HOW THE MEMORABILITY OF EVENTS AFFECTS FREQUENCY JUDGMENTS*

Frederick Conrad
Bureau of Labor Statistics

Norman Brown
University of Alberta

Erin Cashman
Bureau of Labor Statistics

Surveys often ask respondents to report the frequency of their everyday activities. For example, surveys administered by the Bureau of Labor Statistics ask respondents to report the number of expenditures they have made over the last three months, and the number of times they have looked for work in the past four weeks. The responses to these questions have widespread, social implications, and so understanding how they are formulated is valuable. Researchers have traditionally assumed that respondents recall specific episodes of a target behavior remembered as having occurred within the reference time period, and count up those episodes to produce a frequency report. By this view, errors arise from either forgetting or incorrectly dating events (Sudman & Bradburn, 1974). More recently, however, a compelling body of evidence has emerged for the use of alternative strategies, in particular, rate-based strategies (Blair & Burton, 1987; Burton & Blair, 1991; Means & Loftus, 1991). Researchers have assumed that over time, mental event categories become associated with rates of activity and a frequency report can be generated by retrieving such a rate. Errors presumably arise when respondents overlook exceptions to their normal rate of activity for a behavior.

This paper is concerned with the factors that lead people to use different strategies in formulating behavioral frequencies. In particular, are there characteristics of events that predispose respondents to prefer one strategy over another? We describe an experiment that examines two event characteristics, regularity and similarity of occurrence, as possible determinants of strategy selection. In addition, we discuss behavioral data from three sources: reported strategy, reported frequency and response time. These data illuminate the range of different strategies and how they are executed.

*We thank the following people for their thoughtful advice and assistance: Marj Jones and David Cantor from Westat, Inc.; James Doree and Jon Truswell from the University of Alberta; Steve Miller, Paul Mullin and Clyde Tucker from the Bureau of Labor Statistics. The views expressed here are those of the authors alone and do not reflect the opinions of the Bureau of Labor Statistics or the University of Alberta.

Factors Known to Affect Response Strategies

The observation that multiple response strategies are involved in answering frequency questions is of value to survey researchers only if the circumstances under which respondents use particular strategies can be identified. Blair and Burton (1987) and Burton and Blair (1991) point to the importance of frequency in predicting strategy use. They found that episodes tend to be enumerated when frequency is low, *i.e.*, there are relatively few episodes to recall; rate information tends to be used when frequency is higher. Another variable they associate with selecting a strategy is time allocated to respond: more time leads to more episode enumeration.

The regularity with which events occur has been shown to affect how respondents produce behavioral frequencies. Means and Loftus, (1991) found that activities which presumably occur irregularly, like visiting the doctor for an injury, lead to frequency judgments based on recalled incidents more than do activities like visiting the allergist, which presumably occur on a regular schedule. They also showed that similar episodes (minor medical conditions) were recalled less often than distinctive ones (serious medical conditions), although they did not show that such activities lead respondents to use rate knowledge instead of enumerating episodes. They did not measure event regularity or distinctiveness, and did not look at both variables together.

Menon (in press) has systematically explored the effects of event regularity and similarity on strategy choice. She found that events are primarily enumerated when regularity and similarity are low. Other levels of these event characteristics lead to increased use of rate knowledge.

Felcher and Calder (1992) argue that memory structure for different types of events largely determines the strategy used to answer frequency questions. They found that questions about superordinate events, such as buying food, are answered by referring to rate and frequency information about subcategories, like going to the supermarket, the neighborhood market and the convenience store. Individual episodes were mentioned only when the event categories were ad hoc

(using a pay phone) or exemplars of event categories (eating in a sit-down restaurant). One issue raised by such an analysis is why information about individual episodes is stored with some categories and not others. For this reason we looked at the characteristics of events themselves.

Predictions

Regularity and Similarity. Frequency reports can only use the information that is available, whether that is a number of remembered episodes or a retrieved rate for a particular activity. It would seem that for people to acquire rate information for an event, the event must occur on a relatively regular schedule. If someone performs a certain activity every Thursday from 2:00 - 3:30, it is likely they would become aware of this schedule. In contrast, it is unlikely someone would have weekly rate information available for an event that may occur several times on one week, and then, not again for several weeks. Rate information, should therefore be available for regularly occurring events, and, as a result, more likely to be used in a frequency judgment, than for irregularly occurring events.

In contrast to rate information, individual episodes should be most clearly differentiated in people's memories, and so more available, if they occur on an irregular schedule. A regularly occurring event has identical, temporal characteristics each time it takes place, so people may poorly encode the separate episodes. An episode enumeration strategy requires access to episodes, and so we would expect to observe it more often for irregular than for regular events. Our first prediction, then, is rate-based responding should increase as event regularity increases while episode enumeration should increase as event regularity decreases, *i.e.*, for irregular events.

Event similarity, the second characteristic we consider, should vary with response strategy much like regularity is expected to do. Our second prediction, therefore, is that the more similar the individual episodes of an event, the more evidence of rate based responding we would expect; the less similar the episodes, the more we should observe enumeration. The reasoning is that enumeration requires the recall of episodes. The more distinctive, and therefore discriminable remembered events are, the easier they are to recall.

Menon's (in press) study also looked at event regularity and similarity in frequency judgments and her model of how these factors influence respondents is much like ours. However our methodologies differ in several ways, most notably, between the populations we studied (university students in her case, a national sample in ours) and the mode by which the studies were administered (laboratory interviews in her case, telephone interviews in ours). Our studies are also distinguished by the types of analyses we performed on the data. Menon (in press) was concerned with reported frequencies to the extent that their accuracy varied as a function of event regularity and similarity, and she treated decision time solely as a measure of effort. We are concerned with the interrelation of frequency reports and response time as a function of response strategy.

Frequency and Response Time. It is usually assumed that when episodes are enumerated they are retrieved from episodic memory and when rates are invoked they are retrieved from semantic memory (Means & Loftus, 1991). Episodic memory is a temporally-based event memory; semantic memory is permanent knowledge about the world, *e.g.* the sky is blue. In addition to episodic and semantic sources of information, frequency judgments might also rely on imprecise quantitative information, based on a sense of frequency, *e.g.* "I do it a lot."

These different sources of frequency information imply different response times patterns. Presumably it takes a fixed amount of time to retrieve one episode, so the more episodes retrieved the longer the response time. In contrast, rate knowledge is presumed to be associated with particular event categories and directly retrievable. It should take no more time to retrieve a rate of once a day than of once a month. Similarly, for impression based responding, there is no reason, *a priori*, why the time to decide "a lot" should differ from the time to decide "not very often." Our third prediction, therefore, is that response time will increase with frequency for enumeration but will not vary with frequency for rate and impression based responding.

Method

The study was conducted as a telephone survey. Three experienced interviewers administered the questions to 106 respondents, contacted from a sample of 250 telephone numbers. These numbers were randomly drawn from a nationally representative sample, based on rural and metropolitan areas of the country and stratified by census region and population. The sample was used previously in an unrelated survey and numbers were removed from the sample if they were not working or the respondents refused to participate. Each interview was audiotaped with the respondent's consent.

Procedure. The respondents were presented with four tasks. First, they were asked to answer ten behavioral frequency questions (see Table 1). They were told to think silently before answering, and to take as much time as they needed, but to respond as soon as they had made up their mind. Second, after answering each question, the respondents were asked to explain how they came up with the answer, *i.e.* they were asked for a retrospective protocol (Ericcson & Simon, 1984). The interviewers occasionally probed if they felt the protocol was not informative.

In the third task, the interviewers presented the same ten events and asked respondents to rate the regularity of each on a four point scale. Respondents were told that if they engaged in an event on a routine schedule it was *Very Regular* (4); if they never engaged in it on a routine schedule, it was *Very Irregular* (1). The fourth task required respondents to judge the similarity of the activities on a four point scale. They were told that if their experiences were virtually the same each time they engaged in the activity then they should rate it *Very Similar* (4); if each encounter with the event seemed like a totally unique experience, then it should receive a *Very Different* (1) rating.

Items. Three different random orders of the events were generated for each respondent, one for eliciting the frequency reports and retrospective protocols, one for the regularity rating task, and one for the similarity rating task. The behavioral frequency questions are presented in Table 1. A single reference period of one month was used for the frequency questions. Pretesting indicated that respondents found it disruptive to shift reference periods between questions and additional pretesting indicated that zero responses were relatively infrequent across the events for this time frame.

During the last month, how many times did you . . .
1. conduct a transaction with an Automated Teller Machine?
2. perform a transaction with a teller in a bank?
3. shop in a Grocery Store?
4. shop in a Department Store?
5. purchase gas for your car?
6. pay to have your car repaired?
7. receive subscription magazines by mail?
8. receive catalogues by mail?
9. eat ice cream?
10. eat spicy food?

Table 1. Stimulus Questions

Coding. The authors listened to a subset of taped interviews and developed a preliminary taxonomy for the response strategies that were evident in the protocols. Two groups of two coders then classified the complete set of protocols, reaching each coding decision by consensus within a group, and expanding the coding taxonomy as needed. Because the groups consistently disagreed on how to code one event (receiving subscription magazines) it was eliminated from further analysis. Inter-group reliability for the nine remaining items at this point was .75. All discrepancies were then resolved through discussion between the groups. The final set of strategy categories is presented in Table 2.

Response Strategy	Percent
1. Episode Enumeration	27
2. Known Rate	15
3. Estimated Rate	12
4. Rate and Adjustment	9
5. General Impression	18
6. Uncodable Response	18
	n=660

Table 2. Major response categories and percent of observations

Response Times. The duration of each taped response was measured from completion of the question by the interviewer to initiation of the spoken frequency by the respondent. Response times included the durations of extraneous speech prior to the actual frequency, for example, qualifiers such as "I'd say about ..." and verbalizations of respondents' thinking, such as "Well I do it twice a week, so ..." . Observations were excluded if respondents asked for clarification or were disrupted before responding.

Results and Discussion

Response Strategies. Response strategies were classified separately for frequency reports of zero (26%) and all other values (74%). Here we focus exclusively on the non-zero responses. Of these, ninety-six percent of the codable data were assigned to one of five response strategy categories: *episode enumeration*, *known rate*, *estimated rate*, *rate* and *adjustment*, *general impression*. The remaining four percent of the codable protocols were assigned to the category of *item enumeration* (in particular, listing the catalogues typically received in answering the catalogue question). These are not considered further

because they were rare and were concerned with items rather than events.

We expected evidence in the protocols of both *episode enumeration* and of some kind of rate strategies, and both were observed. We subdivided the rate-based responses into *known rate* and *estimated rate*. By "known rate" we are referring to rate information stored some time prior to the interview. "Estimated rate" refers to our observation that respondents compute rate information while answering the question. An *estimated rate* protocol relied on knowledge that the event recurred regularly but with an uncertain frequency; that frequency was somehow constructed in the response process. An example, in response to the gas question, is "Three...Usually it will take about two weeks to burn a tank of gas. On the average, I'd say about three"

Respondents were considered to use the *rate and adjustment* strategy when they relied on rate information (we did not distinguish between known and estimated rates for this strategy) but adjusted up or down to account for exceptions to the rate. An example from the question about grocery shopping is: "Eight...Uhhm...Once a week, regularly, and four more times for ... spontaneous items."

In addition to *enumeration* and the several rate-based strategies, evidence that respondents relied on *general impressions* was also observed in the protocols. The key characteristic of these reports is that respondents use a coarse sense of magnitude to infer a specific frequency. This was usually indicated by magnitude terms which have quantitative implications but are not explicitly numerical, for example *all the time*, or *a lot*. Consider the following report in response to the question about bank teller transactions: "Eight...We're in the process of buying a house so I've been going there *a lot*." Occasionally, respondents justified a response by asserting that "it just felt like that." These reports were also coded as relying on *general impressions*.

Regularity and Similarity. Consistent with findings reported by Menon (in press), regularity and similarity had clear effects on the choice of response strategy. Mean regularity and similarity ratings for the major response strategies are presented in Table 3. Looking first at regularity, this characteristic varied with response strategy, F (4,424) = 17.41, $p < .01$[1]. When respondents enumerated episodes, they judged events to be less regular than when they used known rate information, F (1,424) = 54.85, $p < .01$, confirming our first prediction. Moreover, regularity ratings for *enumeration* were lower than for all other strategies, F (1,434) = 62.58, $p < .01$ by a Sheffé test. This last effect is consistent with the idea that events are enumerated primarily when they occur on an irregular basis. Finally, frequency reports relied on *general impressions* when events were less regular than for the three strategies involving rates, F(1, 424) = 13.54, $p < .01$, by a Sheffé test. This may imply that general impressions are used under circumstances similar to those promoting enumeration. We return to this topic in the next section.

Response Strategy	Regularity*	Similarity**
1. Enumeration	2.39	2.86
2. Known Rate	3.66	3.54
3. Estimated Rate	3.43	3.31
4. Rate and Adjustment	3.65	3.43
5. General Impression	2.91	3.11

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

Table 3. Mean regularity and similarity ratings for the major response strategies

Similarity showed much the same pattern as regularity. Overall, similarity varied with response strategy, F (4,423) = 4.39, $p < .01$, and similarity ratings for *enumeration* were lower than for *known rate*, F (1,423) = 12.54, $p < .01$, confirming the second prediction. Similarity ratings were lower for *enumeration* than for all of the other strategies, F (1,423) = 16.42, $p < .01$ by a Sheffé test, consistent with the idea that enumeration is used primarily when episodes are quite dissimilar and, consequently, differentiated in episodic memory.

Regularity and similarity were correlated across the entire data set (r = .48, $p < .01$, n = 936). Although one can imagine scenarios in which regularity and similarity are independent, it may be a fact about the world that events which recur on a regular schedule are experienced much the same from one episode to the next: even if they differ in all other characteristics, each occurrence of a regularly occurring event is temporally similar to the next.

[1]All ANOVA results reported here are based on models that include a subject term and an item term as classification variables. The subject term is included to account for repeated measures effects and the item term is included to account for item effects. All contrasts reported on the basis of ANOVAs are planned unless otherwise noted.

Frequency and Time. Subjects are more likely to enumerate low frequency than high frequency events in arriving at frequency reports, presumably because the effort of retrieving individual episodes increases with the number of episodes (Blair & Burton, 1987; Burton & Blair, 1991; Means & Loftus, 1991). This is replicated in the current study. Mean frequencies and times for five strategies are presented in Table 4. Overall, frequencies vary with strategy, $F(4,425) = 10.99, p < .01$. In addition, frequency for *enumeration* is lower than for *known rate*, $F(1, 425) = 16.27, p < .01$ and for all other strategies, $F(1,425) = 41.18, p < .01$ by a Sheffé test.

Response Strategy	Frequency	Time (in seconds)
Non-Zero Responses		
1. Enumeration	2.3	4.52
2. Known Rate	7.8	4.36
3. Estimated Rate	11.1	7.28
4. Rate and Adjustment	11.9	6.67
5. General Impression	12.3	6.17

Table 4. Mean frequency and response times for the major response strategies

This effect could be related to the availability of contextual information. In order to retrieve events, there must be some information to distinguish one event from another. Context can play exactly such a distinguishing role. Repeated events, that is high frequency events, become *decontextualized* (Linton, 1982) and so cannot easily be retrieved; the context of low frequency events can be remembered, making them easier to retrieve.

The use of general impressions to report high estimates can be explained if we assume that repeated occurrences allow respondents to directly encode imprecise frequency information (like *a lot*). Respondents may then be able to use this information as the basis of a response. Because respondents do not have precise rate information available (these events receive the second lowest regularity ratings), all they can do is convert their impression that frequency is high to a relatively large number.

Not only do frequency reports vary with response strategies, but they are related to response times in different ways for the various strategies. In the current study, response times provide direct evidence that enumeration processes differ substantially from the other response processes we have identified. In particular, we predicted that as frequency reports based on enumeration increase, the latencies for those reports should increase. In contrast, frequency and time should not be associated for rate or impression based reports. This prediction was confirmed. The regression lines for response time as a function of reported frequency are plotted for five strategies in Figure 1. Notice the steep curve for *enumeration* ($r^2 = .25, p < .01$) and the flat curves for *known rate* ($r^2 = .003$, n.s.) and *general impression* ($r^2 = .04$, n.s.). These results indicate that a noticeable amount of time is associated with the retrieval of each additional episode (about .84 seconds), but time does not change with the size of rates or impressions.

The regression line for *estimated rate* increases in time with an increase in frequency ($r^2 = .14, p < .01$), though the relationship does not appear to be as strong as for *enumeration* (the slope is .15 in contrast to .84 for *enumeration*). One explanation for a slope of intermediate steepness is that rates are estimated by enumerating episodes for a portion of the reference period, for example a week; the respondent then extrapolates to the complete time period much as would be the case with a known rate. It is the extrapolation part of the strategy that would weaken the relationship between frequency and time relative to that observed for pure enumeration: with extrapolation, only a single mental operation (presumably multiplication) leads to a relatively large frequency report. No relationship between frequency and time is evident for *rate and adjustment* ($r^2 = .04$, n.s.).

A homogeneity of slopes test confirms that the relationship between time and frequency varies with response strategy, $F(4,368) = 7.74, p < .01$. In particular, the regression line for *enumeration* was steeper than that for *known rate* and *general impression*, $F(1,368) = 20.56, p < .01$. The curve for *estimated rate* was also steeper than that for *known rate*, $F(1,368) = 6.20, p < .05$, by a Sheffé test, but not as steep as the *enumeration* line, $F(1,368) = 14.05, p < .01$, by a Sheffé test, consistent with the idea that the *estimated rate* strategy is a hybrid of *enumeration* and *known rate* strategies. These results suggest that respondents set a threshold for time spent enumerating and once this is exceeded, they extrapolate.

Regardless of the reported frequency, *known rate* is faster than *enumeration*, $F(1,368) = 12.35, p < .01$. This essentially replicates a pattern found by Menon (in press) in which frequency responses for regular items were faster than those for irregular items. And it is also consistent with the finding of Burton & Blair (1991) that inducing respondents to use more time in responding increases the incidence of enumeration. But the lesson of the current study is that overall time

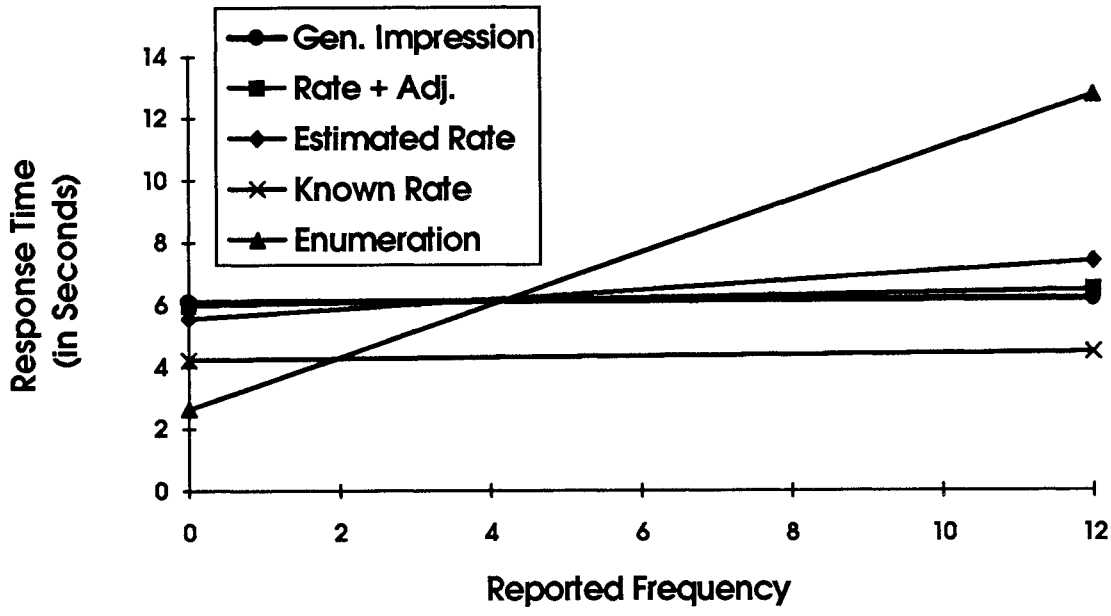

Figure 1. Response Time as function of Reported Frequency and Response Strategy.

tells only part of the story. In arriving at a strategy, respondents may be influenced by their sense of how many discrete, mental steps are required for a response.

Conclusions

The current study underscores the complexity of how behavioral frequencies are produced by respondents. The evidence points to the use of several distinct strategies and at least one hybrid strategy, that have different temporal, event and frequency characteristics. It seems likely that, at some level, respondents consider many of these variables, and perhaps others, in determining how to formulate a frequency.

An area of study that follows from the current work is to better understand the strategies that have not previously been given attention, in particular estimated rate and impression based strategies. Their accuracy under various conditions should be explored and the mechanisms on which they rely should be explicated.

The current study has identified some regularities in a complicated set of processes, but it has primarily framed questions for future work. In that respect, it is a step in the direction of reducing measurement errors in surveys.

References

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

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.

Ericsson, K. A. & Simon, H. A. (1984). *Protocol Analysis: Verbal Reports as Data.* Cambridge, MA: MIT Press.

Felcher, E. M. & Calder, B. J. (May, 1992). The relationship between respondent memory structure and retrieval strategies: A protocol analysis of behavioral frequency data. Paper presented at the 47th Annual Conference of the American Association for Public Opinion Research, St. Petersburg, FL.

Linton, M. (1982). Transformations of memory in everyday life. In U. 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.

Menon, G. (in press). The effects of accessibility of information in memory on judgments of behavioral frequencies. *Journal of Consumer Research.*

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