2000

Decomposition Can Harm the Accuracy of Behavioral Frequency Reports

Robert Belli
University of Nebraska-Lincoln, bbelli2@unl.edu

Norbert Schwarz
University of Michigan - Ann Arbor

Eleanor Singer
University of Michigan - Ann Arbor

Jennifer Talarico
University of Michigan - Ann Arbor

Decomposition Can Harm the Accuracy of Behavioral Frequency Reports

Robert F. Belli, Norbert Schwarz,
Eleanor Singer, and Jennifer Talarico

University of Michigan

Corresponding author — R. Belli

Abstract
In survey research, the use of decomposition can lead to pronounced reporting errors as seen by overreporting and overall reporting error. A total of 87 subjects answered either decomposed or undecomposed questions concerning telephone calls made by them while at work. The questionnaire conditions varied the length of the reference period (1 week or 6 months), and the type of call (local or long-distance). Decomposition conditions introduced either spatial or temporal cues. In all comparisons, decomposed questions increased over- reporting bias relative to undecomposed questions. In addition, undecomposed questions with a 1-week reference period led to increased overreporting bias in compar-ison to undecomposed/ 6-month questions. Results are consistent with a category split estimation model in which smaller categories are predicted to lead to overreporting, and larger categories to under- reporting. Decomposition is not recommended for gaining ret-rospective reports of non-distinctive, frequent events.

Decomposition is a questionnaire design strategy, often advocated in survey research, in which behavioral frequency reports for a category are broken down by asking about the behavioral frequencies for subcategories. For example, instead of asking respondents how often they have eaten at a restaurant during a specified reference period, this ques-tion may be "decomposed" into several more specific ones, pertaining to how often re-spondents have eaten at a fast-food restaurant; an Italian restaurant; a Chinese restau-rant; and so on (e.g. Sudman and Schwarz, 1989; for a review see Sudman *et al.*, 1996). This strategy reliably results in higher-frequency reports, i.e. the sum of the events re-ported in response to the subcategory questions exceeds the number of events reported in response to the general question. It is unclear, however, if these higher-frequency reports reflect higher accuracy. In fact, plausible theoretical arguments can be marshaled for the assumption that decomposition results in erroneous overreporting, as well as for the as-sumption that it increases the accuracy of recall.

Decomposition May Increase Accuracy

In general, accurate recall is facilitated by appropriate retrieval cues. Breaking down the global category into specific subcategories provides such cues, fostering the recall of events that may otherwise be forgotten (Strube, 1987). Moreover, respondents are generally more likely to engage in an enumeration strategy when the number of events is small, but resort to estimation strategies when the number of events is large (e.g. Blair and Burton, 1987; Schwarz, 1990; for a review see Sudman *et al*., 1996). Because the number of events pertaining to each subcategory is smaller than the number of events pertaining to the global category, decomposition may foster the use of enumeration strategies. This should facilitate the accuracy of recall in conditions in which enumeration is helpful. This is the case when the behaviors are irregular and relatively distinct (Menon, 1997).

Empirically, the best evidence that decomposition can improve the validity of retrospective reports in surveys has been provided by Menon (1997). In her laboratory-based work, respondents kept diaries to record their engagement in three regularly or periodically occurring behaviors (e.g. washing hair) and three irregularly occurring behaviors (e.g. drinking water from a fountain) for the period of 1 week. Subsequently, they completed a self-administered questionnaire that did or did not introduce decomposition in assessing the same behaviors. Using the diary entries for validation, Menon found that the decomposition strategy improved the accuracy of reports for irregularly occurring behaviors. Consistent with the observation that low-frequency categories encourage enumeration whereas high- frequency categories encourage estimation (e.g. Blair and Burton, 1987), Menon observed that decomposition increased the use of enumeration strategies. These strategies are beneficial to the recall of irregularly occurring behaviors, particularly if the decomposition introduces cues that facilitate the mnemonic location of specific instances.

Decomposition May Decrease Accuracy of Recall

On the other hand, decomposition may decrease the accuracy of recall for at least two reasons. First, when the behavior is highly regular a rate-based estimation strategy ("I do this once a week, which makes four times a month") results in more accurate reports than an enumeration-based strategy, as Menon (1997) documented in the study described above. Encouraging enumeration through decomposition produces no added benefits and may actually harm accuracy in these conditions, since highly similar multiple episodes are poorly represented in memory.

Second, estimation processes are known to be influenced by category size. For example, Fiedler and Armbruster (1994) presented stimuli (e.g. geometric symbols or letters) at different frequencies and subsequently asked participants for frequency estimates. They observed that the frequency estimates for superordinate categories (e.g. "geometric symbols") were lower than the sum of the estimates for the smaller subcategories (e.g. "triangles", "squares", "circles") that made up the superordinate category. Note that the procedures and results of these category split experiments parallel those of decomposition strategies. In both cases, a larger category ("geometric figures" or "eating out for lunch")

is decomposed into several smaller and mutually exclusive subcategories (specific shapes or types of restaurants) and the sum of the frequency estimates for the subcategories exceeds the frequency estimate for the superordinate category. However, the interpretation of these parallel effects differs in the respective literatures. Whereas survey researchers hope that the higher estimates resulting from decomposition reflect better memory for more specific events, the category split effects observed in laboratory experiments can be traced to information loss, i.e. poorer memory (Fiedler, 1991). The less well actual frequencies are represented in memory, the more estimates regress to the mean, resulting in the underestimation of high frequencies and the overestimation of low frequencies (Fiedler, 1991; Fiedler and Armbruster, 1994). Moreover, such category split effects are not limited to frequency estimates: Decomposing a category into its subcategories has been found to influence a variety of different judgments, with the summed weights given to the subcategories being higher than the weight given to the category as a whole (see Fischhoff *et al.*, 1978; Tversky and Koehler, 1994; Van der Pligt and van Schie, 1991, for examples from the decision literature).

Several processes, which are not mutually exclusive, may contribute to category split effects. First, the more concrete subcategories may facilitate recall of a few pertinent examples relative to the more abstract superordinate categories. If so, respondents may infer from the ease with which examples come to mind that there must be many of them, in line with Tversky and Kahneman's (1973) availability heuristic. Second, as numerous experiments in the tradition of Parducci's (1986) range–frequency model demonstrated, respondents are likely to assign an equal number of items to each of the categories presented to them, resulting in judged frequency distributions that are flatter than the actual frequency distribution. Finally, if the actual frequencies for subcategories are not well represented in memory, estimates may regress to the mean of the overall distribution (Fiedler, 1991; Fiedler and Armbruster, 1994). Any of these processes would result in overestimates of lower frequencies associated with subcategories and underestimates of the higher frequencies associated with more inclusive superordinate categories.

Despite the theoretical appeal of the category split hypothesis, there has been little direct evidence that supports its influence during decomposition in the survey literature. Means *et al.* (1994) did find a tendency for decomposition to decrease accuracy and lead to overreporting. Other validation work has either found no influence of decomposition (e.g. Menon, 1991, unpublished dissertation), or, as noted above, benefits of decomposition on data quality (Menon, 1997; see also Armstrong *et al.*, 1975, on decomposition improving the accuracy of quantitative estimates of trivial facts such as the number of families in the USA in 1970).

Menon (1997) also found, somewhat counter-intuitively, that in the undecomposed condition, the irregularly occurring behaviors were consistently overreported, a finding that replicated earlier work (Menon, 1993). The exact mechanism for this overreporting is unknown, although a reasonable speculation is that subjects are forward telescoping into the 1-week reference period events that had happened earlier (see Menon, 1993). Importantly, Menon (1997) discovered that decomposition was effective in reducing this overreporting. This observation stands in direct contrast to the implications of the category split model (Fiedler and Armbruster, 1994), which predicts that decomposition processes increase overreporting.

Present Research

In sum, several researchers have observed that decomposition increases the number of events reported. As Bradburn *et al*. (1987) noted, survey researchers usually assume that forgetting is the key problem with behavioral reports. Hence, they often rely on a "more-is-better" logic in assessing the usefulness of questioning strategies, despite early warnings that aided recall strategies with events that are not likely to be forgotten may result in overreporting (see Sudman and Bradburn, 1982, p. 38). Accordingly, decomposition has become a popular strategy because it reliably results in higher reports. It is unclear, however, how decomposition affects the accuracy of the obtained reports, as increased as well as decreased accuracy has been observed.

This study addressed these issues by asking employees of a survey research firm to report the number of long-distance and local telephone calls that they had placed from work. We subsequently assessed the accuracy of their reports on the basis of the firm's telephone bills. In addition to varying the presence or absence of decomposition, our experimental design varied the type of recall cues offered as part of the decomposed questions (temporal or spatial), and the length of the reference period (6 months or 1 week). Type of cues was varied to assess whether the effectiveness of decompositional cues would affect the accuracy of reports. Six-month and 1-week reference period lengths are paradigmatic of long and short reference periods in survey research; their introduction was designed to assess the robustness of effects of decomposition across a wide range of reference period lengths.

Long-distance calls

Respondents in the *undecomposed* conditions were asked a single question on how many long-distance calls they had placed during a reference period of 6 months or 1 week. In the decomposed conditions, two different types of cues were used. When asked to report calls during the past 6 months, either temporal or spatial decomposition cues were used. The *temporal cue* condition asked the number of long-distance calls in each of the 6 months. The *spatial cue* condition asked about calls made to particular cities within regions of the United States over the entire 6-month period.

Recall models as well as estimation models predict that decomposition will increase the number of long-distance calls reported, although for different reasons. From an estimation perspective, *any* decomposition procedure results in the decomposition of a larger global category (all calls within a 6-month period) into several smaller categories. Given that large categories are typically underestimated, whereas small categories are overestimated (Fiedler and Armbruster, 1994), it follows that more calls will be reported in the decomposed than in the undecomposed conditions, and that this decomposition effect is independent of the cue used. With regard to the accuracy of recall, we may further expect that calls are underestimated relative to records in undecomposed conditions, but overestimated in decomposed conditions.

In contrast, recall models suggest that the impact of decomposition depends on the specific cues used. Previous research has shown that cues based on time are ineffective with respect to remembering autobiographical events (e.g. Wagenaar, 1986) and we assume this holds true for remembering one's phone calls as well. On the other hand, cues

based on place, on persons, and on what has transpired have been found to facilitate autobiographical recall (e.g. Wagenaar, 1986). Of these three cue types, place cues were introduced in the present research. Decomposing on the basis of person and content cues is rather complicated because of the extent to which they are idiosyncratic to individual respondents. In contrast, we were able to decompose the places in which calls were made in a general way by categorizing on the basis of geographical region.

The place cues used by Wagenaar (1986) pertained to the location of the to-be-recalled event, which for the phone calls assessed in the present study, is always the respondent's office. Nevertheless, it seems highly likely that the location of the person called is a better retrieval cue than the time of call, resulting in more accurate recall. Accordingly, recall models predict that spatial cues (location of the person called) will increase recall accuracy, whereas temporal cues (time of call) will not. Note, however, that estimation and recall processes are not mutually exclusive (see Schwarz, 1990, for a review). Instead, episodic recall may either serve as input into estimation processes, or the result of an estimation strategy may be checked for plausibility against recalled episodes. If so, spatial cues may limit the overestimation resulting from decomposition strategies relative to temporal cues.

With respect to telephone usage over the past week, only a temporal decomposition cue was tested. In this case, respondents were asked to provide estimates of their long-distance calling during each day of the week. Consistent with the detailed predictions derived above, we expected that temporal decomposition would increase reports relative to an undecomposed condition, but would be associated with a loss of accuracy.

Finally, the assumption that estimation shows a regression to the mean (Fiedler and Armbruster, 1994) also bears on the undecomposed conditions. If small categories are likely to be overestimated and large categories underestimated, we should observe overreporting relative to records for the 1-week reference period and underreporting for the 6-month period. Moreover, the number of calls reported for a 1-week period should greatly exceed the number of weekly calls one would expect on the basis of the reports for the 6-month period.

Local calls

For questions about local calls, only temporal decomposition cues were used. In the *undecomposed* condition, respondents were asked a single question on the number of local calls for either a 6-month or a 1-week reference period. For the 6-month reference period, only one temporal decomposition cue was tested, asking respondents about *each of the months*, separately. Two different temporal decomposition cues were tested with the 1-week reference period, one that decomposed the reference period by asking separate questions for telephone usage during *each day of the week*, and another condition in which respondents were asked separate questions about usage during different *times of the day* over the entire week.

We again predicted that the temporal decomposition cues would increase the number of reported calls relative to the undecomposed conditions. Again, this increase in reporting should be associated with a loss of accuracy. Since the telephone billing records did not include a daily tally of local telephone calling, however, we could only assess the accuracy of reports provided in the 6-month reference period.

In sum, the conditions included in this experiment allow for targeted tests of the impact of different types of decomposition cues on the number of reported events, and the accuracy of these reports, for both long and short reference periods.

Method

Subjects

A total of 130 employees of the University of Michigan's Survey Research Center were selected on the basis of whether they had a telephone for their professional use while at work.[1] Altogether, 87 (66.9%) employees responded to the telephone use items on the questionnaire, with the remainder either refusing, failing to answer the telephone use questions, or not being available for contact during the data collection period.

Design

The design was unbalanced. Five self-administered questionnaire versions were developed, and the selected employees were randomly assigned to one of the questionnaires with the constraint that an equal number of employees ($N = 26$) were assigned to each version. Each of the questionnaire versions asked about both local and long-distance telephone calls made at work, although respondents were presented with only one type of decomposition cue (none, month, regional cities, day of week, or time of day) and one length of reference period (6 months or 1 week) for each type of call (long-distance or local). Table 1 illustrates each of the questionnaire versions with regard to the presentation of the types of calls, the absence or presence of decomposition cues, the type of cue, and

Table 1. Questionnaire version by type of call, type of decomposition cues, and length of reference period

Version	Calls	Decomposition cues	Reference period	Completed N	Validated N
6M/U	Long	Undecomposed	6-month	19	15
	Local	Undecomposed	6-month	19	15
6M/D	Long	Month	6-month	16	13
	Local	Month	6-month	16	13
1W/U	Long	Undecomposed	1-week	19	14
	Local	Undecomposed	1-week	19	0
1W/D	Long	Day of week	1-week	16	11
	Local	Day of week	1-week	14	0
Mixed/D	Long	Regional cities	6-month	17	15
	Local	Time of day	1-week	17	0

1. An additional 45 respondents participated by answering one of three questionnaires that contained a level of decomposition that was lower than the decomposed questions reported in this paper. Our intent was to determine whether, consistent with the estimation model, overreporting would be a function of the degree of decomposition. Unfortunately, there was considerable attrition in validated responses in these middle conditions due either to the refusal to examine records ($N = 3$) or to the telephone numbers of these respondents not being located in the database ($N = 17$). The remaining 25 respondents led to insufficient data to allow for the degree of power that would be necessary to determine whether a middle degree of decomposition incrementally induced overreporting, especially given the high levels of variance found with all conditions. The reader is advised that the reported results are based on comparing no decomposition conditions against those conditions with a high level of decomposition, and that the influence of middle levels of decomposition remain as an open issue.

the length of the reference period. All questionnaires were answered during the 2 weeks from January 15 to January 26, 1996.

Six-month period

For the 6-month reference period, respondents were asked to report calls that they made from 1 July 1995 through 31 December 1995. In the *undecomposed* condition, respondents were asked to respond to a single question that covered the entire 6-month reference period. In the *month cue* condition, respondents were asked six questions regarding their telephone usage for each of the months of the reference period. In the *regional cities cue* condition, a total of 12 questions asked about calls made in the United States. Respondents were asked to list the three cities that they call most often in each of three regions: East Coast States, West Coast States, and Other States. They were then specifically asked to report the number of calls made during the past 6 months to each of these nine cities, as well as to any other cities that they had not specified within each of the regions, resulting in twelve specific questions. Both long- distance and local calls were queried in the undecomposed and month cue conditions, whereas the regional cities cue condition was used only for long-distance calls.

One-week period

For the 1-week reference period, respondents were asked to report calls from the Monday through Friday of the week before the week that contained the day on which they answered the questionnaire. In the *undecomposed* condition, respondents answered one question concerning the entire reference period. In the *day-of-week cue* condition, respondents were asked ten questions that asked about calls made during the morning and afternoon on each of the 5 days. In the *time-of-day cue* condition, respondents answered ten questions about calls that they had made during the entire last week that corresponded to ten hourly spells (i.e. before 9 a.m., 9–10 a.m., 10–11 a.m., … after 5 p.m.) of the day. The undecomposed and day-of- week cue conditions were used in queries about both long-distance and local calls, whereas only local calls were tested with the time-of-day cue condition.

Validation

During administration, respondents were informed that the questionnaire was seeking to discover the best way to help people remember their telephone use while at work, and that they were not to consult any records. To ensure that records were not consulted, a research assistant was on hand to collect questionnaires as soon as they were completed. Immediately following the completion of a questionnaire, respondents were asked for their written permission to examine their telephone records, and also asked to provide their telephone number in order to match their responses with the records. Of the 87 respondents who completed the questionnaire, all but one agreed to both requests. Records in a centrally maintained billing database served as the validation data, and matching of records with responses was made on the basis of the telephone number given by the respondent. The telephone numbers of 18 of the 86 respondents who gave permission to examine records were not located in the database as billing charges did not go directly to these numbers. The record data for each respondent consisted of individual listings for each long-distance telephone call that was made. Information concerning each long-dis-

tance call included the date on which the call was made and the city and state of destination. Weekly information concerning long-distance calls was therefore readily available. For local calls, the number of calls per month was determined by taking the total charge for calls made and dividing by the standard per-call charge. Because of the lack of a day-to-day breakdown of local calls, validation for any time period shorter than 1 month was not possible.

In summary, the amount of validated data was compromised both by an inability to locate the telephone numbers of 18 respondents, and by the inability to validate local calls that were reported for the 1-week reference period. Table 1 provides a summary of the conditions presented with each questionnaire version, and the number of respondents who completed the telephone use questions and then were validated for each of the conditions.

Results and Discussion

Table 2 presents the mean number of calls reported in each condition, the number of calls in the telephone records, the signed differences between reports and records (negative numbers indicate underreporting, and positive numbers overreporting), and the absolute values of the differences between reports and records, with lower numbers indicating a greater overall level of accuracy.

Does decomposition increase the number of reported events?

We first examined whether decomposition increases the number of reported calls relative to the respective undecomposed conditions. The relevant data for long-distance (Panel A) and local calls (Panel B) are shown in the third column of Table 2. As expected, more calls were reported in decomposed than in undecomposed conditions and this pattern holds for each of the six possible comparisons. The binomial probability for this result is $p = 0.02$, two-tailed, confirming that, overall, decomposition increased the number of reported events.

More detailed comparisons, using between group t-tests, indicate, however, that the size of the individual decomposition effects is quite variable. Moreover, the variance within each condition is very high, rendering many of the individual comparisons nonsignificant. For long distance calls during a 6-month period, the decomposition by month led to an increase of 59.0%, $t(33) = -0.72$, ns, and the decomposition by regional cities increased the estimates by a whopping 176.8%, $t(36) = -1.58$, $p = 0.06$, one-tailed. For long-distance calls during a 1-week period, day-of-week decomposition increased the number of reported events by 23.0% relative to the undecomposed condition, $t(33) = -0.43$, ns.

The same pattern holds for local calls, as shown in Panel B of Table 2. For the 6-month reference period, decomposition by month increased reports by a whopping 208.4% ($t(33) = -3.24$, $p = 0.001$, one-tailed). For the 1-week reference period, day-of week and time-of-day decomposition led to increases in reports relative to no decomposition by 6.0% ($t(31) = -0.18$, ns) and 66.3% ($t(34) = -1.46$, $p = 0.08$, one-tailed), respectively.

Finally, the effect of decomposition is also apparent in a comparison of the reports obtained in the undecomposed 1-week and 6-month conditions. For long-distance calls,

Table 2. Number of cases, means, and standard deviations of reported calls, recorded calls, and accuracy of reports, by condition

(1) Reference period	(2) Decomposition cues	(3) Reports[a]		(4) Records[b]		(5) Raw differences[b]	(6) Absolute differences[b]
		N	M(s)	N	M(s)	M (s)	M (s)
A: Long-distance calls							
6-month	None	19	61.7 (133.9)	15	82.1 (127.4)	-9.8 (61.0)	34.6 (50.4)**
	Month	16	98.1 (165.2)	13	37.4 (27.8)	30.5 (85.5)	40.8 (80.7)*
	Regional cities	17	170.8 (265.4)	15	85.1 (102.4)	98.6 (199.1)*	124.7 (182.8)**
1-week	None	19	7.4 (9.1)	14	3.7 (3.5)	5.0 (9.6)*	6.4 (8.6)**
	Day of week	16	9.1 (13.6)	11	2.4 (3.3)	7.4 (13.3)*	8.4 (12.7)*
B: Local calls							
6-month	None	19	65.4 (47.3)	15	87.7 (66.5)	-18.2 (75.3)	47.8 (59.8)***
	Month	16	201.7 (176.6)	13	72.3 (74.1)	136.6 (190.5)**	154.3 (175.3)***
1-week	None	19	11.6 (8.7)				
	Day of week	14	12.3 (12.4)				
	Time of day	17	19.3 (21.0)				

Two-tailed t-tests for whether sample mean is significantly different from 0: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

a. The means and standard deviations for those reports that have associated records are presented in footnote 2.

b. The Records, Raw differences, and Absolute differences columns have the same number of respondents per condition.

multiplying the weekly report of 7.4 by 26 weeks provides an estimate of 192.4 calls for the 6-month period, exceeding the report in the 6-month condition of 61.7 calls by 211.8%. For local calls, the weekly average of 11.6 calls results in a 26-week estimate of 301.6 calls, which exceeds the report in the 6-month condition of 64.7 calls by a whopping 366.2%.

In sum, decomposition increased the number of reported long-distance and local calls in all cases, with differences up to 366%, although many of the individual comparisons failed to reach significance. As Bradburn *et al*. (1987) noted, this observation would encourage many researchers to employ a decomposition strategy based on the logic that the key problem is forgetting and hence, "more is better". Yet, the increased reports may not reflect increased accuracy, an issue to which we turn next.

Does decomposition increase or decrease accuracy?

The fourth column of Table 2 shows the mean number of long-distance calls according to the billing database. For some respondents, no records could be located and our analyses of accuracy are based on the subset for which complete data are available (*N*s are shown in the fourth column of Table 2).[2] In addition, since the database contained only records of local calls on a monthly basis, analyses were restricted to the 6-month reference period for local calls. There were no systematic differences between the undecomposed and the various decomposed conditions with regard to the actual number of long-distance and local calls made.

The fifth column shows the mean raw differences for all conditions that allowed validation, with negative numbers indicating underreporting and positive numbers indicating overreporting. As predicted by Fiedler and Armbruster's (1994) category split model, *all* the decomposed conditions *increased overreporting* in comparison to the undecomposed conditions, a pattern associated with a binomial probability of $p = 0.06$, two-tailed.

To test the degree of over- or underreporting, we compared respondents' reports with their records, using pairwise *t*-tests. For the 6-month reference period, the undecomposed questions resulted in underreporting of long-distance calls of 11.9% relative to records ($t(14) = -0.62$, ns). In contrast, the decomposed questions elicited considerable overreporting, both when months [81.6%, $t(12) = -1.28$, ns) and regional cities (115.9%, $t(14) = 1.92$, $p = 0.04$, one-tailed) were used as cues. Moreover, the raw discrepancies between the reports and the records were higher in each of these decomposed conditions than in the un-

2. The means and standard deviations presented in column 3 of Table 2 are based on data from all the available reports. Yet analyses in this section are limited to those reports in which there are associated records. Thus, the mean and standard deviation values of reports are not the same here as those depicted in Table 2, and instead, are reported in this footnote. Limiting data to those reports with associated records, for long-distance calls and the 6-month reference period, respondents in the undecomposed condition (*N* = 15) reported a mean of 72.3 calls (*s* = 149.6), those in the month cue condition (*N* = 13) reported a mean of 67.9 calls (*s* = 95.6), and those in the regional cities cue condition (*N* = 15) reported a mean of 183.7 calls (*s* = 280.4). For long-distance calls and the 1-week reference period, respondents in the undecomposed condition (*N* = 14) reported a mean of 8.7 calls (*s* = 10.1), whereas those in the day-of- week cue condition (*N* = 11) reported a mean of 9.8 calls (*s* = 14.9). Finally, in responses to local calls for a 6-month reference period, respondents in the undecomposed condition (*N* = 15) reported a mean of 69.5 calls (*s* = 51.2), and those in the month cue condition (*N* . 13) reported a mean of 208.9 calls (*s* = 184.8).

decomposed condition. For decomposition by month, $t(26) = -1.45$, p . 0.08, one-tailed; for decomposition by regional cities, $t(28) = -2.02$, $p = 0.03$, one-tailed.

For the 1-week reference period, long-distance calls were overreported by 135.1% in the undecomposed condition to begin with (pairwise $t(13) = 1.94$, $p = 0.07$, two-tailed), and this overreporting increased to 308.3% when the 1-week period was decomposed by day-of-week ($t(10) = 1.86$, $p = 0.045$, one-tailed). However, the difference in the raw discrepancies between reports and records in the decomposed and undecomposed condition failed to reach significance; $t(23) = -0.54$, ns, for the between-groups comparison.

The pattern was repeated with reports of local calls in the 6-month reference period. The undecomposed question led to an underreporting of 20.8% relative to records ($t(14) = -0.94$, ns), whereas the decomposed question led to an 188.9% increase in reported calls ($t(12) = 2.59$, $p = 0.01$, one-tailed). The raw difference between reports and records was significantly higher in the decomposed condition than the undecomposed condition, $t(26) = -2.90$, $p = 0.004$, one-tailed. It is worth emphasizing the results in the undecomposed conditions that have just been reported. Comparing the 1-week and 6-month reference periods, we have shown marginally significant overreporting for the 1-week period and non-significant underreporting for the 6-month period. These results are again consistent with the predictions of the category split model, which assumes that estimations regress to the mean.

Finally, both spatial (regional cities) and temporal (month) decomposition cues led to overreporting. This finding is again consistent with the category split predictions and does not support our assumption that the effects of decomposition may depend on the mnemonic usefulness of the decomposition cues.

The final column of Table 2 shows the absolute amount of error. Interestingly, all conditions produced error rates significantly greater than zero, reflecting that respondents do generally err when reporting about their past telephone use. More importantly, decomposition led to significantly greater error than no decomposition ($p = 0.06$, two-tailed, as a binomial probability). The greater inaccuracy of decomposition in comparison to no decomposition was most pronounced for reports of long- distance calls over a 6-month reference period with regional cities as decomposition cues, $t(28) = -1.84$, $p = 0.08$, two-tailed, and for reports of local calls over a 6-month reference period with months as decomposition cues, $t(26) = -2.21$, $p < 0.05$, two-tailed.

General Discussion

Overall, our results confirm that the use of decomposition strategies increases the number of reported events, and that the use of decomposition strategies is fraught with more problems than has typically been assumed. Assuming that forgetting is the key problem in behavioral frequency reports (Bradburn *et al.*, 1987), questionnaire designers have typically been delighted with the observation that decomposition increases the number of events reported. Yet, more is not always better. Instead, decomposition can exacerbate overall reporting error and, worse, can introduce a systematic bias, namely pronounced overreporting. These observations are consistent with Fiedler and Armbruster's (1994) category split model, which predicts overestimation for small categories and un-

derestimation for large categories. Decomposing a global question into several more specific ones creates smaller categories and hence facilitates overreporting, resulting not only in increased error rates but in systematic bias. Although the overall pattern of our findings provides strong support for this prediction, it is worth emphasizing that many of the individual comparisons did not reach significance.

The detrimental effects of decomposition were most pronounced with a 6-month reference period. Several factors are likely to contribute to this observation. First, longer reference periods contain a larger number of events, and the higher the frequency of events, the more likely respondents are to rely on an estimation strategy (e.g. Blair and Burton, 1987; Burton and Blair, 1991). Second, longer reference periods include events that happened long ago, making recall of specific instances more difficult. Accordingly, corrections of the estimate in light of partial recall may be less likely. In contrast, shorter reference periods encourage enumeration strategies and partial episodic recall may serve to correct any estimates. If so, decomposition may be helpful for short reference periods, provided that effective decomposition cues are used—which was not the case for the 1-week reference period of the present study, which relied on temporal cues.

Importantly, our data provide no support for the assumption that spatial cues may facilitate accurate recall relative to temporal cues and may hence limit the over- reporting bias inherent in estimation strategies. As noted above, the spatial cues used here differed from the spatial cues that have been found beneficial in autobiographical memory research (e.g. Wagenaar, 1986). Whereas the cues typically used pertain to the location of the event, the cues used in the present study pertained to the location of the person called, with all calls originating from the respondent's office. Hence, additional research will be needed to address the potential helpfulness of spatial decomposition cues pertaining to the respondent's location, or of decompositional strategies that can take advantage of person cues or cues directed to remembering what had transpired.

Moreover, we observed a tendency for spatial cues to lead to greater overestimation of long-distance calls made during a 6-month period than was the case for temporal cues. The spatial cues decomposed the global category into more subcategories (12) than did the temporal cues (6), and thus a greater degree of overreporting from the presentation of more subcategories is expected from the category split model. As an additional possibility, spatial cues are more potent pointers to specific events than temporal cues and may hence facilitate the retrieval of a few examples. Relying on the availability heuristic (Tversky and Kahneman, 1973; Schwarz, 1998), respondents may infer from the ease with which examples come to mind that there must be many of them, resulting in overestimation. This raises the interesting possibility that effective retrieval cues may exacerbate overreporting under conditions where they are sufficient to facilitate the easy recall of some examples but insufficient to allow for an enumeration strategy. Note, however, that presumably ineffective temporal decomposition cues also led to overreporting. Hence, overreporting cannot be solely traced to the possible operation of the availability heuristic but is likely to reflect other estimation processes as well, as discussed in the context of the category split model (Fiedler and Armbruster, 1994).

It is worth noting that both our data as well as Menon's (1997) show a tendency for overreporting in the 1-week reference period. Menon (1993) speculated that this overreporting might be due to respondents remembering and including instances from out-

side the reference period as a form of forward telescoping. Although we do not disagree that forward telescoping may occur, this finding may also reflect that small categories are likely to be overestimated, a process that does not require erroneous dating of events. As one possibility, if remembering a few instances from a 1-week reference period is found to be easy, the use of an availability heuristic to infer category size will lead to overestimation.

To date, decomposition has only been shown to be clearly beneficial when the questions pertain to a short reference period, when enumeration strategies are used, and when decompositional cues are effective towards increasing the location in memory of those specific episodes that are contained within the decomposed categories (Menon, 1997). Importantly, Menon has found that the use of enumeration strategies, which have been benefited by effective decompositional cues, reduces overreporting. Apparently, decomposition can be effective in reducing the over- estimation that typically occurs for small global categories, such as short 1-week reference periods, when the decomposition encourages successful enumeration of specific instances. In addition, decomposition may be beneficial for more extended reference periods when the questions provide helpful cues and pertain to distinctive events that are relatively few in number, thus again encouraging enumeration (e.g. Means and Loftus, 1991).

In contrast, decomposition seems unlikely to increase accuracy in conditions that do *not* encourage enumeration, as the present findings indicate. Unfortunately, many events that deserve study are frequent rather than rare, and common rather than distinctive, thus making enumeration unlikely. In these conditions, the use of decomposition strategies does not seem very beneficial – although decomposition will increase the number of reported events, it may do so at the cost of decreased accuracy. Of course, a lucky researcher may occasionally hit on a degree of decomposition that is "just right", for example conditions under which overestimates for small subcategories are compensated by underestimates for larger subcategories. Unfortunately, such lucky circumstances are likely to be rare and, based on the available evidence, decomposition cannot be recommended as a general strategy for the assessment of behavioral frequency reports.

References

Armstrong, J. S., Denniston, W. B. and Gordon, M. M. (1975). The use of the decomposition principle in making judgments. *Organizational Behavior and Human Performance*, **14**, 257–263.

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

Bradburn, N. M., Rips, L. J. and Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science*, **236**, 157–161.

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

Fiedler, K. (1991). The tricky nature of skewed frequency tables: An information-loss account of distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology*, **60**, 24–36.

Fiedler, K. and Armbruster, T. (1994). Two halfs may be more than one whole: Category-split effects on frequency illusions. *Journal of Personality and Social Psychology*, **66**, 633–645.

Fischhoff, B., Slovic, P. and Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance*, **4**, 330–344.

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

Means, B., Swan, G. E., Jobe, J. B. and Esposito, J. L. (1994). The effects of estimation strategies on the accuracy of respondents' reports of cigarette smoking. In N. Schwarz and S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 105–119). New York: Springer-Verlag.

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

Menon, G. (1997). Are the parts better than the whole? The effects of decompositional questions on judgments with frequent behaviors. *Journal of Marketing Research*, **34**, 335–346.

Parducci, A. (1968). The relation of absolute judgments. *Scientific American*, **19**, 84–90.

Schwarz, N. (1990). Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction. In C. Hendrick and M. S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 98–119). Beverly Hills, CA: Sage.

Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review*, **2**, 87–99.

Strube, G. (1987). Answering survey questions: The role of memory. In H. J. Hippler, N. Schwarz and S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 86–101). New York: Springer-Verlag.

Sudman, S. and Bradburn, N. M. (1982). *Asking questions*. San Francisco, CA: Jossey-Bass.

Sudman, S., Bradburn, N. and Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.

Sudman, S. and Schwarz, N. (1989). Contributions of cognitive psychology to advertising research. *Journal of Advertising Research*, **29**, 43–53.

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, **5**, 207–232.

Tversky, A. and Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, **101**, 547–567.

Van der Pligt, J. and van Schie, E. C. M. (1990). Frames of reference, judgment and preference. In W. Stroebe and M. Hewstone (Eds.), *European review of social psychology* (Vol. 1). Chichester: Wiley.

Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology*, **18**, 225–252.