

Retrieving actions from goal hierarchies

NANCY FRANKLIN and GORDON H. BOWER
Stanford University, Stanford, California

We studied how rapidly subjects could retrieve the order of two actions in a goal hierarchy whose actions were temporally ordered. The subjects learned a four-level hierarchy of 27 goals and actions regarding steps for joining a secret club. Then, with respect to this hierarchy, they verified statements of the form "In order for you to accomplish X, you must have already completed Y." Response times were faster the farther X was from Y in the temporal ordering of the actions, which suggests that judgments were derived by referring to the relative locations of X and Y on a fuzzy time line. Yet when asked to indicate the structure of the subgoal relationships, the subjects drew the hierarchy. This result was replicated in a second experiment. Moreover, when asked to judge whether two actions came from the same one of two plans, the subjects responded no faster for close than for distant actions.

When people read stories, they try to understand the characters' actions in terms of their goals and plans. If the reader can see how the action fits into an ongoing goal or plan of the actor, the action can be so explained and understood.

This analysis suggests that actions vary in their comprehensibility according to how easily they can be linked to an actor's goals. Bower (1982) suggested that this ease of connecting an action to a goal might be explicated using the concept of a goal-subgoal hierarchy. In such a hierarchy, a top-level goal is reduced to a set of subgoals, and each of these is reduced to further subgoals and/or actions to be performed. Such goal hierarchies presumably exist in memory for the plans people know for achieving common goals. We suppose that readers will understand an action in light of a goal when they can find a connection between the actor's goal and that action. In terms of such a goal tree, some actions are near a given goal, whereas others are far from it. Correspondingly, the time to find a connection (and understand the action) should be greater for far goal-action pairs than for near pairs. This "distance" hypothesis was tested below.

To avoid the variability of subjects' knowledge of cultural plans, we created an artificial goal-plan hierarchy in which we could exactly specify the content and the pattern of links. We had subjects memorize an artificial goal-plan hierarchy about the nested procedures required to join a social club. (Figure 1 shows the form of the subgoal hierarchy.) Subjects learned that the subgoals and actions must occur in a specific temporal order. For instance, to achieve subgoal 2 you must first do 4, then 5, then 6. Once subjects memorized that hierarchy, we could examine how they used it for various tasks. A task that

resembles comprehension asks subjects to decide whether an action is part of a plan to achieve a specified goal. The general form of such statements in Experiment 1 was: "In order for you to accomplish X, you must have already completed Y." We varied the hierarchical locations of X and Y in the probe statements to realize various distance relationships between the items, where distance is measured by the minimum number of downward and lateral links from the goal X to the action Y.

If memory represents a set of goal-subgoal dependencies according to a tree of connections isomorphic to Figure 1, and if retrieval of each link on the X-to-Y path is needed to verify such a statement, then verification latencies should increase with the X-Y distance. Foss and Bower (cited by Bower, 1982) found some support for this distance effect. However, in their experiment every subject learned the same goal tree, so one cannot generalize beyond that specific tree. Moreover, the Foss and Bower tree was not temporally ordered; subgoals of a given goal could be accomplished in any order. We were interested in examining plan retrieval times from a greater variety of goal trees and from temporally ordered trees.

EXPERIMENT 1

Method

Subjects. Thirty Stanford students enrolled in Psychology 1 were used.

Materials and Procedure. By studying a two-page textual description, the subjects first learned an artificial plan conforming to the goal structure shown in Figure 1. Each paragraph described a single node and its immediate subgoals. The contents of the nodes were varied across subjects in a quasi-random manner.

The subjects studied their passage for 10 min, then were tested for cued recall. Study and recall trials alternated until two errorless recalls occurred. In the judgment task that followed, the subjects received four cycles of 141 plan statements. They made *true/false* judgments using the keyboard of an IBM PC which presented text statements in the form: "IN ORDER FOR YOU TO ACCOMPLISH (X) YOU MUST HAVE ALREADY COMPLETED (Y)." Fifteen percent of the statements were false. Feedback ("right" or "incorrect") was presented after each judgment. Reaction times (RTs) were measured from presentation of the plan statement until the subject's keypress.

This material is based on work supported under a National Science Foundation Graduate Fellowship for the first author and Grant MH 13950-17 from the National Institutes of Mental Health to the second author. The authors' mailing address is: Department of Psychology, Stanford University, Stanford, CA 94305.

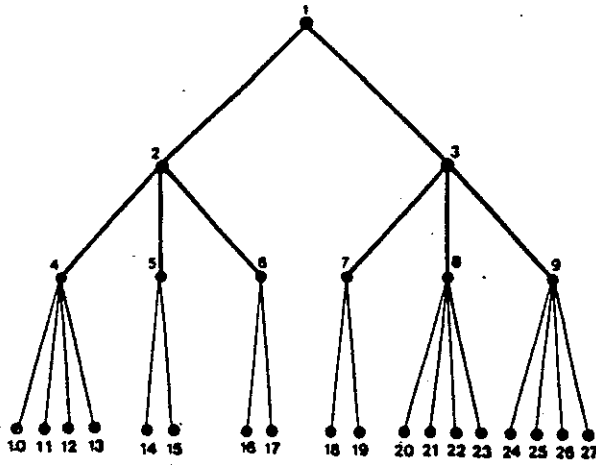


Figure 1. A sample plan hierarchy used in the present experiment, showing the hierarchically ordered plan-goal structure, in which level of detail increases with depth of subgoal. For any goal, all immediate subgoals are ordered linearly along a temporal dimension (from left to right).

After the 564 (4×141) judgment trials, the subjects were asked to draw from memory a tree diagram of the plan on a blank sheet of paper, indicating temporal order of actions from left to right within a level.

Results

All subjects correctly depicted the hierarchical structure of the plan. This fact is important in light of the RT data below. Judgment error data (less than 5%) were eliminated, as were all RTs more than 2.5 standard deviations from each subject's mean RT; this left 94% usable RTs. Overall RTs for *true* judgments averaged 3.08 sec, which is less than the 3.61-sec average for all *false* judgments. We focus below on the *true* RT data.

Hierarchical distance (level). We compared all X-Y pairs for which the upward distance in the tree was one step with those pairs for which the vertical distance was two steps. In Figure 1, for example, Node 12 is one step from Node 4 and two steps from Node 2. The distance hypothesis discussed earlier predicts that one-step pairs will be judged more quickly than two-step pairs. Surprisingly, just the *opposite* was found: two-step pairs were verified more quickly ($M = 3.11$ sec) than one-step pairs ($M = 3.24$ sec), with $t(29) = 3.69, p < .001$.

Serial (linear) distance. The above inverse distance relationship suggests a symbolic distance effect (Moyer & Bayes, 1976). Consequently, we reanalyzed the results in terms of serial distance between the X and Y nodes in the plan statement. In this analysis, distance between plan elements was redefined in terms of the serial order in which the actions occur in real time. If we flatten the tree in Figure 1, this real-time sequence of events becomes 10, 11, 12, 13, 4, 14, 15, 5, 16, 17, 6, 2, 18, 19, 7, 20, 21, 22, 23, 8, 24, 25, 26, 27, 9, 3, 1. By this ordering, for example, the distance from Node 11 to Node 15 is five steps.

Pooling pairs into different distance categories to obtain more stable estimates, we plotted true RT against X-Y

linear distance in Figure 2. Analysis of variance revealed a significant drop in RT as distance increased [$F(5, 145) = 24.4, p < .001$]. A similar analysis was performed on *false* RTs, with distances pooled into three categories. RTs were 3.93 sec (near), 3.48 sec (middle), and 3.24 sec (far), revealing a significant decline with distance [$F(2, 252) = 8.1, p < .001$]. From the above, we may conclude that an inverse distance effect obtains for judgments of what must precede what in the plan order.

Branching effects. The plan hierarchy was designed in part to investigate the influence of associative branching (or "fanning") upon memory search to connect elements. For example, in Figure 1, Node 4 branches into four subgoals, whereas Node 5 branches into two. A statement involving Nodes 4 and 5 represents a four-way followed by a two-way branch. To check for a branching effect, we compared RTs for certain X-Y pairs that exemplified various degrees of branching. If we use only the bottom-level nodes, judging whether Node 15 precedes Node 16 is a two-fan to two-fan decision, whereas judging whether Node 23 precedes Node 24 is a four-fan to four-fan decision. Surprisingly, RTs proved to be insensitive to this branching factor, particularly if serial distance was controlled. None of the relevant comparisons of branching was statistically significant. Of course, this result would be expected if subjects were flattening down the tree of Figure 1 and using it as a simple linear series of actions.

Discussion

The subjects showed an interesting discrepancy with regard to our two different assessments of how they represented the plan. When asked to draw a diagram of the plan, they drew a hierarchy isomorphic to Figure 1. Yet, when repeatedly judging the order of actions, they appeared to transform the hierarchy into the linear order specified above and to respond faster the more distant the plan elements were in the probe statement. The RT data are reminiscent of the symbolic distance effect frequently obtained when subjects make order judgments of pairs of elements arrayed along some ordinal scale such as numerosity, size, and semantic extremity. (For a review, see Banks, 1977.) Explanations of the effect commonly invoke the metaphor of an internal psychophysical judgment, wherein the relative location of each item in the series is somewhat fuzzy, entailing a longer examination to determine the order of two near elements than of two distant elements.

Our symbolic distance effect for judgments about ordinal relationships of elements in an artificial plan brings to mind comparable findings by Galambos and Rips (1982) and Nottenburg and Shoben (1980).

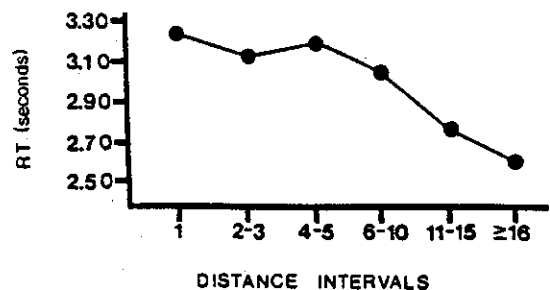


Figure 2. Mean response times (RTs) for true probe statements in Experiment 1. Statements have been grouped into six categories of serial distance.

They investigated subjects' RTs to judge which elements preceded which other elements in routine plans (i.e., scripts, such as changing a flat tire). Both pairs of investigators reported a symbolic distance effect: The more distant two actions were in people's judgment of their temporal order in the script, the faster subjects were to decide which element came first. The memory structures accessed in those studies were generic scripts whose precise contents and organization might vary substantially across subjects. Our experiment carefully controlled the form of the learned plan hierarchy and checked subjects' ability to reproduce the hierarchy exactly, and our subjects still produced the symbolic distance effect when asked to make precedence judgments.

EXPERIMENT 2

In Experiment 2a, we investigated whether memory search of a given tree structure would be sensitive to the type of statement judged. We reasoned that the multiple testing in Experiment 1 on precedence statements might have led our subjects to transform their initial hierarchy into a strict linear order. In Experiment 2a, half the subjects judged statements that we thought might encourage them to maintain their plan hierarchy. Relying on results by Glowalla and Colonius (1982), we presented these subjects with statements of the form "X belongs to the same plan as Y." For this statement to be sensible, the subjects must have learned at least two plans. Consequently, our subjects first learned the two arbitrary plans in Figures 3a and 3b. The subjects would then verify statements such as "Action 5 belongs to the same plan as Action 6" (*true*) or "Action 7 belongs to the same plan as Action 15" (*false*).

Supposing that the subjects' memory structures for these two goal trees were isomorphic to the trees in Figure 3, they might judge whether X and Y belong to the same plan by locating Nodes X and Y and searching along links emanating from them, looking for an intersection. Elements within the same tree have an intersection in either one search step (e.g., Nodes 4 and 6, for which search terminates with an intersection at or near Node 2) or two search steps (e.g., Nodes 5 and 8), whereas elements from different trees (e.g., Nodes 6 and 17) have no intersections. An implication of such a search process is that X-Y pairs for which both actions fall within a subgoal chunk (such as Nodes 4,5 or 18,19) should be judged faster than pairs in which the actions are drawn from two different subgoal chunks (such as Nodes 6,7 or 15,16).

In Experiment 2a, all subjects learned two trees structured as in Figures 3a and 3b. Then half judged statements of the form "X belongs to the same plan as Y," whereas the other half judged statements of the form "X necessarily precedes Y." We were interested in how RTs for searching the same plans might vary with the statement being judged. Experiment 2b replicated the "precedes" judgment condition with slightly different materials to clarify the interpretation of the findings.

Method

The general method was similar to that in Experiment 1. In Experiment 2a, 32 Stanford undergraduates learned the two precedence goal trees diagrammed in Figures 3a and 3b by reading a descriptive text until they could verbally reproduce the tree dependencies and order of

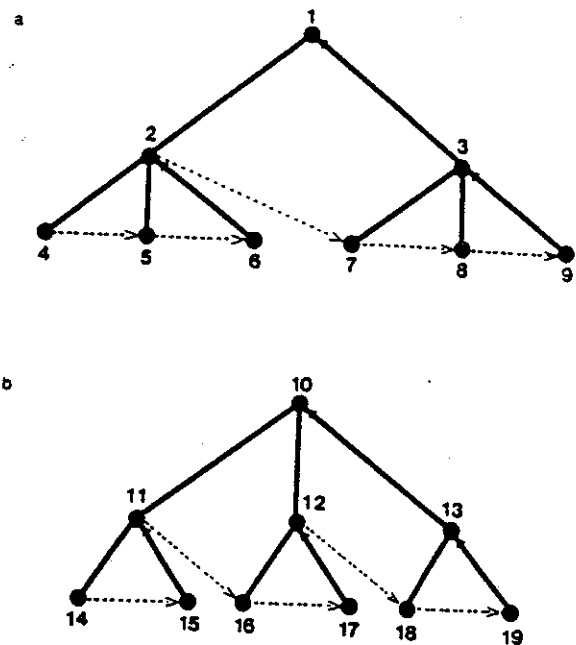


Figure 3. Hierarchical descriptions of the 9-node (3a) and 10-node (3b) plan structures used in Experiment 2. Arrows indicate the corresponding linear order of plan actions.

actions, twice in succession. Each subject received a different assignment of contents to nodes in the trees. Moreover, the subgoals and actions in the two trees were created such that any action could have belonged a priori to either plan. This control was needed to prevent subjects from judging "same plan" simply on the basis of their preexperimental knowledge.

After learning the plans, 16 subjects then received a series of 272 precedence test statements, and another 16 received 272 same-plan test statements. Sixty-eight true statements involving different distance relationships (or within-chunk vs. between-chunk relationships) were mixed among 68 false statements, and the lot was presented twice. Statements were presented on the PC, and the subjects responded as in Experiment 1. The first and last event in the serial listing of nodes (i.e., the end-anchors) were never presented in the judgment phase. Following these test statements, the subjects were given a deck of cards naming actions in the two trees of Figure 3 and were asked to (1) place them in a hierarchical order and (2) arrange them in the appropriate temporal sequence. The order of these latter two tasks was counterbalanced over subjects.

Experiment 2b was a replication of the precedence judgment condition of Experiment 2a with different goal-action materials. Sixteen subjects learned goal hierarchies like those in Figure 3, but with a different set of content phrases. During testing, the proportion of *true* statements was 70%, and both first and last end-anchors appeared in the statement set. The procedure was otherwise identical to the precedence judgment condition of Experiment 2a.

Results

Error data were discarded, as were all RTs more than 2.5 standard deviations from each subject's mean; 93.3% of the data were usable. All subjects performed both sorting tasks perfectly; that is, they correctly depicted both goal hierarchies in Figure 3, as well as the underlying temporal order of actions each implied.

Same-plan statements were judged faster than precedes statements. The mean RTs in Experiment 2a were 2.98 sec (same plan) and 3.70 sec (precedes) (55% correct).

.05]. Such a relation is understandable, since judging precedence requires a finer discrimination than judging a same-plan relationship (Galambos & Rips, 1982; Nottenburg & Shoben, 1980). In both experiments, "false" decisions were slower than "true" decisions.

Within-chunk versus between-chunk. We expected the subjects who judged precedence questions to decide between-chunk statements faster than within-chunk statements, revealing an inverse distance effect. On the other hand, the same-plan subjects were expected to judge the within-chunk statements faster than the between-chunk ones, because memory searches should intersect more quickly in the former case. A 2×2 analysis of variance of RTs on within- versus between-chunk statements crossed with judgment type failed to reveal the expected interaction. Within-chunk and between-chunk pairs (such as Nodes 4,5 and 6,7, respectively) were answered at comparable speeds for both judgment types.

Serial distance. Serial distances for the trees in Figure 3 vary only over the range 1 through 7 when end-anchors are excluded. These distances were pooled into categories 1-2, 3-4, and 5-7 for Experiment 2a. The same-plan judgments varied little over these three categories. For the precedes judgments, mean RTs in Experiment 2a for Distances 1-2, 3-4, and 5-7 were 3.88 sec, 3.97 sec, and 3.75 sec, respectively, which did not represent a reliable trend.

In Experiment 2b, distances were somewhat greater, so the distance categories were 1-3, 4-6, and 7-9. Mean RTs dropped off significantly with distance for precedence judgments. The means were 3.67 sec (near), 3.40 sec (middle), and 3.20 sec (far) [$F(2,30) = 15.76$, $p < .001$]. This inverse distance effect on RT was still significant (though reduced) when statements involving end-anchors were removed from the analysis. Since this analysis involved substantially fewer RTs than the one described above, only two categories were compared: Distances 1-3 (near) and Distances 4-7 (far). Far probes (3.55 sec) were judged more quickly than near probes (3.73 sec) [$t(15) = 2.18$, $p < .05$].

GENERAL DISCUSSION

Experiment 2a failed to find the expected RT difference for within-chunk versus between-chunk pairs in the same-plan condition. This failure occurred even though the subjects were accurate in reconstructing the goal hierarchies of Figures 3a and 3b. If the subjects answered same-plan statements by intersection searches using such hierarchies from memory, then within-chunk statements should have been faster than between-chunk statements. The failed prediction could arise if, in the course of multiple testing, the subjects reorganized their memories of the goal hierarchies of Figure 3 into bi-level structures, with all nodes in Plan A just one link away from their main goal (1), and all nodes in Plan B just one link away from their main goal (10). Such a reorganization would make intersection search more expedient for the same-

plan subjects, reducing all same-plan searches to only one step. Such a reorganization would eliminate the chunk structures of the goal hierarchies depicted in Figure 3.

The symbolic distance effect of temporal order appeared strongly in Experiment 2b but only weakly in Experiment 2a, which made use of a smaller range of distances. We may compare those distances to the extended range in Figure 2 of Experiment 1, where the distance effect on RT was most substantial for larger distances. As noted, our symbolic distance effect with artificial hierarchies is similar to that observed by Galambos and Rips (1982) and Nottenburg and Shoben (1980) for routine plans. This suggests that the retrieval and decision processes for precedence judgments about recently learned plans are similar to those for generic plans abstracted in long-term memory.

The two different measures of subjects' memory provide an interesting contrast. On the one hand, they drew the plans as multilevel hierarchies. Yet they judged precedence statements in two of the three experiments as though searching through a linear list of actions. Memory theorists have typically assumed that a given knowledge structure in a person's memory is fixed, as in a rigid data format, and that different questions evoke different procedures for retrieving relevant data from that structure. What our results seem to suggest is that action plans can have several different organizations simultaneously, such as goal-subgoal dependencies and temporally ordered sequences, and that one or another organization of the plan can be made dominant, depending upon the task.

Upon reflection, such discrepancies should not be an uncommon state of affairs, since they would arise whenever any collection of objects or activities could be ordered differently either on two different linear continua or according to two different lattices. For example, a group of paintings may be ordered independently and differently according to their size, their value, their date, and so on. Clearly, one would obtain different serial position curves and end-anchors across the objects, depending on the statement (dimension) subjects verified about them. An unpublished experiment by Mosteller (cited in Moyer & Dumais, 1978, p. 144) illustrated just this point. Perhaps, then, we have observed another instance of this general principle: Behavioral results depend on the interaction between the features specified in the judgment question and the way the relevant features are stored and retrieved. The results remind us that inferences about the representations of knowledge are typically limited by the type of task being performed.

REFERENCES

- BANKS, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. In G. H. Bower (Ed.), *The Psychology of learning and motivation* (Vol. 11). New York: Academic Press.
- BOWER, G. H. (1982). Plans and goals in understanding episodes. In A. Flammer & W. Kintsch (Eds.), *Discourse processing*. Amsterdam: North-Holland.
- GALAMBOS, J. A., & RIPS, L. J. (1982). Memory for routines. *Journal of Verbal Learning & Verbal Behavior*, 21, 260-281.
- GLOWALLA, U., & COLONIUS, H. (1982). Toward a model of macro-structure search. In A. Flammer & W. Kintsch (Eds.), *Discourse processing*. New York: North-Holland.
- MOYER, R. S., & BAYER, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology*, 8, 228-246.
- MOYER, R. S., & DUMAIS, S. T. (1978). Mental comparisons. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12). New York: Academic Press.
- NOTTENBURG, G., & SHOBN, E. (1980). Scripts as linear orders. *Journal of Experimental Social Psychology*, 16, 329-347.

(Manuscript received for publication June 29, 1987.)