Estimating Frequency: A Multiple Strategy Perspective

Frederick G. Conrad

Bureau of Labor Statistics

and

Norman R. Brown

University of Alberta

In Herrmann, D.J, McEvoy, C., Hertzog, C., Hertel, P. and Johnson, M.K. (Eds.) (1996) *Basic and Applied Memory Research*: *Practical Applications, Volume 2*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers, pp. 166-179.

Introduction

People are often required to judge the frequency of everyday events. For example, we are likely to consider the frequency of particular disasters when deciding whether to purchase insurance, or the number of phone calls we typically make in order to select appropriate telephone service. In addition, policy makers and market researchers rely on survey data based on people's responses to behavioral frequency questions, *e.g.*, "During the last month how many times did you receive treatment from a medical doctor?" Because frequency judgments are a part of everyday life, they have recently been explored by applied researchers. In particular, studies of how survey respondents answer behavioral frequency questions have identified several, distinct strategies, and some of the conditions under which these strategies are used (*e.g.,* Blair & Burton, 1987; Burton & Blair, 1991; Conrad, Brown & Cashman, 1993; Menon, 1993). Based on this research, it is evident that people rely on multiple strategies in order to perform a particular estimation task.

Basic researchers in psychology have also studied the processes by which people estimate frequency. Laboratory research has identified a wide variety of strategies, but each study has tended to examine a single strategy. For example, Barsalou and Ross (1986) provided evidence that subjects retrieve and count category instances when reporting category frequency -- an "enumeration" strategy; Hintzman (1988) modeled frequency judgments as a comparison of a target item's similarity to exemplars in memory -- a similarity principle; Tversky and Kahneman (1973) demonstrated that people use the ease of retrieving instances as a key to frequency -- the availability heuristic. In this chapter, we will argue that people judge the frequency of everyday occurrences by using whatever information their memories make available to them. Such a multiple strategy approach is plain to see when estimation is observed in its natural context; however, subjects in most laboratory studies exhibit relatively uniform behavior that varies with the experimental task. In this respect, everyday memory researchers -- in particular those studying behavioral frequency judgments -- have led the way in developing a coherent picture of a fundamental, human behavior.

In the following pages, we will attempt to characterize frequency estimation from the *Multiple Strategy Perspective*. This view is based on real-world memory research yet it is testable in the laboratory, generalizes across tasks and situations, and provides theoretical insights that have eluded laboratory researchers. As such, it illustrates the value of moving from natural contexts to the laboratory in developing psychological theory, and serves as a response to those who have assailed the everyday memory approach (cf. Banaji & Crowder, 1989).

The Multiple Strategy Perspective

The primary feature of the Multiple Strategy Perspective is that people can estimate event frequency on the basis of several different types of remembered information. The particular types of information available for a given estimate depend largely on the characteristics of the target event. The estimation strategies associated with the various types of relevant information can have a strong effect on the size and accuracy of the resulting frequency judgments. We have constructed a taxonomy of the various strategies used by people to estimate frequencies in both laboratory and survey contexts. This taxonomy is presented in Figure 1.

Insert Figure 1 about here

In Figure 1, the major branch separates numerical from non-numerical strategies. In general, studies of answering behavioral frequency questions have identified two fundamentally different strategies that rely on numerical information: episode enumeration and rate strategies. These appear in the left branch of the figure. When respondents use enumeration strategies, they simply count retrieved episodes; when they use rate based strategies, they retrieve specific, numerical facts about the regularity with which events occur, for example "I go grocery shopping once a week so in the last month I must have gone four times." Enumeration strategies are inherently numerical because they involve counting retrieved instances or episodes of the target category. Rate strategies are numerical in that they are based on quantitative facts about regularity of occurrence.

When episodes cannot be readily retrieved and rates cannot be inferred, one is unlikely to have quantitative information available for estimating frequency. Under these circumstances, the only available information is non-numerical so one must deploy strategies from the right side of the figure. By our view, all of these strategies involve converting a qualitative or relative sense of frequency into a numerical form; what differs is the source of the non-numerical information. Under certain circumstances, one might store such information, for example, that a particular event has occurred "many times" or "more than" some other event, and then recall it (direct retrieval). In contrast, one might infer qualitative frequency information by evaluating the state of ones memory for the event in question (memory assessment). For example, if one can easily recall instances of the event (availability), or if the target seems familiar (familiarity), one might conclude that its frequency is high.

Because each of these estimation techniques operates on a different type of remembered information, the circumstances under which each is used is constrained by the contents of memory. In the next section we discuss characteristics of events that affect what people remember about events and consequently which strategies they are most likely to use in estimating frequency for particular events.

Event Characteristics and Estimation Strategies

The likelihood that one will enumerate remembered episodes depends on the ease of retrieving those episodes in the first place. One factor that might affect ease of retrieval is the distinctiveness of individual episodes. A number of autobiographical memory studies (e.g., Barclay & Wellman, 1986; Brewer, 1988;

Linton, 1982; Neisser, 1986; Wagenaar, 1986) have demonstrated that events judged to be similar to one another are hard to recall and recognize. Highly similar episodes lead people to represent common features of similar episodes, but not features that distinguish one from the next (Neisser, 1981; Strube 1987). Therefore, enumeration should be rare when instances of a target activity closely resemble one another and should be used more often as the episodes become more distinctive, and hence are easier to retrieve. A related factor is the regularity with which episodes are experienced. If one engages in a particular activity on a regular schedule, its temporal characteristics will be similar from one occurrence to the next, reducing the likelihood that each occurrence is remembered distinctly. What's more, if one is aware of those regular, temporal characteristics, it becomes possible to estimate frequency on the basis of rate information.

The evidence in the behavioral frequency literature supports this view. Means and Loftus (1991) found that respondents were more likely to recall distinguishable episodes (serious medical conditions) than episodes which were similar to one another (minor medical conditions) in answering frequency questions. Menon (1993) found that everyday events such as snacking or bathing were more likely to be enumerated when they had received low similarity ratings than when similarity was judged to be high. In addition, both sets of studies provide evidence that event regularity affects people's response strategies. Means and Loftus found that when people reported frequencies of irregularly occurring doctor visits, such as those due to an injury, they were more likely to recall individual visits than when reporting frequency of visits which occur in fixed intervals, such as visits to an allergist. Menon found an increase in reported use of rate knowledge when regularity ratings were relatively high.

To explore this idea for both numerical and non-numerical strategies, we (Conrad, Brown & Cashman, 1993) asked a national, random sample of survey respondents to answer ten behavioral frequency questions, *e.g.* "During the last month, how many times did you purchase gas for your car?" "During the last month, how many times did you pay to have your car repaired?". They were asked to think quietly as they arrived at a frequency and to report the frequency as soon as they had it in mind. After they answered each question, the respondents were asked to describe their thinking while they were determining the frequency of the target event -- an immediate, retrospective, verbal protocol (Ericsson &

Simon, 1993). The estimation and protocol tasks were tape recorded, enabling us to measure response times (discussed in a later section) and to analyze the protocols for evidence of particular estimation strategies. Following the frequency estimation and protocol tasks, each respondent rated the event categories in the stimulus questions on two dimensions: the *similarity* of individual episodes to one another and their *regularity* of occurrence. Respondents registered their ratings using a four point scale: *Very Similar/Regular* (4) to *Very Different/Irregular* (1).

The most common strategies evident in the protocols were episode enumeration, rate retrieval, rate estimation, rate and adjustment, and general impression. The first four of these are numerical strategies: Enumeration and the three rate-based strategies involved counts or stored numbers. We distinguished rate retrieval from rate estimation because for the former class, respondents mentioned using remembered rate information while for the latter they reported constructing a rate for a portion of the reference period by enumerating for that period -- a hybrid strategy. Protocols were assigned to rate and adjustment when respondents modified a rate-based estimate because of an exception to the usual rate. General impression statements, in contrast, are non-numerical. Protocols were assigned to this class when they included "vague quantifiers" (O'Muircheartaigh & Gaskell, 1994) such as "a lot" or "not too often," or relative frequency statements such as "I use an ATM more than I use an actual teller in the bank."

The use of these strategies varied with similarity, $F(4, 423) = 4.39$, MSE = 3.043 and regularity, $F(4,424) = 17.41$, MSE = 12.853. (Note that all statistical test reported in this chapter are significant beyond the $p < .01$ level). Mean ratings for each strategy are presented in Table 1. Episode enumeration was used when events were judged least similar, that is most distinctive, and least regular. Rate retrieval was used under opposite circumstances. The difference between these ratings is significant by a planned comparison for both similarity, $F(1,423) = 12.54$, MSE = 8.696 and regularity $F(1,424) = 54.85$, MSE = 40.493. This is consistent with the idea that one must be able to retrieve episodes in order to enumerate them; non-distinct events are probably remembered as generic or schematic event categories, making enumeration difficult, if not impossible. Rate estimation was used for activities that people rated moderately similar and regular, as one might expect for a hybrid strategy. General impression responses are also associated with events that received moderate ratings. One interpretation of the general impression responses is that people used non-numerical information when individual episodes were not distinctive enough to be recalled and then counted, and not regular enough for people to have encoded rate information.

Insert Table 1 about here

A final characteristic of the target events that seems related to response strategy is their frequency. Blair and Burton (1987) found that respondents were more likely to enumerate low than high frequency events. Our results are also consistent with these: Strategy use varied with estimated frequency, $F(4,425) = 10.99$, MSE = 1153.089. Mean estimated frequencies appear in the third column of Table 1. When our respondents enumerated, their estimates were smaller than when they retrieved rate knowledge, $F(1, 425) = 16.27$, MSE = 1707.399, by a planned comparison. In fact, episode enumeration led to smaller estimates than those produced by any other strategies, $F(1,425) = 41.18$, MSE = 4321.430, by a Sheffé test. The largest frequency estimates were based on general impression statements, consistent with a finding by Bruce and Read (1988). Presumably the frequent occurrence of these events reduces what is remembered about the contexts of individual episodes, further discouraging enumeration. If an event is experienced frequently, there are more opportunities to directly encode this fact than if the event has a lower frequency. However, what is encoded is unlikely to be precise numerical information but rather a sense that frequency is large.

One implication of these results is that under certain circumstances, people cannot rely on the content of their memories to estimate frequency, in particular when episodes are hard to distinguish, occur irregularly and are rare. Enumeration requires that episodes be distinguishable, rate strategies depend on regularly occurring episodes, and general impressions are most likely to be used when episodes occur frequently. By our analysis, when people cannot judge frequency based on the content of memory, they must rely on memory assessment strategies. Using this approach people can generate information regarding ease of retrieval, familiarity of the target event or similarity of the event to what is remembered regardless of the target's characteristics. Therefore, one might consider memory assessment strategies to be available by default.

Accuracy of Estimation Strategies

Of primary concern to survey practitioners is the accuracy of respondents' estimates based on particular strategies. Unfortunately, the factors that seem to influence accuracy in the literature vary between studies, so it is hard to evaluate their effect. For example, it has been suggested that the length of the reference period -- one week, one month, *etc*. -- affects the accuracy of enumeration-based estimates (Sudman & Bradburn, 1974) and this usually varies between studies. Whether it is due to this variable or others, the accuracy of enumeration-based estimates varies widely. For example, instructions to enumerate can both increase (Means, Swan, Jobe & Esposito, 1994) and decrease (Burton & Blair, 1991) accuracy. In addition, conditions that foster enumeration have led to lower accuracy than those which promote the use of rate strategies. Menon (1993) found highest absolute error for activities rated Dissimilar (that is, distinctive) and Irregular, and smallest absolute error for activities rated Similar and Regular. It seems then that in general one needs to know more than the strategy in use to predict its accuracy; one needs to identify and measure the relevant variables. This is most easily accomplished In the laboratory, where it is straightforward to control the relevant variables including actual frequency.

To examine the accuracy of enumeration versus non-numerical strategies, one of us (Brown, in press) conducted a laboratory study. Instead of being asked about categories of autobiographical events, the subjects in this experiment were presented with category names, such as CITY and MAMMAL, and were questioned about their presentation frequency. Each category name was presented at one of six frequencies (0, 2, 4, 8, 12, 16), and each time a category name was presented it was paired with an instance of that category, for example CITY - Boston. The subjects were instructed to study the presentations for a later memory test, but they were not told this would consist of frequency judgments.

The distinctiveness of the presentations was manipulated between groups. In the different-context condition, a unique category exemplar was presented each time a category name appeared, *e.g.* CITY - Boston, CITY - Chicago, CITY - London, *etc.* In the same-context condition, a single exemplar was presented each time a particular category name appeared, *e.g.* CITY - London, CITY - London, CITY -

London. The Different-context condition created relatively distinctive presentation contexts for each category name; the Same-context condition led to virtually identical presentation circumstances. After the presentation phase, the subjects were shown each category name and asked to estimate its presentation frequency. They were instructed to think aloud while producing a frequency response, that is, to provide a concurrent verbal protocol (Ericsson & Simon, 1993). If distinct episodes are needed in order to enumerate them, then the different-context condition should promote the use of enumeration strategies. On the other hand, if we are correct that people rely on non-numerical information when they can neither retrieve episodes nor apply rate strategies, then the same-context condition should lead them to use such information in their estimation strategies.

The context manipulations seemed to affect strategy use in the expected ways. In the different-context condition, subjects enumerated on 57% of the trials and used general impressions in 20% of their protocols; in the same-context condition, they did not enumerate at all, referred to general impressions in 24% of their protocols and produced uninformative reports on the remaining 66% of the trials. Uninformative protocols would be the likely result -- though not a guarantee -- of using memory assessment strategies because such strategies do not necessarily engage working memory which is a requirement for coherent verbal reports (Ericsson & Simon, 1993).

When accuracy in the two conditions was measured in absolute terms, there was little difference between groups. However, the different-context subjects underestimated and the same-context subjects overestimated: when estimated frequency was regressed onto presentation frequency the slope of the regression line was .73 for the different-context subjects and 1.30 for the same-context subjects, $t(38) = 3.0$; perfect accuracy would have produced a slope of 1.0. If subjects primarily enumerated in the different-context condition, this would have produced the observed underestimation: If one makes any error in this situation, it will be an error of omission; one is unlikely to report an episode that was not presented in the study phase. Overestimation is a sensible result for the use of non-numerical strategies because in the absence of numerical guidelines, estimates can range either above or below true values, but to differing degrees: estimates cannot fall below zero at the low end, but are essentially unbounded at the high end and so can be quite large.

Converging Evidence: Response Times and Differential Boundary Effects

Verbal protocols provide a direct report on subjects' strategies, but our analysis of the protocols can be strengthened by converging evidence based on more objective measures. It is possible to derive differential response time predictions for the different strategies. If enumeration involves a series of retrieval operations, and the number of these operations is related to the number of episodes that are ultimately reported, then response time should increase with reported frequency. In contrast, using rate information should require the same amount of mental activity, regardless of the size of the retrieved rate, so response times should be unrelated to reported frequency. Similarly, response times should not vary with frequency when estimates are based on non-numerical information: it should take no longer to retrieve "a lot" than "hardly ever," nor to compute such impressions via some memory assessment process.

We tested this in our telephone study (Conrad, Brown & Cashman, 1993) by measuring response times from the end of the interviewer's question until the respondent uttered a number, and then regressing response times onto estimated frequency for the major strategies evident in the protocols. The regression lines for three of the strategies are presented in Figure 2. It is important to notice is that for enumeration, response times increase sharply with frequency (slope=.84) but there is no relationship between response time and frequency for rate retrieval (slope=.02) and general impression (slope=.01) strategies. These effects, which were predicted on theoretical grounds, were observed in a naturalistic, noisy environment.

Insert Figure 2 about here

A comparable pattern was observed in a second lab study (Brown, in press). Here, response times were measured from the onset of a category name until the subject began to enter a frequency at the keyboard. Figure 3 presents response times plotted against presentation frequency for the different-context and same-context subjects. Response time clearly increases with frequency in the different-context condition, where the conditions required to enumerate are present; the response time curve is flat in the same-context condition where enumeration is not possible. The Context X Presentation Frequency interaction was significant, $\underline{F}(5,240) = 14.62$, $\underline{MSE} = 6.244$.

Insert Figure 3 about here

Another way to substantiate the protocol analyses is to monitor changes in accuracy while manipulating the available information. We have suggested that the process underlying non-numerical strategies involves converting a qualitative sense of frequency into a number. If that is so, then the conversion process should be sensitive to whatever numerical referents are available. This can be explored by providing different types of response range information. Recall that subjects overestimated in the same-context condition above. In a third laboratory study (Brown, in press) these estimates were reduced by providing range information; however range information had no effect for different-context subjects.

In both conditions, one group of subjects was told that no frequencies were greater than 16 (boundary-16), another group was told that no frequencies were greater than 24 (boundary-24), and a final (control) group was given no boundary information. Estimated frequency was regressed onto actual frequency for the three groups in each condition. In the same-context condition, where presumably subjects were relying on non-numerical strategies, control subjects overestimated (slope=1.65), boundary-24 subjects were relatively accurate (slope=1.09) and boundary-16 subjects underestimated (slope=.70). In the different-context condition, where presumably subjects induced range information by enumerating episodes, all three groups underestimated and did so to the same degree (slopes were .70, .68, and .65 for control, boundary-24, and boundary-16 subjects, respectively).

<u>Naturalistic and Laboratory Research</u>

A multiple strategy perspective on frequency estimation emerges when one considers the range of results in the behavioral and experimental frequency literature. The principal components of this perspective are that (1) people use multiple strategies to estimate event frequency (see Figure 1), (2) non-numerical estimation strategies differ from numerical ones because they convert qualitative impressions of frequency into numerical responses, and (3) strategy use is related to event properties, in particular, distinctiveness, regularity and frequency. It appears that distinct events promote enumeration, that regularity fosters the use of rate-based strategies, and that people are most likely to retrieve qualitative impressions when event frequency is high.

This view has emerged by studying survey respondents in their natural context. By controlling event characteristics in the laboratory it was possible to corroborate our observations and explore the implications more deeply. For example, it was possible to examine accuracy and to manipulate the numerical information that was available when subjects were likely to use non-numerical strategies. In this case at least, the complementary relationship of naturalistic and laboratory research has shed more light on a phenomenon than would have either approach by itself.

Banaji and Crowder (1989) have complained that naturalistic research is bankrupt because, among other reasons, it does not generalize beyond the particular situation in which it is carried out. That is certainly not the case here. Our ideas grew out of performance in a natural setting but were replicable, and extendible, in the laboratory. In addition, certain aspects of what we observed about estimating frequencies apply to other forms of quantitative estimation (cf. Brown & Siegler, 1992). Moreover, this approach unifies a number of phenomena identified in disparate laboratory studies that apparently do not generalize themselves. For example, Tversky and Kahneman (1973) observed evidence of availability, not enumeration; Barsalou and Ross (1986) found that their subjects enumerated but found no evidence that they used the familiarity of the stimulus as a key to frequency, *etc.* It appears that this is a case in which theoretical insight has emerged from naturalistic study; the implications of those insights were then explored in detail using the precision of the laboratory.

References

Banaji, M. R. & Crowder, R.G. (1989). The bankruptcy of everyday memory. *American Psychologist 44*, 1185-1193.

Barclay, C. & Wellman, H. (1986). Accuracies and inaccuracies in autobiographical memories. *Journal of Memory and Language*, 25, 93-103.

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

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

Brown, N. R. (in press). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

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

Bruce, D. & Read, J. D. (1988). The how and why of memory for frequency. In Gruneberg, M. M., Morris, P. E. and Sykes, R. N. (Eds.), *Practical Aspects of Memory: Current Research and Issues, Volume 1, Memory in Everyday Life* (pp. 317 - 322). New York: John Wiley and Sons.

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., Brown, N. & Cashman, E. (1993). How the memorability of events affects frequency judgments. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, *Volume 2*, 1058-1063.

Ericsson, K. A & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data, Revised Edition.* Cambridge, MA: MIT Press.

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

Linton, M. (1982). Transformations of memory in everyday life. In Ulrich Neisser (Ed.), *Memory Observed: Remembering in Natural Context*, pp. 77-91. San Francisco: W. H. Freeman.

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

Means, B., Swan, G., Jobe, J. & Esposito, J. (1994). The effects of estimation strategies on the accuracy of respondents' reports of cigarette smoking. In Schwarz, N. and Sudman, S. (eds.), *Autobiographical memory and the validity of retrospective reports*, (pp. 107- 119). New York: Springer Verlag.

Menon, G. (1993). The effects of accessibility of information in memory on judgments of behavioral frequencies. *Journal of Consumer Research*, 20, 431-440.

Neisser, U. (1981). John Dean's Memory: A Case Study. *Cognition,* 9:1-22.

Neisser, U. (1986). Nested Structure in Autobiographical Memory. In Rubin, D. C. (Ed.) *Autobiographical Memory* (pp. 71-81). New York: Cambridge University Press.

O'Muircheartaigh, C. & Gaskell, G. (1994, May). *Vague Quantifiers: Universal or Segmented Norms*. Paper presented at the 49th Annual Conference of the American Association for Public Opinion Research. Danvers, MA.

Strube, Gerhard. 1987. Answering survey questions: The role of memory. In Hippler, H., Schwarz, N. & Sudman, S. (Eds.), *Social Information Processing and Survey Methodology* (pp. 86-101). New York: Springer-Verlag.

Sudman, S. & Bradburn, N. (1974). *Response Effects in Surveys: A Review and Synthesis.* Chicago, IL: Aldine Publishing Company.

Tversky, A. & Kahneman, D. (1972). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, *4*, 207-232.

Wagenaar, W. (1986). My memory: a study of autobiographical memory over six years. *Cognitive Psychology*, 24, 666 - 678.

Table 1. Mean regularity and similarity ratings and reported frequency for the major response strategies

Response Strategy	Similarity*	Regularity**	Frequency
Non-Zero Responses			
1. Episode Enumeration	2.86	2.39	2.3
2. Rate Retrieval	3.54	3.66	7.8
3. Rate Estimation	3.31	3.43	11.1
4. Rate and Adjustment	3.43	3.65	11.9
5. General Impression	3.11	2.91	12.3

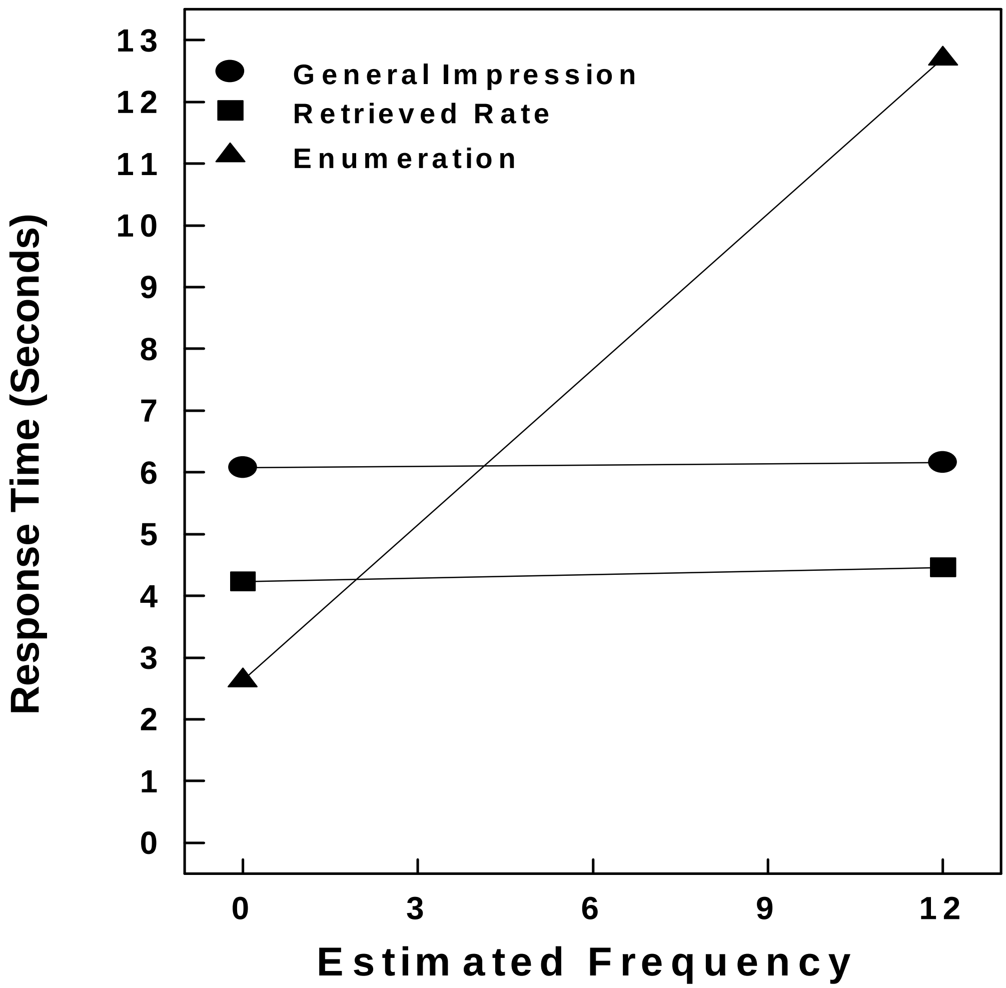
***** 1 = very different, 4=very similar
****** 1 = very irregular, 4=very regular

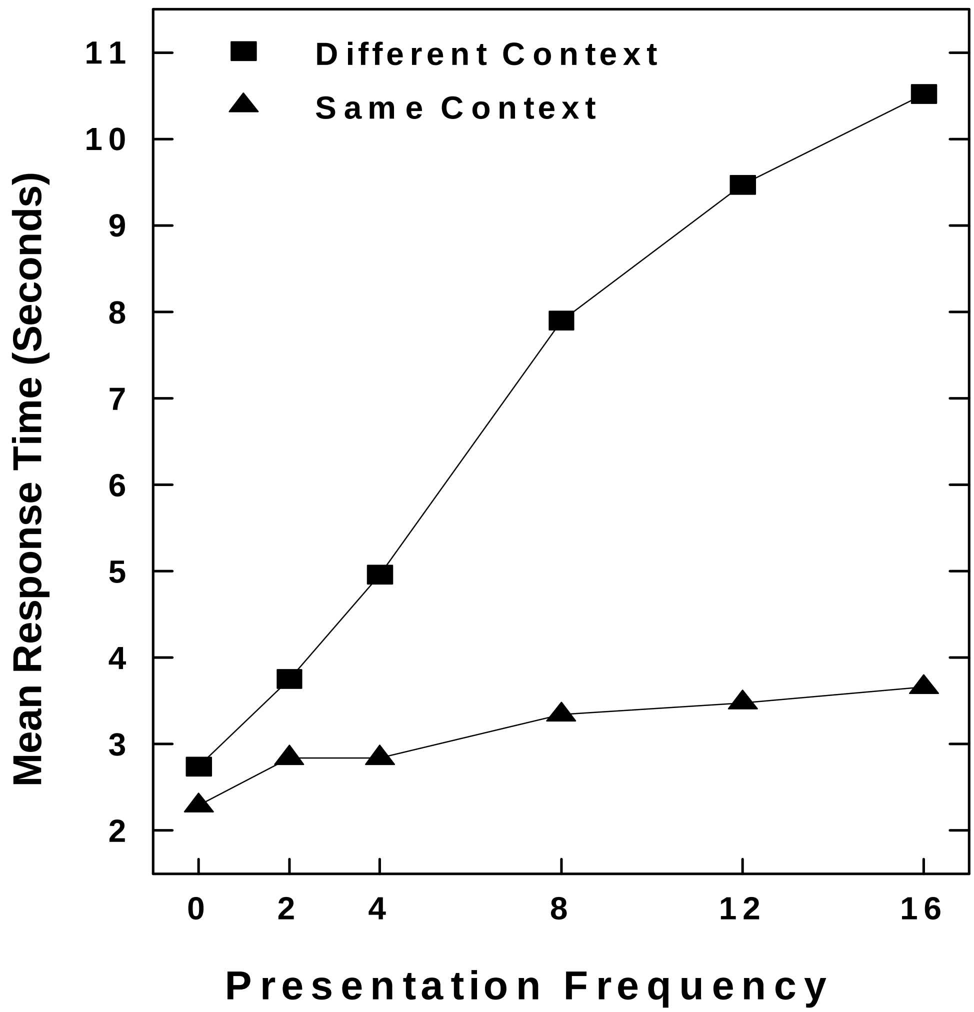
Figure Captions

Figure 1. Taxonomy of strategies implied by multiple strategy perspective.

Figure 2. Regression lines for three most prevalent strategies.

Figure 3. Response times as a function of presentation frequency for same- and different-context subjects.


General Impression
Retrieved Rate
Enumeration
Response Time (Seconds)
0
1
2
3
4
5
6
7
8
9
10
11
12
13
0
3
6
9
12
Estimated Frequency


Different Context
Same Context
Mean Response Time (Seconds)
11
10
9
8
7
6
5
4
3
2
0
2
4
8
12
16
Presentation Frequency