Memory & Cognition
1988, 16 (2), 147-157

Focused search of semantic cases in question answering

MURRAY SINGER, GRANT PARBERY, and LORNA S. JAKOBSON
University of Manitoba, Winnipeg, Manitoba, Canada

Three experiments tested the hypothesis that people can execute focused searches of semantic cases when answering questions about complex facts, such as "the teacher watered the peas, the corn, and the lettuce with the hose." In Experiment 1, answer time varied mainly with the number of concepts in the relevant or focused case, supporting the hypothesis. Experiment 2 indicated that the irrelevant case undergoes some processing. Experiment 3 confirmed that search in Experiments 1 and 2 focused on semantic cases rather than on taxonomic categories. It is proposed that focused memory search has a high probability of retrieving the relevant case and a low probability of retrieving the irrelevant case. In the latter event, the irrelevant concepts receive full processing.

In a series of studies, Anderson (1974, 1976) provided evidence that the time needed to retrieve a test fact increases as a function of the total number of learned facts in which the concepts of the test fact have participated. This result is known as the *fan effect*. Since those studies, several investigators have provided evidence for an interesting qualification of the fan effect: Namely, it has been shown that memory search can focus on one of several subsets of facts related to a concept (Anderson & Paulson, 1978; McCloskey & Bigler, 1980; Reder & Anderson, 1980). In McCloskey and Bigler's (1980) study, the subjects learned one or two sets of facts about characters identified by profession names. For example, they learned that the "architect" liked certain animals and/or certain countries. Corresponding test probes linked the profession name with exemplars from these categories (e.g., "architect Madrid"). A strict interpretation of Anderson's (1976) fan effect is that fact retrieval time in this paradigm will be a function of the total number of facts learned about the architect.[1] Instead, judgment time varied primarily as a function of the number of facts in the category to which the other concept in the probe belonged. Thus, according to McCloskey and Bigler, activation from a test fact concept (e.g., "architect") can focus on a relevant category of concepts.

Similarly, for each of several characters, Reder and Anderson's (1980) subjects learned facts that could be grouped according to either one or two themes, such as "the circus" and/or "skiing." Retrieval time varied as a function of the number of facts in the theme relevant to a test probe, rather than as a function of the total number of facts. Reder and Anderson proposed that thematically related facts can be organized into thematic *subnodes*. Memory search, in turn, can focus on the subnode representing the theme to which the test fact refers. As a result, the number of facts stored at the irrelevant subnode exerts little if any effect on retrieval time.

Consider the question "did the artist clean a cup?" The form of this question designates "artist" as given information, and asks about the accuracy of the new information, "cup" (Clark & Haviland, 1977; Lehnert, 1977, p. 56). Singer (1981, 1984) proposed that the memory search that results when one attempts to answer this question focuses on the semantic case (Fillmore, 1968) of "cup," namely, the patient. The goal of the present study was to make a direct test of this focused search hypothesis.

To accomplish this goal, we asked subjects to learn complex facts that might include several concepts in a given semantic case. For example, the fact "the pilot painted the fence with the brush, the roller, and the spraygun" includes one agent, one patient, and three instruments. Following Reder and Anderson's (1980) proposal concerning theme subnodes, this fact might be represented as illustrated in Figure 1. Consider the related question "did the pilot paint a garage?" This question interrogates the patient case. The focused search hypothesis predicts that answer time will be influenced only by the number of patients in this fact (viz., one), rather than by the total number of concepts linked to "pilot" (viz., four).

The hypothesis that people can execute focused searches of semantic cases is based on the proposal that cases function as cognitive categories. Case analysis was undertaken in an attempt to identify the different semantic roles that a noun may play in relation to its predicate. Linguists have identified numerous cases, including agent, goal, experiencer, instrument, location, and "patient" object of an action (Chafe, 1970; Fillmore, 1968). Experimental studies have indicated that case analysis has considerable psychological validity (Braine & Wells, 1978;

This research was supported by Grant A9800 from the Natural Sciences and Engineering Research Council of Canada. Experiments 1 and 2 were presented at the meeting of the Psychonomic Society, Boston, November, 1985. Experiment 3 was presented at the meeting of the Psychonomic Society, New Orleans, November 1986. Lorna Jakobson is now at the Department of Psychology, University of Western Ontario, London, ON, Canada. Requests for reprints should be addressed to Murray Singer, Department of Psychology, University of Manitoba, Winnipeg, MB R3T 2N2, Canada.

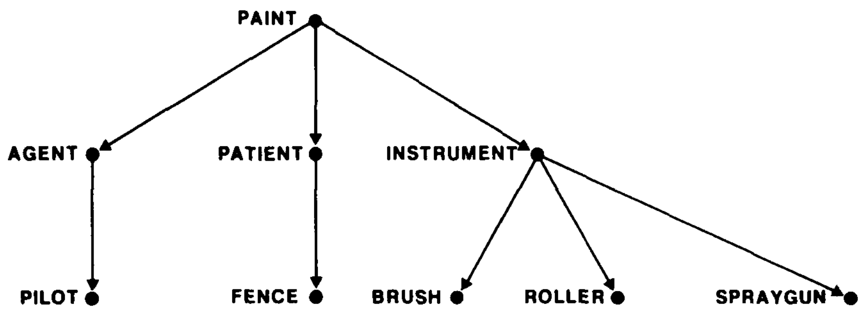

Figure 1. Subnode organization of a complex fact.

Chaffin & Herrmann, 1984; Shafto, 1973). For example, in five experiments, Braine and Wells showed children simple pictures that were described by corresponding sentences. The children readily learned to place tokens, which designated the cases underlying the sentences, on the appropriate pictured objects. Furthermore, the children's responses revealed that they could distinguish between similar cases, such as the agent and experiencer, and could make sensible generalizations between cases. Braine and Wells concluded that the results supported their view that cases function as cognitive categories, which, if available to young children, are likely to be available to adults.

Previous studies of focused memory search have examined taxonomic categories (McCloskey & Bigler, 1980) and schematic themes (Reder & Anderson, 1980). The knowledge that permits people to group concepts according to categories and themes is presumably acquired gradually during childhood. In contrast, semantic cases appear to be crucial perceptual/cognitive categories that are available to children even at the earliest stages of language development (Clark & Clark, 1977, pp. 301-302).

It is likely that there is considerable correlation between membership in taxonomic categories and membership in semantic cases. For example, profession names may function most frequently as agents, carpenter's tools as instruments, and items of furniture as patients. However, this correlation is far from perfect. Profession names can occupy the experiencer and patient cases, carpenter's tools can be patients, and furniture items can play the role of locations. Therefore, it would not be appropriate to assume that the finding of focused search for taxonomic categories will generalize to semantic cases.

Three experiments are presented here. Experiment 1 provided evidence that people can focus memory search on the patient case and the instrument case. Experiment 2 provided additional support for the focused search hypothesis, and indicated that the unfocused or irrelevant case also undergoes some processing. Furthermore, Experiment 2 generalized the findings to the locative case. Experiment 3 was designed to determine whether memory search in Experiments 1 and 2 focused on semantic cases or on taxonomic categories.

EXPERIMENT 1

Experiment 1 was designed to make an initial examination of the focused search hypothesis. In this experiment, the subjects learned 12 complex facts, such as "the teacher watered the peas, the corn, and the lettuce with the hose." Each fact connected a randomly chosen profession name with concepts filling the patient case, the instrument case, or both. Each corresponding test question linked the profession to one concept in one of the cases, such as "did the teacher water some corn?" Accordingly, it was possible to classify each question in terms of the number of concepts that the subject had learned in the case of the interrogated question element (the relevant case) and in the other case (the irrelevant case). For example, "did the teacher water some corn?" asks about the patient case. The "teacher" fact included three patients and one instrument. Therefore, this question is a 3-relevant and 1-relevant, or a 3-1, question.

The experimental conditions corresponded closely to those examined by McCloskey and Bigler (1980, Experiment 1). Each subject learned facts that may be referred to as 1A, 1A-3B, and 4A. Facts in the 1A condition linked a profession to one fact in one case only. For example, the 1A fact "the architect cleaned with the sponge" linked the architect to one concept in the instrument case. Similarly, 4A facts linked a profession to four concepts in one case. Facts in the 1A-3B condition (e.g., "the architect cleaned the cup, the jar, and the plate with the sponge") linked a profession to one fact in one of the cases and three facts in the other.

All test questions linked a profession name to either a patient or an instrument. Each question asked about a case that had been mentioned in the fact about that profession. For example, questions about a 1A (instrument) fact never asked about a patient. The probe type condition of each question is designated by the notation *x-y*, where *x* refers to the number of concepts that were learned in the case that the question asked about (the relevant case), and *y* refers to the number of concepts that were learned in the irrelevant case.

Questions about 1A facts were always in the 1-0 probe type condition (one concept in the relevant case, zero in the irrelevant case). Questions about 4A facts were 4-0 probes. Finally, questions about 1A-3B facts were 1-3 probes if they asked about case A and 3-1 probes if they asked about case B. The probe type variable (1-0, 1-3, 3-1, 4-0) was completely crossed with the semantic case variable.[2]

According to a simple interpretation of the fan effect, response time (RT) should be a function of the total number of concepts linked to a profession. If this were true, RT would be fastest in the 1-0 condition, and equal in the 1-3, 3-1, and 4-0 conditions. The focused search hypothesis (Anderson, 1983; McCloskey & Bigler, 1980; Reder & Anderson, 1980) predicts that RT will be determined mainly by the number of concepts in the case that is the focus of the question. This position predicts that answer times for 1-0 and 1-3 probes will be about equal. Furthermore, 1-3 probes are predicted to be answered faster than 3-1 probes, which, in turn, should be answered faster than 4-0 probes. Experiment 1 was designed to test these predictions.

The test probes in the present study were ordinary questions, such as "did the architect clean the jar?" In contrast, Anderson (1976, chap. 8) used word probes, such as "architect jar," and assertions, such as "the architect cleaned the jar." Ordinary questions were used here because the goal of the present project was to inspect the mental processes of natural question answering.

Table 1
Selection of Probe Types and Corresponding Facts in Experiments 1 to 3*

	Probe Type†	Corresponding Fact
Experiment 1	1-0 (8)	1A (4)
	1-3 (8)	1A-3B (4)
	3-1 (16)	1A-3B‡
	4-0 (16)	4A (4)
Experiment 2	1-0 (8)	1A (4)
	1-1 (16)	1A-1B (4)
	1-3 (8)	1A-3B (4)
	2-0 (16)	2A (4)
	3-1 (16)	1A-3B‡
Experiment 3§	1-3 (8)	1A-3B (4)
	3-1 (16)	1A-3B‡

Note—*ns* are in parentheses. *For all probe types, half the items were in the *yes* condition and half in the *no* condition. †Probe type *x-y* means that there were *x* concepts in the relevant case and *y* concepts in the irrelevant case. ‡The 3-1 probes asked about the same 1A-3B facts as did the 1-3 probes. §In Experiment 3, probe types 1-0, 1-1, and 2-0 acted as fillers.

Method

Materials

The materials were constructed from a master list of 20 profession names and eight pairs of fact frames, which appear in Appendix A. Each of the 16 frames in the eight pairs consisted of a verb, four patients, and four instruments, such as "destroyed the boat/plane/truck/van with the hammer/drill/crowbar/anvil."

A unique set of materials was constructed for each subject. For each set, 12 profession names and six fact pairs were chosen at random. The profession names were then randomly assigned to the facts. Next, two of the six pairs of facts were assigned to each of the conditions 1A, 1A-3B, and 4A. Within each of these three conditions, the patient case was randomly designated as the A term for one of the two pairs of facts, and the instrument case became the A term for the other pair. As a result, each subject learned two facts (i.e., a pair) in each of the following conditions: 1A (one patient), 1A (one instrument), 1A-3B (one patient, three instruments), 1A-3B (one instrument, three patients), 4A (four patients), and 4A (four instruments). The distributions of facts and probe types of Experiments 1-3 are shown in Table 1.

Having made the assignment of master pairs to conditions, we constructed the precise facts that the subject was to learn by choosing the first *n* (1, 3, or 4) of the needed case elements from the list shown in Appendix A. The four concepts in each case in each fact had been randomly assigned to the first through fourth positions in the master frame. For example, if the "destroyed" master frame was linked to "teacher" and assigned to the condition 1A-3B (instrument-patient), then the precise fact to be learned was "the teacher destroyed the boat, the plane, and the truck with the hammer." All 1A-3B facts mentioned the concepts in the patient case first.

For each subject, 48 test questions (probes) were constructed, with one *yes* probe and one *no* probe for each of the four 1A facts, one *yes* and one *no* probe for the A term of each of the four 1A-3B facts, two *yes* and two *no* probes for the B term of the 1A-3B facts, and two *yes* and two *no* probes for each of the four 4A facts. Although more questions could have been constructed for the 4A facts and the B term of the 1A-3B facts, this would have resulted in a strong correlation between fan and the number of questions about a particular fact. Such a correlation results in an attenuated fan effect (Whitlow, 1984). The present distribution of test probes did not completely eliminate this correlation, in that there were more questions about facts with higher fan. The present approach, however, was viewed as superior to having, for example, four *yes* probes and four *no* probes about each 4A fact.

The foil words for *no* probes were randomly chosen from the corresponding case of the other fact in the pair. The facts were constructed in pairs simply to provide a source of foil words that would yield probes in the *no* condition that were not anomalous. The concepts in the patient case in both facts of a pair came from one taxonomic category, and those in the instrument case were selected from another. These procedures yielded sensible foil probes, such as "did the teacher harvest with a hoe?" and "did the dentist plant some spinach?" Conversely, anomalous probes such as "did the architect rob a lettuce?" were avoided. Reder and Anderson (1980, p. 456) constructed their facts in sets of four for comparable reasons. Anderson (1974, 1976) did not need to use such a procedure because he studied only profession names and locations, all of which were mutually interchangeable.

Subjects

The subjects were 11 male and female students of introductory psychology at the University of Manitoba, who participated for course credit. All subjects were native speakers of English.

Procedure

Following previous studies of fact retrieval (Anderson, 1974, 1976), the experiment consisted of a learning session and a test session. Each session took about 45 min. In particular, procedures very similar to those of McCloskey and Bigler (1980, Experiment 2) were used. Each subject was run individually.

Learning. The learning session consisted of several phases. First, the subject was handed a shuffled deck of 12 8×2½ in. index cards. Each card showed one of the 12 facts typed on a single line. The subject was instructed to learn the fact on each card by studying it for as long as he/she wanted, and to then place it face down on the table. Second, after studying the whole deck, the subject received a cued recall task. The experimenter read the profession names in random order, and the subject was requested to state all of the information associated with that profession. The subject was scored as being correct if all of the patients and instruments from the related fact were recalled. The order in which the concepts were recalled was disregarded.

At the end of the cued recall test, the deck was shuffled and handed back to the subject for a second learning trial, identical in procedure to the first. Cued recall again followed learning. Beginning with the second learning-recall sequence, cards were removed from the deck when the subject had recalled the corresponding facts correctly on two consecutive trials. This double-dropout procedure was continued until each fact had been correctly recalled on two consecutive trials.

At the end of this procedure, the entire deck was reshuffled and handed back to the subject for one more self-paced learning trial. The subject then received a cued recall test on all of the facts. If, for any facts, the subject was incorrect, or, in the opinion of the experimenter, was hesitant, then those facts were relearned, following the double-dropout procedure of the original learning-recall sequences.

Test. The reaction time test took place approximately 24 h after learning. When the subject arrived at the laboratory, he/she studied the entire randomized deck of facts once, and then received a cued recall test. If the subject erred or hesitated on any facts, he/she restudied and was retested, using the procedure of the final step of learning on the previous day.

The subjects next received 48 practice RT trials to familiarize themselves with the monitor screen and response panels. On each trial, the word "yes" or "no" appeared on the screen, and the subjects pressed the corresponding button using their index fingers. The assignment of "yes" and "no" to the left and right button was made randomly. The *yes* and *no* trials were randomly interspersed, subject to the restriction that there be 24 of each.

In the timed question-answering test, each subject received three blocks of test probes. During each block, the 48 probes appeared in random order. The probes were presented on one 40-column line of a computer-controlled 12-in. monochrome video monitor. The monitor was 44 cm from the subject. On each trial, a fixation point appeared on the screen for 500 msec. Then the question was displayed. The subject had 3 sec in which to press either the "yes" or "no" button. After a 2-sec intertrial interval, the fixation point reappeared, and the next trial began.

There was a rest period of 2 min between blocks. All experimental events were controlled by an Apple II+ microcomputer, and the subjects' responses and response latencies were automatically recorded on each trial.

Results

Learning

In the cued recall tests, responses were considered correct if the subject produced, in any order, all of the concepts associated with a profession. The subjects required an average of 6.3 (SD = 1.6) learning-recall sequences to complete the double-dropout study procedure. Relearning before the test session required a mean of 1.1 trials (SD = .3). No subject failed to learn the 12 facts.

Test

Mean correct RTs were computed for each subject for each of the 16 response × case × probe type conditions. Analysis of variance (ANOVA) was applied to these scores. Because each subject received a unique set of materials, this analysis takes into account both subject and materials variability (Clark, 1973). An alpha level of .05 was used throughout, unless otherwise indicated. All t tests conducted to compare pairs of conditions used the Bonferroni technique of controlling experimentwise error rate. The mean correct RTs, collapsing across case, and the corresponding error proportions are shown in Table 2.

Table 2
Mean Correct Response Latencies (in msec) and Error Rates, Collapsing Across Case, in Experiment 1

	Response		
Probe Type	Yes	No	Mean
1-0	1656 (.076)	1677 (.091)	1666 (.084)
1-3	1724 (.061)	1804 (.121)	1764 (.091)
3-1	1884 (.114)	1912 (.122)	1898 (.118)
4-0	1902 (.182)	2009 (.243)	1955 (.213)

Note—Error rates in parentheses.

The ANOVA revealed that the probe type effect was highly significant [$F(3,30) = 11.97$, $MSe = 62{,}557$]. The response effect was also significant [$F(1,10) = 5.77$, $MSe = 26{,}717$]. This reflected means of 1,791 msec and 1,850 msec for "yes" and "no" responses, respectively. The only significant interaction was case × response [$F(1,10) = 7.48$, $MSe = 14{,}945$]. This reflected the fact that the RT difference between "yes" and "no" answers was greater for the patient case than for the instrument case. The mean RTs for this interaction were 1,789 msec for the yes-patient condition (error rate = .10), 1,898 msec for the no-patient condition (.19), 1,793 msec for the yes-instrument condition (.12), and 1,802 msec for the no-instrument condition (.09).

To compare the levels of probe type, t tests were carried out. RT for 1-0 probes was not significantly faster than for 1-3 probes [$t(76) = 1.75$], nor were 3-1 probes faster than 4-0 probes [$t(76) = 1.02$]. However, on the crucial comparison of 1-3 probes and 3-1 probes, the difference was significant [$t(76) = 2.39, p < .01$]. Thus, as predicted, RT was primarily influenced by the number of concepts in the relevant semantic case.

The overall error rate in Experiment 1 was 12.6%. The correlation between the RTs and error rates for the 16 conditions was .78. A pattern like this is very familiar in studies of this sort (e.g., Singer, 1984). The positive correlation discounts the possibility of a speed-accuracy trade-off.

An ANOVA applied to the error proportions revealed effects generally consistent with the RT analysis. Both the probe type main effect [$F(3,30) = 6.85$, $MSe = .025$] and the case × response interaction [$F(1,10) = 13.3$, $MSe = .011$] were significant. In addition, there was a main effect of case [$F(1,10) = 5.28$, $MSe = .015$], reflecting error rates of 14.3% and 10.1% for the patient and instrument cases, respectively. The response main effect was not significant.

Discussion

A simple model that assumes that all concepts associated with a probed concept are searched predicts equal response latencies for 1-3, 3-1, and 4-0 probes. Instead, RTs to 1-3 probes were 162 msec faster than the joint 3-1 and 4-0 mean RT of 1,926 msec. Conversely, 1-3 RTs were not significantly slower than 1-0 RTs. In other words, Experiment 1 revealed mean latencies of 1,715 msec, 1,898 msec, and 1,955 msec for facts with one, three, and four relevant concepts, respectively. This outcome is consistent with the focused search hypothesis.

McCloskey and Bigler (1980) considered two classes of focused search model. According to their Relevant Subset Only model, only facts in the relevant category are searched during fact retrieval. In contrast, according to nonselective models, such as their Relevance Filter model, facts in the irrelevant subset are searched but are subjected to less processing than are relevant facts. These notions are readily extendable to the present context. Nonselective models predict that the number of irrelevant concepts will exert some influence on RT, whereas selective models predict that they will not.

The present results have some bearing on this issue. RTs for 1-3 probes were 98 msec longer than those for 1-0 probes. Although this difference was not significant, it does suggest that the number of irrelevant concepts had some impact on RT. However, for 1-0 and 1-3 probes, the number of irrelevant concepts is confounded with the number of cases mentioned in the original fact (see McCloskey & Bigler, 1980, pp. 257-258). Therefore, the comparison of these two probe type conditions does not resolve this issue.

To summarize, Experiment 1 provided preliminary support for the focused search hypothesis. Furthermore, the data suggested that the search was nonselective in nature. Experiment 2 was designed with two goals in mind. First and foremost, it was designed to replicate Experiment 1 and thus to lend additional support to the focused search hypothesis. Second, 1A-1B facts and corresponding 1-1 test probes were added to the materials. The comparison of 1-1 and 1-3 RTs was intended to indicate whether the number of irrelevant concepts influenced RT. In this comparison, the number of irrelevant concepts is not confounded with the number of cases in the corresponding fact, as was true in the comparison between the 1-0 and 1-3 probe types in Experiment 1.

Two features of Experiment 1 were altered in Experiment 2. First, the locative case replaced the instrument case. This was intended to extend the generality of the finding of focused search of semantic cases. Second, to reduce the error rate, the answer time limit was increased to 4 sec from 3 sec, and the subjects were more specifically instructed to avoid errors.

EXPERIMENT 2

Method

Materials

The materials were constructed from a master list of eight fact pairs, shown in Appendix B. Each master fact included a verb, three inanimate patients, and three locations. The locations were randomly selected from a list of 55 locations (Anderson, 1976, chap. 8; Paivio, Yuille, & Madigan, 1968).

As in Experiment 1, a unique set of materials was constructed for each subject. In Experiment 2, all eight fact pairs were used in each set. Each of the 16 facts in each set was randomly linked to one of the 20 profession names. Each subject's set of materials included two fact pairs in each of the conditions 1A, 1A-1B, 2A, and 1A-3B. Each subject learned one pair of facts in the conditions 1A (one patient), 1A (one location), 2A (two patients), 2A (two locations), 1A-3B (one patient, three locations), and 1A-3B (one location, three patients) plus two pairs of facts in the condition 1A-1B (one patient, one location) (see Table 1). In all facts that included both patients and locations, the patients were mentioned first.

The test materials for each subject consisted of 64 probes. Each test set included one *yes* and one *no* probe for each 1A fact (total equals 8), one *yes* and one *no* probe for each of the A and B terms in each 1A-1B fact (16), two *yes* and two *no* probes for each 2A fact (16), one *yes* and one *no* probe for the A term of each 1A-3B fact (8), and two *yes* and two *no* probes for the B term of each 1A-3B fact (16).

Facts in the 2A condition, such as "the author sewed the shirts and the pants," were used because they seemed less contrived than the 4A facts of Experiment 1, such as "the author sewed the shirts, the pants, the dresses, and the socks." Furthermore, the corresponding 2-0 test probes had the merit of having the same number of total concepts but more relevant concepts than did the 1-1 probes. The focused search hypothesis predicted that 1-1 RTs would be faster than 2-0 RTs.

All foils in the *no* probes of a fact were selected from the concepts in the same case of the test probe's pair mate. No concept appeared as a foil in more than one probe. For the two *no* probes about the B term of the 1A-3B facts, two of the three different concepts in the B case were selected at random to serve as foil words.

Subjects

The subjects were 16 individuals from the same pool that was used in Experiment 1.

Procedure

The learning procedure was identical to that of Experiment 1. In the RT test session, the subjects received a practice block of 48 *yes* and *no* trials, followed by three blocks of the 64 test probes. In Experiment 2, the answer time limit was 4 sec rather than 3 sec. To encourage the subjects to be accurate, the written instructions regarding error avoidance were underlined. The session was controlled by an Apple IIe computer. The test probes were presented on a single 80-column line, and the monitor was 22 cm from the subject. In all other regards, the procedure of RT testing was the same as in Experiment 1.

Results

Learning

The mean number of trials required to learn the 16 facts was 7.2 (SD = 3.5). Following the retention interval of 24 h, relearning to criterion was accomplished in an average of 1.4 trials (SD = 1.0).

Test

The mean correct response latencies and accompanying error rates for Experiment 2 are shown in Table 3. An ANOVA applied to the RTs showed the probe type effect to be significant [$F(4,60) = 13.5$, $MSe = 147{,}232$]. RTs for 1-3 probes were faster than 3-1 RTs, as revealed by a t test [$t(124) = 3.10$]. RTs for 1-1 probes were also

Table 3
Mean Correct Response Latencies (in msec) and Error Rates, Collapsing Across Case, in Experiment 2

	Response		
Probe Type	Yes	No	Mean
1-0	1796 (.037)	1862 (.042)	1829 (.040)
1-1	1884 (.032)	1872 (.042)	1878 (.037)
1-3	1996 (.065)	2134 (.068)	2065 (.067)
2-0	1932 (.047)	2052 (.047)	1992 (.047)
3-1	2228 (.112)	2322 (.120)	2275 (.116)

Note—Error rates in parentheses.

marginally faster than 2-0 RTs [$t(124) = 1.68, p < .05$, one-tailed]. Finally, 1-1 RTs were faster than 1-3 RTs [$t(124) = 2.61$].

There was a significant main effect of response [$F(1,15) = 7.86$, $MSe = 67{,}331$], with mean "yes" and "no" latencies of 1,967 msec and 2,048 msec, respectively. The main effect of case also reached significance [$F(1,15) = 7.80$, $MSe = 141{,}376$], with mean latencies for the patient and location cases of 2,066 msec and 1,949 msec, respectively. No interactions reached significance.

The mean error rate in Experiment 2 was 6.0%, and the correlation between error rates and RTs was .76. The only significant effect revealed by the error ANOVA was the probe type main effect [$F(4,56) = 6.44$, $MSe = .011$].

Discussion

The results of Experiment 2 provided further support for the focused search hypothesis: RTs were once again faster for 1-3 probes than for 3-1 probes, and 1-1 RTs were marginally faster than 2-0 RTs. In addition, the results of Experiment 2 bore more clearly on the distinction between selective and nonselective search than did the results of Experiment 1. Answer times were faster for 1-1 probes than for 1-3 probes, indicating that the number of facts in the irrelevant case affected response time. This means that the search of memory was nonselective.

It is useful to compare the present findings and analysis with those reported by McCloskey and Bigler (1980) and Reder and Anderson (1980).[3] All three studies have revealed that retrieval time varies mainly as a function of the number of concepts or facts in the relevant category, the basic focused search finding. However, the studies differed in the apparent impact of the number of irrelevant concepts or facts. RTs for 1-3 probes in the present experiments exceeded those for 1-1 probes (see Table 3). In contrast, Reder and Anderson (1980) reported no RT difference when there was 1 and when there were 3 irrelevant facts, although the joint mean of these conditions significantly exceeded the mean for no irrelevant facts. Finally, McCloskey and Bigler found that the number of irrelevant facts affected RTs for "no" responses, but not for "yes" responses.

How can the differences among these studies be explained? Consider first Reder and Anderson's (1980) study. Suppose that a subject has studied two themes about Alan, such as the circus and skiing. According to Reder and Anderson, upon the presentation of the test item, "Alan rode the chair lift," there is some probability of retrieving the wrong theme subnode, namely, circus. However, a direct comparison of the test fact either with the subnode or with only one of the predicates attached to it (e.g., "Alan watched the clowns") would permit the subject to reject the subnode and proceed to the correct one (Reder & Anderson, 1980, pp. 467-468). Therefore, RT is slower with some irrelevant facts (e.g., 1 or 3) than with none, but the precise number of irrelevant facts has no impact.

Reder and Anderson's (1980) explanation applies to McCloskey and Bigler's (1980) taxonomic categories as well. That is, if upon presentation of the probe "editor bears," the subject retrieved the category "cities," the comparison of "bears" and "cities" or of "bears" and one of the city names, would permit rejection of the "cities" subnode. However, this argument does not extend to the materials of the present study. The reason for this is that particular concepts can fill different cases. For example, a direct comparison of a test concept, such as "gallery," and the patient subnode does not definitively indicate that the incorrect subnode has been retrieved. As a result, the subject would have to inspect the concepts stored below a subnode before rejecting it. In turn, the number of irrelevant concepts would affect RT, as was found in Experiment 2.

This analysis does not explain why the number of irrelevant concepts affected the "no" RTs in McCloskey and Bigler's (1980) study. Those authors proposed (p. 260) that, when the search of the relevant category failed, the subjects sometimes examined the irrelevant category, in order to avoid errors. This hypothesis was supported by the fact that the effect of irrelevant facts was greatest in Experiment 3, in which McCloskey and Bigler strongly encouraged their subjects to be accurate.

In spite of the general clarity of the present results, it is possible to offer a somewhat different interpretation of the present focused search findings. Consider one of the master frames of Experiment 1, "repaired the car/bus/train/wagon with the jack/wrench/pliers/vise." The four alternatives in the patient case belong to the taxonomic category "vehicle," and the instruments are all "tools." Therefore, it might be argued that the subjects' memory search in Experiments 1 and 2 focused on taxonomic categories rather than on cases. In fact, such an interpretation would be quite consistent with McCloskey and Bigler's (1980) findings concerning the focused search of categories.

The next experiment was designed to address this alternative interpretation. Experiment 3 examined the patient and instrument cases. However, for those facts that included both cases, all of the patient and instrument concepts came from the same category. Thus, subjects learned facts such as "the lawyer used the missile to destroy the bazooka, the cannon, and the bomb" and "the dancer used the nitrogen to contaminate the helium, the oxygen, and the hydrogen." Suppose that people group the concepts of these facts according to their taxonomic category rather than their semantic case. Then, the answer times for 1-3 and 3-1 probes should be equal, because in both conditions four relevant concepts would have to be considered. If, in contrast, the representation of these facts reflects their semantic case organization, then 1-3 probe latencies should be faster than 3-1 probe latencies, as was the case in Experiments 1 and 2.

EXPERIMENT 3

Method

Materials

A unique set of materials was constructed for each subject. Each set consisted of 16 complex facts plus 64 corresponding test probes. Both the facts and the probes appeared in precisely the same conditions as in Experiment 2. However, only the 1A-3B facts and the related 1-3 and 3-1 probes served as experimental materials. The other materials were fillers.

The experimental materials were based on the eight master frames that are shown in Appendix C. The master frames referred to categories for which it was possible to link members to one another in an instrument-patient relationship. Eight categories of this sort were identified from the list of 56 categories studied by Battig and Montague (1969). For each of the eight categories, four members were selected at random from the 10 most frequent members shown in the Battig and Montague norms. This random selection was subject to the restrictions that the member name consist of a single word and that only one member of a pair of synonyms (e.g., *car*, *auto*) be included.

The experimental materials were constructed as follows: For each subject, four master frames were selected at random. Two 1A-3B (one patient, three instruments) and two 1A-3B (one instrument, three patients) facts were constructed from these four master frames (see Table 1). For example, suppose that "destroy—bomb, cannon, missile, bazooka" was assigned to appear as a 1A-3B (one patient, three instrument) fact. First, a profession name, such as "lawyer," was selected randomly from the list of 20 professions. Then, the four weapons were randomly ordered. The final result was a fact such as "the lawyer used the bazooka, the missile, and the bomb to destroy the cannon." If the same frame had been assigned to the 1A-3B (one instrument, three patient) condition, the resulting fact would have been "the lawyer used the bazooka to destroy the missile, the bomb, and the cannon."

In each subject's materials, six test probes were constructed for each of the four experimental 1A-3B facts, for a total of 24 experimental test probes. There was one *yes* probe and one *no* probe about the A term of each fact. There were two *yes* probes and two *no* probes about the B term.

The foil words for the *no* probes in the experimental materials were selected from the alternate case in the same fact. For the latter "lawyer" fact, an example of a patient *no* probe is "did the lawyer destroy a bazooka?" and an example of an instrument *no* probe is "did the lawyer use a cannon?" This procedure for choosing foil words differed from that of the other experiments, in which the foil words were selected from the companion fact in a fact pair. The new procedure made it unnecessary to pair the master frames, as had been done in Experiments 1 and 2.

The experimental facts always mentioned the cases in the instrument-patient order. It was observed that the use of the opposite order yielded ambiguous questions. Consider, for example, the fact "the tailor contaminated the hydrogen, the oxygen, and nitrogen with the helium." If one then presented the question "did the tailor use some hydrogen?" it seemed conceivable that the subject might answer "yes," following the rationale that the tailor was indeed using the hydrogen that subsequently became contaminated. In contrast, the instrument-patient order, "the tailor used the helium to contaminate the hydrogen, the oxygen, and the nitrogen," appears to assert that the tailor intentionally contaminated the gases, and so draws an especially clear distinction between the concepts in the instrument and the patient case.

The 1A, 2A, and 1A-1B facts and their corresponding test probes constituted filler materials in Experiment 3. Only one set of filler materials was constructed, and it was added to the unique set of experimental materials constructed for each subject. The filler materials were constructed by randomly assigning pairs of facts from the master list of Experiment 1 (see Appendix A) to the appropriate conditions. From the subjects' perspective, the experiment was virtually identical to Experiment 2.

Subjects

The subjects were 15 individuals from the same pool that was used for the previous experiments.

Procedure

The procedure was identical to that of Experiment 2.

Results

During learning, the subjects needed a mean of 5.4 trials (SD =2.2) to meet the double-dropout criterion. No subject failed to reach criterion. Relearning after 24 h required a mean of 1.3 trials (SD =.7).

Only the data of the experimental trials were examined. The RTs and error rates for these trials are shown in Table 4. An ANOVA applied to the mean correct RTs revealed, most importantly, that the probe type effect was significant [$F(1,14) = 31.8$, $MSe = 33{,}638$], reflecting the fact that 1-3 latencies were 189 msec faster than 3-1 latencies. "Yes" RTs were 104 msec faster than "no" RTs [$F(1,14) = 5.16$, $MSe = 62{,}654$]. The case effect was also significant [$F(1,14) = 6.52$, $MSe = 32{,}262$], with means of 2,008 msec and 2,092 msec for the patient and instrument cases, respectively.

The 1-3 versus 3-1 latency difference was considerably larger for "yes" responses than for "no" responses, resulting in a significant response × probe type interaction [$F(1,14) = 6.02$, $MSe = 36{,}660$]. A test of simple main effects indicated that, even in the *no* condition, 3-1 RTs significantly exceeded 1-3 RTs [$F(1,14) = 6.75$, $MSe = 33{,}638$]. The only other interaction to reach significance was that of case × response [$F(1,14) = 5.50$, $MSe = 18{,}134$].

The overall error rate on the experimental trials was 6.3%, approximately the same as in Experiment 2. The

Table 4
Mean Correct Response Latencies (in msec) and Error Rates in Experiment 3

		Probe Type	
Case	Response	1-3	3-1
Instrument	Yes	1915 (.022)	2106 (.039)
	No	2156 (.144)	2189 (.100)
Patient	Yes	1806 (.011)	2164 (.072)
	No	1944 (.033)	2118 (.089)
Mean		1955 (.053)	2144 (.075)

Note—Error rates in parentheses.

correlation between error rates and RTs was .81. The only significant effects revealed by an error ANOVA were the main effect of response [$F(1,14) = 7.77$, $MSe = .012$] and the case × probe type interaction [$F(1,14) = 11.4$, $MSe = .003$].

Discussion

The results were very similar to those of Experiments 1 and 2. Most important, answer latencies for 3-1 probes exceeded those of the 1-3 probes by 189 msec, a magnitude similar to those measured in the previous experiments. In Experiment 3, all of the patients and instruments in each experimental fact were drawn from a single category. Furthermore, each experimental fact included a total of exactly four patients and instruments. Therefore, had the subjects been carrying out a search of the taxonomic categories rather than of semantic cases, answer times for 1-3 and 3-1 probes would have been the same. This clearly was not the case. Accordingly, it is concluded that the subjects of Experiment 3 and Experiments 1 and 2 executed focused searches of the semantic cases included in the representation of the learned facts.

Unlike the previous experiments, the response × probe type interaction reached significance. This reflected a larger probe type effect for "yes" than for "no" responses. McCloskey and Bigler (1980, Experiment 3) detected a similar trend in their data. They argued that subjects may carry out extra searches of the irrelevant category when search of the relevant category fails, resulting in smaller probe type effects in the *no* condition than the *yes* condition. Consistent with these observations, a reexamination of the data of Experiments 1 and 2 revealed that the mean difference between 1-3 and 3-1 RTs was 196 msec in the *yes* condition and only 148 msec in the *no* condition, across the two experiments. However, the response × probe type interaction did not reach significance in Experiments 1 and 2.

Finally, it is noted that patient questions were answered significantly faster than instrument questions in Experiment 3. In contrast, the patient versus instrument difference did not reach significance in Experiment 1.

GENERAL DISCUSSION

The present experiments revealed that people can execute focused searches of semantic cases in question answering. Furthermore, it was found that the number of irrelevant concepts exerted a small but significant effect on RT. However, before a precise account of processing can be proposed, it is necessary to distinguish between two interpretations of nonselective focused search: In particular, it is possible that the irrelevant case is always processed, or that it is processed only occasionally.

Suppose, first, that the irrelevant case is always searched. Then there are two ways to account for the finding that the effect of irrelevant fan was smaller than that of relevant fan: (1) Individual irrelevant concepts receive full processing, but only a subset of them are examined. In other words, the search of the irrelevant case is subject to some sort of self-terminating rule. (2) All of the irrelevant concepts are processed, but to a lesser degree than are the relevant concepts.

There are several reasons to reject alternative 1. First, McCloskey and Bigler (1980, Experiment 3) encouraged their subjects to maintain a high degree of accuracy. This manipulation apparently induced their subjects to search the irrelevant case in the *no* condition, because McCloskey and Bigler detected an irrelevant-fan effect for "no" responses only. Most important, McCloskey and Bigler (1980, Experiment 3) found that the relevant- and irrelevant-fan effects for "no" responses were approximately the same size. This indicates that, when subjects search the irrelevant case, they do so exhaustively.

Second, Naus, Glucksberg, and Ornstein (1972) had subjects memorize concepts that could be grouped according to either one or two categories. Their data indicated that the search of a category was always exhaustive, not self-terminating (see also Sternberg, 1966).

Third, if the search of the irrelevant cases was self-terminating, it would be reasonable to expect to find other evidence of self-terminating search in this paradigm. For example, one might expect search of the relevant case to terminate upon the location of the test concept. In that event, search for *yes* 3-1 questions would, on the average, terminate at the second concept. In contrast, search for *no* 3-1 questions would examine all three relevant concepts. As a result, the mean difference between RTs in the 1-3 and 3-1 conditions would be greater for "no" responses than for "yes" responses. However, as discussed earlier, both our data and those of McCloskey and Bigler (1980) consistently revealed the opposite trend.

Consider alternative 2, which stated that all of the irrelevant concepts are processed, but less completely than are relevant ones. The difficulty with this position is that it is not clear what "processed less" means (see McCloskey & Bigler, 1980, p. 257). It is possible that an incomplete analysis of a concept would consist of an evaluation of a subset of its semantic features (Smith, Shoben, & Rips, 1974). However, because concepts may appear in different semantic cases, such an evaluation would not reveal a clear inconsistency between a concept and a retrieved case subnode.

The second interpretation of nonselective focused search is that the irrelevant case is examined only occasionally. According to this alternative, there is a small probability that search for the relevant case will result instead in the retrieval of the irrelevant case. In the latter event, all of the irrelevant concepts receive complete processing. Then the impact of irrelevant-fan represents the averaging of a large proportion of trials in which the irrelevant case is not examined with a small proportion of trials in which it is examined. This results in a measurable, though small, irrelevant-fan effect.

That there is a small probability of retrieving the irrelevant case is consistent with both the data and analyses of Reder (1982; Reder & Anderson, 1980). Reder and Anderson reported that RTs were smaller with no irrelevant facts than with one or three irrelevant facts. This indicates that subjects sometimes retrieved the incorrect theme subnode. Reder and Anderson proposed that the system can tell if it is at the wrong subnode by directly comparing the test concept with the subnode or with one of its subordinates. Reder (1982) presented a model that addressed the role of a probabilistic element in the retrieval of learned facts and the selection of question answering strategies. Her analyses revealed a satisfactory fit between the predictions of the model and her observations.

These considerations suggest the following account of processing of question answering in the present tasks: (1) The test question is propositionally encoded, and the case of the new information is identified as the relevant case (Clark & Clark, 1977; Singer, 1984). (2) The search for the relevant case associated with that particular fact results in the retrieval of that case with high probability, and of the irrelevant case with low probability. (3) Retrieved cases are searched exhaustively. Whether, after searching the irrelevant case, the subject proceeds to search the relevant one, would not exert much impact on the present findings. This is true because of the hypothetically small proportion of trials in which the irrelevant case is searched.

This analysis has some speculative elements. The present experiments, which were designed to test the basic focused search hypothesis, cannot definitively distinguish among the alternative analyses of nonselective focused search considered in this section. However, the joint findings of Experiments 1, 2, and 3 and of other examinations of focused memory search lend some support to the present proposal.[4]

The present results suggest a considerable agenda of research concerning focused memory search. First, it will be important to determine what conditions favor focused search over the sort of alogical search (Anderson, 1976, p. 279) that is associated with the classic fan effect. One possibility is that the given-new structure of questions guides focused search more effectively than do word probes (e.g., "editor bears") or sentence probes (e.g., "the editor likes bears"). A second possibility is that the learning procedure used in studies of focused memory search (McCloskey & Bigler, 1980; Reder & Anderson, 1980) resulted in the topicalization of the profession names. This may have facilitated the focus of memory search, proceeding from the profession name to its related concepts.

Second, the present materials included a certain degree of confounding between semantic case, on the one hand, and the categories of linguistic surface structure and deep structure, on the other. For example, in Experiment 1, the concepts in the patient case were surface- and deep-structure direct objects, and those in the instrument case were surface- and deep-structure objects of the preposition. This problem is partly addressed by the fact that, in the 1A-3B facts of Experiment 3, both patients and instruments appeared as surface-structure direct objects (e.g., "the dancer used the nitrogen to contaminate the helium, the oxygen, and the hydrogen"). By examining cases such as the experiencer and the goal, it might, in further studies, be possible to completely unconfound these three variables (e.g., Healy & Levitt, 1978). The confounding that existed in the present materials suggests the need for some caution in attributing the present findings to case organization.

Third, it also will be important to study the subnode organization of concepts within a fact (see Figure 1). Many substantive issues will need to be examined: for example, do semantic cases differ in their relative status within a fact (Segalowitz, 1982)? Are the representations of "the pilot painted with the brush" and "the pilot painted something with the brush" identical (Turner & Greene, 1978)? Likewise, are there differences between the representations of "the pilot painted with the brush and the roller" and "the pilot painted with the brush and the latex" (Charniak, 1981)? In our research, we have already begun to address some of these issues.

REFERENCES

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, **4**, 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., & Paulson, R. (1978). Interference in memory for pictorial information. *Cognitive Psychology*, **10**, 178-202.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. *Journal of Experimental Psychology Monographs*, **80**(3, Part 2).

Braine, M. D. S., & Wells, R. S. (1978). Case-like categories in children: The actor and some related categories. *Cognitive Psychology*, **10**, 100-122.

Chafe, W. L. (1970). *Meaning and the structure of language*. Chicago: The University of Chicago Press.

Chaffin, R., & Herrmann, D. J. (1984). The similarity and diversity of semantic relations. *Memory & Cognition*, **12**, 134-141.

Charniak, E. (1981). The case-slot identity theory. *Cognitive Science*, **5**, 285-292.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior*, **12**, 335-359.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. Freedle (Ed.), *Discourse production and comprehension*. Hillsdale, NJ: Erlbaum.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart & Winston.

Healy, A. F., & Levitt, A. G. (1978). The relative accessibility of semantic and deep-structure syntactic concepts. *Memory & Cognition*, **6**, 518-526.

Lehnert, W. (1977). Human computational question answering. *Cognitive Science*, **1**, 47-73.

McCloskey, M., & Bigler, K. (1980). Focused memory search in fact retrieval. *Memory & Cognition*, **8**, 253-264.

Naus, M. J., Glucksberg, S., & Ornstein, P. A. (1972). Taxonomic word categories and memory search. *Cognitive Psychology*, **3**, 643-654.

Paivio, A., Yuille, J. C., & Madigan, S. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, **76**(1, Part 2).

Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *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.

Segalowitz, N. S. (1982). The perception of semantic relations. *Memory & Cognition*, **10**, 381-388.

Shafto, M. (1973). The space for case. *Journal of Verbal Learning & Verbal Behavior*, **12**, 551-562.

Singer, M. (1981). Verifying the assertions and implications of language. *Journal of Verbal Learning & Verbal Behavior*, **20**, 46-60.

Singer, M. (1984). Toward a model of question answering: Yes-no questions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 285-297.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, **81**, 214-241.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, **153**, 652-654.

Turner, A., & Greene, E. (1978). Construction and use of a propositional text base. *JSAS Catalog of Selected Documents in Psychology*, **8**, 58. (Ms. No. 1713)

Whitlow, J. W., Jr. (1984). Effects of precuing on focused search in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 733-744.

NOTES

1. Retrieval time should also vary as a function of the fan of the category examplar (e.g., "Madrid") (Anderson, 1976). To control the effect of exemplar fan, the fan of all category exemplars was held constant at two (McCloskey & Bigler, 1980). In the present study, all concepts in the manipulated cases had fans of one.

2. The present *x-y* (e.g., 1-3) probe type designation (see also McCloskey & Bigler, 1980) needs to be distinguished from that of Anderson (1976). For Anderson, a probe such as "a lawyer is in the park" was a 1-3 probe if there were fans of one and three from "lawyer" and "park," respectively. In the present study, this same probe is called 1-3 if the original "lawyer" fact in the learned set included one concept in the relevant or interrogated case (e.g., the locative case for "was a lawyer in the park?") and three concepts in the irrelevant case (e.g., patient, instrument).

3. For the present purposes, only Reder and Anderson's (1980) related foil condition is of interest.

4. In a previous draft of this paper, we also considered the possibility that the order of retrieval of the relevant and irrelevant cases corresponded to the order of the cases in the learned facts (see Whitlow, 1984). To address this possibility, we randomized the order of the cases during learning in a replication of the present Experiment 1. Consider a sample fact from Experiment 1, "the teacher painted the door, the steps, and the closet with the brush." In the control experiment, this fact was separated in two parts, "the teacher painted the door, the steps, and the closet" and "the teacher painted with the brush." During learning, the subject always received the two cards showing these parts consecutively, but the order of presentation of the two cards was reversed on every learning trial.

Twelve subjects participated in the control experiment. The results were very similar to those of Experiment 1. In particular, there was a significant effect of probe type [$F(3,33) = 15.3$, $MSe = 94{,}257$]. This reflected mean latencies of 1,838 msec, 1,851 msec, 2,019 msec, and 2,208 msec in the 1-0, 1-3, 3-1, and 4-0 probe type conditions, respectively. Therefore, it does not appear that the probe type effect was due to the fixed order of cases within the facts in Experiments 1 through 3.

APPENDIX A
Master Lists of Experiment 1

Profession Names Used in Experiments 1, 2, and 3

1. actor	11. lawyer
2. artist	12. mayor
3. banker	13. merchant
4. chemist	14. plumber
5. dancer	15. preacher
6. dentist	16. sailor
7. doctor	17. scholar
8. farmer	18. tailor
9. fireman	19. teacher
10. grocer	20. writer

Master Frame Pairs of Experiment 1

1. sewed the shirt/pants/dress/coat/with the needle/thread/scissors/thimble
   ripped the socks/blouse/skirt/tie with the nail/pick/switchblade/pin
2. painted the door/steps/closet/porch with the brush/roller/spraygun/cloth
   stripped the wall/floor/ceiling/window with the mop/solvent/broom/bucket
3. repaired the car/bus/train/wagon with the jack/wrench/pliers/vise
   destroyed the boat/plane/truck/van with the hammer/drill/crowbar/anvil
4. harvested the peas/corn/lettuce/spinach with the rake/pitchfork/shovel/scythe
   planted the carrots/beans/potatoes/celery with the spade/hoe/trowel/pitchfork
5. mapped the ravine/glacier/cliff/volcano with the radar/compass/sextant/transit
   studied the valley/canyon/plain/lake with the binoculars/camera/chart/telescope
6. robbed the bank/pharmacy/restaurant/bar with the gun/bomb/knife/grenade
   defended the church/house/school/hospital with the sword/cannon/missile/club
7. cut the oak/pine/elm/walnut with the axe/chainsaw/hatchet/wedge
   carved the maple/birch/spruce/hickory with the chisel/file/lathe/awl
8. stirred the tequila/gin/vodka/brandy with the spoon/ladle/straw/toothpick
   brewed the scotch/rum/bourbon/rye with the beaker/flask/tube/pipe

APPENDIX B
Master Frame Pairs of Experiment 2

1. sewed the shirt/pants/dresses at the gallery/stable/apartment
   ripped the socks/blouses/skirts at the laundromat/beach/clinic
2. painted the doors/steps/closets at the shack/racetrack/factory
   stripped the walls/floors/ceilings at the office/hotel/stadium
3. repaired the cars/buses/trains at the nursery/tower/greenhouse

APPENDIX B (Continued)

destroyed the planes/boats/trucks at the dormitory/barn/hospital
4. stirred the tequila/gin/vodka at the tent/church/motel
brewed the scotch/rum/bourbon at the university/temple/station
5. wrote the letter/story/book at the mansion/warehouse/mall
copied the notes/instructions/essays at the bank/rink/school
6. broke the vases/televisions/windows at the theatre/clubhouse/grocery
ordered the mirrors/dishes/frames at the trailer/garage/arena
7. cleaned the shoes/golfballs/clubs at the jail/bar/cottage
packed the swimsuits/towels/goggles at the library/cave/laboratory
8. listened to the speech/opera/sermon at the playground/college/synagogue
watched the play/choir/debate at the farm/agency/winery

APPENDIX C
Master Experimental Frames of Experiment 3

1. cut—ruby, sapphire, emerald, opal
2. harden—tin, zinc, iron, copper
3. destroy—bomb, cannon, missile, bazooka
4. contaminate—hydrogen, oxygen, nitrogen, helium
5. disrupt—clarinet, trombone, drum, violin
6. buy—francs, rubles, marks, pesos
7. exterminate—beetles, spiders, bees, wasps
8. follow—submarine, destroyer, tugboat, yacht

(Manuscript received January 30, 1987;
revision accepted for publication July 20, 1987.)