Journal of Experimental Psychology:
Learning, Memory, and Cognition
1997, Vol. 23, No. 4, 898–914

0278-7393/97/$3.00

Context Memory and the Selection of Frequency Estimation Strategies

Norman R. Brown
University of Alberta

When people estimate event frequency, they sometimes retrieve and count event instances. This study demonstrates a direct relation between the use of these enumeration-based strategies and the contents of memory. In 3 experiments, participants studied target–context word pairs, estimated presentation frequency for target words, and recalled context words. Study time, target–context relatedness, and study-phase instructions were manipulated, producing large differences in memory for context words. When context memory was best, estimation time increased sharply with presentation frequency, and the steepness of this estimation time–presentation frequency function decreased with context memory. These results indicate that enumeration was common only when context memory was good, that encoding factors determine how frequency is represented, and that the contents of memory restrict strategy selection.

People use a variety of strategies to estimate event frequency. Under some conditions, *enumeration* is common; enumeration-based estimates are produced when a person retrieves and counts relevant event instances and uses the derived count as the basis for a frequency judgment (Barsalou & Ross, 1986; Begg, Maxwell, Mitterer, & Harris, 1986; Blair & Burton, 1987; Brown, 1995; Bruce, Hockley, & Craik, 1991; Burton & Blair, 1991; Conrad & Brown, 1994; Conrad, Brown, & Cashman, in press; Greene, 1989; Williams & Durso, 1986). When people do not enumerate, they rely on *memory assessment* and *direct retrieval* strategies. A memory assessment strategy has two components: First, some aspect of memory is evaluated (e.g., similarity between a probe and the contents of memory, see Hintzman, 1988; Jones & Heit, 1993; and Nosofsky, 1988; ease of retrieval, see Tversky & Kahneman, 1973; trace strength, see Hintzman, 1969, 1970; and Morton, 1968) and used as an index of relative event frequency; second, this index is converted to a numerical value by mapping it onto a subjective response range (Brown, 1995; Conrad et al., in press; Hintzman, 1988). When people use enumeration-based and memory assessment strategies, information about frequency is inferred from some aspect of memory that represents frequency in an indirect manner (e.g., number of retrievable traces, similarity, etc.). People can also represent frequency information directly by storing and updating facts that code frequency of occurrence in a numerical or nonnumerical format (e.g., "The word *BOOK* appeared on the prior list six times," or "The word *BOOK* appeared on the prior list quite a few times"; Alba, Chromiak, Hasher, & Attig, 1980; Brooks, 1985; Blair & Burton, 1987; Conrad et al., in press; Jonides & Jones, 1992; Menon, 1993; Menon, Raghubir, & Schwarz, 1995; Underwood, 1969; Watkins & LeCompte, 1991). When such facts are retrieved and evaluated, people are said to be using a direct retrieval strategy.

Strategy selection is not arbitrary, nor is it without consequence. It is obvious that the contents of memory restrict strategy selection. For example, direct retrieval strategies are ruled out when facts about event frequency have not been stored in memory. Likewise, enumeration seems unlikely when memories for individual event instances are inaccessible. It is also becoming clear that event properties and encoding factors determine the nature of the frequency-relevant contents of memory. In particular, there is now evidence that direct coding of event frequencies is more common when event instances occur at regular intervals than when they do not (Conrad et al., in press; Menon, 1993), and that retrievable traces are more common when event instances are distinctive than when they are similar to one another (Brown, 1995; Conrad et al., in press; Menon, 1993). Finally, strategy selection can affect the effort required to generate a frequency estimate and the magnitude and accuracy of the estimate produced. This can be seen when enumeration-based responses are compared with non-enumerated responses; other things being equal, the former take longer to produce and may be smaller in magnitude and more accurate than the latter (Brown, 1995).

In brief, it is likely that event properties and encoding factors determine how frequency is represented, that the nature of the frequency representation influences strategy selection, and that strategy selection is related to the speed and accuracy of a response. Taken together, these claims constitute a *multiple-strategy perspective* on event frequency. The three experiments reported below were designed to provide evidence for this perspective by documenting a direct connection between encoding variables, the frequency-relevant contents of memory, strategy selection,

This research was supported by Operating Grant 138076 from the Natural Sciences and Engineering Research Council of Canada. I thank Aaron Kucher and the Freqs for their assistance in conducting this study. In addition, I thank Ann Bostrom, Fred Conrad, Alinda Friedman, and Judy Goodman for their helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Norman R. Brown, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9. Electronic mail may be sent via Internet to nbrown@psych.ualberta.ca.

and estimation performance. Specifically, this research demonstrates that stimulus and instructional factors that make event instances more memorable also increase the use of enumeration-based strategies, and that use of these strategies affects both response times and the magnitude of the estimates produced.

The relation between event memory and enumeration has previously been investigated, though indirectly (Barsalou & Ross, 1986; Bruce et al., 1991; Greene, 1989; Lewandowsky & Smith, 1983; Tversky & Kahneman, 1973; Williams & Durso, 1986). Typically, researchers have approached this issue by exposing people to a word list comprising exemplars drawn from distinct taxonomic categories. After the study phase, these people judge how many members of the test categories were included in the study list and are then asked to recall as many of the previously presented items as they can. This procedure is often extended by manipulating factors known to affect cued recall (e.g., study time, typicality, generation, blocking, spacing). If people do enumerate, and if the magnitude of a frequency estimate is proportional to the number of items retrieved, then conditions that produce the best memory for the contents of the study list should also produce the largest frequency estimates. As predicted, the largest and, in some cases, the most accurate estimates are often observed when event memory is the best. This relation between memory and estimation has been taken as evidence for the use of enumeration-based strategies.

There are two problems with this conclusion. First, as Watkins and LeCompte (1991) pointed out, the fact that event memory and judged frequency are affected in the same way by the same variables does not mean that people recall event instances when estimating event frequency. There are alternative ways to interpret this finding. For example, one could adopt the view that people never enumerate, but that factors that promote accurate memory also create robust, stable memory traces. Other things being equal, these memorable traces should be more accessible, feel more familiar, and/or resemble probe items more closely than the weak or degraded traces encoded under less advantageous conditions and, hence, should yield larger estimates. The second problem concerns the assumption that larger estimates necessarily reflect greater success in retrieving relevant event instances. Recent findings have indicated that enumeration-based estimates are smaller, sometimes much smaller, than nonenumerated estimates (Brown, 1995; Burton & Blair, 1991; Conrad & Brown, 1994). Thus, in the absence of an independent indication that instances are retrieved, a relation between memory performance and estimate size is ambiguous; it may indicate that people in one group were more successful at retrieving relevant event instances or that they tended to rely on nonenumeration strategies.

As were the studies cited above, the present research was concerned with the issue of event memory and enumeration and used a research strategy that involved manipulating memory variables and observing their effect on judged frequency and cued recall. However, this study differed from earlier ones in two important respects. First, participants were timed as they generated their estimates because enumeration is characterized by a unique, readily identifiable response time profile (Brown, 1995; Conrad et al., in press). This makes it possible to determine whether a given condition promotes enumeration, without having to accept questionable assumptions concerning the relation between cued recall, estimate size, and strategy selection. Second, the prediction here was not that "any variable that facilitates cued recall should lead to higher categorical frequency estimates" (Green, 1989, p. 235), or that "any factor that affects the memorability of instances of an event will have an effect on frequency estimations of that event" (Williams & Durso, 1986, p. 388). Rather, this research evaluated how memory affects strategy use. In particular, it investigated the hypothesis that people are more likely to use enumeration-based strategies when event memory is good than when it is poor and are more likely to rely on nonenumeration strategies when event memory is poor than when it is good. This hypothesis gave rise to two specific predictions: first, that the function relating response times to presentation frequency should be steeper when event memory is good than when it is poor; second, that participants should tend to underestimate event frequency, except when event memory is so poor that enumeration is rarely attempted.

Previous research (Brown, 1995) provided the basis for both predictions. In this study, verbal protocols (Experiment 1) and response times (Experiments 2 and 3) were collected as participants judged the frequency of target words presented in either variable or consistent contexts. In both conditions, participants studied word pairs consisting of a target word (always a category label; e.g., *MAMMAL*) and the context words (always a category exemplar; e.g., *dog*). The context word changed from one presentation of the target word to the next in the variable-context condition (e.g., *MAMMAL–dog, MAMMAL–tiger, MAMMAL–horse, . . .*) but not in the consistent-context condition. Instead, each target word was paired with the same context word on each presentation (e.g., *MAMMAL–dog, MAMMAL–dog, MAMMAL–dog, . . .*).

The protocol data indicated that participants in the variable-context group often enumerated (i.e., retrieved and counted context words) and that participants in the consistent-context group did not. Studies by Conrad and Brown (1994) and by Marx (1985) provided evidence that the relation between context variability and enumeration demonstrated by these verbal reports is a general one. Conrad and Brown (1994) replicated these protocol results by using a study list that differed in both content and presentation frequency from the one used by Brown (1995). In particular, the target words, which never appeared more than 12 times, always referred to different classes of consumer products, and the context words always referred to brand names (e.g., *NEWSPAPER–New York Times, AUTOMOBILE–Chevrolet,* etc.). Marx (1985) collected retrospective written protocols rather than concurrent verbal reports and manipulated the nature of the encoding task rather than the nature of the context word. Participants in one group were required to generate a unique sentence each time a repeated target word appeared, whereas participants in a second group were required to judge whether the repeated target words were common or uncommon. Consistent with the results reported in Brown (1995) and Conrad and Brown (1994), 77% of the generation participants claimed that they enumerated, whereas only 11% of the judgment participants made the same claim. In other words, enumeration was common only when the semantic context differed from one presentation of a target word to the next. Given the range of stimuli, encoding tasks, and response modes involved in these studies, it is highly unlikely that the relation between context variability and enumeration they illustrate is an artifact of some particular experimental detail (e.g., the target words assigned to different levels of presentation frequency or the specific context words selected to accompany the target words). Rather, these studies support the conclusion that context variability, regardless of how it is achieved, is at least a necessary condition for enumeration (also see Blair & Burton, 1987; Conrad et al., in press; Menon, 1993).

Response times reported in Brown (1995; Experiments 2 and 3) provided converging evidence for this conclusion. Specifically, response times increased steeply with presentation frequency in the variable-context condition but not in the consistent-context condition. In other words, participants in the variable-context group responded more rapidly to target words that had been presented only a few times than to those that had been presented many times, whereas participants in the consistent-context group responded rapidly to all target words regardless of how often they had been studied. This Context × Presentation Frequency interaction for response times was observed in two experiments and was interpreted as follows. First, it was assumed that participants in the response time experiments used the same strategies as their counterparts in the protocol experiment and, hence, that variable-context participants often enumerated. Second, it was assumed that enumeration involves the serial retrieval of event instances and that the number of instances retrieved prior to a response was related to the number of instances studied. Given that the retrieval process can access only one event instance at a time (Bousfield & Sedgewick, 1944; Gruenewald & Lockhead, 1980; Indow & Togano, 1970), it follows that response times should increase with presentation frequency when participants rely on enumeration-based strategies.

The relatively flat response time functions observed in the consistent-context condition were explained by assuming that participants who received the consistent-context study list relied on nonenumeration strategies. In general, people who do not enumerate either retrieve a value from memory (i.e., use a direct retrieval strategy) or determine the target item's relative frequency and then convert this qualitative value to a numerical one (i.e., use a memory assessment strategy). Because these strategies do not involve the iterative retrieval of event instances, they can be executed rapidly, and the time required to arrive at an estimate should be unrelated to presentation frequency. Thus, when people rely on nonenumeration strategies, response times should be fast and little affected by presentation frequency.

This earlier research provided good evidence that people use both enumeration-based and nonenumeration strategies and demonstrated that conditions that foster enumeration yield steep response time functions and that conditions that hamper it yield very shallow ones. In the present study, the steepness of the response time function was expected to be directly related to memory for event instances, with conditions leading to the best event memory producing the steepest function, and those leading to the worst producing the shallowest. This prediction assumed that strategy selection is influenced by memory for relevant event instances and that enumeration-based strategies would be selected less often as instances become more difficult to retrieve. In other words, the percentage of slow enumeration-based responses should be the highest when event memory is best and should drop off as it worsens. This would produce the predicted relation between response time and context memory.

Event memory was predicted to be related to the magnitude of the frequency estimates, as well as to the time taken to produce them, with participants underestimating event frequency when event memory is moderate to good, but not when it is poor. Like the response time prediction, results reported in Brown (1995) provided a basis for this somewhat counterintuitive prediction (cf. Greene, 1989; Williams & Durso, 1986). In this earlier study, variable-context participants, who often enumerated, tended to underestimate event frequencies, and consistent-context participants, who relied on nonenumeration strategies, tended to overestimate them (at least when an upper bound for the response range was not provided).

Underestimation observed in the variable-context condition was considered to be a direct consequence of enumeration. When people enumerate, they often forget or fail to retrieve all relevant instances, and they are prone to extrapolate conservatively when they attempt to adjust enumeration-based counts to compensate for retrieval failure. These tendencies work together, yielding estimated frequencies that are typically smaller than actual frequencies. Moreover, even when people do not always enumerate, the use of enumeration on some trials may lead to underestimation. As mentioned above, most nonenumeration strategies involve the mapping of a qualitative or relative value onto a response range. When no range information is provided, people appear to use numbers generated by enumeration-based responses to induce its properties (Brown & Siegler, 1993, 1996). Because enumeration-based estimates tend to be conservative, the upper bound of the inferred response range is likely to be as small or smaller than the upper bound of the stimulus range. In turn, estimated frequency should be as small if not smaller than actual frequency when people map a relative value (produced by a nonenumeration process) onto a restricted response range. As a result, people who enumerate on some trials, even a few, should also tend to underestimate event frequencies when they use nonenumeration strategies. Thus, if context memory and enumeration are related, and enumeration-based responses are used to induce the properties of the response range, then conditions that promote moderate to good recall of event instances should produce a pronounced and equivalent underestimation bias. This bias should be mitigated only when context memory is so poor that people rarely, if ever, enumerate.

Overview of the Experiments

The experiments described below were designed to determine whether factors that increase the memorability of event instances would also increase the use of enumeration. In all three, participants studied lists of word pairs. Each pair consisted of a target word and a context word. The target words were repeated a variable number of times across the list, whereas each context word appeared only once. As a result, each word pair represented a unique stimulus combination. The study phase was followed by a frequency test, which was followed by a cued-recall test. During the frequency test, participants were timed as they estimated how many times each target word had been presented, and during the (untimed) cued-recall test, they attempted to recall all of the context words that had been paired with each of the target words.

To assess the relation between context memory and enumeration, target–context relatedness and study times were varied across experiments and study-phase instructions were manipulated within experiment, but between subjects. Specifically, in Experiment 1, participants studied *related context* word pairs for 6.0 s each; in Experiment 2, participants had only 2.0 s to study each related-context word pair; and in Experiment 3, participants studied unrelated word pairs for 6.0 s each.[1] In the related-context list, each target word was a category label, and each context word was a category exemplar. (This list was identical to the one studied by people in the variable-context groups in a previous study; Brown, 1995.) The unrelated list contained the same target words as the related list, but the context words were selected so that there was no obvious association between a given context word and the target word with which it was paired (e.g., *MAMMAL–lid, MAMMAL–paper, MAMMAL–ring,* . . .).

Data were collected from three groups of participants in each experiment: a context memory group, a frequency group, and a general memory group. Prior to the study phase, participants in the context memory group were informed that a cued-recall test would follow the study phase and were advised to commit each target–context pair to memory. Frequency group participants were told about the frequency test and were encouraged to pay careful attention to how often each target word was presented. General memory participants were instructed to study the word pairs in preparation for a memory test, but the nature of the test was left unspecified.

Study time and context type were expected to affect memory for context words. Specifically, memory should be better for 6-s lists than for 2-s lists because participants have more time to develop, elaborate, and rehearse associations between the target and context words (Bugelski, 1962; Cooper & Pantle, 1967). Likewise, context memory should be better for related context words (i.e., category exemplars) than for unrelated context words because participants can take advantage of the prior association between target and context words to facilitate encoding, and because they can use categorical knowledge to guide retrieval and to generate candidate responses (e.g., Bower, Clark, Lesgold, & Winzenz, 1969; Mandler, 1967). Data consistent with both expectations have been reported in prior frequency studies. Lewandowsky and Smith (1983) and Williams and Durso (1986) found that recall for category exemplars improved as study time increased, and Begg et al. (1986) found that same-category context words were recalled better than unrelated context words.

Instructions were also expected to affect context memory. This expectation was based on the assumption that memory for context words would be related to the effort exerted to encode them. In addition, it was assumed that context memory participants would use their study time attempting to relate the context word to the target word in a memorable manner, that frequency participants would spend much of their effort monitoring target-word frequency, and that general memory participants might divide their attention, in some cases focusing on the target word and in others on the target–context pair. It follows that memory for context words should be best in the context memory condition and worst in the frequency condition. Of interest, the current study appears to be the first to compare directly the memorial consequences of studying a categorical frequency list under general memory and frequency instructions, and the only one to use context memory instructions.

It is worth noting that frequency judgments produced by participants receiving general memory instructions are often similar to those produced by participants receiving frequency instructions (Attig & Hasher, 1980; Flexser & Bower, 1974; Greene, 1984; Howell, 1973b; Naveh-Benjamin & Jonides, 1986; Rose & Rowe, 1976; Zacks, Hasher, & Sanft, 1982).[2] Some researchers (e.g., Hasher & Zacks, 1979, 1984) have taken this as evidence that frequency information of some sort accrues when people attend to repeated events and that this information reflects presentation frequency with the same degree of faithfulness regardless of the type of study-phase instructions they receive. However, it is also true that different strategies operating on different representations can, under some conditions, yield identical estimates (e.g., Brown, 1995, Experiment 3). Thus, the failure to find an instruction-type effect does not preclude the possibility that study-phase instructions affect how people process repeated events, how they represent event frequency, or how they generate their frequency judgments.

In summary, study time, target–context relatedness, and study-phase instructions were varied with the expectation that these factors, alone and in combination, would affect memory for context words. It was necessary to generate different levels of context memory to evaluate the hypoth-

[1] The low level of recall observed in Experiment 3 made it unnecessary to conduct the obvious fourth experiment, one in which participants studied unrelated pairs for 2.0 s each.

[2] But see Alba et al., 1980, Experiment 2; Hasher & Chromiak, 1977, Experiment 1; and Williams & Durso, 1986, Experiment 2. In these experiments, participants receiving frequency (i.e., "intentional") instructions judged event frequencies more accurately than did those receiving general memory (i.e., "incidental") instructions.

esis that memory for event instances (which, in this case, is equivalent to memory for context words) is related to the use of enumeration-based estimation strategies. If this hypothesis is correct, the steepness of the response time function should be directly related to context memory; this function should be steepest when context memory is best and shallowest when it is worst. In addition, there should be a relation between estimation bias and context memory, with participants tending to underestimate event frequencies, except when context memory is very poor.

Experiment 1

In this experiment, participants received context memory, general memory, or frequency instructions. They next studied a list of target–context word pairs, estimated the presentation frequency of the target words, and finally attempted to recall the context words. With the exception of the final cued-recall test, the general memory condition is identical to the variable-context condition used in an earlier study (Brown, 1995). Verbal protocols described in that study (Experiment 1) indicated that participants often used enumeration-based strategies (57%) but that nonenumeration strategies were also quite common (43%). If instructions would affect context memory as expected (i.e., context memory participants would recall more context words than would general memory participants, who would recall more than frequency participants), and if context memory and enumeration were related as expected, then context memory participants should enumerate the most and frequency participants should enumerate the least. This implies that the function relating response times to presentation frequency should be steepest in the context memory condition and shallowest in the frequency condition. If all participants enumerate on some trials, event frequencies should typically be underestimated and the degree of underestimation should be unrelated to the study-phase instructions.

Method

Materials. Participants studied a list of 260 word pairs. The first four word pairs presented in the study list served as a primacy buffer and the last four served as a recency buffer. Each buffer pair consisted of a target item (always a category label) and a context item (always a category exemplar) that were not repeated elsewhere in the list and were not used as probes during frequency and cued-recall tests.

Each of the 252 word pairs allocated to the body of the study list also consisted of a target word that identified a taxonomic category and a context word that identified a member of that category (e.g., *SPORT–swimming, COUNTRY–Cuba, FISH–guppy*). Because each target item was paired with a different context item on each presentation, and each context word was presented only once, each target–context pair was unique. Target words appearing in the body of the list were presented either 2, 4, 8, 12, or 16 times, with six target words assigned to each level of presentation frequency.

The words used as target and context items were drawn from published category norms (Battig & Montague, 1969; McEvoy & Nelson, 1982). These norms made it possible to identify categories that could be described by a single noun (e.g., *METAL, FLOWER, OCCUPATION*) and that included a reasonable number of frequently listed category members. These one-word category names served as target items, and frequently listed one- and two-syllable category members as context items. Each category was assigned to a single level of presentation frequency depending on the availability of suitable context items.

A unique study list was created for each general memory participant and was presented to that participant and to yoked participants from the context memory and frequency groups. In all cases, the target–context pairs were assigned positions such that the repeated presentations of a given target word were evenly distributed across the list, and such that target items repeated at a given level of presentation frequency were not over or underrepresented in any portion of the list. In addition, each study list had the target items arranged in a unique random order, and each used a different random ordering of the context items (for details, see Brown, 1995).

Test lists were composed of the 30 target items and six category labels that did not appear in the study list. The latter served as 0-frequency catch trials. A different test list was constructed for each yoked triad in the following manner. First, the list was divided into six blocks, with one target item from each frequency level (0, 2, 4, 8, 12, 16) randomly assigned to each block. Then, the target words were randomly ordered within blocks. A similar scheme was used to create the test booklets for the cued-recall test.

Procedure. Prior to the presentation of the study phase, participants in all groups were told that they would see 260 word pairs, that each pair would consist of a target word and a context word, and that the target words but not the context words would be repeated. In addition, all participants were instructed to study the pairs for a later memory test, though only context memory and frequency participants were informed of the type of test they would receive. Specifically, the cued-memory task was described to participants in the context memory group, and these participants were advised to commit each target–context pair to memory. Participants in the frequency groups were informed that knowledge of target-word frequency would be tested and were advised to pay close attention to how often each target word appeared in the study list.

During the study phase, the target–context pairs were displayed one at a time, for 5.5 s on a computer-controlled video monitor. The target item always appeared in the center of the screen in uppercase letters, and the context item always appeared two lines beneath in lowercase letters. After the 5.5-s study interval, the screen was erased and remained blank (except for markers indicating the screen positions of the target word and context word, and a trial counter) for 0.5 s.

Following the study phase, participants were instructed that they would be presented with 36 category names and that they would be required to estimate as accurately as possible the number of times each had appeared in the previous list. Participants were informed that decision times would be collected. They were also warned that some test items did not appear during the study phase and in this way were informed of the lower bound of the response range. The upper bound of the response range, however, was left unspecified.

The participant initiated each trial by pressing the enter key on the computer keyboard. This caused a target word to appear in the center of the computer display. The participant was required to decide how many times it had appeared in the study list and to press the keyboard's space bar just as soon as he or she had "a single numerical response in mind," but not before. When the space bar was pressed, a response field appeared two lines beneath the test word. At this point, the participant entered an estimate at the keyboard and pressed the enter key. The test word, response field, and frequency estimates were then erased and replaced by a message prompting the participant to initiate the next trial.

Each trial yielded a *decision time*, an *initiation time*, and an *entry time*. The decision time interval began with the presentation of the test word and ended when the participant pressed the space bar. This interval was particularly important because it indicated how long it took the participant to formulate an estimate. The initiation interval began as soon as the space bar was pressed and ended as soon as the first digit was typed, and the entry interval began when the initiation interval terminated and ended when the enter key was pressed. In combination, initiation and entry time indicated the time required to enter an estimate once it had been formulated.

The first six trials were treated as practice trials. During these trials, the experimenter sat with the participant and made sure he or she understood the task and the test procedure.

A cued-recall test followed the frequency test. Each participant was given a booklet that listed the 36 target words, 12 to a page. Participants were reminded that most of the target words had been paired with multiple context words during the study phase, and they were instructed to recall as many of these context words as they could. Specifically, participants were asked to work through the booklet one target word at a time at their own pace and to respond by listing all the context words that were paired with it during the study phase.

Participants. Ninety undergraduates were recruited from the University of Alberta subject pool. Thirty of these were randomly assigned to each group. Participants were tested individually in sessions lasting about 50 min, and they received course credit for their cooperation.

Results

Cued recall. For each participant and cuing target word, cued recall was scored by using both lenient and strict criteria. When the lenient criterion was applied, all responses were summed to produce a *total recall* score. When the strict criterion was applied, only words that had been paired with the cuing target word during the study phase were summed into a *correct recall* score. Only total recall scores are reported below. There were two reasons for this. First, the two measures produced an identical pattern of effects. Second, total recall seemed a better indicator of subjective list knowledge because it reflects both true and false beliefs about the list, whereas correct recall reflects only true beliefs (Bruce et al., 1991; Williams & Durso, 1986).

A mean total recall measure was computed for each participant and each level of presentation frequency. These means were submitted to an Instructions (context memory vs. general memory vs. frequency) × Frequency (0, 2, 4, 8, 12, 16) analysis of variance (ANOVA), where the former was treated as a between-subjects variable and the latter as a within-subjects variable.

Mean total recall is plotted against presentation frequency in Figure 1 (left panel), and percentage recall for each group is listed in Table 1. As predicted, study-phase instructions affected context memory, $F(2, 87) = 5.12$, $MSE = 4.65$. (Unless otherwise noted, the significance level was set at .05 for all analyses reported in this article.) On average, the context memory participants recalled 115.3 words (equal to 46% of the 252 context words associated with the target words); the general memory participants recalled 100.8 words (40%); and the frequency participants recalled 89.1 words (35%). In addition, there was a main effect of frequency, $F(5, 435) = 437.21$, $MSE = 0.83$, indicating that more words were recalled when the target word was presented with many context words than when it was presented with few. Finally, there was a reliable Instructions × Frequency interaction, $F(10, 435) = 4.18$, $MSE = 0.83$, indicating that the effect of instruction type was most pronounced at higher presentation frequencies.

Decision times. During the frequency task, each response period was divided between a decision time, an initiation time, and an entry time. These three measures were added together to produce a *total time* measure, which

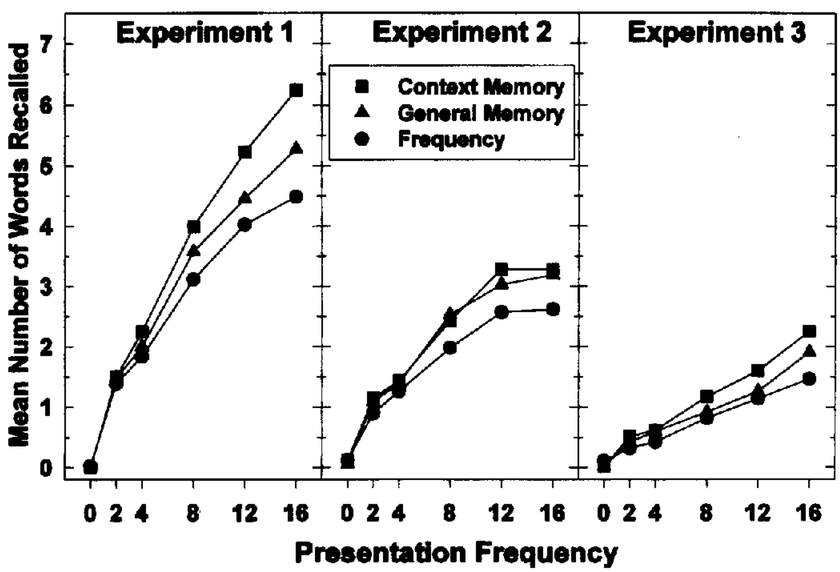

Figure 1. Mean number of words recalled as a function of presentation frequency for context memory, general memory, and frequency groups, in Experiments 1, 2, and 3.

Table 1
Mean Percentages Recalled, Response Times (in Seconds), Regression Slopes, and Correlations for Experiments 1–3

	Percentage recalled			Response time			Slope[a]			Correlation[b]		
Instructions	Exp. 1	Exp. 2	Exp. 3	Exp. 1	Exp. 2	Exp. 3	Exp. 1	Exp. 2	Exp. 3	Exp. 1	Exp. 2	Exp. 3
Context	46	28	15	6.3	4.9	3.7	0.59	0.55	0.81	.93	.86	.83
General	40	27	12	4.8	5.4	4.5	0.64	0.55	0.96	.93	.83	.83
Frequency	35	23	10	4.3	4.0	4.2	0.86	0.76	1.08	.93	.84	.88
M	40	26	12	5.1	4.8	4.1	0.70	0.62	0.95	.93	.84	.85

Note. Exp. = experiment.
[a]Estimated against actual frequency. [b]Back-transformed; estimated with actual frequency.

indicated how long it took to generate and enter a frequency estimate. Thus, there were 4 response time measures per trial. Because the first 6 trials were treated as a practice block, response times and estimates collected during these trials were discarded. The remaining 30 trials included five items at each level of presentation frequency. For each participant and each response time measure, a mean and a median were computed for each level of presentation frequency. These measures were submitted to separate Instructions × Frequency ANOVAs, although only analyses performed on decision time means are presented in detail below. There are, however, a few points concerning these auxiliary analyses that are worth noting. First, for all measures, means and medians were similar in magnitude and displayed the identical pattern of effects. Second, participants typically initiated and entered their responses rapidly. In the present experiment, the mean initiation and entry times were 0.8 s and 0.7 s, respectively. Third, both initiation and entry time increased slightly (no more than 0.6 s) across the range of presentation frequencies. Fourth, the total time analysis displayed the same pattern of effects as the decision time analysis. Finally, the preceding points are valid not only for the current experiment, but for two prior experiments (Brown, 1995, Experiments 2 and 3) and for the two experiments that follow.

Mean decision time is plotted against presentation frequency in Figure 2 (left panel). As this figure suggests, response times increased with presentation frequency, $F(5, 435) = 53.48$, $MSE = 9.14$, and both the main effect of instruction type, $F(2, 87) = 3.34$, $MSE = 56.69$, and the Instruction × Frequency interaction, $F(10, 435) = 3.25$, $MSE = 9.14$, were reliable. This interaction is particularly noteworthy because it indicates that the response time function was the steepest when context memory was best (i.e., in the context memory condition) and the shallowest when it was the worst (i.e., in the frequency condition). This is consistent with the hypothesis that enumeration-based estimates become less common as context retrieval becomes more difficult.

Frequency estimates. As with the response times, the first 6 estimates produced by a participant were discarded, and statistics were derived from the remaining 30 responses. For each participant and level of target frequency, a mean frequency estimate and a mean absolute error (i.e., |esti-

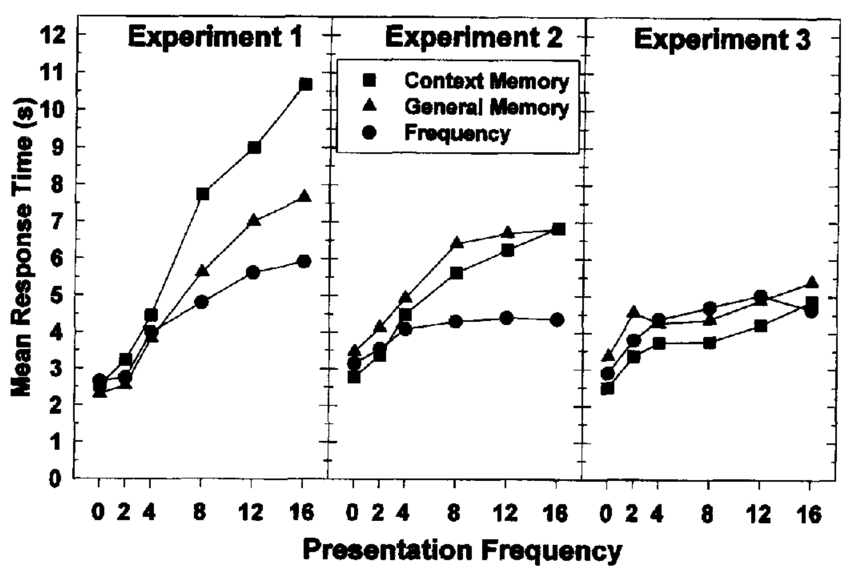

Figure 2. Mean decision time as a function of presentation frequency for context memory, general memory, and frequency groups, in Experiments 1, 2, and 3.

mated frequency − actual frequency|) were computed.[3] Two additional measures were computed for each participant. The first was the slope obtained by regressing estimated frequency against actual frequency. These slopes provided a measure of estimation bias, with slopes greater than 1.0 indicating a tendency to overestimate event frequencies and slopes less than 1.0 indicating a tendency to underestimate them. The second was the rank order correlation between estimated and actual frequency. These correlations provided a measure of how well participants discriminate between the frequency levels, independent of their ability to accurately induce the metric properties of the response range (Brown & Siegler, 1993; Flexser & Bower, 1974; Naveh-Benjamin & Jonides, 1986). The estimate means and absolute error means were submitted to separate Instructions × Frequency ANOVAs, and the slopes and correlations to separate one-way (Instructions) ANOVAs.[4]

Estimated frequency means are presented in Table 2. These data indicate that participants in all groups tended to underestimate event frequencies, but not to the same extent. Specifically, estimates produced by frequency participants were less biased than those produced by the context memory and general memory participants. Consistent with this observation, the Instructions × Frequency interaction was significant, $F(10, 435) = 5.50$, $MSE = 5.09$; as were the two main effects for instructions, $F(2, 87) = 6.00$, $MSE = 17.09$, and for frequency, $F(5, 435) = 327.70$, $MSE = 5.09$. Regression slopes reflected the same pattern of relative underestimation; the mean regression slope was 0.86 for the frequency condition, 0.64 for the general memory condition, and 0.59 for the context memory condition, $F(2, 87) = 6.26$, $MSE = 0.103$. A set of least significant difference tests confirmed that the frequency condition slopes differed from the context memory and general memory conditions, which did not differ from each other.

Estimates produced by context memory and general memory participants were more biased than those produced by participants in the frequency group, but not less accurate. Evidence for this can be found in Table 1, which lists the mean rank order correlations, and in Table 2, which lists absolute error means. The mean correlations listed in Table 1 (.93 for each group) make it clear that participants had an accurate understanding of relative event frequency and that this understanding was unaffected by study-phase instructions. Instructions also had no effect on absolute error (see Table 2), as neither the main effect of instructions, $F(2, 87) = 1.30$, $MSE = 4.53$, nor the Instructions × Frequency interaction, $F(10, 435) = 1.06$, $MSE = 1.69$, was significant. There was, however, a reliable main effect of frequency, $F(5, 435) = 378.81$, $MSE = 1.69$, indicating that estimates for high-frequency items were less accurate than estimates for low-frequency items.

Discussion

In the current experiment, study-phase instructions affected participants' memory for the context words, the speed with which they estimated event frequencies, and the magnitude of the estimates they produced. As expected, cued recall was best following context memory instructions and worst following frequency instructions. Likewise, the response time function was steepest in the context memory condition and shallowest in the frequency condition. These results are consistent with the notion that people use both enumeration-based and nonenumeration estimation strategies and that they are more likely to enumerate when event memory is good than when it is not. The general tendency for participants to underestimate event frequencies in this experiment is also consistent with the notion that strategy selection and context memory are related (Brown, 1995). As noted above, underestimation appears to be a necessary consequence of enumeration both because forgetting may lead to conservative enumeration-based estimates and because these estimates may be used to infer a conservative response range.

Although underestimation was expected in this experiment, the tendency for frequency participants to produce less biased estimates than context memory and general memory participants was not. There are two plausible explanations for this: an averaging explanation and a frequency-coding explanation. The averaging explanation assumes that some frequency participants ignored the context words during the study phase. As a result, these participants would not be able to enumerate during the frequency test and, instead, would have to rely on nonenumeration strategies. Brown (1995) has found that participants who do not enumerate often adopt a response range that is considerably wider than the stimulus range, which leads to overestimation. Assuming some participants enumerated and others relied on nonenumeration strategies, the relatively steep regression slope observed in the frequency condition could have been produced by averaging the shallow slopes associated with the former and the steep slopes associated with the latter.

The second explanation assumes that frequency participants, unlike (or to a greater extent than) context memory and general memory participants, attended to event frequency during the study phase. Although it seems unlikely that they would have been able to maintain an accurate running tally for all target words, they may have attempted to keep track of a few of the most commonly presented items (Alba et al., 1980; Flexser & Bower, 1974; Zacks et al., 1982). If so, these participants could have used these counts to accurately define an upper bound for the response range and may have extrapolated more aggressively when they did enumerate.

There is evidence that argues against the averaging position and that favors the frequency-coding position. First, the averaging position predicts that estimated frequency should be inversely related to response time: On average, partici-

---

[3] In this experiment and following ones, medians were also computed and analyzed. These data are not reported because, in all cases, means and medians were similar in size and displayed the identical pattern of effects.

[4] All correlations were transformed by using Fisher's r-to-z method before being submitted to statistical tests, and a back-transformation was used to compute the correlation means reported below (Silver & Dunlap, 1987).

Table 2
Mean Estimated Frequencies and Absolute Errors for Levels of Presentation Frequency for Experiments 1–3

	Experiment 1			Experiment 2			Experiment 3		
Freq.	Con.	Gen.	Freq.	Con.	Gen.	Freq.	Con.	Gen.	Freq.
				Estimated freq.					
0	0.1	0.0	0.1	0.1	0.1	0.2	0.2	0.3	0.2
2	2.0	1.9	2.0	1.8	2.0	2.2	3.0	3.0	3.3
4	3.6	3.3	3.8	3.2	4.2	4.2	6.2	5.6	5.9
8	5.8	5.7	6.9	6.1	6.3	7.5	9.8	10.3	9.5
12	7.9	7.6	9.7	7.3	6.7	10.5	11.3	12.5	15.2
16	9.7	10.8	14.5	9.1	9.7	12.1	13.7	15.9	17.3
M	4.8	4.9	6.2	4.6	4.9	5.2	7.4	7.9	8.6
				Absolute error					
0	0.1	0.0	0.1	0.1	0.1	0.2	0.2	0.3	0.2
2	0.5	0.7	0.6	1.1	1.4	1.7	2.2	2.2	2.3
4	1.0	1.3	1.4	2.0	2.9	3.0	3.8	3.5	3.5
8	2.5	3.0	2.9	3.9	3.9	3.9	5.4	5.6	4.3
12	4.8	5.1	4.6	6.2	5.9	5.4	6.2	7.0	7.8
16	6.7	7.5	6.4	8.4	8.8	7.0	8.4	8.1	7.2
M	2.6	2.9	2.7	3.6	3.8	3.5	4.4	4.5	4.2

Note. Freq. = frequency; con. = context; gen. = general.

pants who enumerate should respond slowly and underestimate, and those who rely on nonenumeration strategies should respond rapidly and overestimate. The data provide no support for this prediction; the correlation between mean estimated frequency and mean decision time computed over the 30 frequency participants was .12. Comparable correlations computed over the context memory and general memory participants were .13 and −.06, respectively (in all cases, $p > .1$). Second, the averaging hypothesis predicts a bimodal distribution of slopes in the frequency condition, with slopes obtained from the enumerating participants forming the lower mode and those obtained from nonenumerating participants forming the higher one. There was also no support for this prediction because a Lilliefors test indicated that this distribution of slopes did not differ significantly from normality. Evidence in support of the frequency-coding position came from a questionnaire administered to all participants following the cued-recall test. Of the frequency participants, 93% claimed that they attempted to keep track of how frequently the target words were presented, whereas only 43% of the general memory participants and 30% of the context memory participants made the same claim. Experiment 2 provides additional evidence for the frequency-coding explanation.

Experiment 2

Experiment 2 was identical to Experiment 1, except that participants had only 2 s to study each word pair rather than 6 s. Because recall is often related to study time (Bugelski, 1962; Cooper & Pantle, 1967; Williams & Durso, 1986), it was expected that context memory would be worse in this experiment than in the last and that as a result, enumeration-based estimates would be less common and response time functions less steep. As in Experiment 1, participants were expected to underestimate event frequencies, and on the basis of the results of Experiment 1, this tendency was expected to be greater in the context memory and general memory conditions than in the frequency condition. This prediction assumed that differences in frequency coding explained between-group differences in estimation bias observed in Experiment 1. However, if the averaging explanation was correct and context memory was worse in this experiment than the last, then participants in all groups should be less biased to underestimate event frequencies than their Experiment 1 counterparts.

Method

Design, materials, and procedure. With one exception, the design, materials, and procedure were identical to those used in Experiment 1. In this experiment, each word was displayed for 1.5 s, followed by a 0.5-s intertrial interval. Thus, participants had 2.0 s, rather than 6.0 s, to study each word pair.

Participants. Ninety participants were recruited from the University of Alberta subject pool and randomly assigned to the three study-phase instruction groups. Participants were tested individually in sessions lasting about 35 min, and they received course credit for their participation.

Results

Recall data, response times, and estimation data were summarized and analyzed as in Experiment 1.

Cued recall. The total recall data are presented in Figure 1 (center panel) and in Table 1. In some respects, these data resemble those obtained in Experiment 1, though as expected, reducing study time impaired memory for context words. As before, the main of effect of instructions was significant, $F(2, 87) = 3.12$, $MSE = 2.30$, and frequency participants recalled the least ($M = 23\%$) and context memory participants the most (M = 28%). However, unlike Experiment 1, recall following general memory instructions (M = 27%) was almost as good as recall following context memory instructions. Frequency had a large effect on the number of words recalled, $F(5, 435) = 273.50$, $MSE = 0.45$, and interacted marginally with instructions, $F(10, 435) = 1.73$, $p < .1$, $MSE = 0.45$.

Decision times. Assuming that context memory and enumeration are related, the cued-recall data imply that response time functions should be shallower in this experiment than in the last, that the shallowest function should appear in the frequency condition, and that response times should be similar in the context memory and general memory conditions. Means plotted in Figure 2 (center panel) are consistent with these implications. It is obvious that participants in this experiment responded more rapidly than their counterparts did in Experiment 1 (cf. Figure 2, left panel; also see Table 1). It is also clear that response time functions produced by context memory and general memory participants were similar to one another and that they were steeper than the one produced by the frequency participants. This pattern of means produced a reliable Instructions × Frequency interaction, $F(10, 435) = 2.41$, $MSE = 4.78$. The main effect of frequency, $F(5, 435) = 26.30$, $MSE = 4.78$, was also significant, though the main effect of instructions was not, $F(2, 87) = 1.88$, $MSE = 50.80$.

Frequency estimates. As in the last experiment, estimated frequency tended to be smaller than actual frequency, particularly for high-frequency items, and this tendency was greater for context memory and general memory participants than for frequency participants. This difference was reflected in the regression slopes that were steeper in the frequency condition (M = 0.76) than in either the context memory condition (M = 0.55) or the general memory condition (M = 0.55), $F(2, 87) = 5.16$, $MSE = 0.08$ (see Table 1). This difference is also apparent in the pattern of estimate means displayed in Table 2. The ANOVA performed on these means indicated that the Instructions × Frequency interaction, $F(10, 435) = 3.48$, $MSE = 5.87$, the instructions effect, $F(2, 87) = 3.12$, $MSE = 37.82$, and the frequency effect, $F(5, 435) = 229.33$, $MSE = 5.87$, were all reliable.

Although study-phase instructions affected response times and estimation bias, they had only a limited effect on absolute accuracy and no effect on relative accuracy. Specifically, there was a reliable Instructions × Frequency interaction for absolute error, $F(10, 435) = 2.42$, $MSE = 3.08$, indicating that frequency participants estimated the values of high-frequency items more accurately than participants did in other groups (see Table 2). As before, the main effect of frequency, $F(5, 435) = 249.82$, $MSE = 3.08$, was significant, indicating that absolute error increased with event frequency, and the main effect of instruction type, $F(2, 87) < 1.0$, was not significant. In this experiment, participants were good at discriminating levels of presentation frequency, though not as good as in Experiment 1; the correlation means were .86, .83, and .84, for the context memory, general memory, and frequency groups, respectively, $F(2, 87) < 1.0$.

Discussion

Reducing study time from 6 s to 2 s affected both context memory and response times. As expected, participants recalled fewer context words in this experiment than in the last, and they responded more rapidly. Study-phase instructions also affected these measures; as in Experiment 1, frequency participants recalled fewer context words and responded more rapidly than context memory and general memory participants. These relations between encoding factors (study time and study-phase instructions), context memory, and response time provide support for the view that participants used both enumeration-based and nonenumeration estimation strategies and that strategy selection is related to the availability of relevant event instances.

The pattern of estimation bias was the same in this experiment as in the last. Again, participants tended to underestimate event frequencies, and this tendency was more pronounced for context memory and general memory participants than for frequency participants. In addition, participants in the present experiment produced estimates that were as small if not smaller than those produced by their counterparts in Experiment 1 (Table 2). As in Experiment 1, it appears that context memory and general memory participants enumerated at least some of the time and that they used their conservative enumeration-based estimates to establish a relatively narrow response range. It also appears that frequency participants rarely enumerated and that they kept count of the frequency of some items and used these counts to select a boundary for the response range that was close to the upper bound of the stimulus range.[5] Note that these results are incompatible with the averaging proposal described above that incorrectly predicted that regression slopes should have been steeper in the present experiment than in the last.

The current experiment produced one unexpected result. In contrast to Experiment 1, cued recall was no better in the context memory condition than in the general memory condition. This suggests that context memory participants in this experiment did not have the time required to elaborate or rehearse target–context associations in a way that would produce superior memory performance.

Experiment 3

Experiment 3 was identical to Experiment 1 except that target items were paired with random nouns rather than category exemplars during the study phase. It was expected that context memory would be much worse in this experiment than in Experiment 1. After all, participants were given only one trial to learn more than 250 unrelated paired associates. Poor context memory should compel participants

[5] The pattern of moderate underestimation combined with a very shallow response time function is identical to one observed in a previous experiment (Brown, 1995, Experiment 3). In this experiment, participants studied consistent-context word pairs and were informed by the experimenter that the upper bound of the response range was 16.

to rely on nonenumeration strategies. If so, response times should be fast, response time functions flat, and underestimation less pronounced in this experiment than in Experiments 1 and 2.

Method

Design, materials, and procedure. With one exception, the design, materials, and procedure were identical to those used in Experiment 1. In this experiment, category exemplars were not used as context words. Instead, each context word was replaced by a one- or two-syllable word that had no obvious relation to its target word (e.g., *INSTRUMENT–fence, CURRENCY–claw, APPLIANCE–jury,* . . .) and that was not an exemplar of any other category described by a target word.

Participants. Ninety participants were recruited from the University of Alberta and randomly assigned to the three study-phase instruction groups. Participants were tested individually in sessions lasting about 50 min, and they received course credit for their cooperation. One person in the general memory group was eliminated because his mean estimate was 3 *SD*s greater than the mean for his group, and his regression slope was 5 *SD*s greater.

Results

Recall data, response times, and estimation data were summarized and analyzed as in Experiments 1 and 2.

Cued recall. As expected, participants had difficulty remembering the context words. Mean total cued recall was 15%, 12%, and, 10% for context memory, general memory, and frequency conditions, respectively. Although, the main effect of instructions was only marginally significant, $F(2, 86) = 2.39$, $p < .1$, $MSE = 1.90$, both the main effect of frequency level, $F(5, 430) = 128.50$, $MSE = 0.31$, and the Instructions × Frequency interaction, $F(10, 430) = 2.37$, $MSE = 0.31$, were reliable. These effects indicate that participants recalled more context words when cued with frequently presented target words than with rarely presented target words and that this tendency was most pronounced in the context memory condition and least pronounced in the frequency condition (see Figure 1, right panel).

Decision times. The cued-recall data indicate that all participants had limited access to relevant event instances. It follows that enumeration-based estimates should be uncommon in all conditions and, hence, that response time functions should be shallow and should resemble one another. The decision time means plotted in Figure 2 (right panel) bear out these predictions. In contrast to the prior experiments, neither the main effect of instructions, $F(2, 86) = 1.01$, $MSE = 24.68$, nor the Instructions × Frequency interaction, $F(10, 430) = 1.03$, $MSE = 2.24$, was significant. There was, however, a reliable main effect of frequency, $F(5, 430) = 20.57$, $MSE = 2.24$. Averaging across groups, mean decision times for the six levels of presentation frequency (i.e., 0, 2, 4, 8, 12, 16) were 2.9 s, 3.9 s, 4.1 s, 4.3 s, 4.7 s, and 5.0 s. In other words, decision time increased by 1.1 s as presentation frequency increased from 2 to 16. By comparison, response times increased by 7.5 s across the same range in the context memory condition in Experiment 1. Thus, although response times increased slightly, the modest size of this increase, in conjunction with the difficulty participants had recalling context words, make it unlikely that enumeration-based estimates were common.

Frequency estimates. As in Experiments 1 and 2, participants tended to underestimate event frequency, and this tendency was more pronounced in the context memory and general memory conditions than in the frequency condition. Estimates produced in the current experiment diverged from this pattern in two respects. First, estimated frequencies tended to be as large if not larger than actual frequencies. Second, study-phase instructions did not have a consistent effect on estimation bias. The tendency to produce relatively unbiased estimates is evident in the means presented in Table 2. Averaging across conditions, estimate means for the six frequency levels (0, 2, 4, 8, 12, 16) were 0.2, 3.1, 5.9, 9.9, 13.0, and 15.6, $F(5, 430) = 179.97$, $MSE = 17.34$, and the average regression slope was .95. Although the Instructions × Frequency interaction (see Table 2) was significant, $F(10, 430) = 1.91$, $MSE = 112.80$, it was considerably smaller in this experiment than in Experiments 1 and 2. Of more importance, instructions did not have a reliable effect on the regression slopes, $F(2, 86) = 1.74$, $MSE = 0.31$ (see Table 1), or on the overall magnitude of the estimates, $F(2, 86) < 1.0$ (see Table 2).

In the current experiment, only presentation frequency had a reliable effect on absolute error (see Table 2), $F(5, 430) = 79.25$, $MSE = 9.48$; for instructions, $F(2, 86) < 1.0$; for Instructions × Frequency, $F(10, 430) = 1.0$. Although study-phase instructions did not affect absolute accuracy, they did have a reliable effect on relative accuracy. Specifically, the mean rank order correlation between estimated and actual frequency was larger in the frequency condition ($M = .88$) than in the context memory or general memory conditions ($M = .83$, in both cases); $F(2, 86) = 3.43$, $MSE = 0.08$.

Discussion

In this experiment, cued recall was very poor, response time functions were shallow, frequency judgments were relatively unbiased, and study-phase instructions had little if any effect on response times or context memory. This pattern indicates that enumeration-based strategies were rarely, if ever, used and is consistent with the position that instance memory and strategy selection are related. In this case, it appears that context memory was so poor that most participants were unable or unwilling to enumerate on most trials. This suggests that context memory must exceed some threshold before enumeration is perceived to be an effective estimation strategy and that below this threshold, event retrieval may be too difficult or unreliable to support enumeration. The current experiment demonstrates that subthreshold recall levels can occur even when people have studied a list of unique event instances and, in so doing, indicates that it is the accessibility to event instances rather than their distinctiveness per se that determines whether an enumeration-based strategy will be used (cf. Brown, 1995; Conrad et al., in press; Menon, 1993).

It is interesting that the mean rank order correlation between estimated and actual frequency was highest in the frequency condition, indicating that frequency participants developed a better sense of relative event frequency than did context memory or general memory participants. This finding is important because it provides additional evidence that warning people about an upcoming frequency test can, under some conditions, affect their knowledge of event frequency (cf. Hasher & Zacks, 1984). Of course, study-phase instructions and relative accuracy were not related in Experiments 1 or 2. Apparently, deliberate encoding strategies can only be used effectively when people have sufficient time to execute them (cf. Experiment 2), and the effects of intentional encoding can only be observed when performance is not at ceiling in the incidental-learning condition(s) (cf. Experiment 1).

General Discussion

Context Memory and Response Times

The primary aim of the present study was to establish a relation between context memory and strategy selection. Specifically, this research was designed to determine whether people are more likely to enumerate when instance memory is good than when it is not. To evaluate this hypothesis study-phase instructions, study time, and target–context relatedness were manipulated, and response times, frequency estimates, and cued-recall data were collected. The expectations were that instructional differences and stimulus properties would affect context memory, that enumeration-based estimates would be most common when context memory was the best, and that such estimates would become less common as encoding factors reduced participants' memory for context words. These expectations, in conjunction with prior findings that related response times to strategy usage, gave rise to the prediction that the function relating response times to presentation frequency would be the steepest when context memory was the best and that these functions would flatten out as context memory declines.

The analyses presented above demonstrate that response times and context memory are related. However, because only study-phase instructions were manipulated within experiment, it was not possible to examine this relationship across the full range of recall performance. Nonetheless, data plotted in Figures 1 and 2 suggest that the steepness of the response time functions and level of context memory were related across experiments as well as within experiment. This point is made more clearly in Figure 3. In this figure, a measure of the steepness of the response time function, called the *response time difference,* is plotted against mean probability of recall for each group and each experiment. The response time difference was computed by subtracting the average decision time for the low-frequency items (i.e., target words presented 2 and 4 times) from the average decision time for the high-frequency items (i.e., target words presented 12 and 16 times). This measure controls for incidental between-group differences, while

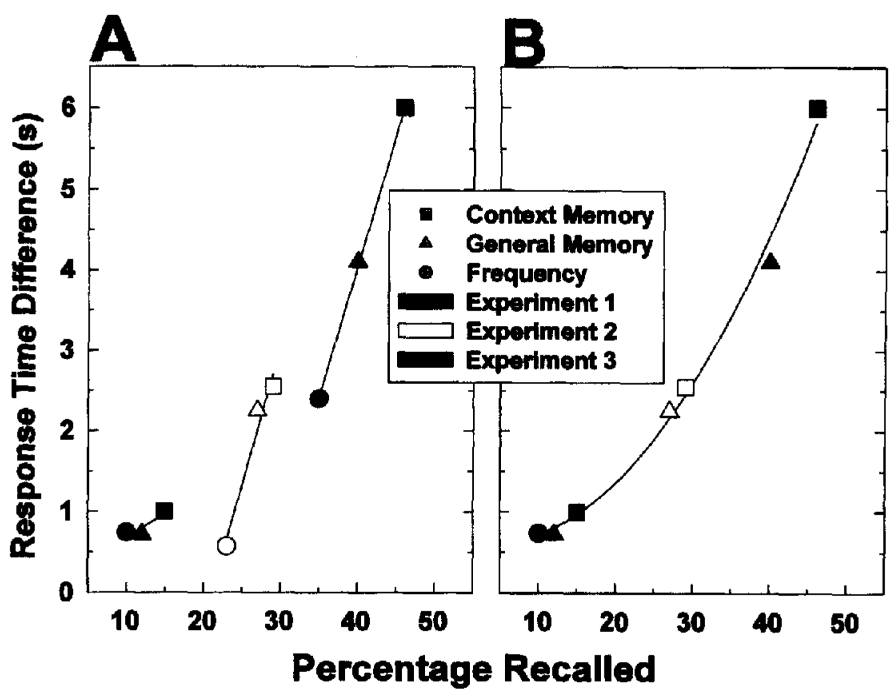

Figure 3. Response time difference versus percentage recalled for each study-phase instruction group and each experiment. The straight lines plotted in Figure 3A indicate the best within-experiment linear fit. The curved line plotted in Figure 3B indicates the best quadratic fit across experiments, excluding data from the frequency groups in Experiments 1 and 2.

providing an indication of the degree to which estimates for high-frequency items tended to be slower than estimates for low-frequency items. Regardless of the exact shape or intercept of the response time function, the response time difference should be large when the function increases steeply with presentation frequency, and small when it does not.[6]

A number of things should be noted about Figure 3. First, there were large differences in both context memory and in the steepness of the response time function, with probability of recall ranging from 10% to 46% and response time differences ranging from 0.6 s to 6.0 s. Second, within experiment, there was a strong tendency for response time differences to increase with recall, particularly in Experiments 1 and 2, in which the linear relation between recall and response time difference was virtually perfect (see Figure 3A). Third, across experiments, recall was a very good predictor of response time differences; when the response time differences were regressed against mean recall, a linear model accounted for 83% of the variance, and a quadratic model accounted for 94%.

An examination of the data plotted in Figure 3A suggests that one reason fits were not better was that the response time differences obtained for the frequency conditions in Experiments 1 and 2 were smaller than might be expected given the availability of context words. This can be seen by comparing the Experiment 1 frequency condition with the Experiment 2 context memory and general memory conditions, and by comparing the Experiment 2 frequency condition with all conditions in Experiment 3. In both cases, recall was considerably better in the frequency condition than in conditions yielding comparable response time differences. The relatively flat response time functions observed in the frequency conditions in Experiments 1 and 2 can be explained as follows. In principle, retrieving a single number directly from memory should be easier and faster than retrieving and counting multiple instances (Menon, 1993). As a result, people should prefer direct retrieval to enumeration-based strategies when they have access to both a prestored frequency count and retrievable event instances. In other words, it may be that the use of enumeration-based strategies not only increases with the availability of retrievable event instances but that it also decreases with the availability of retrievable frequency counts. Thus, holding context memory constant, response times should be faster when people have access to prestored frequency counts than when they do not.

Differences in the magnitude of the frequency estimates observed in Experiments 1 and 2 indicated that participants in the frequency groups directly coded and later retrieved frequency counts, at least some of the time. This implies that the response time functions observed in these conditions should be shallow relative to the functions produced by participants who did not directly encode frequency counts, but who had a reasonably good memory for event instances (i.e., those in the context memory and general memory groups in Experiments 1 and 2). It also correctly implies that recall should be more predictive of enumeration in the context memory and general memory conditions than in the frequency condition because prestored counts should only be available to compete with retrieval traces in the frequency conditions. When data from the frequency conditions in Experiments 1 and 2 were excluded and response time differences were again regressed against recall, the linear fit accounted for 95% of the variance, and the quadratic fit accounted for 99% (see Figure 3B).

In brief, the present analysis demonstrates the existence of a strong systematic relation between context memory and response time. This relation is interpreted as indicating that people are more likely to enumerate when event instances are readily available than when they are not. In addition, the nonlinear aspect of this function suggests that there may be a threshold for enumeration and that enumeration is rarely, if ever, attempted when event memory falls below this threshold. Finally, relatively small response time differences observed for the frequency conditions in Experiments 1 and 2 suggest that people may directly code frequency information when they are expecting a frequency test and that direct retrieval strategies compete with enumeration strategies when such information is available.

Conclusion

Psychologists have long speculated that multiple formats may be used to encode information about event frequency and that multiple strategies may be used to generate frequency judgments (e.g., Hasher & Zacks, 1984; Hintzman, 1976; Howell, 1973a; Johnson, Raye, Wang, & Taylor, 1979; Jonides & Naveh-Benjamin, 1987). Yet, it is only recently that researchers have demonstrated the existence of multiple frequency representations and estimation strategies (Blair & Burton, 1987; Brown, 1995; Bruce & Van Pelt, 1989; Burton & Blair, 1991; Conrad & Brown, 1996; Conrad et al., in press; Means & Loftus, 1991; Menon, 1993). The multiple strategy perspective that has emerged from this

[6] The response time difference was one of two enumeration indices evaluated. The other was the rank order correlation between estimated frequency and response time. This correlation should be strongest when people rely heavily on enumeration-based strategies and weakest when enumeration is never used. One correlation was computed for each participant over the 30 target words. Then a mean correlation was taken for each condition over all participants. Not surprisingly, these means differed widely across the nine experimental conditions, with the context memory group in Experiment 1 producing the strongest average correlation ($M = .69$), and the frequency participants in Experiment 2 producing the weakest ($M = .20$). Of more importance, these means correlated almost perfectly with the response time differences ($r = .96$). Thus, either index could have been used to demonstrate a relation between strategy selection and context memory. In this context, it is worth noting that the rank order correlation between response time and percentage recall was also computed for each participant over the 30 target words and then averaged over all participants, within conditions. These averages ranged from .62 for the context memory group in Experiment 1 to .09 for the frequency group in Experiment 2 and correlated almost perfectly with the average correlation between estimated frequency and response time ($r = .98$). In other words, when response times were related to estimated frequency, they were also related to context memory.

research holds that encoding factors influence the way information about event frequency is represented, that strategy selection is restricted by the task-relevant contents of memory, and that different estimation strategies have distinctive behavioral consequences.

The research described in this article provides support for the multiple strategy perspective by demonstrating that study time, target–context relatedness, and study-phase instructions affected context memory and that context memory and response time were related. As predicted, conditions yielding the best context memory also produced the steepest response time functions, and those yielding the worst produced the shallowest. Taking the steepness of these functions as an index of enumeration, this pattern indicates that enumeration-based strategies were executed most often when event instances were most accessible and that the tendency to enumerate decreased as event instances became more difficult to retrieve. Conversely, these findings indicate that participants relied on nonenumeration strategies when context memory was poor, but not when it was good. Estimated frequency and event memory were also related in a way that is consistent with a multiple strategy interpretation. As predicted, participants tended to underestimate event frequencies, except when context memory was very poor. An underestimation bias has been associated with enumeration in prior studies (Brown, 1995; Burton & Blair, 1991; Conrad & Brown, 1994) and is thought to occur because people often fail to retrieve all relevant events and because conservative enumeration-based estimates may lead people to induce a response range that is as narrow or narrower than the stimulus range.

In brief, the pattern of response times and estimation bias observed in this study indicates that people enumerate some of the time, that the tendency to enumerate is determined, in part, by the availability of relevant event instances, and that the availability of event instances is influenced by a variety of encoding factors. The present study also indicates that access to event instances is not the only factor that determines whether people enumerate. The results from the frequency conditions in Experiments 1 and 2 are of particular relevance to this claim. In both cases, the response time functions were shallower than would have been expected if context memory were the only factor determining a preference for enumeration, and in both cases, frequency estimates were less biased than those produced by participants receiving context memory or general memory instructions. These results can be explained by assuming that participants in the frequency group were more likely to keep track of event frequency than were participants in other groups and that people may prefer to rely on prestored frequency information, when it is available, rather than to retrieve and count available event instances.

Availability of event instances and the presence of a competing source of frequency-relevant information are not the only factors that determine whether an enumeration-based strategy will be used. For example, there is evidence in the *behavioral frequency* literature that time restrictions reduce the use of enumeration.[7] In one study (Burton & Blair, 1991, Study 2), undergraduate business majors were given from 10 s to 70 s to estimate the number of courses taken outside of the business school. Retrospective written protocols were used to determine how these estimates were generated. Students in the 10-s group (61%) were much less likely to enumerate than students in the 70-s group (92%), and they provided less accurate estimates (also see Williams & Durso, 1986, Experiment 2). This indicates that the presence of retrievable instances does not automatically trigger enumeration; if it did, enumeration-based responses would have been as common in the 10-s condition as in the 70-s condition.

Taken together, these findings provide the basis for informed speculation about the nature of strategy selection in this task. In particular, it seems that people prefer concrete information (in the form of instance counts or directly coded frequency facts) to vague information (in the form of intuitions supplied by memory assessment processes) and that they are willing to work quite hard to accumulate the former, provided they have enough time to do so. However, when two sources of information are equally credible (e.g., enumerated counts and directly coded frequency facts), people prefer the more convenient source. Thus, perceived accuracy and convenience may be evaluated when an estimation strategy is selected, with accuracy given more weight when one source of information is clearly more credible and convenience given more weight when competing sources are considered equally credible, or when the more credible strategy is deemed to be too demanding.

This article has focused on response times and, to a lesser extent, on patterns of estimation bias because these data provide information about people's estimation strategies. Measures of absolute and relative accuracy were also reported above but were not discussed in detail because they reveal little about strategy selection. Nonetheless, it is worth noting that encoding factors did affect estimation accuracy. Specifically, relative accuracy (i.e., the rank order correlation between estimated and actual frequency) was better in Experiment 1 (related context, 6-s presentation) than in Experiment 2 (related context, 2-s presentation) or Experiment 3 (unrelated context, 6-s presentation), and study-phase instructions did affect relative accuracy in one experi-

[7] Researchers who study behavioral frequency estimation are concerned with understanding how people estimate the frequency of real-world events or activities (e.g., "How many times have you gone shopping for groceries in the past month?"). This research indicates that people estimate behavioral frequencies and list frequencies in much the same way. For example, enumeration-based estimates are common when people are asked to estimate the frequency of distinctive, hence memorable, real-world events, and general impression strategies are often used when the to-be-estimated events are mundane, and hence difficult to recall. People also use direct-retrieval strategies when responding to behavioral frequency questions, though these take the form of rate-based responses (e.g., "I buy groceries twice a week, so I have gone shopping eight times in the past month"). Rate-based strategies are most commonly used when the target event happens on a regular basis (Blair & Burton, 1987; Bruce & Van Pelt, 1989; Burton & Blair, 1991; Conrad & Brown, 1996; Conrad et al., in press; Means & Loftus, 1991; Menon, 1993; Menon et al., 1995).

ment (Experiment 3).[8] These differences provide evidence that knowledge of event frequency can be affected by encoding factors and instructional manipulations (cf. Barsalou & Ross, 1986; Greene, 1984, 1986; Hanson & Hirst, 1988; Hasher & Zacks, 1979, 1984; Jonides & Naveh-Benjamin, 1987; Naveh-Benjamin & Jonides, 1986; Rose & Rowe, 1976; Rowe, 1974; Williams & Durso, 1986; Zacks, Hasher, & Sanft, 1982). However, they do not indicate that superior knowledge of relative event frequency was always related to the nature of the study-phase instructions, or to the time allotted for studying word pairs, or to the strength of the target–context association. Moreover, neither context memory nor the response time difference measure was strongly predictive of relative accuracy.

Although no simple generalization captures the pattern of differences displayed in Table 1, it is interesting that a high level of accuracy was achieved under all conditions and that the differences across conditions were quite modest. These facts, in conjunction with the evidence that different conditions promote the use of different estimation strategies, indicate that a variety of mechanisms are capable of generating accurate frequency judgments and that different strategies, operating on different types of task-relevant information, can produce comparable levels of estimation accuracy (Brown, 1995). In other words, one estimation strategy is not necessarily better than another, and each is capable of producing an accurate assessment of relative frequency, provided the available information accurately reflects the relative frequency of the target items.

In retrospect, it is not surprising that estimated and actual frequency were highly correlated in the present study and in a previous one (Brown, 1995). After all, in both studies and in all conditions, the word pairs were studied under full attention, the study phase was followed immediately by the test phase, and presentation frequencies increased from 2 to 16 in a roughly geometric progression. It is likely that participants would have been less accurate in their assessment of event frequency if any one of these factors had been different. Both divided attention during the study phase (Greene, 1984, 1986; Naveh-Benjamin & Jonides, 1986) and delayed testing have been shown to reduce the accuracy of frequency estimates (Hintzman & Stern, 1984; Underwood, Zimmerman, & Freund, 1971). Apparently, full attention facilitates the encoding of frequency-relevant information, regardless of its form, and forgetting caused by delay makes this information less reliable, accessible, or predictive (Hintzman, 1988). There is also evidence that people are better able to discriminate between levels of event frequency when the levels are widely spaced than when they are close together (Hintzman & Gold, 1983; Hintzman, Grandy, & Gold, 1981). This implies that one could affect the correlation between estimated and actual frequency by manipulating the frequency levels represented in the study list; other things being equal, relative accuracy should be good when the distance between levels is large and poor when it is not. As yet, the effects of divided attention, study–test delay, and list composition on strategy selection have not been investigated. However, to the extent that these factors interfere with the encoding, retrieval, and/or the evaluation of individual event instances or frequency counts, memory assessment strategies should be used more frequently and enumeration-based and direct retrieval strategies less frequently.

In summary, the present research indicates that people tend to enumerate when they have ready access to retrieval event instances, and it suggests that people use direct retrieval strategies when facts about event frequency are available and that they fall back on memory assessment strategies when enumeration and direct retrieval strategies cannot be used effectively. This research has also begun to identify the laboratory conditions that produce good context memory and that promote the encoding of frequency-relevant facts. As a result, it is possible to anticipate when people are likely to use enumeration-based and direct retrieval strategies. Moreover, to the extent that conditions associated with enumeration and direct retrieval are well understood, it should be possible to create conditions in which people rely exclusively on memory assessment strategies. This is of interest for two reasons. First, memory assessment appears to be the most widely applicable of the frequency estimation strategies, yet the most difficult to understand. Second, it is possible that the processes that inform the memory assessment strategies may be the ones that mediate frequency effects in perception, language processing, and judgment and decision making. Thus, there are good reasons for wanting to study memory assessment strategies in isolation, and there are now also good reasons for believing this can be done.

[8] It is also true that there were between-experiment differences in absolute error. Overall, absolute error was smaller in Experiment 1 ($M = 2.7$) than in Experiment 2 ($M = 3.6$), and smaller in Experiment 2 than in Experiment 3 ($M = 4.4$). However, measures of absolute accuracy reveal less about knowledge of event frequency than do measures of relative accuracy because absolute error is determined jointly by beliefs about the response range and by beliefs about the relative frequency of the target items (Brown, 1995; Brown & Siegler, 1993).

References

Alba, J. W., Chromiak, W., Hasher, L., & Attig, M. (1980). Automatic encoding of category size information. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 370–378.

Attig, M., & Hasher, L. (1980). The processing of frequency of occurrence information by adults. *Journal of Gerontology, 35,* 66–69.

Barsalou, L. W., & Ross, B. H. (1986). The role of automatic and strategic processing in the sensitivity of superordinate and property frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 116–134.

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, Pt. 2).

Begg, I., Maxwell, D., Mitterer, J. O., & Harris, G. (1986). Estimates of frequency: Attribute or attribution? *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 12, 496–508.

Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research, 14,* 280–288.

Bousfield, W. A., & Sedgewick, C. H. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology, 30,* 149–165.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorical word lists. *Journal of Verbal Learning and Verbal Behavior, 17,* 573–587.

Brooks, J. E. (1985). Judgments of category frequency. *American Journal of Psychology, 98,* 363–372.

Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1539–1553.

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review, 100,* 511–534.

Brown, N. R., & Siegler, R. S. (1996). Long-term benefits of seeding the knowledge-base. *Psychonomic Bulletin & Review, 3,* 385–388.

Bruce, D., Hockley, W. E., & Craik, F. I. M. (1991). Availability and category-frequency estimation. *Memory & Cognition, 19,* 301–312.

Bruce, D., & Van Pelt, M. (1989). Memories of a bicycle tour. *Applied Cognitive Psychology, 3,* 137–156.

Bugelski, B. R. (1962). Presentation time, total time, and mediation in paired-associate learning. *Journal of Experimental Psychology, 63,* 409–412.

Burton, S., & Blair, E. (1991). Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly, 55,* 50–79.

Conrad, F. G., & Brown, N. R. (1994). Strategies for estimating category frequency: Effects of abstractness and distinctiveness. In *American Statistical Association, Proceedings of the section on survey methods research* (pp. 1345–1350). Alexandria, VA: American Statistical Association.

Conrad, F. G., & Brown, N. R. (1996). Estimating frequency: A multiple strategy perspective. In D. Herrmann, M. Johnson, C. McEvoy, C. Hertzog, & P. Hertel (Eds.), *Basic and applied memory: Research on practical aspects of memory* (pp. 167–178). Hillsdale, NJ: Erlbaum.

Conrad, F. G., Brown, N. R., & Cashman, E. (in press). Strategies for estimating behavioral frequency in survey interviews. *Memory.*

Cooper, E. H., & Pantle, A. J. (1967). The total time hypothesis in verbal learning. *Psychological Bulletin, 68,* 221–234.

Flexser, A. J., & Bower, G. H. (1974). How frequency affects recency judgments: A model for recency discrimination. *Journal of Experimental Psychology, 103,* 706–716.

Greene, R. L. (1984). Incidental learning of event frequency. *Memory & Cognition, 12,* 90–95.

Greene, R. L. (1986). The effects of intentionality and strategy on memory for frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 489–495.

Greene, R. L. (1989). On the relationship between categorical frequency estimation and cued recall. *Memory & Cognition, 17,* 235–239.

Gruenewald, J. P., & Lockhead, G. R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 225–240.

Hanson, C., & Hirst, W. (1988). Frequency encoding of token and type information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 289–297.

Hasher, L., & Chromiak, W. (1977). The processing of frequency information: An automatic process? *Journal of Verbal Learning and Verbal Behavior, 16,* 173–184.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39,* 1372–1388.

Hintzman, D. L. (1969). Apparent frequency as a function of frequency and the spacing of information. *Journal of Experimental Psychology, 80,* 139–145.

Hintzman, D. L. (1970). Effects of repetition and exposure duration on memory. *Journal of Experimental Psychology, 83,* 139–145.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10, pp. 47–91). New York: Academic Press.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multi-trace memory model. *Psychological Review, 95,* 528–551.

Hintzman, D. L., & Gold, E. (1983). A congruity effect in the discrimination of presentation frequencies: Some data and a model. *Bulletin of the Psychonomic Society, 21,* 11–14.

Hintzman, D. L., Grandy, C. A., & Gold, E. (1981). Memory for frequency: A comparison of two multiple-trace theories. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 231–240.

Hintzman, D. L., & Stern, L. D. (1984). A comparison of forgetting rates in frequency discrimination and recognition. *Bulletin of the Psychonomic Society, 22,* 409–412.

Howell, W. C. (1973a). Representation of frequency in memory. *Psychological Bulletin, 80,* 44–53.

Howell, W. C. (1973b). Storage of events and event frequencies: A comparison of two paradigms in memory. *Journal of Experimental Psychology, 98,* 260–263.

Indow, T., & Togano, K. (1970). On retrieving sequence from long-term memory. *Psychological Review, 77,* 317–331.

Johnson, M. K., Raye, C. L., Wang, A. Y., & Taylor, T. H. (1979). Fact and fantasy: The role of accuracy and variability in confusing imaginations with perceptual experiences. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 220–240.

Jones, C. M., & Heit, E. (1993). An evaluation of the total similarity principle: Effects of similarity on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 799–812.

Jonides, J., & Jones, C. M. (1992). Direct coding for frequency of occurrence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 368–378.

Jonides, J., & Naveh-Benjamin, M. (1987). Estimating frequency of occurrence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 230–240.

Lewandowsky, S., & Smith, P. W. (1983). The effect of increasing memorability of category instances on estimates of category size. *Memory & Cognition, 11,* 347–350.

Mandler, G. (1967). Organization and memory. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1, pp. 327–372). New York: Academic Press.

Marx, M. H. (1985). Retrospective reports on frequency judgments. *Bulletin of the Psychonomic Society, 23,* 309–310.

McEvoy, C. L., & Nelson, D. L. (1982). Category names and instance norms for 106 categories of various sizes. *American Journal of Psychology, 95,* 581–634.

Means, B., & Loftus, E. F. (1991). When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology, 5,* 297–318.

Menon, G. (1993). The effects of accessibility of information in memory on judgments of behavioral frequencies. *Journal of Consumer Research, 20,* 431–440.
Menon, G., Raghubir, P., & Schwarz, N. (1995). Behavioral frequency judgments: An accessibility–diagnosticity framework. *Journal of Consumer Research, 22,* 212–227.
Morton, J. (1968). Repeated items and decay in memory. *Psychonomic Science, 10,* 219–220.
Naveh-Benjamin, M., & Jonides, J. (1986). On the automaticity of frequency coding: Effects of competing task load, encoding strategy, and intention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 378–386.
Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54–65.
Rose, R. J., & Rowe, E. J. (1976). Effects of orienting task and spacing of repetitions on frequency judgments. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 142–152.
Rowe, E. J. (1974). Depth of processing in a frequency judgment task. *Journal of Verbal Learning and Verbal Behavior, 13,* 638–643.
Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's *z* transformation be used? *Journal of Applied Psychology, 72,* 146–148.
Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 4,* 207–232.
Underwood, B. J. (1969). Attributes of memory. *Psychological Review, 76,* 559–573.
Underwood, B. J., Zimmerman, J., & Freund, J. S. (1971). Retention of frequency information with observations on recognition and recall. *Journal of Experimental Psychology, 87,* 149–162.
Watkins, M. J., & LeCompte, D. C. (1991). Inadequacy of recall as a basis for frequency knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1161–1176.
Williams, K. W., & Durso, F. T. (1986). Judging category frequency: Automaticity or availability? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 387–396.
Zacks, R. T., Hasher, L., & Sanft, H. (1982). Automatic encoding of event frequency: Further findings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 106–116.

Received January 2, 1996
Revision received October 4, 1996
Accepted October 19, 1996 ■