

Strategy use in question-answering: Memory strength and task constraints on fan effects

LYNNE M. REDER

Carnegie-Mellon University, Pittsburgh, Pennsylvania

and

CYNTHIA WIBLE

Johns Hopkins University, Baltimore, Maryland

The time taken to recognize a studied fact increases as a function of the number of other previously studied facts sharing concepts with the test fact. The phenomenon, known as the "fan effect," has been shown to disappear and sometimes even reverse itself when the set of facts are thematically related. The shift from interference toward facilitation occurs only when subjects can use a plausibility-like strategy. In this experiment, subjects learned variously sized sets of related facts about fictitious people. Subjects were asked to make either recognition judgments ("Did you study this fact?") or consistency judgments ("Is this fact consistent with what you studied?"). Subjects made these judgments both the day the material was acquired and 2 days later. The research reported here supports the hypothesis that, with delay, there is a shift in tendency toward more use of the plausibility strategy, away from the careful strategy of searching for an exact match that produces the fan effect. The plausibility strategy produced either a speedup with greater fan or an increase in error rates when the strategy was inappropriate. Plausibility effects were larger at a delay, in both reaction time and error patterns, regardless of whether subjects were asked to make consistency judgments or recognition judgments. Also as predicted, response times became faster as the tendency to adopt the plausibility strategy without first trying direct retrieval increased.

The time taken to answer a question has been shown to depend on a number of factors besides the nature of the question itself. For example, some experiments have shown that the time taken to recognize that a statement has been studied depends on the number of other studied statements that share some of its concepts (e.g., Anderson, 1974, 1976; Anderson & Bower, 1973; Lewis & Anderson, 1976; Thorndyke & Bower, 1974). One explanation of this phenomenon is that the time taken to recognize a fact depends on the amount of activation received by the underlying propositional representation. This interference effect was dubbed the "fan effect" by Anderson (1974, 1976, 1983) because of the assumed underlying propositional representation in which facts are stored as a set of links between concepts, and facts that share the same concepts all "fan out" from the concept node. Concepts are activated by the words in the question, and activation spreads through

the network in inverse proportion to the number of links that must share the activation.

The number of links fanning out of a concept node is not the only variable that affects the time taken to answer a question. The thematic relatedness of the facts fanning out of a node affects reaction time (RT) (e.g., Moeser, 1979; Smith, Adams, & Schorr, 1978). It has also been shown that the effect of the memory representation on question-answering is sensitive to the strategy invoked (e.g., Keenan & Brown, 1982; Reder, 1982; Reder & Anderson, 1980; Reder & Ross, 1983). For example, Reder and Ross (1983) found that the same memory representation for the same subjects produced radically different "fan functions," depending on whether or not subjects could use a plausibility strategy to judge whether a statement was consistent with other facts studied.¹ RTs were slower with increased fan when subjects were asked to make recognition judgments that required discriminating targets from thematically similar foils. Conversely, RTs were faster with increased fan when subjects were in the plausibility condition and had to accept thematically related but not-studied sentences. It is reasonable to find facilitation with increased fan, since there is more knowledge consistent with the judgment to be made.

In a set of studies that required subjects to answer questions about stories they had read, Reder (1982)

This work was supported in part by Grant BNS-03711 from the National Science Foundation to the first author and by a Sloan Foundation grant to support cognitive science research at Carnegie-Mellon University. We would like to thank J. R. Anderson, C. Dennler, B. H. Ross, and M. Taylor for commenting on the manuscript. Requests for reprints should be sent to the first author at the Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA 15213.

found that RTs were faster at delayed tests than at immediate tests when the required task was to make plausibility judgments. Conversely, RTs for recognition judgments increased with delay. This result was unexpected because most memory research that had manipulated delay between study and test had produced decrements in memory performance with longer delays (e.g., Ebbinghaus, 1885/1964; Luh, 1922). Indeed, in the same experiment, RT for retention of gist information did not decline as fast (Sachs, 1967), but neither did it improve. In RT experiments that have tested for understanding of a passage (e.g., McKoon, 1977), latencies have been longer at delayed test.

The explanation offered for the speedup found in Reder (1982) was that, at short delays, subjects had a tendency to search memory for an exact match to the query (direct retrieval strategy), although they had been asked only about its plausibility. When the query involved a statement that had not been presented earlier, this strategy was ineffective and had to be followed by a computation of plausibility (plausibility strategy). At longer delays, verbatim traces are not so available, reducing the tendency to first try the direct retrieval strategy, and thereby reducing the overall RT.

The present experiment tested this theory of direct retrieval and plausibility judgments by looking for convergence of three factors that previously had not been manipulated simultaneously: amount of relevant knowledge that bears on the question, type of judgment officially required of subjects, and delay between acquisition and test. The fan paradigm offered the opportunity to test the explanation that faster responses with delay are due to a shift in strategy preference, by looking at changes in the slope of the fan function with delay. The assumption was that the direct retrieval strategy would produce a positive fan slope, whereas the plausibility strategy would produce a negative fan slope. In situations in which only one strategy could produce a correct response, the fan function would reflect strongly the required strategy—negative for consistency and positive for recognition. On the other hand, for those test items that could be answered correctly with either strategy, we hypothesized shallower fan slopes to reflect the mixture of the two strategies over different test trials. More important, we hypothesized that the preferences for a particular strategy would change, thereby changing the mixture of strategies and the overall slope, for those items for which either strategy could work.

This experiment also provided an additional test of the theory proposed by Reder (1982): Just as responses become faster with longer delays after a story has been read when subjects are making plausibility judgments for not-studied items, responses should become faster with delay for not-studied plausible items in a fan experiment such as this. The logic is the same for both types of experiments: When the task requires a plausibility judgment and the queried fact has not been

studied previously, responses will be slow if the direct retrieval strategy is tried first (or will produce more errors when subjects stop after it has been tried). Since that strategy will not work, the plausibility strategy must then be used to get a correct response. As the propensity to use the plausibility strategy first increases with delay, correct RTs get faster in those situations (and the number of errors decreases).

The present experiment used the same type of methodology that had been used by Reder and Anderson (1980) and Reder and Ross (1983). Although there were a number of changes in design that could have changed the pattern of results, the only modification that was predicted to affect the fan function was the addition of later delays of testing. In both of these earlier papers, the same subjects produced different RT and error-rate functions plotted against fan, depending on the task requirements of a particular test block. Here we expected to find, for the same subjects in the same task, different fan functions for both RT and accuracy, depending on the delay between study and test.

METHOD

Procedure and Overview

There were three major phases in the experiment: initial presentation, learning, and test. In the initial presentation phase, the subjects were shown sets of facts about various characters. Table 1 gives examples of the types of facts used.

All facts about a given character were presented together on a CRT screen for a fixed amount of time for the subject to study. Then all facts about another character were displayed. Facts were organized by theme, with titles of themes presented as well

Table 1
Examples of Facts to Learn

Condition Abbreviation	Facts
	Tennis Facts
(4-2)	Julie reserved a court for the afternoon. Julie practiced against a backboard. Julie had trouble while serving. Julie lost the first two sets.
	Beach Facts
	Julie rented an umbrella and chair. Julie found a parking spot near the beach.
	Laundry Facts
(2-0)	Mark sorted the clothes in colors and whites. Mark asked the laundry attendant for change.
	Circus Facts
(4-4)	Alice thought the clowns were funny. Alice liked listening to the circus music. Alice nervously watched the tightrope walker. Alice ate cotton candy and popcorn.
	Beach Facts
	Alice was covered thoroughly with suntan lotion. Alice tried to keep sand out of the food. Alice spread the beach towel on the sand. Alice watched the kids make sand castles.

as the facts. The duration of presentation of the facts about a character was proportional to the number of facts that were to be learned about that character: Facts stayed on the screen for $(j+1) \times 10$ sec, where j is equal to the number of facts to be learned. In order to facilitate learning, it was suggested to the subjects that they try to associate the facts with a person whom they knew who had the same name as the fictitious character.

In the second phase, the subjects had to demonstrate that they knew which facts were associated with each character. This involved a dropout cued-recall test. The screen displayed the name of a character, the name of one of the themes associated with that character, and the number of facts that had been studied for that particular character and theme. For example, if a subject had studied four beach facts about Alice, the prompt would be "Alice, beach, 4." The subjects were required only to type in the key words of each sentence so that the gist of the sentence was clear and could be distinguished from other sentences associated with that theme. After the key words had been typed in, the correct sentences were displayed on the screen so that the subjects could rate their performance and type in the number recalled correctly. The subjects' recall protocols and self-scoring were stored in a file so that the experimenter could exclude any subjects who might have cheated. (One did, and his data were discarded.) When all facts had been correctly recalled by a subject for a theme for a character, that character-theme combination dropped out of the test. After all character-theme combinations had dropped out, this dropout procedure was repeated one more time. In this way, each fact had to be recalled correctly at least twice before the third phase of the experiment. It is worth noting here that, unlike in prior studies in this paradigm, the subjects were not required to memorize the association between the characters' names and themes, since the theme name was always part of the prompt.

In the critical third phase, the test phase, we collected the subjects' RTs for making judgments about the studied material. In this phase, one fact at a time was displayed for the subjects to verify. One group of subjects was asked to determine if that fact had been studied; the other group was asked to determine if that fact was consistent with what had been studied about the particular character queried. (Unlike in Reder & Ross, 1983, type of task was a between-subjects variable here.) The subjects were instructed to indicate their responses by typing a "D" for no or a "K" for yes. They were told to keep their index fingers on the "D" and "K" keys at all times. There were four 3-min breaks during this test phase.

Feedback was given to the subjects on trials on which they gave the wrong response. If a subject in the consistency task called a fact consistent when it was unrelated to either theme studied with the character in the test probe, the screen displayed "Wrong, this is not a consistent sentence." If the subject responded negatively to a fact that had been presented or was consistent with a presented fact, the feedback was "Wrong, this is a consistent sentence." For subjects asked to make recognition judgments, the two feedback sentences for errors were "Wrong, this is not a sentence you studied" and "Wrong, this is a sentence you studied."

Design and Materials

All study and test items were statements about a person performing some activity. There were several variables that defined the condition within which a particular probe was tested: relevant fan, irrelevant fan, and testing condition. Relevant fan refers to the number of facts studied with the probed character that are related to the theme of the probe. There were two levels of relevant fan: two or four facts. A test probe can also be described in terms of the number of facts studied with the probed character but irrelevant to the theme of the probe. Therefore, irrelevant fan could be zero, two, or four facts. The six possible fan conditions are described by two digits (see Table 1). For a test-probe statement, the digit on the left refers to number of facts studied with the character on the theme that is relevant to the theme of the probe (relevant fan). The other digit refers to the amount of fan for the other theme associated with that probed character (irrelevant fan). "Themefan" refers to the number of themes associated with the probed character. When irrelevant fan was zero, themefan was one; otherwise, it was two. Consider the first example test probe in Table 2. The subject has studied two facts about Julie at the beach and four facts about her playing tennis. The test probe concerns Julie's playing tennis, so the probe would be in the 4-2 condition. If the probe were about the beach theme, it would be a probe for the 2-4 condition.

The test variables are orthogonal to the fan conditions described above. A studied fact can be tested with three different types of probes: a stated probe, a consistent (but not stated) probe, or an inconsistent probe. The subjects were assigned randomly to one of two judgment tasks, either recognition or consistency. The correct responses for consistent probes vary, depending on the judgment task assigned. Since the first example in Table 2 is a studied fact, the correct response would be positive, regardless of whether the subject was asked to make recognition judgments or plausibility judgments. The unrelated statement at the bottom of the table (unrelated to anything studied with Julie) should be rejected, regardless of the task required of the subject. However, the middle, consistent statement has not been studied and, therefore, should be rejected by subjects asked to make recognition judgments, but accepted by subjects asked to make consistency judgments.

The design had several further constraints. Each predicate was studied with two characters. Test items that had not been studied (inconsistent and consistent statements) were always new combinations of studied character names and studied predicates. In this way, word familiarity could not serve as a decision-making cue. Therefore, for every predicate studied with a character, we needed a thematically related predicate studied with another character to use as a related distractor (foil). So, for example, note that the related foil (consistent probe) in Table 2 about making sand castles uses the predicate studied with Alice (see Table 1). It is considered a 2-4 foil because two facts are studied about Julie that are relevant to the beach theme and four facts are studied about her that are unrelated to the queried theme.

The subjects learned 64 facts about 16 characters. There were

Table 2
Test Sentences Based on Facts in Table 1

Type of Probe	Response	
	Recognition Task	Consistency Task
Studied (Julie lost the first 2 sets.)	Yes	Yes
Consistent (Julie watched the kids make sand castles.)	No	Yes
Inconsistent (Julie thought the clowns were funny.)	No	No

four different themes and 24 character-theme pairings. There were 512 test trials representing 72 conditions, defined by the factorial combination of probe type (studied, consistent, or inconsistent) x relevant fan (two vs. four) x irrelevant fan (zero vs. two vs. four) x delay of test (same day or 2 days later) x task type (recognition or plausibility). The test trials were divided equally between the two testing sessions, with the same distribution or configuration of trials to conditions in both sessions. Trials also were divided equally between positive and negative intended responses. Therefore, there were 128 trials of stated probes in either session for the recognition subjects, and 64 trials of consistent (but not-studied) foils and 64 inconsistent foils. For subjects asked to make plausibility judgments, there were 64 stated probes, 64 consistent probes, and 128 inconsistent probes per session. A given set of 64 trials was distributed as follows: 12 trials (6 characters) for 2-0, 16 trials (4 characters) for 2-2, 4 trials (2 characters) for 2-4, 8 trials (2 characters) for 4-2, 8 trials (2 characters) for 4-0, and 16 trials (2 characters) for 4-4.

For each subject, the computer program randomly assigned themes and facts within a theme to conditions, characters to themes and conditions, and order of presentation of trials. The subjects were assigned randomly to task. By randomizing separately for each subject, effects due to materials were part of the error term, obviating an analysis of variance for materials (see Clark, 1973).

Subjects

There were 20 subjects assigned randomly to the recognition condition and 36 to the plausibility condition.² The subjects could receive compensation partly in money and partly in credit toward a class requirement. Depending on the amount of credit received, some subjects were paid nothing and some earned \$7.50. Day 1's session lasted between 1.5 and 2.5 h, depending on speed of learning. The second session lasted about 0.5 h. One subject was excluded from the analyses because of cheating during the learning phase.

RESULTS AND DISCUSSION

Correct RTs greater than 6.5 sec were truncated to 6.5 sec. Data from missing cells, caused by a subject's erring on all trials in a condition, were replaced with this truncation time or the subject's longest RT, whichever was less. (Fewer than 2% of the data were truncated, and fewer than 2% of the data had missing cells.) After this truncation, analyses of variance (ANOVAs) were performed that used subjects' mean correct RTs in each condition. One ANOVA was performed using subjects assigned to both tasks (task conditions). Separate ANOVAs were also performed for each task and for each delay of test for two of the three types of probes. The error term used was always the interaction of subjects with the effect(s) of interest. By performing ANOVAs on subsets of the data, we obtained the most appropriate error terms for the contrasts of interest.

Relevant Fan

To simplify the presentation of the results, we first describe the data for relevant fan, collapsed over irrelevant fan. Figure 1 displays the correct RTs to make judgments for both tasks for each type of question.³

The top left panel of Figure 1 displays the recognition data for the immediate test; the top right panel

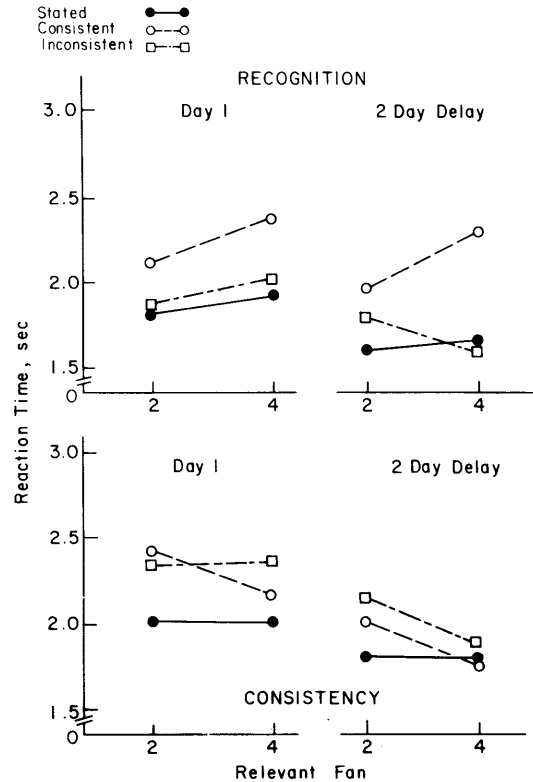


Figure 1. Mean RTs for correct responses as a function of relevant fan for each probe type. The left quadrants are for the Day 1 test, and the right quadrants are for the delayed (48-h) tests. The top half is for subjects in the recognition task, and the bottom graphs plot responses for subjects asked to make consistency judgments.

plots the data for these subjects at the delayed (48-h) test. The bottom panels display the equivalent data for subjects asked to make consistency judgments. Stated facts were verified slightly faster in the recognition task than in the consistency task. This is because stated facts were repeated twice in each session for the recognition task to balance trues and falses. Each fact was tested only once per session for subjects in the consistency task. The data from the first session showed the same general pattern that had been found by Reder and Ross (1983), despite the fact that different subjects were tested in the two tasks and that both types of foils used in recognition were tested in the same test block.

For the first session, there was a significant positive fan effect for (not-stated) consistent probes in the recognition test [$t(38) = 4.05, p < .01$] and a significant negative fan effect for these same probes in the consistency task [$t(70) = 3.73, p < .01$]. The interaction of task x relevant fan was highly significant [$F(1,54) = 25.4, p < .01$]. The other types of probes did not produce significant contrasts for relevant fan on Day 1. The interpretation is that in situations in which only one strategy would produce the correct response, the fan

functions corresponding to that strategy were quite pronounced. In situations in which either strategy would work, there were attenuated functions reflecting the mixture of strategies. Note that when recognition was the "official" task, the slope was somewhat positive for stated (+120 msec) and inconsistent (+150 msec) statements. Although neither slope was significant by itself, the increasing trend over the two was significant [$t(38) = 3.04$, $p < .01$]. In the consistency task, the slope for the same items was much flatter (+12 msec) and nonsignificant.

The right-hand panels of Figure 1 present the data for the second session, 48 h later, and the critical test of the theory. The data showed the same general pattern, with some important adjustments. One salient change is that the subjects were significantly faster on Day 2 [$F(1,54) = 10.3$, $p < .01$]. There are several reasonable explanations for this, including uninteresting ones such as a fatigue effect at the end of Day 1 and/or a speedup with practice. A second, important adjustment is that there was a significant shift in the fan effect with delay [$F(1,54) = 4.9$, $p < .05$], such that for both tasks the size of the overall relevant fan effect decreased with delay. In the recognition task, the fan functions got flatter; in the consistency task, these same functions changed from being relatively flat to more negative in slope. Note that the shift in the slope of the fan function occurred only for the stated and inconsistent probes, because, for consistent probes, only one strategy would produce correct RTs. A mixture of strategies would not.

The data for the second session were consistent with the explanations given for prior results. More important, the change in pattern from the early to the late test was expected from the theory. Both the pattern of speedup effects and the change in the fan effects can be explained by a shift in strategy selection at longer delays toward a greater reliance on plausibility as the preferred question-answering strategy. As discussed earlier, the consistent probes were not expected to change in size of fan effect, since only one strategy would work with those facts in a given task. For these probes, the fan effect stayed very positive over delay in the recognition task and stayed very negative over delay in the consistency task (consult Figure 1). When strategy shifts were possible, namely, for stated and inconsistent probes, the size of the fan effect in the recognition task went from +135 msec on Day 1 to -65 msec at the delay (stated went from +120 to +60 msec, and inconsistencies from +150 to -190 msec). The interaction of delay \times relevant fan for these two probe types was significant [$F(1,19) = 6.23$, $p < .05$]. The greater shift in strategy use for inconsistencies was also significant [$F(1,19) = 4.47$, $p < .05$]. For the plausibility group, the fan effect went from +12 to -146 msec, which was a highly significant change in slope [$F(1,35) = 7.96$, $p < .01$]. Here, stated probes changed from a fan effect of -9 to -30 msec, and

inconsistents went from +33 to -260 msec. In this case, the greater shift with delay for inconsistencies (as contrasted with stated probes) was not significant [$F(1,36) = 2.96$]. This is true even though the negative fan slope was significantly different from zero for inconsistencies at a delay [$t(35) = 2.68$, $p < .05$] and was not for stated probes at either delay.

Differential Speedup

The assumption that direct retrieval is the strategy of choice on Day 1 also accounts for differences in speedup over delay for the different types of probes. Figure 2 plots the speedup (Session 1 - Session 2) for correct RTs as a function of probe type and task. The greatest speedup was for not-stated consistent facts in the consistency task. Conversely, facilitation was smallest for these same items in the recognition task. (Later, we will argue why this second case really exhibits interference.) It is important to note that this big difference in speedup for the consistent items across tasks is mathematically unrelated to the result that their fan functions are so strongly discrepant in slope.

During the first session, the subjects first tried the direct retrieval strategy most often in either task. Using the direct retrieval strategy first slowed down RTs in the consistency task, since the subjects had to subsequently use the other, appropriate strategy for not-stated items. On the other hand, for the recognition task, the subjects responded after having failed to find the fact using the direct retrieval strategy. Therefore, the direct retrieval strategy was appropriate and fast on Day 1 for recognition subjects, but slow for plausibility subjects.

The situation was reversed 48 h later. The subjects tended to adopt the plausibility strategy first more often in both tasks; therefore, the consistent items that should be rejected in the recognition task were either accepted

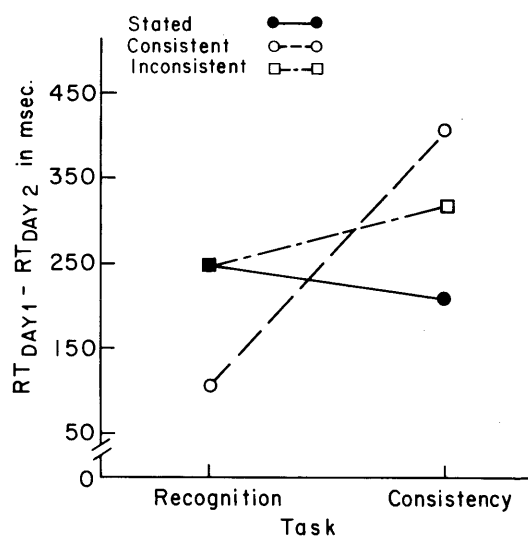


Figure 2. Mean facilitation or speedup from first session to second session, plotted as a function of task and probe type.

erroneously or responded to slowly (since two strategies must be employed to produce the correct response). Those same statements, in the consistency task, benefited from the shift in strategy preference; that is, there was no fruitless first search to slow responses.

Stated facts did not speed up much with delay and might be considered a baseline of no true facilitation or interference caused by strategy shift. (This facilitation or interference should not be confused with the fan-effect interference or facilitation.) For stated facts, we expected a change in slope with delay as strategy preference changed; however, we did not expect an overall change in mean RT, since either strategy worked for either task. It is also the case that either strategy will produce a correct response for inconsistent probes. However, if subjects used the direct retrieval strategy in the consistency task and failed to retrieve the fact from memory, they could not be sure that their answer would be correct. As a result, during the first session, when direct retrieval failed in the consistency task, the subjects sometimes then tried the correct strategy, making inconsistent statements slower than stated statements.

The inconsistent-statement responses did not speed up as much as responses to the consistent probes, because of the fast "gambles" tried on Day 1. When the plausibility subjects tried direct retrieval as an initial strategy on Day 1 and decided to respond from the outcome, their fast responses (fast because only one strategy had been used) were correct for the inconsistent probes but erroneous for the consistent probes. These correct responses for inconsistent probes made their mean RTs faster. These fast times did not get averaged into the mean RTs for the plausible statements, so it was easier to produce a speedup with a strategy shift.

Error Analysis

The pattern of errors in these data is also consistent with the theoretical explanation given to account for the RT pattern. Figure 3 plots percentage correct as a function of relevant fan for each probe type. These data are plotted separately for both tasks and each level of delay, with quadrant organization the same as in Figure 1. In conditions in which the plausibility strategy did not produce the wrong response, accuracy was at least as good at delayed test. There was a significant increase in accuracy with delay for the consistency task [$F(1,35) = 4.46, p < .05$] but not for the recognition task, and the interaction of delay \times task was significant [$F(1,54) = 5.18, p < .01$].

The accuracy improvement was most marked for the consistent not-stated facts in the consistency task. That is, the interaction of probe type with delay in the consistency task was significant [$F(2,35) = 12.94, p < .01$]. During the first session, in which the subjects tended to use the direct retrieval strategy for the consistency task, accuracy was poor for consistent not-stated probes, since it produced the wrong responses. Accuracy improved 10% for both levels of relevant fan.

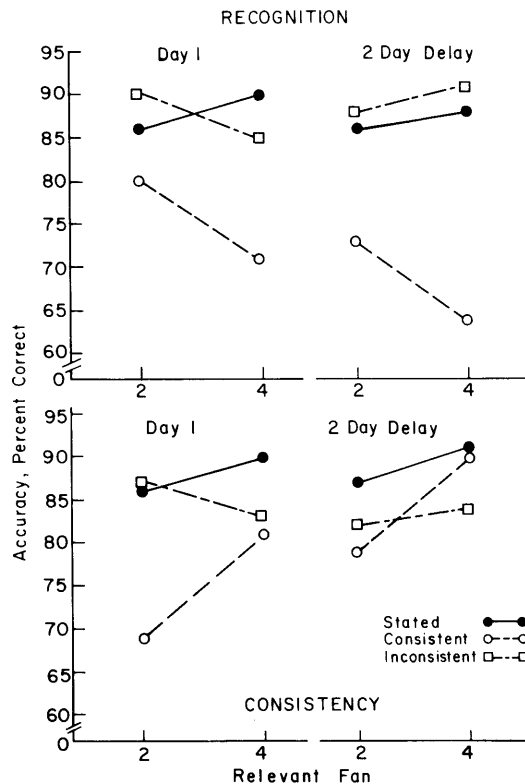


Figure 3. Mean percentage correct for judgments in the recognition task (top) and consistency task (bottom), plotted as a function of relevant fan for each probe type. The short-delay data are displayed in the left quadrants, and the long-delay test data are plotted on the right.

Furthermore, accuracy was 10% better when there were twice as many facts relevant to the theme (four fan vs. two fan). In the consistency task, there was a significant positive advantage for higher levels of relevant fan [$F(1,35) = 30.04, p < .01$] and a significant interaction of relevant fan with probe type [$F(2,70) = 12.13, p < .01$], such that the advantage was greatest for those items that could be answered correctly only when the plausibility strategy (consistent not-stated items) was used.

The accuracy pattern in the recognition task presents a different picture. Across the two tasks, there were highly significant interactions of task \times relevant fan [$F(1,54) = 22.59, p < .01$] and probe type \times task \times relevant fan [$F(2,108) = 17.11, p < .01$]. The consistent statements produced the mirror-image accuracy pattern in the recognition task: These were the only probes that got worse with delay. (They were the only items that would produce an error when the plausibility strategy was invoked.) The interaction of day \times probe type \times task was significant [$F(2,108) = 14.74, p < .01$]. Furthermore, these probes were 10% worse in accuracy for the high-fan condition, that is, when there was more information consistent with the theme of the nonstudied probe.

Summary of Supportive Results

There are three independent sets of results that support the theory of an increased tendency to employ the plausibility strategy with longer delays between acquisition and test. First, we showed that, with delay, there was a shift in the RT slope when plotted against relevant fan. There was a shift toward a more negative function for inconsistent statements, a more attenuated shift for stated facts, and no shift for consistent non-stated facts. We explained above why one would not expect any change in slope for the consistent probes. Below, we will discuss why this shift was attenuated for the stated probes.

The second set of supportive results is the differential speedup with delay for the three types of probes in the two types of tasks. We found, as the theory predicted, the greatest speedup for consistent not-stated probes in the consistency task and the least speedup for these same items in the recognition task. It is worth noting that a similar pattern of differential speedup was found by Reder (1982), although the paradigm and materials were quite different.

The third set of converging results are the shift in error rates with delay. Complementing the relevant fan RT data and the differential speedup data, the error rates indicate a greater tendency with delay to use the plausibility strategy where appropriate and where inappropriate. In the recognition task, more errors were made at a delay to the not-stated consistent facts, but not to any other probes; in the consistency task, fewer errors were made at a delay to these same probes.

The result to be explained is why the shift in slope was less for stated probes than for inconsistent probes. Although there were slight changes in the slope of the fan effect in the correct direction for stated probes, these changes did not approach significance. In fact, the stated probes produced nearly flat fan functions for both delays for both tasks. (Despite these almost flat functions, there is some evidence for the use of the plausibility strategy with stated probes, since accuracy was higher in the four-fan condition than in the two-fan condition.) We suspect that the reason that the effects of strategy shifts are so small for stated probes is that these facts are overlearned. That is, they do not qualify as memory traces "too weak to be reliably retrieved." In the Reder (1982) experiment, subjects read a story once and were asked questions about it once. There, accuracy dropped off considerably for stated probes after a delay of 2 days. In the present experiment, accuracy did not decline for stated probes. Unlike in the story experiment, subjects were drilled for hours on the probes until they could recall them several times. Then, they were tested multiple times on each probe. Later, we will discuss why overlearning of a statement should affect which strategy is selected.

Irrelevant Fan

In addition to manipulating the number of statements studied about an individual for the theme relevant to

the test probe, we also manipulated the number of facts studied about an individual for the theme not relevant to the theme of the test probe. The latter independent variable is called irrelevant fan. We had expected variations in irrelevant fan (or themefan) to replicate prior research and to be unaffected by delay. Some prior research (Reder & Anderson, 1980; Reder & Ross, 1983) has suggested that irrelevant fan has an effect only when one contrasts zero with greater than zero levels. That is, the critical variable is the number of themes associated with the character, not the number of facts studied with the irrelevant theme. However, the research of McCloskey and Bigler (1980) and Zaragoza and McCloskey (1983) has qualified these findings. Zaragoza and McCloskey's data suggested that there is an effect of irrelevant fan for not-stated probes when the second theme is semantically related to the first.

Figure 4 plots correct RTs to judgments made as a function of irrelevant fan and delay, collapsing over relevant fan, using the same quadrant organization as Figure 1. The error percentage is also given below each data point. At first inspection, these patterns seem random; however, there are some sensible systematic trends. For example, with increasing irrelevant fan, the subjects who had been asked to make consistency judgments

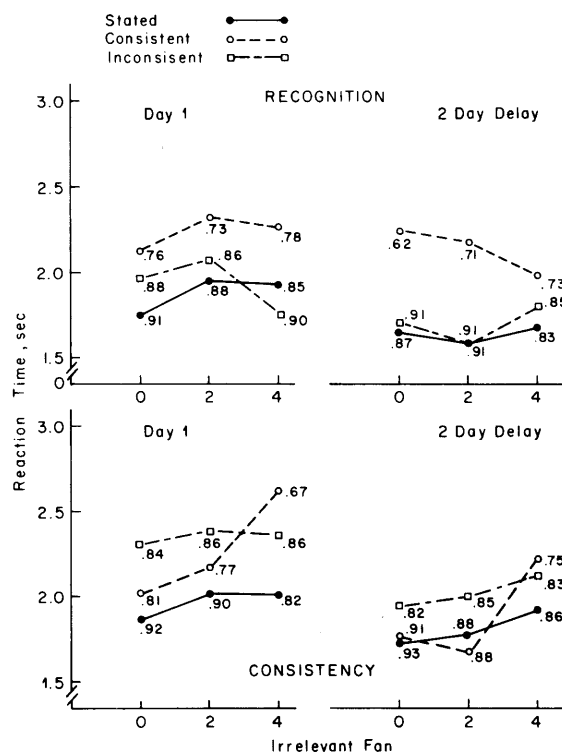


Figure 4. Mean correct RTs and accuracy plotted as a function of irrelevant fan for each probe type. The left quadrants are for the Day 1 test, and the right quadrants are for the delayed (48-h) tests. The top half is for subjects in the recognition task, and the bottom graphs plot responses for subjects asked to make consistency judgments.

were less accurate and slower to respond to the consistent probes. Conversely, the subjects in the recognition task (group) tended to respond more accurately and quickly to these same items as irrelevant fan increased. These patterns were more extreme at delay than at the initial test.

The pattern described above can be explained with two assumptions. First, connections (or links) between "person-nodes" and thematic subnodes get stronger as more facts about that person are learned about that theme. This means that the strength of the connection to the irrelevant or the "wrong" theme will be greater for test probes in the four-irrelevant-fan condition than in the two-irrelevant-fan condition. Second, subjects sometimes respond before both thematic subnodes are active, that is, have received activation from concepts already activated (cf. Anderson, 1983). The thematic subnode with the greater fan will have a stronger link to the person-node and therefore will receive activation from it faster.

Not-stated consistent facts are slower in the four-irrelevant-fan condition of the plausibility task because the irrelevant theme takes most of the activation, slowing the activation of the relevant one. On those trials on which the subject responds before the relevant theme is activated, the subject errs and says "implausible" by inspecting the inconsistent (irrelevant) theme. This accounts for the high error rates with greater irrelevant fan.

These assumptions also account for the opposite trend obtained in the recognition task: On those trials on which subjects do not wait for both themes to become active and the irrelevant theme is found first, they reject the test probe as unrelated (and are correct, without having waited to inspect specific facts). There are more errors in the two- than in the four-irrelevant-fan condition in the recognition task because the consistent theme is more likely to be activated, and sometimes subjects adopt the plausibility strategy in the recognition task. For consistent not-stated probes, activating the relevant theme would be harmful in those instances. These patterns may have been more extreme at the delayed test if subjects adopted an even earlier "deadline" for responding. An earlier deadline is consistent with the general speedup found at delayed test.

One problem left unexplained is why we found an effect of strength of links from the person node to the thematic subnode when our prior studies had not. As mentioned in the Method section, the subjects were not required to memorize the association between character and theme, as they had been in the past experiments. This time, the subjects were always given the character name and theme label as prompts. Therefore, the name-to-theme connections may not have been strengthened as equivalently as in previous experiments. Stated probes are less affected by the differential strength of the person-to-theme links. This is because activation will go from the predicate links to the correct thematic

subnode token. For consistent not-stated probes, activation spreads to a subnode attached to the wrong individual and therefore must rely on the person-to-subnode link.

GENERAL CONCLUSIONS

This experiment extended the previous finding that the effect of relevant fan varies in particular ways depending on the task asked of the subject. The pattern of response times varies not only as a function of task, but also as a function of delay between acquisition and test. Presumably, the change in the shape of the functions with delay results from a change in strategy, rather than from a change in the memory structure. Properties of memory storage do play a role in the sense that, as memory traces become too weak to be efficiently or reliably retrieved, subjects adopt other strategies to answer questions.

This study also indicated that probe type, as well as task and delay, affects the slope of the fan function. The weaker fan effect for stated probes than for consistent probes has been found previously (Reder & Anderson, 1980; Reder & Ross, 1983), and has been explained by the mixture of strategy use. Above, we argued that stated probes shifted less with delay because they were "overlearned." Why would degree of learning affect strategy selection?

There must be a mechanism that allows people to select a particular question-answering strategy that contains "strength-estimation heuristics"; variables such as "delay between study and test" would not often be applicable. We speculate that the strategy-selection stage is influenced by variables such as activation level of concepts in the test probe and activation of the proposition-node of the probe itself. When the retrieved propositions seem weak, the plausibility strategy is invoked. Conceivably, the overlearned stated probes retained enough strength so that they received the critical amount of activation 2 days later, making the strategy shift less likely.

In conclusion, our research contributes to the resolution of the paradox of interference, first addressed by Smith et al. (1978), in which knowing more about a topic hurts a person's ability to answer a question on that topic. Knowing more about a concept hurts memory performance only if a specific fact must be retrieved. In most situations, a careful discrimination or specific retrieval is not required. The ubiquity of judging plausibility as the preferred mode of thinking is becoming apparent in light of recent research. Interference with RTs can occur with "real world" materials (e.g., Lewis & Anderson, 1976; Peterson & Potts, 1982); however, this problem is mitigated by our tendency to use plausibility as a question-answering strategy. Chiesi, Spilich, and Voss (1979) showed that an expert in a domain can acquire new, domain-relevant information better than a novice. On the other hand, Arkes and Freedman (1984)

showed that this is true only if the memory task allows the use of a plausibility-like strategy. If a plausibility-like strategy produces the wrong response, experts are worse.

Finally, the preference for a particular verification strategy may depend on variables other than official task requirements, thematic-relatedness of foils, and the strength of the memory traces. The manner in which the information is acquired may also influence strategy-selection preferences. For example, Clifton and Slowiaczek (1981) found that subjects verified studied (fictitious) facts about a famous person faster if the facts were consistent with prior knowledge about the person; however, this was true only when the material had been studied as part of a minibiography and not when studied as a list of sentences. Clifton and Slowiaczek suggested that the minibiography is more likely to facilitate integration of the new facts with the old memories. The minibiography may also increase the tendency to use the plausibility-consistency strategy. This view would predict that the difference between the minibiography and list conditions would grow larger at longer delays.

REFERENCES

- ANDERSON, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 5, 451-474.
- ANDERSON, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.
- ANDERSON, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- ANDERSON, J. R., & BOWER, G. H. (1973). *Human associative memory*. Washington, DC: Hemisphere Press.
- ARKES, H. R., & FREEDMAN, M. R. (1984). A demonstration of the costs and benefits of expertise in recognition memory. *Memory & Cognition*, 12, 84-89.
- CHIESI, H. L., SPILICH, G. J., & VOSS, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 18, 257-274.
- CLARK, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.
- CLIFTON, C., JR., & SLOWIACZEK, M. L. (1981). Integrating new information with old knowledge. *Memory & Cognition*, 9, 142-148.
- EBBINGHAUS, H. (1964). *Memory*. (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover. (Originally published as *Über das Gedächtnis*, 1885)
- KEENAN, J. M., & BROWN, P. (1982, November). *Fanning: When does it hurt and when does it help?* Paper presented at the meeting of the Psychonomic Society, Minneapolis.
- LEWIS, C. H., & ANDERSON, J. R. (1976). Interference with real world knowledge. *Cognitive Psychology*, 7, 311-335.
- LUH, C. W. (1922). The conditions of retention. *Psychological Monographs*, 31(5, Whole No. 142).
- MCCLOSKEY, M., & BIGLER, K. (1980). Focused memory search in fact retrieval. *Memory & Cognition*, 8, 253-264.
- McKoon, G. (1977). Organization of information in text memory. *Journal of Verbal Learning and Verbal Behavior*, 16, 247-260.
- MOESER, S. D. (1979). The role of experimental design in investigations of the fan effect. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 125-134.
- PETERSON, S. B., & POTTS, G. R. (1982). Global and specific components of information integration. *Journal of Verbal Learning and Verbal Behavior*, 4, 403-420.
- REDER, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for question-answering. *Psychological Review*, 89, 250-280.
- REDER, L. M., & ANDERSON, J. R. (1980). A partial resolution of the paradox of interference: The role of integrating knowledge. *Cognitive Psychology*, 12, 447-472.
- REDER, L. M., & ROSS, B. H. (1983). Integrated knowledge in different tasks: The role of retrieval strategy on fan effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 55-72.
- SACHS, J. D. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2, 437-442.
- SMITH, E. E., ADAMS, N., & SCHORR, D. (1978). Fact retrieval and the paradox of interference. *Cognitive Psychology*, 10, 438-464.
- THORNDYKE, P. W., & BOWER, G. H. (1974). Storage and retrieval processes in sentence memory. *Cognitive Psychology*, 5, 515-543.
- ZARAGOZA, M., & MCCLOSKEY, M. (1983). *An iterative search model for retrieval from long-term memory*. Unpublished manuscript, Johns Hopkins University, Baltimore, MD.

NOTES

1. In this paper, we use the terms "plausibility," "consistency," and "thematic relatedness" to mean similar things. Plausibility is the name given to the generic strategy described in Reder (1982); consistency is the name used to describe the task that most closely matches this process; thematic relatedness is what is really involved in both constructing the stimulus materials and making the judgments of "consistency" or "plausibility." In general, we will refer to the "plausibility strategy" and "plausibility judgments," and we will refer to the task as the "consistency task."

2. The reason that there were more subjects in the plausibility condition is that these data represent subjects from two experiments, whereas the recognition data are from one study. These two experiments differed only in that test items for subjects assigned to the recognition task of the other experiment were blocked such that foils that are consistent were tested in one block and inconsistent foils were tested in another block. The results mirrored those obtained in prior studies and those that will be reported here; for example, consistent foils were much slower and less accurate than inconsistent foils. The reason this paper describes one study rather than two is that the additional complexity of yet another factor and the additional space required did not seem justified by the small increase in new information. The unblocked foil presentation version was chosen to be reported because the design was comparable to the design for the plausibility task, making a factorial design.

3. For inconsistent probes, the distinction between relevant and irrelevant fan is not very meaningful. This is because, when the probe is irrelevant to all facts studied with the probed character, one cannot tell which theme or themes a subject will inspect for that character. For the 2-0, 4-0, 2-2, and 4-4 fan conditions, assignment is obvious; for the 2-4 and 4-2 conditions, data were split randomly by computer into two or four relevant fan.

(Manuscript received October 14, 1983;
revision accepted for publication March 9, 1984.)