Center for Demography and Ecology

University of Wisconsin-Madison

Open and Closed Questions About the Frequency of Events: Replication, Extension, Paradox

Nora Cate Schaeffer
Jennifer Dykema

CDE Working Paper No. 94-18

Open and Closed Questions About the Frequency of Events: Replication, Extension, Paradox

Nora Cate Schaeffer
Jennifer Dykema

University of Wisconsin - Madison

November 1994

Center for Demography and Ecology Working Paper 94-18

This research was supported in part by National Institutes of Health grant HD31042-01 to Nora Cate Schaeffer and Judith A. Seltzer. Computing was provided by the Center for Demography and Ecology, which receives core support from the Center for Population Research of the National Institute for Child Health and Human Development (HD-05876). New data were provided by the Letters and Sciences Survey Center, a unit of the College of Letters and Science at the University of Wisconsin-Madison, which is supported by the College and the Graduate School. National Opinion Research Center graciously provided data and documentation for the secondary analysis of the Leisure Activity Study; Ed Blair answered our questions about the study; Charles N. Halaby gave his usual patient answers to many questions we had in the course of the analysis; we incorporated observations Robert M. Groves made during a presentation of these results; and I-Fen Lin ably and cheerfully accomplished the difficult task of creating the original analysis files. An earlier version of this paper was presented at the 1994 meetings of the American Association for Public Opinion Research, Danvers, MA, where Ben Rowe provided a helpful discussion.

Open and Closed Questions About the Frequency of Events: Replication, Extension, Paradox

Abstract

Previous research comparing open and closed questions has focused on the mean frequency among those who report engaging in a behavior. Our experiment extends this research by examining the effects of both filtering (ever vs never) and question form (open vs closed) on reports about the frequencies of events. We examine the effects of question structure on two measures: the proportion saying they ever engaged in the behavior and the mean level of the behavior among those who engaged in it. As predicted, closed questions result in a higher proportion reporting that they engaged in the behavior, but the mean among those engaging in the behavior is (sometimes) higher with the open question. The latter difference may be partially due to how question structure affects the composition of the sample reporting events and partially due to more frequent reporting of high frequencies when an open question is used. Both effects appear across a variety of threatening and nonthreatening behaviors in two surveys, and effects may appear regardless of whether a filter is used. The difference in the proportion reporting a behavior appears more consistently than does the effect on the means. Results for the use of a filter are more ambiguous. New data are provided by a national RDD sample of over 1200 cases collected in 1991, and we also reanalyze approximately 1100 cases from the 1975 Leisure Activity Study collected by NORC.

Open and Closed Questions About the Frequency of Events:

Replication, Extension, Paradox

In their 1974 review of the literature to that date, Sudman and Bradburn concluded that open and closed questions about the frequency of behaviors obtained comparable results in face-to-face interviews except for socially desirable or threatening behaviors (Sudman and Bradburn 1974, pp. 35-36, 47-48). Closed questions appeared to be associated with higher reports of desirable behaviors and lower reports of threatening behaviors; open questions were associated with higher reports of threatening behaviors. The patterns discovered in this meta-analysis were then examined in a set of complex experiments included in a national survey, the Leisure Activity Study (LAS). The published report of these experiments concluded that, compared with closed questions, open questions obtained higher reports of the amount of beer, wine, or liquor usually drunk (by those ever drinking or drinking in the past year), and the frequency of three sexual behaviors (by those engaging in the behaviors) (Blair, Sudman, and Bradburn 1977, pp. 318-319, reprinted in Bradburn, Sudman, and Associates 1979). They also concluded that "nonthreatening items elicited virtually flat reporting across questionnaire manipulations," including the open-closed manipulation (Blair et al. 1977, p. 318).

This line of research can be interpreted as recommending the use of open questions when asking about the frequency of events and behaviors (see discussion in Schaeffer and Charng 1991). Using open frequency questions relieves the investigator of having to determine in advance which behaviors are threatening or socially desirable and of having to design closed response categories, which respondents may interpret as informative in ways that the investigator did not intend (see, for example, Schwarz, Hippler, Deutsch, and Strack 1985, who consider closed categories that give ranges).

But the published analyses of the experiments in the LAS address only one issue posed by the comparison between open and closed frequency questions: the mean frequency reported by those who say they engaged in a behavior. Because the key frequency questions in their analysis were preceded by filter questions that asked if the event had occurred, their analysis does not address the prior question of how question structure affects reports of whether an event has occurred at all.[1] Blair et al. (1977) judged that "question structure was not amenable to manipulation" for "yes-no" questions (1977, p. 318).[2] But how open and closed questions affect reports of whether or not an event occurred can also be assessed. And this examination can be augmented by considering the impact of using a filter question, as these influential experiments did.

We report here an experiment that varies the use of a filter question (yes or no) and question form (open vs closed) on reports about the frequency of events. The events asked about vary in level of threat and in whether social desirability effects - if they occur - might lead to overreporting or underreporting. Our focus is on

---

[1]We use "question form" to refer to variations such as that between open and closed questions and "question structure" more generally to encompass variations in question form as well as such variations as inserting a filter question before a frequency question.

[2]The open and closed questions that Blair et al. report were preceded by filter questions (details given below). Bradburn (1983, pp. 301-302) later summarized the results of these analyses as showing that "open-ended response categories reduced underreporting of the frequency or intensity of threatening behaviors, but not whether the respondent had ever engaged in the activity....the question form affected only the frequency or amount of behavior reported, not reports of whether or not the behavior had ever had [sic] engaged in." Our experiment and a reanalysis of the Leisure Activity Study suggest a different conclusion.

reports of whether the activity occurred at all, but we also examine the mean frequency among those engaging in the activity. We supplement our analysis of this experiment with results from the Leisure Activity Study that were not reported by Blair et al. (1977).

Question Structure and the Reporting of Events

Social-information-processing models of the response process posit four stages to answering a survey question: understanding the question, retrieving or constructing an answer, editing the candidate answer, and mapping it onto response categories (Cannell, Miller, and Oksenberg 1981; Tourangeau 1984; Schwarz 1990). Researchers have given more attention to the first two stages than to the last (see, for example, Eisenhower, Mathiowetz, and Morganstein 1991). But to examine how question structure affects the reporting of events we must consider how a respondent who has an answer to map onto the response categories might do so.[3]

Ever vs Never. One model for examining how respondents map answers onto response categories suggests that respondents act as though the response categories divide a hypothetical response continuum into regions (Schaeffer and Charng 1991; see also Cliff 1959). The model, originally applied to intensity rating scales, describes how the boundaries between response categories divide the response continuum. Other things being equal, a respondent will be more likely to select the category within whose boundaries her or his true answer falls. A survey

[3] Not all respondents will go through this process. For example, those who "construct" answers using information in the response categories (e.g., Schwarz et al. 1985) probably do not "map" answers onto response categories in this way.

question may be more or less explicit in indicating where the boundaries that establish answer "regions" are located on the continuum. We can speculate about the effects of using a closed frequency question or a filter question by using this model.

The model suggests that the response continuum for a **closed frequency question** could be illustrated in this way:

```
0 times    x_____¦_____x__________¦_________x_____________________>    (1)

           never       less than        every few
                       every few        months ....
                       months
```

"X" indicates that the category label establishes a characteristic value for a category; the boundaries between categories are indicated by "¦." A true answer that falls between a pair of boundaries is "closer" to that characteristic value than to the neighboring characteristic value(s). Thus, for example, a respondent with a very low frequency to report will decide whether to answer "never" or "less than every few months" by determining which category is closer to his true answer. Depending on how he interprets "less than every few months," someone whose answer is "1" might well select "never" as closer to his true answer. The diagram indicates that the labeling of the lowest frequency category (i.e., the lowest frequency greater than zero) contributes to defining the "meaning" of "never." If the lowest frequency category is "once a year," the categories imply that "never" is "zero times" within the 12-month reference period. But if the lowest frequency category is "a few times a year," "never" could include frequencies greater than zero.

A "yes-no" **filter question** that asks whether an event has ever occurred implicitly offers two response categories and establishes only a vague boundary between them, as illustrated by the following diagram:

0 times ____________!____> (2)
no yes

Depending on where she perceives the boundary between "no" and "yes" as being, a respondent whose true answer is "every few months," could consider "no" as closer to her true answer than "yes," even though she had actually engaged in the behavior. Because a filter question leaves the exact location of the boundary between "no" and "yes" ambiguous, respondents may feel relatively free to use their impressions about the point of the question, the level of accuracy requested, the difficulty of retrieving the requested information, or the desirability or threat of the requested information to locate the threshold between "no" and "yes."[4]

Two possible interpretations of an **open question** may have different consequences. One interpretation is that an open question implicitly asks the respondent to distinguish between whether an event has never occurred and whether it has occurred even once. A second interpretation is that an open question, like a filter question, permits the respondent to establish a threshold for determining whether the number of events is reportable at all. Some respondents may use the same threshold for reporting whether an event has occurred when answering an open question and a filter question. But other respondents may consider that open

[4] "Never" can also mean more than zero times (Pohl 1981, p. 238), and "no" is likely to be less restrictive than "never."

questions request reports of even a single occurrence, and thus be more likely to report a low frequency with an open question than they would be to report "yes" to a filter question.

This model suggests that, when different types of questions are compared, the label of the lowest frequency category determines whether or not the proportion who report engaging in a behavior will differ for closed questions as compared to either a filter question or an open question. Other things being equal, the more the lowest frequency category restricts the meaning of "never," the greater will be the proportion of respondents reporting that the event occurred to a closed question - and the greater the difference will be between the results obtained with the closed question and the results obtained with an open question or a filter question.[5]

Our approach yields several predictions about the effects of question structure on the proportion who report that an event occurred. If the label of the lowest frequency category in the closed question restricts the meaning of "never," the proportion who report engaging in a behavior should be (1) larger when a closed question is used than when an open question is, and (2) larger when a filter question is not used than when one is. When a filter question is used, however, the respondent encounters the filter before the open or closed frequency question; this means that the effect of the filter takes precedence. Thus, the proportion reporting an event will be larger with an unfiltered closed question and smaller with a filtered question, whether open or closed. Our prediction is indeterminate for the remaining

---

[5]This prediction assumes that some respondents have values in the portion of the response continuum between "never" and the lowest frequency category.

combination, an unfiltered open question. In this case, the proportion reporting an event should fall between the proportions obtained by an unfiltered closed question and by a filter question, if respondents vary in the threshold they use when an open question is used (as discussed above).

Frequency of Events. According to the model, other things being equal, a respondent whose frequency is greater than her threshold value for reporting to an open question or a filter question would report a similar frequency to the two questions.[6] The model suggests that if the means differ for open and closed questions or filtered and unfiltered questions, the difference is partly due to the composition of the sample who report the behavior: Some respondents with low frequencies will report they did not engage in the behavior when asked a filter question or an open question. This should raise the mean among those engaging in the behavior for the filtered and open questions. Overall, however, the different composition of the samples reporting events to open and closed or filtered and unfiltered questions makes prediction difficult.

Other Factors Influencing Reports. Obviously, question structure may influence responses by affecting the other stages of cognitive processing. Either rate-based or grouped response categories may suggest that a rounded answer is acceptable, and open questions may suggest that accuracy is preferred. The labeling or order of the response categories may communicate information about the

---

[6]This may be difficult to determine, however, because collapsing answers to open questions for comparison with a closed question introduces errors. Also, there may be global effects (such as social desirability effects) on the thresholds between categories, as discussed above, which may cause differences between open and closed questions. And other things may not be equal, as we discuss in the next section.

desirability of the behavior or about the possible distribution of the behavior in the population (see, e.g., Schwarz 1990). Processing the response categories in a closed question may distract respondents from retrieval; or the reading of the response categories may provide additional time for retrieval. Filter questions may act as "primes" and begin retrieval, so that respondents who pass a filter question remember more events than they would have remembered without the filter; or filter questions may suggest that respondents need not engage in serious retrieval, even when the subsequent request for a number of events is given. In addition, questions occur in a larger context. A series of brief questions that use the same format (a "battery" of questions) lets respondents learn the structure of the questions. But embedding each question in an extended section of an instrument that explores a single topic may encourage retrieval (see, for example, Kessler and Wethington 1991; Lessler, Tourangeau, and Salter 1989).

The literature does not provide much guidance in choosing among these possibilities,[7] and, indeed, these processes may all influence answers, in varying degrees, with different respondents, for some topics. So we focus here on replicating the effects reported previously and on identifying features of answers to open and closed, filtered and unfiltered questions that must be considered in future theoretical developments.

---

[7]There is little systematic discussion of how question structure might affect the various stages of cognitive processing. Schwarz et al. (1985) describe how respondents may use response categories that present ranges as informative. Schwarz, Hippler, and Noelle-Neumann (1994) discuss how open and closed questions might affect recall of which events were engaged in; like the attitude questions discussed by Schuman and Presser (1981) these open questions ask respondents to choose or name objects.

Data

The analysis uses two data sources: A set of questions included in a national survey by the Letters and Science Survey Center (LSSC) at the University of Wisconsin, and items from the Leisure Activity Study (LAS) collected by NORC.

LSSC Data. This experiment crosses two factors. The first varies the use of a filter question that asks whether the respondent "ever" engaged in the behavior in the last 12 months. The second factor varies whether frequencies are obtained using an open question or closed response categories. The format for the open question permits answers to be recorded as rates (e.g., times per day, week, or month) or yearly totals. The closed response categories, modeled after the LAS, are: every day, several times a week, once a week, every few weeks, once a month, every few months, less than every few months, or never.[8] The categories were presented in order of decreasing frequency.

The behaviors we asked about potentially present different levels of threat or social desirability: "have dinner in a restaurant, other than for business" (neutral), "exercise for half an hour or more" (socially desirable), "play cards for money" (moderate threat), "have beer, wine, or hard liquor to drink" (moderate threat), "drink a little too much beer, wine, or liquor, so that [you] act differently than usual" (high threat).[9] Some of the behaviors we selected (and some portions of the question

---

[8]Particularly when read in this order, the lowest frequency category, "less than every few months," could be interpreted as including zero. Our results suggest that few respondents interpreted the category in this way.

[9]Blair et al. (1977, p. 317) report the proportion of the sample who say they would be made "very uneasy" and "not at all uneasy" by questions about various topics. Their data, using their labels for the topics, suggests the following increasing order of threat for the activities we ask about: sports activity, general leisure, drinking, gambling, and intoxication.

wording) are also modeled on the LAS. Figure 1 illustrates the experimental design using one of the behaviors. (See Appendix A for exact question wording.)

Data were collected in February through August 1991 in CATI interviews conducted by the Letters and Science Survey Center (LSSC) of the University of Wisconsin, Madison. The national telephone sample is designed to represent currently working household numbers in the continental United States. The modified RDD design discards blocks of numbers with no listed residential number. An independent national sample is drawn for each day of the week. Up to ten attempts are made for each telephone number. The interview schedule is modified every few weeks, and these experiments combine data collected using several different versions of the interview schedule. The respondent is randomly selected from among those 18 or older. The response rate is estimated to be between 50 and 60 percent. This response rate is of less concern for the type of experiment reported here than for, say, estimation of population parameters.[10]

LAS Data. We also present results from the Leisure Activity Study (Response Effects in Surveys, 1975) that were not reported, or not reported in the form given here, in Blair et al. (1977). The items we report about from the LAS were used in experiments that compared open and closed questions about social activities (giving a party, going to a bar, and spending an evening with relatives, with friends in the neighborhood, and with friends outside the neighborhood) and an item about intoxication from a series of questions about drinking. In some cases, this factor was crossed with one or two additional factors (such as question length). Blair et al.

---

[10]Respondents were randomly assigned to treatments, but the algorithm gives unequal cell sizes.

(1977, p. 320) report that there was a single (uninterpretable) interaction among the experimental factors, and we pool across the other factors to report results only for the factor of interest here.

None of the LAS experiments varied whether or not a filter question was used; the LAS questions about drinking and sexual activity all followed filters. For drinking behaviors, the filters asked if the respondent had ever engaged in the behavior and then whether the respondent had engaged in the behavior in the past year; for sexual activities, the filter asked if the respondent had engaged in the behavior in the past month.[11] The closed questions begin with the highest frequency category for questions about social activities (almost every day, once or twice a week, several times a month, about once a month, several times a year, about once a year, never) and sexual behavior (daily, 4-6 times a week, 2-3 times a week, once a week, 2-3 times a month, once a month or less), and with the lowest frequency category for the questions about drinking and intoxication [(never,) once a year or less, every few months, once a month, every few weeks, once a week, several times a week, daily].[12] (Appendix B gives the wording of these questions.)

The LAS used a national area probability sample with quotas at the last stage (see Bradburn et al. 1979 for a description of these data). As do other publications

[11]The analyses of "all drinkers" in Blair et al. (1977) included those who had engaged in the three drinking behaviors but not in the last year; respondents who had never engaged in the behaviors were omitted from the "all drinkers" analysis.

[12]Thus, the LSSC and LAS categories differ in several ways. For example, the LAS items about social activities use "about once a year" as the lowest frequency category; the LSSC items use "less than every few months" as the lowest frequency category. The response categories are ordered from highest to lowest for all five questions in the LSSC experiment. By extending Sudman and Bradburn's suggestion that categories for attitude questions begin with the "least desirable" (1982, p. 156), one might propose that beginning with the highest frequency category would reduce underreporting of threatening behaviors.

using these data, we report significance tests as though the data were from a simple random sample.

Results: The Proportion Engaging in a Behavior

The model, like the literature reviewed above, suggests examining the effects of question structure in two stages: effects on the proportion saying they ever engaged in the behavior and effects on the mean level of the behavior among those who engaged in it.[13] Table 1, Panel A (LSSC Behaviors) shows the proportion in each treatment who report that they engaged in the specified behavior at least once. (Appendix C shows the full distributions of the items by treatment; the distribution provides the number of cases for subsequent tables.) Significance tests are from logistic regressions that include the experimental factors. We describe differences with a probability of .05 or less as statistically significant. The models exclude interactions that were not statistically significant.

When questions are presented in a series, as these were, respondents may learn the structure of the questions; in the LSSC experiment, any effects of such learning are confounded with increasing question threat. The effects of learning could be important in this experiment. In principle, answers to a filter question should not be affected by whether the filter is followed by an open question or a closed question. In our experiment, however, over 95 percent of respondents said "yes" to

[13]The respondents who pass the filter are no longer a random subset of the sample, nor are the respondents who give frequencies greater than 1 to the open question. This problem would not be removed if answers to the filter and the followup question were combined, although it would be less apparent. The problems posed by an analysis of these items are similar to those that arise in analyzing "unfolded" items (e.g., Groves 1989, pp. 464-468; Krosnick and Berent 1993).

the filter for the first behavior (having dinner in a restaurant), and over 80 percent said "yes" to the filter for the second behavior (exercise). Thus, almost all respondents could learn whether an open question or a closed question followed the filter question; for this reason, the analysis does not pool the two groups of respondents presented with the filter question. That is, we allow the question form to affect answers even after a filter question.

The data generally support the prediction about the effects of open vs closed questions: Although the differences are sometimes trivially small, in nine of the ten comparisons within filter-no filter condition, the closed question produces a larger estimate than the open question; the difference is significant in two of the five replications, playing cards and drinking, and borderline ($p < .10$) for dining in a restaurant and drinking too much. The behavior for which the evidence of an effect is weakest, exercising, appears to have very few respondents in the portion of the response continuum likely to be affected; that is, those who report exercising at all, say they do so more than once or twice a year (see Appendix C). The significant interaction for the first activity, having dinner in a restaurant, reflects the fact that a large proportion report the activity to the unfiltered closed question. The interaction may appear only for this question because respondents administered the filter have not yet experienced the open or closed followup question.

A substantively similar result is shown in Panel B (LAS Behaviors) for the LAS questions that ask about social activities. These LAS questions used "about once a year" as the lowest frequency category and ordered the categories as did the LSSC, from most to least frequent. We also show results for the LAS item that asked respondents who reported drinking in the last year how often they had become intoxicated. The proportion who report that they engaged in the behavior is larger for the closed question for all five social activities ($p < .05$ in four of five comparisons). The difference for the LAS question about intoxication (which followed several filters about ever drinking) is in the predicted direction, but is trivially small.[14]

This analysis presents two strong results that contrast with the suggestions of previous research comparing open and closed questions: First, it is the closed questions that appear to obtain higher levels of reporting, that is, larger proportions engaging in a behavior. Second, the closed questions obtain these higher levels of reporting for most of the more threatening activities we examine (playing cards for money and drinking in the LSSC data and going to a bar in the LAS), but also for the nonthreatening activities (giving a party, spending the evening with friends).

In addition, the comparison between the LAS social activities questions and the LSSC results offers indirect support for our model of how closed response categories divide the continuum. For the LAS questions about social activities, the lowest frequency category is "about once a year," and the comparable category in the LSSC is "less than every few months." Thus, "never" should be defined more restrictively by the closed question in the LAS (where a person whose answer is even "one" should choose a frequency category) than by the closed LSSC question (where a

[14]Results from the LAS also suggest that learning about question form may occur across items. We examined the proportion saying to the relevant filter question they ever drank beer, drank wine, drank liquor, petted or kissed, had intercourse, or masturbated. Considering only respondents who received the short standard question (which is most similar to ours), the proportion who report engaging in the behavior is greater (3 to 6 percent) for five of the six items when the filter is followed by the closed question. The lowest frequency category for the closed question implied that even one time should be reported. The difference is significant for masturbation.

person whose answer is "one" might think "never" is a better answer than "less than every few months"). Consequently, the difference between open and closed questions in the proportion saying they engaged in the behavior should be larger for the LAS questions about social activities than for the LSSC questions. And that is what we find.

The results for the use of a filter are more ambiguous. In the significant interaction for having dinner in a restaurant, the filter has the predicted effect for the closed question, that is, the proportion reporting the activity is greater when no filter is used. For drinking too much, the simple main effect is significant and also in the predicted direction. But even the direction of this effect is not replicated consistently by the other items.

Results: Mean Frequency Among Those Engaging in a Behavior

We next examine the effects of question structure on reports of frequency by those who say they engage in the behavior. The analysis replicates the general approach of Blair et al., that is, we assign a numeric value to characterize each category, use the characteristic values to categorize answers to the open question and to compute means within experimental treatments, and then test the difference in the means. This procedure immediately poses the problem of what characteristic values and boundaries to use. Conclusions about the effects of open vs closed questions are necessarily conditioned on these decisions. The characteristic value determines which answers to the open question are classified in a given category, and it influences the mean that is computed when the open and closed questions are compared. For conclusions about the effects of question structure to be robust, however, they should persist with reasonable changes to the method for categorizing answers to the open question.[15]

Categorizing Answers to Open Questions. We experimented with two different methods of categorizing answers to the open questions in the LSSC. The methods are based on different models of how respondents might interpret and use the categories for the closed question. In Method A, we assigned a characteristic value to each category based on the label in the closed question. Characteristic values were easy to identify for labels like "once a week," but ambiguous labels like "several times a week" required judgments. We set boundaries halfway between each category, as suggested by the model described above.[16]

In Method B, we took an "intuitive" approach. Characteristic values were assigned to the unambiguous category labels (e.g., "once a week"), and only cases with these values were included in that category. Cases with other values were grouped with the adjacent category with a vague label (e.g., "several times a week"); these ambiguous category labels suggest that they can include more variation than do the unambiguous categories, and respondents may have interpreted them in that way.

Note that for both methods, the characteristic value may not be the midpoint

---

[15]The problems that arise in categorizing answers to open questions for comparison with closed questions have received too little consideration. There is some discussion in Blair's (1982) comment on Bachman and O'Malley (1981), but these consider a closed question with category labels that use ranges rather than rates. (See related discussion in LaMotte and Blair 1988 and Blair and Ganesh 1991.) In our analysis we used two explicit models of how respondents might interpret category labels to categorize answers to the open questions.

[16]The exception is the lowest frequency category, which begins with 1. That is, "never" is considered an "absolute zero."

for its category. The two versions give slightly different results. We report results for Method B in Table 2, because it is similar to the method used in the earlier analyses of the LAS,[17] and we comment on differences in the results given by the two methods. (Appendix D gives details about the coding of the answers to the open questions.)

Results. Table 2, Panel A (LSSC Behaviors) shows the mean number of times a behavior was performed by those who reported doing it, for the LSSC questions. Because none of the interaction terms was significant, we show results for the reduced models.[18] In nine of the ten comparisons within filtering condition the mean is greater for the open question than for the closed question. But the difference is clearly significant ($p < .05$) only for having dinner in a restaurant (a nonthreatening activity) and exercising (a desirable activity). The difference is borderline ($p <.07$) for drinking and is negligible and not significant for drinking too much. When Method A is used to categorize answers to the open question (not shown), the difference between open and closed questions is significant only for exercising and is borderline ($p <.07$) for having dinner in a restaurant. Table 2, Panel B (LAS Behaviors) shows the comparable results for the LAS items about social activities and intoxication. The mean obtained with the open question is greater for the items about going to a bar, spending the evening with relatives, spending the

---

[17]Information from Ed Blair (conversation, 1994). We assigned values to the categories of the sexual behaviors in the LAS using similar reasoning, and have obtained results almost identical to those reported in Blair et al. (1977). Because Blair et al. report annual amounts rather than frequencies for drinking, we did not have enough information to replicate results for these questions.

[18]The interaction term is borderline ($p < .08$) for playing cards; we consider this bad luck and do not interpret it.

evening with friends outside the neighborhood, and intoxication, but only in the first case is the effect of even borderline significance ($p < .07$).[19]

In only one case - having dinner in a restaurant - do the means differ depending on whether or not a filter question is used. In that case, the means are higher when no filter is used, which is counter to our expectation. But the direction of the effect is not consistent across the other four items. It is possible that the effect of the filter occurs only for this item because it is the first item in the series, and that the effect disappears after respondents experience the structure of the questions.

Taken together, these results suggest that, among those who engage in a behavior, using an open question will often produce a higher mean than using a closed question. This pattern is so consistent across a variety of behaviors in both surveys that it seems reasonable to conclude that, even when the effect is small enough that it is not statistically significant, the effect is real. The effect occurs when a filter question is used (as Blair et al. did) and without a filter. The effect is not restricted to items for which question structure does (or does not) affect the proportion saying they engaged in the behavior at all. Substantial and significant effects are also not restricted to threatening behaviors and may not occur with threatening behaviors. Moreover, the size of the effect, and hence its significance given a fixed sample size, is somewhat sensitive to the method used to categorize

[19]Blair et al.'s comment about nonthreatening items quoted above (1977, p. 318) presumably refers to this analysis. Although none of the differences is statistically significant, the direction of the observed difference is consistent with the pattern Blair et al. find for threatening questions, and that we find for some nonthreatening behaviors, and so we report it in more detail here.

answers to the open question.[20]

Our finding that question form can affect the proportion who report they engage in a behavior indicates that whatever other factors might contribute to the pattern of means among those who engage in the behavior, the effects of sample composition cannot be ignored. That is, the groups of respondents for whom means are calculated are no longer strictly comparable. The compositional differences between the samples are increased by the differential susceptibility of open and closed question to "don't know" answers. Less than one percent gave a "Don't know" or "Can't remember" answer one or more times to the five LSSC items when a closed question was used; the proportion increased to 2.3 percent for the open-filtered condition and to 4.4 percent for the open-no filter condition. Although the differences are too small to be statistically significant, the pattern for all missing data suggests that using a filter and a closed question each increase the likelihood that the respondent will answer with a recordable frequency. These patterns are consistent with past comparisons between open and closed questions (see Schaeffer and Charng 1991 for frequency questions, Schuman and Presser 1981 for attitude questions), including those in the LAS. For example, for the LAS drinking activities, the percent missing at least one answer is 3.9 percent for the open questions and 1 percent for the closed questions. For sexual activities (counting those who "prefer not to answer" as missing), the corresponding percentages are 15 and 9.6. In neither case is it clear whether this is because of the behavior of respondents or

[20]Our data do not allow us to assess the possible impact of the order of the response categories, because the item for which the LSSC and the LAS presented the categories in different orders, intoxication, does not show the effect in either survey.

interviewers.[21] Whatever the source of this missing information, however, the resulting compositional differences between the samples contribute to the pattern of means for those who report engaging in a behavior.

Locating the Effects of Question Form

Although the use of the filter question does not appear to have consistent consequences, varying whether a question is open or closed sometimes affects both the proportion who say they engage in a behavior and the mean among those who do. To address questions posed by previous research, we have considered portions of the response distribution separately, but locating where the effects of open and closed questions occur requires considering the entire distribution of responses. Our model suggests that among respondents with low frequencies to report, some would report a behavior to a closed question but not to an open question. Thus, if the two lowest categories are combined, the difference between the two question forms should diminish or disappear. A second distributional issue is suggested by speculation that closed questions result in depressed reporting of socially undesirable behaviors (e.g., Blair et al. 1977, p. 316); thus, one might expect the effect to be concentrated at the highest (least desirable) levels. Table 3 addresses these issues

[21]Because this was the first time the LSSC used this format for recording answers to open questions, the interviewers' lack of experience may have led to some problems in obtaining frequency answers.

Mangione, Fowler, and Louis (1992) report the counterintuitive finding that open-sensitive items were less susceptible to interviewer effects (measured by the intraclass correlation coefficient) and open-nonsensitive items were more susceptible to interviewer effects; they also find that interviewer variability is greater when interviewers engage in followup behaviors such as probing. These findings suggest that interviewers may be more likely to follow up a refusal or "don't know" for a nonsensitive than for a sensitive question. But if interviewers followed up all "don't know" or "refusal" answers, the pattern here suggests that all the open questions may be more susceptible to interviewer effects.

by showing the proportions in the two lowest categories, the middle categories, and the two highest categories.[22] The table also presents likelihood ratio chi-square tests that partition the table into two contrasts, that comparing the two lowest categories to the middle categories and that comparing the two highest categories to all lower categories (Agresti, 1990, pp. 50-54).

If open questions and closed questions provide different thresholds for reporting low frequency events, the form difference observed in Table 1 should be mitigated or eliminated when the two lowest-frequency bins are combined. In all three panels,[23] with a few exceptions (exercising, spending the evening with friends in the neighborhood, petting and kissing, and masturbating), there is no difference between forms when the two lowest categories are combined. Even for the exceptions, however, the difference in the use of the lowest categories is small, and it is the middle categories that are affected by question form.

In Table 3, an open question sometimes results in more use of the highest categories, but again, the effect does not appear consistently for all potentially threatening questions in the two surveys. In the LSSC (Panel A), this pattern is present for having dinner in a restaurant and exercising (for which the mean among

---

[22]For the LSSC, Table 3 combines the filtered and the unfiltered conditions. Likelihood ratio tests on the tables with the collapsed frequency distributions (two lowest categories, middle categories, two highest categories) indicate that the filter-no filter distinction is significant only for having dinner in a restaurant ($p < .02$) and drinking too much ($p < .05$). In neither of these cases, however, is the difference substantial, and so we give the simplified presentation in Table 3. When the filtered and unfiltered conditions are not combined, the contrast between the lowest and two middle categories is not significant for exercising, and the contrast between the two highest categories and the middle categories for playing cards is significant only for the unfiltered condition. Other relationships are as shown in Table 3.

[23]This table includes the LAS drinking and sexual activity items, because this analysis was not included in Blair et al. These questions all used filters; "no" answers to the filter questions are treated as the lowest frequency category in this analysis.

those engaging in the behavior differed significantly by form); the pattern also appears for drinking (for which the mean difference was of borderline significance). (For playing cards, use of the highest categories is greater when the closed question is used.) The difference is in the predicted direction for three of the six LAS social activities questions in Panel B, but the differences are often trivially small and are not significant. However for the LAS questions about drinking and sexual activities (Panel C), the proportion in the two highest categories is consistently larger for the open question, and the contrast is significant for five of the six items. When the forms differ, it is due to greater use of the highest category with the open question.

It is at least plausible that the larger context of the LAS questions about drinking and sexual activities - each of the items was embedded in a series of five (for drinking) or three (for sexual activities) questions about each activity - stimulated retrieval in a way that the "batteries" of LSSC questions and LAS social activity questions did not. Perhaps more important, however, is the larger context established by the LAS instrument. Experiments in the instrument tested whether the use of long questions, open questions, or questions using familiar words elicited higher reports of threatening behaviors. Thus, respondents were exposed to treatments in addition to the open-closed manipulation. Even though Blair et al. found that most of the interactions among treatment conditions were not significant, the overall context was probably more stimulating and promoted greater engagement in the task than was true for the batteries of LAS and LSSC items. In addition, we cannot evaluate the impact of the way coders categorized answers to the open LAS questions on the results for that survey. The results reviewed here also raise the possibility that, when it occurs, the difference in means for those engaging in an activity is partially driven by a relatively small number of respondents who give large values to the open question.

Discussion

Our analysis contrasts the previously reported finding that open questions may result in higher means among those engaging in a behavior with the paradoxical result that closed questions may yield a higher proportion reporting that they engage in the behavior. Moreover, the effect on the proportion appears more consistently and for a wider range of behaviors than does the effect on the means. Any explanation of why open and closed questions (sometimes) give different results must consider how closed questions affect the composition of the subsample who report ever engaging in a behavior, and why the use of the highest categories (sometimes) differs by question form. Our results also indicate that such an explanation must consider the possibility that open questions produce higher estimates of means for nonthreatening as well as threatening behaviors, and do not always do so for threatening questions. Blair et al. (1977), for example, make the reasonable argument that because threatening behaviors are underreported, the higher mean obtained by an open question about a threatening behavior is likely to be more accurate than the mean produced by a closed question. But our results suggest that accepting this argument requires one to also accept that the higher mean (sometimes) produced by an open question is also preferable for nonthreatening behaviors, and Burton and Blair (1991) find no difference in accuracy between open and closed questions (about writing checks and using an automatic teller machine).

The results for the manipulation of the filter are somewhat more ambiguous. Significant effects involving this treatment occurred for the first item in our series, both for the proportion saying they ever had dinner in a restaurant (an interaction), and for the mean frequency among those who do. This type of question series is common in survey interviews, and the disappearance of the interaction between question form and filtering after the first item suggests that learning may quickly overcome an effect of filtering. The effect of filtering in reducing the proportion reporting that they drank too much in the LSSC may also result from that learning - respondents may use the filter to close discussion of a threatening behavior. These few significant effects are not further supported by the pattern of results for the other items (in contrast to the consistent patterns we observed for the open-closed manipulation). Survey interviews use filters both in the type of question series that we used and as introductory questions for longer explorations of a topic, as did the LAS. A design to more fully explore the effects of filtering on reports of behaviors must consider both respondents' learning and how sets of questions are structured.

Although it does not have consistent effects on the dependent variables, the use of the filter has some indirect consequences on respondents' behavior. The method for recording answers to open questions provides information about how respondents reported their answers. Table 4 presents the proportion of the sample reporting the frequency of each behavior as a rate or as a total and shows significance tests comparing these overall proportions. The table also provides detail on the proportion of the sample using different rates and the proportion of annual totals that were low frequencies (less than 12 times a year) and high frequencies. The response formats reflect the different social patterning of these behaviors (see Menon 1994), but may also provide evidence of how respondents arrived at their answers. Reporting a total frequency could reflect either the use of episodic enumeration (common with low frequencies) or a decomposition strategy that results in a total rather than a rate (see Blair and Burton 1987).

The proportion giving a yearly total is higher when the open question is unfiltered than when it is, and the difference is significant in four of the five cases. Furthermore, in every case but one (exercising) the majority of those giving yearly totals give low frequency answers. The proportion of both low frequency and higher frequency yearly totals is consistently higher when no filter is used. These findings suggest that the filter question may affect two different types of respondent. First, the greater incidence of reports of low frequency yearly totals when the open question is unfiltered provides indirect supporting evidence that the filtered form of the question establishes a higher "threshold" for reporting low frequencies of a behavior than does the unfiltered form. Second, when there is no filter, respondents with higher frequencies to report may be more likely to use a decomposition strategy; this strategy would probably result in more answers being recorded as yearly totals. If the unfiltered question is considered "more open" than the filtered question, this finding is compatible with Burton and Blair's report of a nonsignificant tendency for respondents to be more likely to use episodic enumeration in answering open than closed questions (1991, p. 70).

The model we presented receives only mixed support here. The predictions about effects on the proportion engaging in a behavior receive fairly strong support for open vs closed questions, but only some support for the filtered vs unfiltered manipulation. Sterngold, Warland, and Herrmann (1994), however, find an effect of filtering that would be predicted by this model for opinion questions; that is, the proportion expressing concern about an issue is larger when no filter is used. Our prediction that the means for open and closed questions are affected by the different composition of the samples saying they engaged in the behavior also receives some support. However, the higher mean for dining in a restaurant in the unfiltered than the filtered form is in direct contrast to the compositional prediction of the model. It is, however, compatible with Sterngold et al.'s finding that the category which indicates high concern is used more when the question is unfiltered. In addition, the model does not account for the greater use of the highest categories in the open than closed question which we sometimes observe, particularly for the threatening questions in the LAS. Although the interpretive mapping processes on which this model focuses probably play a role in some respondents' answers, the model does not address how question structure might affect the process of retrieving or constructing an answer and subsequent effects on the actual frequency reported.

We can rule out some possibilities, however. It seems unlikely that the higher proportion who report engaging in a behavior to a closed question results because the closed question takes more time to read and the respondent uses this time to retrieve information. If that were the case, we would expect that closed questions would also result in higher means among those engaging in the behavior (assuming that the behaviors are underreported).[24] Similarly, if filter questions acted to stimulate retrieval among those who engage in a behavior, we would expect higher means for the filtered than unfiltered forms, which we do not observe.

The model emphasizes the close relationship between the respondent's interpretation of the question - construed broadly - and the process of mapping an answer onto the response categories. It is possible that closed questions send respondents a mixed message that encourages more precision at the threshold for reporting any behaviors but less precision about the frequency itself (again, assuming that the higher mean obtained by open questions is more accurate). But within this broad framework, other processes probably contribute to causing the effects we (sometimes) observe. However these effects are ultimately explained, the explanation must distinguish between effects on reporting the behavior at all and effects on means. An explanation must also consider that the effects can also occur for questions that are not about threatening behaviors - and that they do not appear consistently for threatening behaviors.

[24]Furthermore, the time required to administer the four sets of questions did not differ significantly by experimental treatment (table not shown). The mean ranged from 99 seconds for the open-filtered questions to 102.9 seconds for the closed-filtered question. The standard deviations ranged from 23.4 for the closed-unfiltered questions to 35.2 for the open-filtered questions.

References

Agresti, Alan. 1990. *Categorical Data Analysis.* New York: Wiley.

Bachman, Jerald G., and Patrick O'Malley. 1981. "When Four Months Equal a Year: Inconsistencies in Student Reports of Drug Use." *Public Opinion Quarterly* 45:536-548.

Blair, Edward. 1982. "On 'When Four Months Equal a Year'." *Public Opinion Quarterly* 46:288-290.

Blair, Edward, and Scot Burton. 1987. "Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions." *Journal of Consumer Research* 14:280-288.

Blair, Edward A., and Gopala K. Ganesh. 1991. "Characteristics of Interval-based Estimates of Autobiographical Frequencies." *Applied Cognitive Psychology* 5:237-250.

Blair, Ed, Seymour Sudman, and Norman M. Bradburn. 1977. "How to Ask Questions about Drinking and Sex: Response Effects in Measuring Consumer Behavior." *Journal of Marketing Research* 14:316-321.

Bradburn, Norman. 1983. "Response Effects." Pp. 289-328 in Peter Rossi, James D. Wright, and Andy Anderson, eds., *Handbook of Survey Research*. Orlando: Academic.

Bradburn, Norman M., Seymour Sudman, and Associates. 1979. *Improving Interview Method and Questionnaire Design.* San Francisco: Jossey-Bass.

Burton, Scot, and Ed Blair. 1991. "Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys." *Public Opinion Quarterly* 55:50-79.

Cannell, Charles F., Peter Miller, and Lois Oksenberg. 1981. "Research on Interviewing Techniques." *Sociological Methodology 1981*:389-437.

Cliff, Norman. 1959. "Adverbs as Multipliers." *Psychological Review* 66:27-44.

Eisenhower, Donna, Nancy A. Mathiowetz, and David Morganstein. 1991. "Recall Error: Sources and Bias Reduction Techniques." Pp. 127-143 in Paul Biemer et al., eds., *Measurement Errors in Surveys.* New York: Wiley.

Groves, Robert M. 1989. *Survey Errors and Survey Costs.* New York: Wiley.

Kessler, Ronald C. and Elaine Wethington. 1991. "The Reliability of Life Event Reports in a Community Survey." *Psychological Medicine* 21:723-738.

Krosnick, Jon A. and Matthew A. Berent. 1993. "Comparisons of Party Identification and Policy Preferences: The Impact of Survey Question Format." *American Journal of Political Science* 37,3 (August):941-964.

LaMotte, L. and E. Blair. 1988. "On Estimating Distributional Characteristics from Categorized Data." Proceedings of the American Statistical Association Section on Survey Research Methods., pp. 286-288.

Lessler, Judith, Roger Tourangeau, and William Salter. 1989. "Questionnaire Design in the Cognitive Research Laboratory: Results of an Experimental Prototype." *Vital and Health Statistics* (DHHS Publication No. PHS 89-1076 - Series 6, no. 1). Washington, DC: U.S. Government Printing Office.

Mangione, Thomas W., Floyd J. Fowler, and Thomas A. Louis. 1992. "Question Characteristics and Interviewer Effects." *Journal of Official Statistics* 8(3)293-307.

Menon, Geeta. 1994. "Judgments of Behavioral Frequencies: Memory Search and Retrieval Strategies." Pp. 161-172 in *Autobiographical Memory and the Validity of Retrospective Reports*, edited by Norbert Schwarz and Seymour Sudman. New York: Springer-Verlag.

Pohl, Norval F. 1981. "Scale Considerations in Using Vague Quantifiers." *Journal of Experimental Education* 49(4) 235-40.

Response Effects in Surveys: 1975 National Study (machine-readable data file. 1980. Norman Bradburn and Seymour Sudman, principal investigators, Chicago: National Opinion Research Center. (One data file [1,172] logical records] and SPSS data definition commands, plus accompanying documentation.

Schaeffer, Nora Cate, and Hong-Wen Charng. 1991. "Two Experiments in Simplifying Response Categories: Intensity and Frequency Categories." *Sociological Perspectives* 34(2):165-182.

Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys*. Orlando: Academic.

Schwarz, Norbert. 1990. "Assessing Frequency Reports of Mundane Behaviors." Pp. 98-119 in Clyde Hendrick and Margaret S. Clark, editors, *Research Methods in Personality and Social Psychology*, Volume 11, *Review of Personality and Social Psychology*. Newbury Park: Sage.

Schwarz, Norbert, Hans-Jurgen Hippler, B. Deutsch, and Fritz Strack. 1985. "Response Categories: Effects on Behavioral Reports and Comparative Judgments. *Public Opinion Quarterly* 49:388-395.

Schwarz, Norbert, Hans-Jurgen Hippler, and Elisabeth Noelle-Neumann. 1994. "Retrospective Reports: The Impact of Response Formats." Pp. 187-199 in *Autobiographical Memory and the Validity of Retrospective Reports*, edited by Norbert Schwarz and Seymour Sudman. New York: Springer-Verlag.

Sterngold, Arthur, Rex H. Warland, and Robert O. Herrmann. 1994. "Do Surveys Overstate Public Concerns?" *Public Opinion Quarterly* 58:255-263.

Sudman, Seymour, and Norman M. Bradburn. 1974. *Response Effects in Surveys*. Chicago: Aldine.

Sudman, Seymour, and Norman Bradburn. 1982. *Asking Questions*. San Francisco: Jossey-Bass.

Tourangeau, Roger. 1984. "Cognitive Sciences and Survey Methods," pp. 73-100 in *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines,* edited by Thomas Jabine, Miron Straf, Judith Tanur, and Roger Tourangeau. Washington, DC: National Academy Press.

Appendix A
LSSC Question Wording

Below, we list the questions asked of LSSC respondents in each of the four experimental treatments. Filtered questions use a yes/no filter. The response options in the closed category treatments are: every day, several times a week, once a week, every few weeks, once a month, every few months, less than every few months, or never. These categories are referred to as the Closed Category Format. The open-ended responses are recorded in the manner illustrated in C1a below. These are designated as Open Response Format. Categories for don't know and not ascertained/refused appear on the CATI screen for both the open and closed questions, but are not shown here. The question list below does not include the full set of instructions given to interviewers. The question on drinking too much alcohol was asked regardless of whether or not the respondent reported drinking alcohol in the last 12 months.

Version A: No Filter, Closed Category Format

A1. The next questions are about things that people sometimes do for recreation or to relax. I'm going to ask how often you've done each of these activities in the last 12 months, that is since (THIS MONTH) 1990. First, having dinner in a restaurant. In the last 12 months, did you have dinner in a restaurant, other than for business ...

... EVERY DAY, SEVERAL TIMES A WEEK, ONCE A WEEK, EVERY FEW WEEKS, ONCE A MONTH, EVERY FEW MONTHS, LESS THAN EVERY FEW MONTHS, or NEVER?

A2. Some people like to exercise, for example, by walking, jogging, or playing a sport. In the last 12 months, did you exercise for half an hour or more ... ?

A3. The next activity is playing card games, such as poker, blackjack or 21, bridge, or pinochle for money. In the last 12 months, did you play cards for money ... ?

A4. The next activity is drinking beer, wine or hard liquor. In the last 12 months, did you have beer, wine or hard liquor to drink ... ?

A5. Sometimes people drink a little too much beer, wine, or liquor, so that they act differently than usual. In the last 12 months, did this happen to you ... ?

Version B: Filter, Closed Category Format

B1. The next questions are about things that people sometimes do for recreation or to relax. I'm going to ask how often you've done each of these activities in the last 12 months, that is since (THIS MONTH) 1990. First, having dinner in a restaurant. In the last 12 months, did you ever have dinner in a restaurant, other than for business? (YES/NO)

B1a. (IF YES) Did you have dinner in a restaurant, other than for business ...

... EVERY DAY, SEVERAL TIMES A WEEK, ONCE A WEEK, EVERY FEW WEEKS, ONCE A MONTH, EVERY FEW MONTHS, LESS THAN EVERY FEW MONTHS, or NEVER?

B2. Some people like to exercise, for example, by walking, jogging, or playing a sport. In the last 12 months, did you ever exercise for half an hour or more? (YES/NO)

B2a. (IF YES) Did you exercise for half an hour or more ... ?

B3. The next activity is playing card games, such as poker, blackjack or 21, bridge, or pinochle for money. In the last 12 months, did you ever play cards for money? (YES/NO)

B3a. (IF YES) Did you play cards for money ... ?

B4. The next activity is drinking beer, wine, or hard liquor. In the last 12 months, did you ever have beer, wine or hard liquor to drink? (YES/NO)

B4a. (IF YES) Did you have beer, wine, or liquor to drink ... ?

B5. Sometimes people drink a little too much beer, wine, or liquor, so that they act differently than usual. In the last 12 months, did this ever happen to you? (YES/NO)

B5a. (IF YES) Did you have too much beer, wine, or liquor to drink ... ?

Version C: Filter, Open Response Format

C1. The next questions are about things that people sometimes do for recreation or to relax. I'm going to ask how often you've done each of these activities in the last 12 months, that is since (THIS MONTH) 1990. First, having dinner in a restaurant. In the last 12 months, did you ever have dinner in a restaurant, other than for business? (YES/NO)

C1a. (IF YES) How often did you have dinner in a restaurant, other than for business?

______ TIMES PER ______ DAY
______ WEEK
______ MONTH
______ YEAR/TOTAL } GO TO C2

______ OTHER VERBATIM RESPONSE: Specify what R said } GO TO C1b

C1b. How many times per day, per week, per month, or in total would you say that was?

______ TIMES PER ______ DAY
______ WEEK
______ MONTH
______ YEAR/TOTAL } GO TO C2

C2. Some people like to exercise, for example, by walking, jogging, or playing a sport. In the last 12 months, did you ever exercise for half an hour or more? (YES/NO)

C2a. (IF YES) How often did you exercise for half an hour or more?
C2b. How many times per day, per week, per month, or in total would you say that was?

C3. The next activity is playing card games, such as poker, blackjack or 21, bridge, or pinochle for money. In the last 12 months, did you ever play cards for money? (YES/NO)

C3a. (IF YES) How often did you play cards for money?
C3b. How many times per day, per week, per month, or in total would you say that was?

C4. The next activity is drinking beer, wine, or hard liquor. In the last 12 months, did you ever have beer, wine, or hard liquor to drink? (YES/NO)

C4a. (IF YES) How often did you have beer, wine, or liquor to drink?
C4b. How many times per day, per week, per month, or in total would you say that was?

C5. Sometimes people drink a little too much beer, wine, or liquor, so that they act differently than usual. In the last 12 months, did this ever happen to you? (YES/NO)

C5a. (IF YES) How often did you have too much beer, wine, or liquor to drink?
C5b. How many times per day, per week, per month, or in total would you say that was?

Version D: No Filter, Open Response Format

D1. The next questions are about things that people sometimes do for recreation or to relax. I'm going to ask how often you've done each of these activities in the last 12 months, that is since (THIS MONTH) 1990. First, having dinner in a restaurant. In the last 12 months, how often did you have dinner in a restaurant, other than for business?

_____ TIMES PER _____ DAY
_____ WEEK
_____ MONTH
_____ YEAR/TOTAL } GO TO D2

_____ OTHER VERBATIM RESPONSE: Specify what R said } GO TO D1b

D1a. How many times per day, per week, per month, or in total would you say that was?

_____ TIMES PER _____ DAY
_____ WEEK
_____ MONTH
_____ YEAR/TOTAL } GO TO D2

D2. Some people like to exercise, for example, by walking, jogging, or playing a sport. In the last 12 months, how often did you exercise for half an hour or more?

D2a. How many times per day, per week, per month, or in total would you say that was?

D3. The next activity is playing card games, such as poker, blackjack or 21, bridge, or pinochle for money. In the last 12 months, how often did you play cards for money?

D3a. How many times per day, per week, per month, or in total would you say that was?

D4. The next activity is drinking beer, wine, or hard liquor. In the last 12 months, how often did you have beer, wine, or hard liquor to drink?

D4a. How many times per day, per week, per month, or in total would you say that was?

D5. Sometimes people drink a little too much beer, wine, or liquor, so that they act differently than usual. In the last 12 months, how often did this happen to you?

D5a. How many times per day, per week, per month, or in total would you say that was?

Appendix B
Leisure Activity Study (NORC, 1977) Question Wording

Social Activities

This root question was used in an experiment that compared open versus closed and long versus short question forms. The short version of the question is:

Q22. *We have some questions about various kinds of social activities. As I read each type of social activity, please tell me how often you have done it during the past year.*

- *a.* *Given a party for 5 or more people?*
- *b.* *Gone to a bar, tavern, or lounge?*
- *c.* *Spent a social evening with relatives?*
- *d.* *Spent a social evening with friends who live in your (neighborhood/area)?*
- *e.* *Spent a social evening with friends who live outside your (neighborhood/area)?*

The closed question used the following categories on a card: *almost every day, once or twice a week, several times a month, about once a month, several times a year, about once a year, never.* Answers to the open question were coded into these categories. The coding manual does not provide characteristic values, boundaries, or other details about how the coders were to make this classification.

Intoxication

This question was asked of those who reported drinking beer/ale, wine/champagne or liquor in the past year and was used in several experiments. The short, standard wording of this question is:

Q26. *In the past year, how often did you become intoxicated while drinking any kind of alcoholic beverage?*

The closed question presented the following response categories on a card: *never, once a year or less, every few months, once a month, every few weeks, once a week, several times a week, daily.* Answers to the open question were coded into these categories. The coding manual does not provide characteristic values, boundaries, or other details about how the coders were to make this classification.

Drinking

A series of questions were asked about drinking for relaxation. The questions on drinking beer/ale and wine/champagne were subjected to the open versus closed and long versus short question treatments. The item asking about liquor was also manipulated for familiar versus unfamiliar wording. The short, standard wording of these questions is:

Q23. Did you ever drink, even once, beer or ale? (Yes; No)

- *A.* *IF YES: Have you drunk any beer or ale in the past year? (Yes; No)*
- *[1].* *When you drank beer or ale on the average, how often did you drink it?*

Q24. Did you ever drink, even once, wine or champagne? (Yes; No)

A. IF YES: Have you drunk any wine or champagne in the past year? (Yes; No)
[1]. When you drank wine or champagne on the average, how often did you drink it?

Q25. Did you ever drink, even once, liquor? (Yes; No)

A. IF YES: Have you drunk any liquor in the past year? (Yes; No)
[1]. When you drank liquor, on the average, how often did you drink it?

The interviewer presented the following response categories on a card for the questions at [1]: *once a year or less, every few months, once a month, every few weeks, once a week, several times a week, daily.* Answers to the open question were coded into these categories. The coding manual does not provide characteristic values, boundaries, or other details about how the coders were to make this classification.

Sexual Activity

A series of questions were asked about sexual activity. The question on petting/kissing was tested in both the open and long question forms. The intercourse and masturbation items were also tested for familiar wording. The short, standard wording for each of these questions is:

Q32. In the past month, since (ENTER DATE 1 MONTH AGO TODAY), have you engaged in petting or kissing? (Yes; No; Prefer not to answer)

A. IF YES: On the average, how often did you engage in petting or kissing during the month?

Q33. In the past month, have you engaged in sexual intercourse? (Yes; No; Prefer not to answer)

A. IF YES: On the average, how often did you engage in intercourse during the month?

Q34. In the past month, have you masturbated? (Yes; No; Prefer not to answer)

A. IF YES: On the average, how often did you masturbate during the past month?

The closed question A used the following response categories on a card: *daily, 4-6 times a week, 2-3 times a week, once a week, 2-3 times a month, once a month or less.* Answers to the open question were coded into these categories. The coding manual does not provide characteristic values, boundaries, or other details about how the coders were to make this classification.

Appendix C
LSSC Behavioral Variables

Percentage Distribution of the Frequency of Each Behavior in the Last 12 Months, All Respondents

	Have Dinner				Exercise				Play Cards				Drink				Drink Too Much			
	Filter		No Filter		Filter		No Filter		Filter		No Filter		Filter		No Filter		Filter		No Filter	
Value	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed
0.0	3.8	6.0	4.1	1.9	18.0	15.1	17.4	14.2	81.8	79.7	82.1	74.5	33.0	28.1	36.8	29.1	81.4	73.9	72.7	70.5
1.5	3.8	5.6	4.8	5.8	0.7	0.0	2.1	2.6	6.2	8.6	6.9	10.2	2.8	9.9	5.5	9.1	9.7	14.8	13.1	14.4
7.5	11.7	12.8	16.9	12.7	1.7	1.5	4.9	1.2	4.8	4.9	6.2	5.6	9.5	8.7	12.4	7.9	5.5	5.7	9.3	7.5
12.0	17.2	16.5	12.8	19.4	1.7	0.4	0.3	2.8	1.7	1.5	2.4	3.9	10.9	7.6	7.2	8.4	1.4	3.0	2.1	3.1
32.0	25.8	21.4	23.1	20.8	2.4	5.3	7.3	6.0	2.1	1.1	0.7	2.1	6.3	13.7	5.5	9.6	0.7	1.5	1.0	2.6
52.0	22.3	27.8	16.2	24.3	9.0	16.2	4.9	16.3	2.4	3.0	1.7	1.4	11.6	9.9	10.3	16.1	0.7	0.4	0.7	1.4
208.5	15.5	9.0	20.3	14.1	44.6	41.9	39.0	36.0	0.7	1.1	0.0	0.9	20.7	14.8	16.8	15.6	0.7	0.8	1.0	0.2
365.0	0.0	0.8	1.7	0.9	21.8	19.6	24.0	20.9	0.3	0.0	0.0	1.4	5.3	7.2	5.5	4.2	0.0	0.0	0.0	0.2
N	291	266	290	432	289	265	287	430	292	266	291	432	285	263	291	429	290	264	289	424
Mean	55.10	45.91	67.40	55.50	178.44	169.24	174.44	162.41	5.24	4.95	1.97	9.43	72.47	68.52	64.17	61.04	2.74	3.27	4.00	4.07
S.D.	67.84	61.92	85.59	72.18	130.14	127.53	137.27	133.13	28.64	23.79	7.54	47.30	105.82	108.83	104.48	96.05	18.01	18.80	21.79	21.80

Appendix D
Coding Answers to the Open Questions

Categorizing answers to the open questions and assigning values to the closed categories in order to estimate annual means for the comparison between open and closed questions follows the same general approach for the both the LSSC and LAS experiments. For the closed questions, a numeric value thought to characterize the category label is assigned to each category (e.g., a characteristic value of twelve for "once a month"). However, unlike answers to the open questions in the LAS (which are available only as coded into categories),[1] the answers to the open questions in the LSSC data had to be categorized so that they could be compared with the answers to the closed questions.

The format for the open questions in the LSSC experiment permits responses to be recorded as both rate-based frequencies (e.g., per day, per week, per month) and yearly totals (e.g., in total in last year). For example, if a respondent reports exercising three times a day, the interviewer records a "3" as the frequency and "per day" as the metric. In contrast, a respondent offering an annual total such as "37 times a year" is recorded as giving "37" for the frequency and "in total in last year" as the metric. Respondents offered answers that were not codeable in this format in less than two percent of the cases. An example of this type of response is reporting the frequency of drinking alcohol as "on holidays or special activities." These verbatim answers are recorded as offered. The interviewer then follows the initial response with a question that asks the respondent to convert his answer into an absolute frequency. All responses to the open questions were converted into yearly totals for analysis: "per day" answers were multiplied by 365, "per week" and "per month" answers were multiplied by 52 and 12, respectively.

For the comparison between open and closed questions, we group the answers to the open questions into the closed question categories. Because one cannot be sure how respondents (in the LSSC and LAS) or coders (in the LAS) actually interpreted the closed categories, this process is uncertain at best. Thus, we compared two approaches.

Method A assigns a characteristic value to each category based on the label in the closed question (see discussion of Method B and table below). These values were more easily assessed for the concrete categories. Once we established characteristic values, we set boundaries halfway between each category. We refer to this method as "probabilistic" because cases that were on the boundary between two categories were randomly assigned to either the higher or lower category. The exception was "never" which was treated as an "absolute zero."

In Method B, the "intuitive" method, we modelled a second way respondents and coders might have interpreted the categories of the closed questions by focusing on the meaning of the labels. We think that this method approximates the way the LAS coders understood the categories (conversation with Blair, 1994). An example illustrates the difficulties of transforming the closed categories to absolute frequencies. A respondent may interpret the category label "once a month" as referring to a characteristic value of twelve. While assigning a value to this category is relatively straightforward, interpreting the category "several times a week" is more difficult. Here, the category explicitly refers to a range of frequencies and may also have a wide range of interpretations among respondents. Method B takes the variability of the response category labels into account. Categories that correspond to relatively concrete frequencies such as every day, once a week, once a month, and never are assigned only characteristic values and not ranges; the categories thus include only cases with the characteristic value. In contrast, ambiguous categories such as several times a week, every few weeks and so on include all the cases between the concrete categories.

The following table presents the coding systems used in this analysis. Similar characteristic values, not shown in detail here, were used for the reanalysis of items from the LAS.

[1]The coding documentation for the Leisure Activity Study is incomplete and does not include directions given to coders for deciding how to code answers to the open questions into the categories. It appears that LAS interviewers recorded respondents answers verbatim; then coders coded these answers directly into the categories used in the closed question.

	Method A: Probabilistic			Method B: Intuitive		
Category Label	Characteristic Value	Boundary	Method	Characteristic Value	Boundary	Method
Every day	365.0		65*1	365.0		365*1
		206.5			364.5	
Several times a week	156.0		52*3	208.5		
		104			52.5	
Once a week	52.0		52*1	52.0		52*1
		34.65			51.5	
Every few weeks	17.3		52/3	32.0		
		14.65			12.5	
Once a month	12.0		12*1	12.0		12*1
		8			11.5	
Every few months	4.0		12/3	7.5		
		3			2.5	
Less than every few months	2.0			1.5		
		0.5			0.5	
Never	0.0		0	0.0		0

Table 1. Proportion Who Report Ever Engaging in the Behavior in the Last 12 Months or Past Year, by Question Form and Filter Treatment[a]

Panel A: LSSC Behaviors

	Have Dinner		Exercise		Play Cards		Drink		Drink Too Much	
	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed
Filter	.96	.94	.82	.85	.18	.20	.68	.72	.19	.26
No Filter	.96	.98	.83	.86	.18	.25	.63	.71	.28	.29
Form[b]	W=3.19, p<.07		W=2.05, p<.15		W=4.46, p<.03		W=5.35, p<.02		W=2.80, p<.09	
Filter	W=0.06, p<.81		W=0.09, p<.76		W=1.54, p<.21		W=1.17, p<.28		W=5.28, p<.02	
Form * Filter	W=4.65, p<.03		---		---		---		---	

Panel B: LAS Behaviors

Become Intoxicated[c]		Given a Party		Gone to a Bar		Spent Evening with Relatives		Spent Evening with Friends-Neighborhood		Spent Evening with Friends-Outside	
Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed
.41	.42	.59	.65	.53	.59	.88	.91	.67	.74	.79	.85
t=0.29, p<.77		t=2.13, p<.03		t=1.94, p<.05		t=1.45, p<.15		t=2.49, p<.01		t=2.59, p<.01	

[a] Ns for the LSSC and LAS behaviors are shown in Appendix C and Table 3, respectively.
[b] The Wald test statistic and significance levels are based on the results of logistic regression analysis.
[c] The question on becoming intoxicated is filtered for drinking any alcohol (beer, wine, or liquor) in the past year.

LSSC Proportions and tests: FREQ22.COM
LAS Proportions and tests: LSRA14.COM

Table 2. Annual Means and Standard Deviations for Behaviors Among Respondents Engaging in the Behavior, by Question Form and Filter Treatment[a]

Panel A: LSSC Behaviors

	Have Dinner		Exercise		Play Cards		Drink		Drink Too Much	
	Open	**Closed**	**Open**	**Closed**	**Open**	**Closed**	**Open**	**Closed**	**Open**	**Closed**
Filter	57.3 (68.3)	48.8 (62.7)	217.6 (110.1)	199.3 (114.6)	28.9 (62.4)	24.4 (48.4)	108.1 (113.4)	95.4 (118.0)	14.7 (39.9)	12.5 (35.4)
No Filter	70.3 (86.2)	56.5 (72.4)	211.2 (122.6)	189.3 (124.8)	11.0 (14.9)	37.0 (88.4)	101.5 (116.2)	86.1 (104.2)	14.6 (39.9)	13.8 (38.5)
Form	F=7.32, *df*=(1,1229), p<.01		F=7.58, *df*=(1,1065), p<.01		F=2.27, *df*=(1,266), p<.13		F=3.37, *df*=(1,865), p<.07		F=0.10, *df*=(1,324), p<.75	
Filter	F=5.80, *df*=(1,1229), p<.02		F=1.28, *df*=(1,1065), p<.26		F=0.00, *df*=(1,266), p<.98		F=1.10, *df*=(1,865), p<.30		F=0.03, *df*=(1,324), p<.87	

Panel B: LAS Behaviors

	Become Intoxicated[b]		Given a Party		Gone to a Bar		Spent Evening with Relatives		Spent Evening with Friends-Neighborhood		Spent Evening with Friends-Outside	
	Open	**Closed**	**Open**	**Closed**	**Open**	**Closed**	**Open**	**Closed**	**Open**	**Closed**	**Open**	**Closed**
	20.0 (49.5)	16.4 (40.9)	9.3 (17.6)	10.0 (19.0)	37.2 (57.2)	29.8 (47.5)	34.0 (44.5)	32.8 (46.2)	44.1 (62.0)	47.0 (65.9)	31.2 (45.7)	28.4 (47.4)
Form	F=0.59, *df*=(1,353), p<.44		F=0.29, *df*=(1,722), p<.59		F=3.26, *df*=(1,653), p<.07		F=0.16, *df*=(1,1035), p<.69		F=0.43, *df*=(1,813), p<.51		F=0.85, *df*=(1,943), p<.36	

[a] Ns for the LSSC and LAS behaviors are shown in Appendix C and Table 3, respectively.
[b] The question on becoming intoxicated is filtered for drinking any alcohol (beer, wine, or liquor) in the past year.

LSSC Means and tests: FREQ16.COM
LAS Means and tests: LSRA06.COM (intoxication); LSRA05.COM (social activities)

Table 3. Proportion Who Report Behavioral Frequency in the Lowest, Middle and Highest Categories by Question Form[a]

Panel A: LSSC Behaviors

	Have Dinner		Exercise		Play Cards		Drink		Drinking Too Much	
Frequency	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed
Lowest	.08	.09	.19	.16	.89	.86	.39	.38	.88	.86
Middle	.73	.78	.16	.25	.11	.12	.37	.41	.11	.13
Highest	.19	.13	.65	.59	.01	.02	.24	.21	.01	.01
N	581	698	576	695	583	698	576	692	579	688
G_3^2 = [b]	7.88*		15.93***		5.59		3.30		2.00	
G_1^2 = [c]	0.04		11.04***		0.46		1.13		1.64	
G_1^2 = [d]	7.84**		4.88*		5.13*		2.18		0.35	

Panel B: LAS Intoxication and Social Behaviors

	Become Intoxicated		Given a Party		Gone to a Bar		Spent Evening with Relatives		Spent Evening with Friends-Neighborhood		Spent Evening with Friends-Outside	
Frequency	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed
Lowest	.73	.76	.58	.55	.53	.50	.18	.15	.39	.33	.26	.26
Middle	.25	.23	.41	.42	.34	.39	.62	.65	.40	.45	.59	.61
Highest	.02	.01	.01	.02	.12	.11	.20	.20	.21	.22	.15	.13
N	414	445	584	585	580	584	574	585	572	584	573	584
G_3^2 = [b]	1.28		2.46		2.72		1.64		5.56		0.96	
G_1^2 = [c]	0.34		0.61		2.13		1.64		5.00*		0.00	
G_1^2 = [d]	0.95		1.85		0.60		0.00		0.56		0.96	

(Table continues)

Table 3 (continued)

Panel C: LAS Drinking and Sexual Behaviors

	Drank Beer/Ale		Drank Wine/Champagne		Drank Liquor		Engage in Petting/Kissing		Engage in Intercourse		Masturbated	
Frequency	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed	Open	Closed
Lowest	.42	.44	.47	.50	.41	.40	.30	.27	.38	.37	.94	.93
Middle	.34	.41	.43	.44	.46	.51	.30	.37	.49	.54	.04	.07
Highest	.24	.15	.10	.06	.13	.09	.40	.36	.14	.09	.02	.00
N	580	584	581	584	583	585	549	558	522	546	540	547
G_3^2 = [b]	16.38***		8.15*		5.10		6.17*		6.10*		15.09***	
G_1^2 = [c]	1.07		0.09		0.86		3.88*		0.84		3.84*	
G_1^2 = [d]	15.30***		8.06**		4.24*		2.29		5.26*		11.25***	

[a] The Lowest and Highest categories are formed by combining the two lowest and highest response categories, respectively. The Middle category contains the remaining four middle categories.
[b] For Two Lowest vs. Middle vs. Two Highest responses.
[c] For Two Lowest vs. Middle responses only.
[d] For Two Lowest + Middle vs. Two Highest responses.
*** $p < .001$. ** $p < .01$. * $p < .05$

LSSC Behaviors FREQ33.COM
LAS Intoxication and Social Behaviors LSRA16.COM
LAS Drinking and Sexual Behaviors LSRA17.COM

Table 4. Proportion Who Use a Rate Based or Yearly Total Response Format Among LSSC Respondents Who Report Ever Engaging in the Behavior by Filter Treatment

	Have Dinner		Exercise		Play Cards		Drink		Drink too Much	
Response Format	**Filter**	**No Filter**	**Filter**	**No Filter**	**Filter**	**No Filter**	**Filter**	**No Filter**	**Filter**	**No Filter**
Rate Based Total	**.74**	**.59**	**.96**	**.81**	**.39**	**.21**	**.80**	**.63**	**.17**	**.14**
Per Day	.00	.01	.27	.28	.02	.00	.08	.08	.00	.00
Per Week	.37	.33	.64	.49	.17	.08	.47	.37	.07	.04
Per Month	.37	.25	.05	.05	.19	.14	.26	.19	.09	.10
Yearly Total	**.26**	**.41**	**.04**	**.19**	**.62**	**.79**	**.20**	**.37**	**.83**	**.86**
Answers < 12	.16	.23	.03	.09	.60	.73	.18	.29	.82	.82
Answers >= 12	.10	.18	.01	.10	.02	.06	.02	.08	.02	.04
n	275	277	236	236	52	52	188	182	54	79
χ^2	14.57***		26.04***		3.72*		13.43***		0.19	

NOTE: Statistical test compares the collapsed distribution Rate Based Total with Yearly Total.
*** p < .001. ** p < .01. * p < .05.

Figure 1. Experimental Design and Sample Question Wording

Filter	Question Form: Closed		Question Form: Open	
No	A1.	The next questions are about things that people sometimes do for recreation or to relax. I'm going to ask how often you've done each of these activities in the last 12 months, that is since (THIS MONTH) 1990. First, having dinner in a restaurant. In the last 12 months, did you have dinner in a restaurant, other than for business ... EVERY DAY, SEVERAL TIMES A WEEK, ONCE A WEEK, EVERY FEW WEEKS, ONCE A MONTH, EVERY FEW MONTHS, LESS THAN EVERY FEW MONTHS, or NEVER?	D1.	The next questions are about things that people sometimes do for recreation or to relax. I'm going to ask how often you've done each of these activities in the last 12 months, that is since (THIS MONTH) 1990. First, having dinner in a restaurant. In the last 12 months, how often did you have dinner in a restaurant, other than for business? ______ TIMES PER ______ DAY ______ WEEK ______ MONTH ______ YEAR/TOTAL
Yes	B1.	The next questions are about things that people sometimes do for recreation or to relax. I'm going to ask how often you've done each of these activities in the last 12 months, that is since (THIS MONTH) 1990. First, having dinner in a restaurant. In the last 12 months, did you ever have dinner in a restaurant, other than for business? (YES/NO) (IF YES) Did you have dinner in a restaurant, other than for business ... EVERY DAY, SEVERAL TIMES A WEEK, ONCE A WEEK, EVERY FEW WEEKS, ONCE A MONTH, EVERY FEW MONTHS, LESS THAN EVERY FEW MONTHS, or NEVER?	C1.	The next questions are about things that people sometimes do for recreation or to relax. I'm going to ask how often you've done each of these activities in the last 12 months, that is since (THIS MONTH) 1990. First, having dinner in a restaurant. In the last 12 months, did you ever have dinner in a restaurant, other than for business? (YES/NO) (IF YES) How often did you have dinner in a restaurant, other than for business? ______ TIMES PER ______ DAY ______ WEEK ______ MONTH ______ YEAR/TOTAL

Center for Demography & Ecology
University of Wisconsin
1180 Observatory Drive, Rm. 4412
Madison WI 53706-1393
U.S.A.
608/262-2182
FAX 608/262-8400