We expect a fan effect for the false questions since falses with higher branching factors (downward from the node mentioned in the first phrase) are more likely to contain goal phrases that are distant from the terminal nodes. Thus, for both the trues and the falses, as the branching factor increases, so does the difficulty of deciding how the goal and action are related. As the number of subgoals branching off the goal increases, retrieval time increases, which is consistent with the fan effect found in other fact retrieval studies (e.g. Anderson, 1974; Hayes-Roth, 1977; Thorndyke & Bower, 1974).

EXPERIMENT 3
In Experiment 2 verification time increased almost linearly with the number of subgoals needed to be inferred in order to link the action with the goal. The purpose of Experiment 3 was to extend these results to a situation where the information about the goal–action relationships had been acquired in a more naturalistic situation. In this experiment subjects casually read a folktale, much as they would do in an everyday situation. By reading a story, without instructions emphasizing the character’s goals and actions, subjects are expected to arrive at a “naturalistic” representation of the goal relationships. For this naturalistic representation, we asked whether the subjects would still verify the “in order to” questions by recalling the connecting subgoals mentioned in the story, thus causing them to take more time to verify the far questions.

This follow-up study was conducted using a Grimm fairy tale, which was read in a natural manner. Rather than memorizing the material extensively, the subjects in this experiment simply read the fairy tale and wrote a short summary of it. We did not emphasize memorizing or paying special attention to the goal–action relationships in the story.

This experiment retained several of the methodological advantages of Experiment 2. A goal hierarchy could be inferred from the story, since the story described the characters’ plans and actions carried out in service of their top goals. Thus, distance between goals and actions could be computed, as with the artificial goal structure experiment. Also, the story had a novel goal structure (e.g. the wolf, who wanted to eat the baby goats, disguised himself as their mother by eating chalk to soften his voice). This unusual content controls for pre-experimental associations and knowledge about goals.

A second aim of Experiment 3 was to examine more closely the rejection of false sentences. In Experiment 2 the near and far false targets were varied, while the context sentence was fixed. In this experiment, we held constant the false target phrase and varied the context sentences to be near or far. If subjects use a downward search strategy, then the far context sentence is closer to the near context sentence to the bottom of the goal tree and so we should observe a reverse distance effect for near versus far false targets.
Method

The goal structure used as a basis for questions in this experiment was extracted from a Grimm fairy tale, *The wolf and the seven little kids*. The story describes a mother goat trying to protect her seven kids from being eaten by a wolf. The goal structures, which differ for the two main characters (the wolf and the mother), are depicted in Fig. 6. Only actions which were explicitly stated in the story appear in the goal hierarchy. The actions in the hierarchy are a subset of all the story events. The wolf’s goal structure consisted of 15 nodes, representing actions, nested along four branches and the mother’s consisted of 10 nodes along 4 branches.

The method of constructing questions closely resembles that used in Experiment 2. Thirty true (15 near, 15 far) and 30 false (10 near, 10 far, 10 lateral) “in order to” questions, containing a goal and an action phrase, were constructed from a base set of triplets sampled from the goal hierarchies. An example of a nested triplet, which as before consists of three randomly spaced actions nested along one branch of the tree, is: (1) disguise self; (2) make voice soft; (3) eat chalk. The true questions were constructed as in Experiment 2, by designating the lowest element in the triplet the action and the two higher members as near and far goals. Thus a true–near question from this triplet is: “In order to make his voice soft, the wolf ate some chalk”, and a true–far question is: “In order to disguise himself, the wolf ate some chalk”.

Recall that in Experiment 2 false questions had been constructed by reversing the order of the two clauses in the question, thus altering the truth of the decision. The lowest element in the triplet was designated as the goal, the middle element as a near action and the highest element as a far action. Thus, false–near and false–far questions were created in Experiment 2 by holding the goal phrase constant and following it by near or far actions. In contrast, in Experiment 3, false questions were created in the same way as the true questions: false–near and false–far questions were produced by holding the action phrase constant and preceding it by a near or far goal. For true questions, the action phrase is below the goal phrase in the tree, but for false questions, the action phrase is above the goal phrase. A false–near question from the triplet described above is: “In order to make his voice soft, the wolf tried to disguise himself”, and a false–far question is: “In order to eat chalk, the wolf tried to disguise himself”. In both cases the action phrase, “The wolf tried to disguise himself”, is above the goal phrase in the tree.

The probability of a question being true was not related to the height of its first phrase. The questions were randomized for each subject. Distance between goal and action phrases varied from 1 to 5, with 1.12 the average distance for nears and 2.48 the average distance for fars.

As in Experiment 2, this experiment had two phases, a training and a testing phase. The training phase was simplified considerably. During this phase, subjects read the 1112-word fairy tale, wrote a two-sentence summary without referring back to the story, then re-read the fairy tale. This process was paced by the subjects and usually lasted about 15 minutes. None of the subjects were familiar with the fairy tale prior to the experiment.

After completing the training, the subjects read preliminary instructions like those in Experiment 2, which familiarized them with the “in order to” wording of the questions. The testing proceeded as in Experiment 2. Subjects were presented with 4 practice questions and 60 test questions for which decision time was measured. After this, the 60 questions were re-
randomized and presented a second time, making a total of 120 test questions per subject. Fifteen Stanford University undergraduate students participated for credit in their Introductory Psychology class. Each subject was tested individually and each session lasted about 45 minutes.

Results and discussion
The error rate was slightly higher in this experiment than in Experiment 2 (10.6% versus 7% in Experiment 2). The higher error rate was probably a result of the short training period since in an earlier pilot study, nearly identical except for a more elaborate training sequence, errors were 5%. There were no reliable differences in error rate across types of questions. As in Experiment 2 there were no performance differences between the two matched sets of questions, so they were combined in these analyses. To clean up the data, all reaction times less than 0.5 sec. and greater than 10 sec. were removed. This procedure reduced the data set by less than 4%. All means reported below are based on errorless trials.

Our first question is whether subjects used the organization of the goal hierarchies that were implicit in the story to verify the “in order to” questions. The results suggest that this was the case, even with minimal training. The results, collapsed across both runs, are depicted in Fig. 7. Each point is based on approximately 450 observations for the trues and approximately 300 observations for the falses. The true-far questions took about 100 msec. longer to verify than the true-near questions, a 7% increase. This difference was significant in subjects by distance ANOVA on the verification time for the true questions ($F(1, 14) = 5.8, p < 0.03$).

An overall subjects by run by distance by question type (true versus false) ANOVA showed the expected practice effects. Run 2 decisions were faster (mean = 1.67 sec.) than Run 1 decisions (mean = 1.77 sec., $F(1, 14) = 7.32, p < 0.02$). As discussed earlier, the faster verification times in Run 2 would result if the subjects memorized some of the goal–action linkages they computed in Run 1, thus circumventing the need to recompute them in Run 2.

We next analyzed the data for the link distance between the goal and action. Unlike the aggregated near versus far comparisons in Fig. 7, this link distance analysis is not at all controlled for the target phrase that is being timed. Consequently, the results do not plot as an orderly function (see Fig. 8). In this uncontrolled comparison, the true means do not plot as a linear

![Fig. 7](image-url) - Mean verification time of the action phrase when preceded by near, far, or lateral goal phrases in Experiment 3.

![Fig. 8](image-url) - Mean verification time for the action phrase as a function of the distance separating the goal and action phrases in Experiment 3.
function of link distance. We expect a weakening of the distance effect when the data are regrouped by distance, since recall that the target sentence is controlled for (i.e., held constant as distance is varied) only in the near versus far comparisons displayed in Fig. 7. In addition, as will be discussed later, some unexpected factors resulting from using a natural story are probably responsible for the weakened distance effect found here.

Let us now turn to the false data. The downward search model assumes that a question is verified by entering the tree at the goal phrase and then searching down the branch for the action phrase. For false questions, the action phrase is always above the goal phrase, so the search stops at the terminal node, before intersecting the action phrase. There was no distance effect for the falses in Experiment 2 because the goal phrase was identical for both near and far false questions, so the tree was entered at the same point for both types of questions. Recall that the model assumes a question is disconfirmed if only terminal nodes are reached without finding the action phrase node. Thus, there was no distance effect for the falses in Experiment 2 because the distance to the terminal node was the same for near and far versions of a given question.

In contrast, false in Experiment 3 were constructed so that a reverse distance effect would be produced. We expected false-nears to take more time to disconfirm than false-fars. The action (target) phrase was held constant while the goal (context) phrase was moved down the hierarchy to make it either near or far. Since the goal phrase was always below the action phrase for the falses, the closer the goal phrase was to the terminal node, the more distant it was from the action phrase. False-fars should thus take less time to disconfirm than the false-nears since they were always closer to the terminal node.

When the verification times for the falses are grouped according to distance, we see a reliable reverse distance effect. As displayed in Fig. 8, verification time decreases linearly as distance between the goal and action phrase increases ($F(1, 42) = 26.84, p < 0.001$). This confirms the downward search model.

The primary result of this experiment is that time to verify questions about goal–action relationships depends on the organization of the goal hierarchy, which was inferable from reading the story. It appears that even with minimal training on a natural folktale, subjects link the goal and action phrases by successively inferring subgoals that were described in the story. The pattern of results is similar to those in Experiment 2, although less striking. Although verification time for the trues did not increase linearly with distance when target materials were uncontrolled, subjects generally took longer to verify true–far questions than true–near questions when the crucial materials were controlled.

Some unexpected factors arising from use of a natural story are probably responsible for the weaker distance effect found in this experiment. Inherent in the use of natural materials are various extraneous factors that can affect the outcome of an experiment. A goal hierarchy is surely only one of many possible determinants of the memory a reader has of a story. Other factors such as the importance of an action, its interestingness, salience, and unusualness contribute to how story events are processed.

Suppose, for instance, that a reader pays greater attention to the more important elements of the story. If so, then it would be possible for some of the near questions to be actually more difficult than the fars, because the near items require knowledge of small details of the story; and such details would normally be discounted as unimportant to the main story line. To illustrate, one of the true–near questions used in the experiment was: “In order to have some meal strewn on his feet, the wolf went to the miller”, and its matched true–far question was: “In order to disguise himself as the mother, the wolf went to the miller”. In this example, the true–near question may actually have been more difficult to verify or disconfirm because it tapped some easily confusable story details. It is hard to remember whether the wolf went to the miller, the baker, or the shoemaker to get the meal to disguise his feet. These details can be ignored when answering the true–far question because the only knowledge needed is that the wolf went to a series of proprietors to buy items for his disguise. It is not necessary to know that the wolf bought salt from the storekeeper, dough from the baker, and meal from the miller. Given these factors that could affect the inferencing processes, it is encouraging that any distance effect was produced at all in this experiment.

GENERAL DISCUSSION

Taken together the results of the experiments described in this chapter indicate that goal relationships are understood by inferring subgoals which connect the action with the goal. A consistent finding was that as distance increased between the goal and action, ease of understanding decreased. Psychological distance between goals and actions was characterized as the number of intervening subgoals in a goal-reduction tree.

Each experiment required subjects to comprehend goal–action sequences, based on their knowledge of goal relationships. Experiment 1 tapped the subjects’ general knowledge about goal relationships by measuring comprehension time for episodes dealing with everyday events. Ease of comprehension was affected by the “intuitive” distance separating the goals and actions in the episodes. In Experiments 2 and 3 subjects used newly acquired knowledge about goal relationships, which enabled a better measure of psychological distance between the goals and actions that appeared in the verification tasks. Again distance effects were produced, indicating the necessity of inferring the connecting subgoals when comprehending goal–action sequences.

Decision time did not increase with distance for the false questions in either experiment. In Experiment 2 there was no relationship between distance and verification time, and in Experiment 3 verification time decreased with distance. A downward search model was proposed to account for this pattern of results. The model described how subjects searched their recently learned goal hierarchies to verify goal–action rela-
Future extensions

Of the several issues suggested by our results, we choose for discussion possibly fruitful directions for future research on plans and goals in understanding actions. We find that the “artificial hierarchy” task of Experiment 2 is most easily extended into new research questions. Let us consider several of these research questions.

In Experiment 2, our subjects judged the well-formedness of small plans in the true–false format “in order to achieve goal G, actor performed action A”. Time to decide whether the action in the second clause was a descendant of the goal stated in the first clause increased linearly with the distance (G–A) in the goal hierarchy. A first simple extension of this task is to look at the reading time for the action clause in the context of the prior goal clause. Presumably the action should be read and comprehended more readily the closer it is to the goal just stated for the actor. This outcome could be examined by having the subject memorize a goal tree and then either read a series of short vignettes about different characters going through parts of the plan, or by measuring reading time directly for a number of unrelated goal-action pairs.

Another converging measure would examine the priming of recognition memory judgements. After memorizing a goal-tree, the subject would be tested by presentation of a number of goal (or action) phrases, some from the tree, some from outside it. The subjects judge whether the phrase mentions an element of the tree they studied. A distance effect would be expected in a shorter time to recognize as “old” a target action of the tree when it closely follows in the test sequence a test of a goal that is close to the target action in the tree.

An interesting variation to the plan judgment task of Experiment 2 would alter the question, so that one asks for “sufficiency” judgments rather than “necessity” judgments. To elaborate, the goal tree of Experiment 2 asserted that each action in the tree was necessary in order to achieve the goal (or subgoal) above it. The implicit question that subjects answered during testing was: “Is action X one of the necessary acts I must do in order to achieve goal G?” One could alter the task so as to get at sufficiency judgments, asking, “Are acts X and Y (and Z) jointly sufficient to achieve goal G?” Depending on the branching of the tree, more or less acts would be required to achieve the goal. True sufficiency judgments would surely be slower the more necessary acts required to achieve the goal, whereas false judgments would be slower the closer the set of test actions came to a sufficient set.

The tree studied in Experiment 2 was an “unordered-AND” tree. That is, the subgoals listed under a given node were all necessary and jointly sufficient to attain the goal; however, the subgoals did not have to be attained in any specified order—any order was admissible. Clearly, there are many other kinds of goal-action relationships and goal trees. One modification is to have several alternative ways to achieve a goal; this is called an “OR tree”, in that the goal can be achieved by actions A or B or C. An example was the “steal money” tree shown in Fig. 1. In such a tree, an action is sufficient but not necessary to achieve its parent goal. A given goal tree may have any pattern of AND nodes and OR nodes. Presumably OR nodes are searched in memory in the same fashion as AND nodes—they differ mainly in the criteria for judgments regarding goal-action questions. However, that conjecture awaits experimental testing.

An interesting task to examine in AND/OR trees is how the person propagates through the tree the implications of temporary disenablement information. Suppose that a person believes that a goal can be achieved by plans A or B or C (each with its subtrees), then he learns that a minor but necessary component of plan C is temporarily unattainable. He should now disqualify plan C, and only choose between A and B to achieve the goal. Presumably, it would require some time for the subject to infer that plan C is temporarily disabled. And this disenablement inference time should vary directly with the distance between the goal of plan C and the sub-action that is said to be temporarily disabled. For example, in the “stealing” hierarchy of Fig. 1, a person should more quickly judge that robbery (a top goal) is unattainable if told that “choosing a site to rob” is unattainable than if told that “getting maps to find a robbery site” is unattainable. In an experiment to realize these arrangements, the subject would first memorize an AND/OR goal tree and would then be timed as he was tested repeatedly with a variety of temporarily disabled actions, at varying levels. Disenablement affects AND and OR branches differently. Each subgoal of an AND node is necessary to goal attainment, so that disabling one part will disable its immediate subgoal and all the remote ancestors which depend on that subgoal. Disabling a subgoal of an OR node has less impact on higher goals except in cases where all the subgoals are disabled.

Another type of goal tree we are investigating has ordered subgoals, at least in some parts. Goal nodes which have two or more subgoals which must be attained in a specified order will be called AND-THEN trees. AND-THEN trees can have numerous levels, with each level requiring an ordered set of subgoals to be achieved. Thus, a top goal may require attainment of subgoal A and the subgoal B and the subgoal C, where each subgoal such as A requires attainment of ordered subgoals such as a₁ and then a₂ and then a₃ (and similarly for subgoals B and C).

A number of interesting questions arise in considering how people retrieve parts of AND-THEN trees and make judgments about them. For
example, the subject may be asked to judge whether subgoal X must be attained (at some time) before subgoal Y is attained. The simplest search algorithm would be to locate element X and search for Y among its successors (THEN links) and simultaneously locate element Y and search for X among its ancestors (inverse-THEN links). This algorithm implies that the time to verify that $a_1$ precedes $a_3$ within a subgoal chunk (A) would be shorter than the time to verify that $a_3$ precedes $b_2$ across a chunk boundary (from A to B). There is some doubt whether subjects would learn a lengthy serial list of AND-THEN as a series of one-step associations (which by itself would imply strict distance effects in verifying “X must precede Y” questions). A more plausible hypothesis is that learning subjects will convert a series of AND-THEN nodes into an internal location on a time-line, so that the relative times of two subgoals (events) would be assessed by a kind of internal psychophysical judgement. In such judgments, subjects more readily verify the ordering of two subgoals that are very far apart (on the internal time line) then subgoals that are close together in time. This “reverse distance” is a well-known phenomenon that arises in judgments of stimuli that people have encoded and ordered along a one-dimensional continuum (e.g. Holyoak & Walker, 1976; Moyer & Bayer, 1976; Wooster, Glass, & Holyoak, 1978). Probably the subgoals in our hypothetical AND-THEN tree would be so encoded during learning, which would imply that the subject would show a reverse distance effect when asked to verify statements of the form “X must be achieved before Y”. An interesting question is whether order information about higher-level nodes in the tree can be retrieved directly, or whether intervening subnodes must be visited during the search. In terms of our earlier illustration, in verifying that subgoal A must precede C, can memory search directly from A through B to C, or must our memory fruitlessly search all the subparts of A and B before it arrives at C and verifies the test statement?

Clearly, many interesting variations of the experimental task yield both meaningful goal tree structures as well as practical questions which probe different kinds of information from the tree and judgment criteria. The spreading activation theory is a useful place to start theorizing about how pieces of the goal tree are retrieved and combined to make judgments. We hope that continued exploration of these tasks will provide us with a fuller description and more complete understanding of the process by which memorized plans are retrieved, guide comprehension, and are used to answer questions.

ACKNOWLEDGEMENTS
This research was supported by a research grant, MH-13950, to Gordon Bower from NIMH. Address correspondence to Carolyn Foss, Department of Psychology, Jordan Hall Bldg. 420, Stanford University, Stanford, CA 94305, USA.

REFERENCES


