PREDICTING OTHERS' BEHAVIORAL FREQUENCIES: THE ROLE OF JUDGMENT STRATEGY, KNOWLEDGE, AND REGULARITY

BARBARA BICKART Rutgers University-Camden

JOAN M. PHILLIPS University of Notre Dame

GEETA MENON New York University

> JOHNNY BLAIR Abt Associates

We gratefully acknowledge the contributions made by the late Seymour Sudman during the early stages of this project. We also acknowledge the financial support of the University of Illinois, Rutgers University School of Business-Camden, New York University, University of Kentucky, and the University of Maryland in completing this project and appreciate the helpful comments of Fred Conrad on an earlier draft of this manuscript.

PREDICTING OTHERS' BEHAVIORAL FREQUENCIES: THE ROLE OF JUDGMENT STRATEGY, KNOWLEDGE, AND REGULARITY

Abstract

In this paper, we examine the influence of knowledge, characteristics of a behavior (i.e., regularity) and judgment strategy on the accuracy of reports of another person's behavior. In Experiment 1 we find that when people are asked for reports about others, they tend to underestimate actual behavior and are less accurate when asked to count occurrences of the behavior ("How many times...?") versus estimate ("How often...?"). In addition, they tend to under-report behavior when the level of discussion with the partner about the behavior is limited. Protocol analyses suggest that people are more likely to recall episodes of the behavior learned via discussion with their partner when they are asked to count and when discussion is extensive. We replicate and extend these findings in Experiment 2. In particular, we show that respondents' ability to estimate the frequency of another person's behavior accurately increases with the regularity of the behavior and the closeness of the relationship. We discuss the theoretical implications of the findings for understanding how memory structures about others affect prediction judgment processes, as well as the practical implications of the findings for survey design.

Peoples' ability to report accurately about their own and other household members' past behavior is an important issue for consumer researchers. In the context of survey research, questionnaires often rely on "proxy" reports, where one member of the household answers questions about the behavior of other household members. The rationale behind this strategy is that this method saves time and money. Therefore, marketing research surveys often ask one person in the household to report about the purchase behavior and product usage of other household members. In government surveys, proxy-reports are frequently used to collect information about household participation in the labor force in the Survey of Income and Program Participation (SIPP) and the Current Population Survey, and about household expenditures in the Consumer Expenditures Study. For example, on any one wave of the SIPP survey, another household member provides information for approximately 45% of the sample persons (Kasprzyk & McMillen, 1987). In the context of consumer decision-making, one person in the household frequently makes consumption decisions or purchases products for others in the household or for joint consumption.

However, the mechanism by which one person answers questions about another is not well understood (see Moore, 1988 for a review). The primary objective of this paper is to examine factors that might moderate the accuracy of such reports of behavior. Specifically, we examine the effects of three theoretical variables on the accuracy of behavioral reports about other people: (a) how information about the other person is stored in memory; (b) how information is retrieved from memory; and (c) the amount and structure of information consumers have about other household members.

Previous research has focused on identifying variables that moderate the level of agreement between reports about self and others. For example, Menon, Bickart, Sudman, & Blair (1995) show that increased discussion or level of participation between dyad members leads to a greater correspondence between reports about self and a partner's report about them on behaviors and attitudes (see also Kojetin & Miller, 1993; Kojetin & Mullin, 1995; Kojetin &

Jerstad, 1997; Sudman, Bickart, Blair, & Menon, 1994). In this paper, rather than examining the agreement between self and proxy reports as in these past studies, we compare reports about others to actual behavior (either recorded via a computer or diaries), while manipulating variables of theoretical and practical interest. We report the results of two experiments in which we use different measures of knowledge about a partner's behavior (extent of discussion, observation, and closeness of the relationship), and examine the effects of knowledge on the accuracy of reports about the partner. Further, we examine the effects of question wording on these responses. We examine these issues over a range of behaviors varying on frequency, as well as in a laboratory setting (Experiment 1) and in a real-life everyday setting (Experiment 2) and find convergent validity for our results.

Based on the research on how events are stored in memory and how people answer behavioral frequency questions, we develop a conceptual model that suggests the processes by which respondents should answer questions about others under various conditions, as well as the accuracy of reports about others under such conditions. We discuss the implications of our findings for marketing and consumer research.

EXPERIMENT 1

Our conceptual framework for understanding the processes used to answer questions about others' behavior and the accuracy of such reports is organized around three theoretical constructs that we believe will drive the processes used to answer behavioral frequency questions about others as well as the accuracy of these reports. These factors include the way in which behavioral information is stored in memory, the judgment strategies used to construct a report, and the extent of knowledge about other people.

Structure of event memory

Information about past behaviors is stored in memory in several types of representations. The autobiographical or "episodic" memory representation of a particular experience may include

such information as the emotions experienced as the person engaged in the behavior, the date of the behavior, and its specific characteristics. There is consensus that the frequent experience of similar events decreases the likelihood that events are stored with specific time or location indicators. Such decontextualization turns episodic information into semantic knowledge (Strube, 1987; Tomkins, 1978). Generalized knowledge has its roots in episodic experiences. Over time, the identity of these roots is lost, together with the context in which such information was acquired (Brewer, 1986).

Thus, the episodic representation of an event is more likely to be accessible for events that occur infrequently, at irregular intervals (Means & Loftus, 1991; Menon, 1993) or are distinctive from other occurrences of the event (Menon, 1993). The "semantic" representation is more likely to be accessible when events occur regularly and are quite similar across occurrences (Conrad, Brown, & Cashman, 1998; Menon, 1993). This representation contains more general information about the behavior, such as a rate-of-occurrence or general impression of how frequently the behavior occurs.

Judgment Strategies for Constructing Behavioral Frequency Reports

Previous research has shown that the judgment strategy used to construct behavioral frequency reports depends on how the relevant information is stored in memory (see Blair & Burton, 1987; Burton & Blair, 1991; Conrad, et al. 1998; Menon, 1993, 1997). For example, when the episodic representation of an event is accessible in memory (e.g., when the behavior is irregular and distinct), respondents are more likely to count specific occurrences of the behavior. When the semantic representation is more accessible (e.g., when the behavior is regular and similar), respondents can use one of a number of estimation strategies. For example, they can recall a rate-of-occurrence for the class of behaviors or recall a general, non-numeric impression of how frequently the behavior occurs (e.g., "a lot" or "not often"; see Menon 1993).

Blair & Burton (1987) showed that question wording affected the strategy used to answer a behavioral frequency question. They found that asking a "How many times...?" question prompted an episodic recall strategy, while asking a "How often...?" question prompted an estimation strategy. As discussed next, the effectiveness of these strategies in producing accurate reports is likely to vary as a function of what kind of information is accessible in memory.

The relative accuracy of using a count versus an estimation strategy will depend on the availability and quality of rate information (or other relevant estimation heuristics) in memory. For example, Menon (1997) shows that the match between the frequency question and the information accessible in memory affects the accuracy of self-reports of behavioral frequencies. In particular, she found that for self-reports of frequent behaviors, encouraging episodic recall when it is *not* the normal strategy (i.e., for frequent, irregular behaviors) increases the effort that is required to answer the question and reduces accuracy. In contrast, encouraging episodic recall "matches" the way that information about irregular events is stored in memory.

Effects of Knowledge on Reporting Strategies and Accuracy

Previous research suggests that increased knowledge (attained via either discussion or joint participation in an event) (a) improves peoples' ability to make inferences about their partner's behavior, (b) increases the likelihood that people recall specific episodes of their partner's behavior, and (c) results in strategies more similar to those used for self-reports (Menon et al., 1995). Likewise, in a study of empathic accuracy, Ickes, Stinson, Bissonnette, & Garcia (1990) found that the accuracy of inferences about a partner's thoughts and feelings increased as dyad members elicited more information from each other. Thus, not surprisingly, increased knowledge improves peoples' ability to predict other people's behaviors, thoughts, and feelings. This appears to be true regardless of whether the knowledge base is comprised of episodic representations of specific events or more rate-based semantic representations.

When episodic or semantic representations are not accessible, people may use other inputs to construct judgments about other people. For example, they could infer another person's behavior based on general knowledge about that person (Bickart & Felcher, 1996; Menon et al. 1995) or they could construct a judgment by anchoring on their own behavior or preferences and adjusting (Davis, Hoch, & Ragsdale, 1986). A number of studies have shown that people are more likely to use general knowledge structures to construct reports about another person's behavior (Bickart, Menon, Schwarz, & Blair, 1994; Dashen 2000; Schwarz & Wellens, 1997). For example, Schwarz & Wellens present the results of several experiments showing that when answering behavioral frequency questions about another person, respondents are more likely to focus on more general impressions or dispositional information, relative to self-reports.

There is some evidence suggesting that how knowledge about others is attained affects their ability to later recall this information. Specifically, conversation differs from other forms of learning (i.e., observation or joint participation in a behavior) in that it requires the establishment of a common ground. Therefore, the speaker takes the knowledge of the listener into account (Fussell & Krauss, 1992), facilitating comprehension and memory. Several studies provide evidence that learning via conversation enhances memory. For example, Schober & Clark (1989) found that recall was better for participants in a conversation than for those who overheard the same conversation. In a survey context, Kojetin & Jerstad (1997) examine the relationship between modes of learning about another person's behavior and agreement between reports about self and others. They found that learning about another person's behavior via conversation was associated with greater agreement between self and others' reports of consumer expenditures.

Conceptual Framework

To summarize, information about behaviors can be stored as both episodic and semantic representations. The effectiveness of various judgment strategies used to recall behaviors varies with the accessibility of these representations. In general, the episodic representations of one's

own behavior are more accessible than those of others, which make it easier to recall specific instances of one's own behavior. In contrast, when constructing behavioral frequency reports about others, people must rely more on other types of knowledge, such as dispositional information about their partner, their own behavior in the same domain, or other heuristics. The extent to which one must rely on general information should depend on the level of discussion or the extent to which the behavior was observed and the way in which information is elicited from memory.

We are interested in the factors that will improve peoples' ability to accurately predict others' behavioral frequencies. In the first experiment, we examine the effects of two factors: the wording of the behavioral frequency question and the *level of discussion* about the behavior. The wording of the behavioral frequency question should influence the extent to which a respondent attempts to recall episodic information versus using other inputs to construct a judgment. The *level* of discussion about the behavior should influence the likelihood that an episodic representation is available in memory.

In Experiment 1, we focus on reports of an irregular, infrequent behavior, for which participants have little opportunity to develop an accurate rate-of-occurrence. Hence, we expect that encouraging episodic recall will result in a better match between how information is stored in memory and how information is elicited (Menon 1997), and thus should result in *more* accurate reports than if an estimation strategy was encouraged. Further, because specific episodes of another person's behavior are not generally easily accessible, when respondents are asked to count incidents to form reports about others, they should be more likely to miss some episodes and therefore under-report their partner's behavior relative to when they are asked to estimate (Conrad et al. 1998). Finally, we expect that overall reporting accuracy should be higher as the level of discussion (i.e., knowledge) increases. Because increasing discussion enhances all possible inputs to a judgment about another person, we expect that increased discussion will improve accuracy regardless of whether respondents are asked to count or estimate.

These hypotheses are now tested in a laboratory experiment, in which reports of behavior about others are compared to actual behavior. In this experiment, the wording of the behavioral frequency question and the level of discussion about the focal behavior are manipulated.

<u>Method</u>

The design was developed based on three criteria. First, we needed a mechanism for recording a relatively frequent behavior that could later be recalled. The recorded behavior could then be compared to the respondents' reports of their own and their partner's behavioral frequencies. Second, we needed an involving task in which participants would be motivated to perform the behavior and discuss it with their partner. Finally, we wanted to be able to manipulate the extent to which people discussed their behavior with another person.

A laboratory experiment in which couples (individuals in a long-term relationship) searched for information about a vacation they could win met these criteria. In this experiment, participants individually searched for information about the vacation in a simulated online setting. Then, couples discussed the vacation options they had examined and selected a vacation. We varied the amount of time couples were allowed to discuss their vacation with their partner. Finally, participants completed a questionnaire asking them about the number of vacation options both they and their partner had examined during the information search task, along with other measures.

<u>Sample</u>

A total of 102 couples in three cities (two in the Northeast and one in the Midwest) were recruited to participate in a study on "decision-making" via newspaper ads and flyers.¹ To encourage participation, couples were paid \$25 and were entered in a \$500 drawing to win a vacation that they would plan. Entry in the drawing also increased the relevance and the

¹ Since there are no differences across the three cities, the results are aggregated across them.

realism of the task (see Kleppe, 1990 for a similar example). Participants' ratings of the realism of the study and their interest in the task were quite high (on a 7 point scale, with 7 = very realistic/interesting; <u>M</u>'s: realism = 5.02; interest = 5.92). Computer files were lost or damaged for four dyads and members of an additional seven dyads did not complete all of the frequency questions. Thus, the results are based on the remaining 90 dyads.

Eighty-eight percent of our sample had completed at least some college and 51% were college graduates. The mean age of the sample was 32.6 years (median = 27). Partners reported that they have known each other an average of 10.9 years (median = 6). Finally, 63% of the study participants were white, 19% were Asian/Island Pacificer, 8% were black, and 10% were some other race.

<u>Design</u>

Two factors were manipulated in the study. First, we manipulated the wording of the behavioral frequency question. Participants were asked to either *count* or *estimate* the number of vacation options that they/their partner examined during the information search task. The exact wording of these questions is included in the Appendix. As per Blair & Burton (1987), we assumed that asking subjects to count would encourage episodic recall. Making a similar assumption about the recall strategies used by participants who were asked to estimate, however, is more difficult since a number of estimation strategies are possible (Conrad et al., 1998; see also Bickart & Felcher, 1996). Therefore, while we were less sure about the exact type of estimation strategy used, we believed that in this condition respondents would use a broader range of information sources and strategies to answer the behavioral frequency questions.

Second, we varied the *amount of time* participants discussed the vacation with their partner. In the <u>low discussion</u> condition, participants spent five minutes thinking about the vacation individually, and five minutes discussing it with their partner. In the <u>high discussion</u> condition, participants did not spend any time thinking about the vacation on their own, but

spent ten minutes discussing it with their partner. Thus, in these two conditions we attempted to hold constant the total amount of time that an individual thought about their search behavior, while varying the time devoted to individual elaboration versus group discussion. Note that any effect of the high versus low discussion condition could be attributed to either the increased group interaction in the high discussion condition or the opportunity for individual elaboration in the low discussion condition. Therefore, we included a third condition in which participants spent five minutes thinking about the vacation individually, and ten minutes discussing the vacation with their partner. This condition, which we refer to as the <u>control</u> condition, allowed us to disassociate the effects of individual elaboration versus increased discussion.²

Crossing the two manipulated variables resulted in a 2 (*question-wording:* count versus estimate) x 3 (*discussion:* low discussion versus high discussion versus control) full factorial design. Couples in each of the three locations were randomly assigned to one of the six experimental conditions.

Procedure

There were three phases in the experiment. In the first phase, participants individually examined vacation options via an interactive computer program, which was developed using the SEARCH MONITOR software package (Brucks, 1988; Brucks & Neebe, 1988). This menudriven program allows participants to search for information about a particular topic. The software records a variety of information about this task, including the information screens that the subject examined.

Participants used this program to search for information about vacation options. Participants could look at information about vacation options in four beach locations and three cities. For each beach or city location menus were available describing three categories of

² Although some couples finished planning their vacation early, they were required to stay together for the remaining discussion time. Audiotapes of the discussions suggest that if couples had time remaining, they typically spent this

information: accommodations, restaurants, and activities. The descriptions also included the prices of the options. The behavior of interest (i.e., the behavior to be recalled later) was the total number of options examined by an individual in each of the three categories (accommodations, restaurants, and activities), a task analogous to asking consumers about the number of stores, outlets, or websites visited to look for goods and the alternatives they considered. After preliminary instructions, each *individual* was seated at a personal computer and was allowed twenty minutes to examine information that would later be used to plan their vacation. They were encouraged to examine a variety of options within each of the three information categories and to memorize as much information as possible, but they were not allowed to take notes.³

In the second phase of the experiment, *couples* planned a vacation they could win in the drawing. They were given a list of the names and prices of all available options but were given no other descriptive information. Using a structured worksheet, they selected vacation accommodations, restaurants, and activities and computed the total cost of the vacation, which could not exceed \$500. Couples were told that if the selected vacation was less than this amount, the remaining money could be used for transportation or spending money. In the low discussion and control conditions individuals filled out a worksheet with their preferred vacation separately and then completed a second worksheet together with their partner. Couples in the high discussion condition completed only the final worksheet with their partner.

Each individual then completed several questionnaires. Participants first answered a set of questions about the decision-making process, including questions on how much they agreed with their partner on specific options, perceived influence in the decision-making task, perceived importance of vacation attributes, and perceived knowledge of self and partner regarding

time discussing topics related to the vacation or the search task itself.

vacation-planning tasks. Since the primary purpose of these measures was to support the cover story that we were investigating decision-making processes, they will not be discussed further.

In the second questionnaire, participants were asked to recall the total number of accommodations, restaurants, and activities that both they and their partner examined during the information search task (see Appendix A for question wording). Self-reports for the three categories of options (accommodations, restaurants, and activities, always in that order) were elicited first, followed by reports about the partner's behavior for the same three categories. We also included a retrospective protocol question to help us understand response processes.

The final questionnaire obtained demographic information and ratings of the study, along with some supplementary measures that were intended to assess the accuracy of self-reports and are not of direct relevance to the issues discussed here.

Dependent Measures

To measure *reporting levels and accuracy*, respondents' estimates of their partners' behavior were compared to actual behavior, as recorded by the SEARCH MONITOR software. Initial analyses were conducted at the aggregate level. Then, we calculated three measures of the accuracy of reports about others at the individual level (see Burton & Blair, 1991; Menon, 1993, 1997): (1) the *raw* differences, or *raw error*, which indicates the extent of under versus over-reporting and is therefore of practical importance; (2) the absolute value of differences, or *absolute error*, which gives an indication of the size of individual-level error; and (3) the *correlation* between actual behavior and reports about others, which also provides an indication of the direction of individual-level error.

³ Only one person searched the vacation information for less than 19 minutes, with the remainder of the participants searching for the entire 20 minutes.

In calculating the difference scores (mean raw error and mean absolute error), we first examined the correlation between dyad members' absolute error scores. These correlations are significant (p's < .05), indicating that dyad members' accuracy scores are not independent. Therefore, the accuracy scores are analyzed at the dyad level, using each pair's average score as the dependent measure. For all measures, we present the results summed across the three categories of behavior (accommodations, restaurants, and activities). Repeated measures analysis of variance showed no differences in the results across the three behavior categories. For exposition, we report the summed scores.

<u>Results</u>

Manipulation Check

We conducted a check to verify the effectiveness of the discussion manipulation. Dyads' decision-making discussions were audio-recorded and transcribed.⁴ For each discussion protocol two coders, blind to the experimental conditions and hypotheses, recorded whether each accommodation, restaurant, and activity was mentioned in the transcript. Agreement between coders is 94%, so we used one coder's data. For each dyad, we calculated the total number of options mentioned at least once during the discussion.

We expected that dyads in the high discussion and control conditions would mention more options during their conversations than dyads in the low discussion condition. The results are consistent with this expectation. An analysis of variance, with discussion condition and question wording as the independent variables, shows a significant main effect of discussion ($F_{1,61} = 18.12$, p < .01). Follow-up contrasts show that the mean number of options mentioned does not differ in the high discussion and control conditions (means: high discussion = 13.79 and control = 11.10), while the mean number of options is significantly lower in the low

⁴ We were able to record 76 of the decision-making discussions. In 17 cases (all at one location), the tape recorder malfunctioned, and in 9 cases the quality of the tape was too poor to transcribe.

discussion condition (mean = 6.23, p < .05) than in the other groups. Thus, the discussion manipulation worked as intended.

<u>Aggregate Analyses</u>

In order to get an overall sense of the adequacy of self-reports versus reporting about others compared to actual behavior we first examined the mean aggregate reports about self and others and actual behavior by question wording and level of discussion. The first three columns of Table 1 present the mean actual and reported behavior for both self and others calculated across the individuals in each condition. This information was used to calculate aggregate levels of reporting error, presented in the fourth and fifth columns. Aggregate reporting error was calculated by taking the difference between the average reported and actual behavior and dividing by the average actual level of behavior. Negative reporting error indicates under-reporting at an aggregate level, while positive error indicates over-reporting at an aggregate level.

Insert Table 1 About Here

There are several interesting findings in Table 1. First, at an aggregate level, the average levels of self and other-reporting are quite similar in most conditions and are representative of actual behavior. The relative error associated with self-reports is always positive, indicating over-reporting, while the relative error associated with reports for others is more likely to be negative, indicating under-reporting. As would be expected, the mean levels of actual behavior and self-reports do not vary across conditions (p's > .25).

The level of *reports about others* does vary significantly by question wording ($F_{1,174} = 6.70$, p < .01) and is in the expected direction (but non-significant) for discussion ($F_{2,174} = 1.95$, p < .15). No other effects are significant. As predicted, the level of reports about others is lower when respondents are asked to count versus estimate. In addition, as expected, the level of reports about others is lower in the low discussion condition relative to the high discussion

condition ($F_{1,115} = 4.17$, p < .04 for this contrast). The control condition does not differ significantly from the high or low discussion conditions (p's > .18).

It is interesting to note that while at the aggregate level reports about self and others are equally accurate indicators of actual behavior, this is not the case at the individual level. We examined whether self-reports are more highly correlated with actual behavior than reports about others using a t-test for dependent correlations (see Cohen and Cohen 1983, p. 57). These correlations are shown in the last two columns of Table 1. In all but one condition, the correlations between self-reports and actual behavior are significantly higher than the correlations between reports about others and actual behavior (p's < .05). The one exception occurs when respondents were asked to count, in which case the correlations between reports and actual behavior one's partner. Possibly, when respondents are asked to count they are relying on the same (shared) memory structure that was created based on discussion. Hence, reports may be more likely to be similar under these conditions. To summarize, when relative reporting error is considered, in most cases self-reports are significantly more predictive of actual behavior than are reports from others.

Individual-Level Analyses

Table 2 shows the individual level error measures and correlations by question wording and discussion conditions. The error indices were analyzed using analysis of variance, with question wording and discussion as between subjects variables. Because the size of the difference scores is related to the level of actual behavior, the dyad's average level of search behavior was included as a covariate in the ANOVA models. All significant effects (other than the covariate) are reported below.

Insert Table 2 Here

<u>The effects of question wording on the *level* of reports about others</u>. We predicted that reporting levels would be lower when respondents are encouraged to count episodes of the

behavior versus estimate. The mean raw error measure is used to test this hypothesis. We expected mean raw error to be negative and lower when respondents are asked to count (versus estimate). The findings are consistent with our prediction, with respondents under-reporting more when asked to count versus estimate ($F_{1.83} = 5.94$, p < .02).

<u>The effects of question wording on the accuracy of reports about others</u>. We predicted that reports about others would be more accurate when respondents are asked to count episodes of the behavior than when they are asked to estimate. We examine two measures of relative accuracy—the mean absolute error measure and the correlation between actual behavior and reports about others. We expected that the mean absolute error measure would be smaller in the count than in the estimate condition. The results are consistent with this expectation. The difference for mean absolute error is marginally significant ($F_{1,83} = 2.49$, p = .06, one-tailed). Further, the correlation between actual behavior and reports about others is greater when respondents are asked to count versus estimate (t = 2.39, p < .01). As described previously, it is in this condition that the correlation between reports about others and actual behavior does not differ significantly from that of self-reports and actual behavior. Therefore, question wording does appear to affect the accuracy of reports about others.

<u>The effects of discussion on the *level* of reports about others</u>. We predicted that reporting levels would be lower when discussion is limited. The findings are consistent with this prediction, with respondents under-reporting their partner's behavior in the low discussion condition and slightly over-reporting in the high discussion condition, with the mean for the control group falling in the middle. The main effect of discussion on mean raw error is not significant (p > .27). However, since we had predicted a difference between high and low discussion groups, we conducted a planned contrast, which reveals a significant difference (F_{1,54} = 3.05, p < .05, one-tailed). The mean raw error in the control group does not differ from either the high or low discussion groups (p's > .22). These results are generally consistent with our hypotheses.

<u>The effects of discussion on the accuracy of reports about others.</u> We predicted that the accuracy of reports about others would increase with the level of discussion. Therefore, the mean absolute error should be smaller and the correlation between reports about others and actual behavior should be larger when discussion is high. Mean absolute error does not differ significantly across the three conditions (p's > .57). While the correlation between reports about others about others and actual behavior is slightly higher in both the high discussion and control conditions compared to the low discussion condition, the difference is not significant (p's > .15). Thus, at the individual level, discussion did not affect the accuracy of reports about others.

Analyses of Retrospective Protocols

In order to gain more insight into the processes used by people to answer behavioral frequency questions about others and if these processes differ from those used to construct self-reports, we included a retrospective protocol question. Immediately after answering the question about the number of *activities* examined by self/partner during the search task, participants were asked to describe how they came up with this response. The exact wording was "Describe how you answered the question about the number of ACTIVITIES that YOU/YOUR PARTNER looked at information about." To develop a coding scheme, we first examined a subset of the responses and identified the main judgment strategies reported in the protocols. These strategies are listed in Table 3. Note that three of the strategies were used only for reports about others.

Insert Table 3 Here

After establishing a set of coding categories, each protocol was divided into thought segments. Then two coders, blind to the experimental condition, assigned a code to each thought segment in a protocol. Therefore, the coding scheme captured the use of multiple strategies. Agreement between the coders averaged 88% for self-reports and 90% for reports about others. Discrepancies were resolved by a third person.

Comparison of processes used to formulate self-reports and reports about others. Table 3 shows the use of judgment strategies for self and reports about others. As would be expected, respondents are more likely to use strategies that involve episodic recall when constructing self-reports than when constructing reports about others (e.g., Bickart et al., 1990; Menon et al., 1995). Specifically, respondents are more likely to recall a specific activity when constructing a self-report (32%) versus a report about a partner (7%), McNemar χ^2 = 20.02, p < .01.

The effects of question wording on processes used to construct responses about other people. The bottom three rows of Table 3 show strategies that were used to construct reports about others. Across both question-wording conditions, almost one third of the respondents (28.9%) mention using information learned during the conversation as a judgment cue in constructing their report about their partner. We examined whether the percent of respondents using information from the conversation varied depending on how behavioral frequencies were elicited. Respondents were more likely to mention using information from the conversation to answer questions about their partners' behavior when asked to count (39%) versus estimate (21%; χ^2 = 7.0, p < .01). This finding supports the idea that when asked to count, respondents attempted to think about specific episodes of their partner's behavior that had been discussed while planning the vacation. Further, because when asked to estimate, partners should rely less on specific events or activities in constructing a judgment, we might expect that the use of general knowledge about one's partner would be greater when respondents were asked to estimate versus count. Although the overall use of general knowledge in constructing reports about others is low (6.7%), respondents are more likely to mention the use of general information in the estimate condition (10%) than in the count condition (2%; χ^2 = 4.5, p < .03).⁵

⁵ We compared the use of the various estimation strategies across the question wording condition for proxy-reports. Respondents are more likely to report that they estimated their partner's behavior using the relative importance of

To summarize, asking respondents to count increased the likelihood that reports about others were based on conversational information and decreased the likelihood that reports about others were based on general knowledge.

The effects of discussion on processes used for formulating responses about other people. We also looked at the protocols for evidence of the effects of discussion on the judgment strategies used for reporting about other people. Two significant findings emerged. First, respondents in the high discussion condition are more likely to mention using information learned during the conversation (43%) than are respondents in the low discussion (22%) and control conditions (23%; $\chi^2 = 8.4$, p < .02). This finding is consistent with the idea that increased discussion makes the episodic representation of another person's behavior more accessible. It is intriguing, however, that respondents in the control condition (where discussion was also high) did not rely as much on information learned via the conversation in constructing a report about a partner. Possibly, individual elaboration prior to the discussion in the control condition constrained the conversation to some extent.

Second, respondents in the high discussion and control conditions were more likely to mention that they estimated their partner's behavior using the relative importance of activities in selecting a vacation (high/control = 17% versus low = 8%, χ^2 = 3.6, p < .06). The discussion appears to have improved respondents' ability to estimate the relative importance of different vacation attributes to their partner, and they used this information as a heuristic in estimating responses.

Discussion

The findings of Experiment 1 suggest that at an aggregate level, reports about others are almost as accurate as self-reports as an indicator of actual behavior. At an individual level,

activities when asked to estimate (15%) than when asked to count (5%), χ^2 = 5.368, p < .02. No other differences are significant.

however, reports about others are not as accurate as self-reports. In fact, across the conditions the correlation between self-reports and actual behavior is .64, compared to a correlation between reports about others and actual behavior of .34. Therefore, for any study that focuses on comparing individual differences, the accuracy of reporting about others becomes an issue.

We also find that two factors affect the level and accuracy of reports about others' behavior. First, when reporting about others, people tend to underreport behavior when they are asked to count (versus estimate) and there is some evidence that counting results in more accurate reports. As expected, when respondents are asked to count another person's behavioral frequency they tend to rely more on episodic information (learned via conversation) and less on general information about their partner. Second, discussion also affects peoples' ability to report about other people's behavior. Increased discussion is related to higher levels of reports about others, regardless of the judgment strategy used in recalling behaviors. When discussion is limited people are more likely to underreport their partner's behavior. Finally, when discussion is limited, respondents appear to rely less on specific episodes learned during the conversation and to rely more on general information when constructing an answer.

The effects of question wording on response accuracy did not appear to be moderated by respondent knowledge. We examined the interactions between discussion and question wording for all dependent measures, and none of these interactions are significant (p's > .96). One possibility is that response accuracy is a function not only of how information was elicited from memory (response strategy), but also depends on characteristics of the behavior and respondent knowledge. Previous research has shown that the accuracy associated with a recall strategy for *self-reports* depends on the match between that strategy and characteristics of the behavior, particularly the regularity of the behavior (Menon 1997). Specifically, for regular behaviors respondents have a rate-of-occurrence available in memory, so an estimation strategy is more accurate. In contrast, no rate-of-occurrence is available for irregular behaviors, so respondents must rely on counting specific occurrences of the behavior. Thus, for irregular

behaviors a counting strategy is more accurate. We might expect similar results for judgments about others. In Experiment 2, we examine whether the effects of question wording on the accuracy of proxy reports varies with the regularity of the behavior. In addition, we examine whether such effects are moderated by the respondents' level of knowledge about their partner.

EXPERIMENT 2

Experiment 2 differed from Experiment 1 in three important ways. First, we use a more natural setting in that participants recorded the frequency of a set of common every-day behaviors. Second, we examine how the regularity of behavior affects the accuracy of reports about others. We examine different kinds of behaviors varying on regularity. Third, we use two different measures of knowledge about one's partner: *(i)* We measure the extent to which the respondent has *observed* individual episodes of his/her partner's behavior. We expect that observation will affect the accuracy of counting strategies, but not necessarily the accuracy of estimation. *(ii)* We measure the closeness of the relationship. Presumably, a higher level of closeness should be related to greater general knowledge about one's partner and thus, an enhanced level of knowledge about rates-of-occurrence. Thus, closeness of the relationship may affect the accuracy of estimation strategies.

To summarize, in this study we examine how judgment strategy (question wording) moderates the effects of regularity of the behavior on reporting accuracy. Further, we explore the effects of two types of knowledge (observation and closeness of the relationship) on response accuracy. Finally, in this study we attempt to replicate the results of Experiment 1 using a different methodology, a different measure of knowledge, and a more diverse set of behaviors.

Method

<u>Overview</u>

Pairs of student roommates participated in this study. For seven days, participants maintained individual daily diaries in which they recorded the frequency with which they engaged in ten behaviors each day. Then, both members of the dyad completed a follow-up survey in which they were asked to report on their roommate's behavior during the previous seven days, as well as their own behavior.

Participants

Forty pairs of roommates at two universities participated in the study. Participants received either \$25 in cash at one university or course credit and \$10 in cash at the other university.⁶ Participants ranged in age from 18 to 30, with a median age of 21 years, and half are male and the other half female. Participants reported knowing their partner for an average of 4.55 years (the median was 3 years), and reported that they had lived with their partner for an average of 1.87 years (the median was .75). Most of the participants (76%) reported that they were "close friends" with their partner, 11% reported that they were "casual friends", and 10% reported that they were "siblings."

Design and Procedure

Participants first met with an experimenter who explained how to complete the diaries. Each participant was given seven diary booklets, one for each day that he/she was to record behaviors. The frequency of ten behaviors was recorded. We inadvertently omitted the regularity measure for the behavior "washing hair", so this behavior has been dropped from analyses.

In addition, respondents noted whether or not their roommate was present each time they engaged in the behavior. We chose behaviors which were likely to occur at least once

⁶ There were no differences between universities.

during the diary period and that varied in regularity. Table 4 includes a list of the behaviors, as well as the average actual frequency of the behavior (from the diaries) and the regularity rating.

Insert Table 4 Here

After seven days, the participants returned to the laboratory together. At this time, they each completed a follow-up questionnaire. Each dyad was assigned one of two versions of the follow-up questionnaire. In the "count" version, the respondents were asked, "How many times in the past seven days has your roommate (BEHAVIOR)?" for each of the ten behaviors. They were then asked the same set of questions regarding their own behavior. In the "estimate" version, the respondents were asked, "On average, in a typical seven-day period, how many times does your roommate (BEHAVIOR)?" for each of the ten behaviors, followed by the same set of questions about their own behavior. Behaviors were queried in the same random order for each respondent.

Respondents then estimated how often the "typical student" would engage in each of the behaviors during a seven-day period and how often they observed their roommate engaging in each behavior during the past seven days. They then rated the regularity of the behaviors on a seven point scale (1 = irregular, 7 = regular).

The final section of the survey included demographics and a set of questions taken from Berscheid, Snyder, & Omoto's (1989) Relationship Closeness Inventory. We included questions pertaining to two of the three components of their scale. The first component asked respondents to estimate the amount of time spent alone with their roommate during the past week in the morning, afternoon, and evening. The second set of questions asked respondents to check which of 35 activities they participated in with their roommate without others present during the past seven days. After completing the follow-up survey, the exact purpose of the study was described.

To summarize, we manipulated the wording of the behavioral frequency question between subjects. In addition, we measured the regularity of each behavior, the extent to which the respondent observed the roommate's behavior, and the closeness of the relationship.

<u>Results</u>

For both self-reports and reports about others, we constructed two error measures for each of the behaviors—a raw error index [(self (other) report – actual behavior)/actual behavior] and an absolute error index [|self (other) report – actual behavior|/actual behavior].⁷ In contrast to Experiment 1, dyad members' accuracy indices were not correlated, so analyses were conducted at the respondent level (versus the dyad level).

We also constructed two measures of respondent knowledge. The first measure of knowledge was based on the percent of times that respondents observed their partners' behavior, which was recorded in the partners' diaries. The number of times that partners' reported the respondent observed the behavior was highly correlated with respondents' retrospective self reports of how many times they observed their partner engaging in each behavior (*r*'s ranging from .257 to .628, p's < .02). Second, we constructed a measure of relationship closeness using an adapted measure of the Berscheid, et al.'s (1989) Relationship Closeness Inventory. Specifically, we used two components of their scale—(a) the amount of time that the partners had spent together over the past week and (b) the number of activities that partners participated in together alone over the past seven days. The subscales were correlated (r = .517, p < .001). The sub-scales were each converted to a 10-point scale and summed to form the closeness index (range = 2 to 20; mean = 11.56; standard dev. = 4.89).

⁷ Note that in contrast to Experiment 1, for the *individual level* analyses in this study we use an index for the dependent measures. We used an index in this study in order to control for the frequency of the behaviors, which varied significantly across behaviors.

Our analyses focus on three research questions: (1) whether self-reports are more accurate than reports about others and whether this varies with question wording; (2) whether accuracy of reports about others varies as a function of regularity and question wording; and (3) how knowledge about one's partner affects reporting accuracy.

Accuracy of Self-Reports versus Reports about Others

To examine the relative accuracy of self-reports versus reports about others, we ran a repeated measures ANOVA with the absolute error index as the dependent measure, question wording (count versus estimate) as a between-subjects independent variable and type of report (self versus other) and behavior as within-subjects independent variables.

Table 5 shows the mean absolute error index for self-reports and reports about others by question wording condition. A smaller absolute error index reflects greater reporting accuracy. Notice that the mean absolute error indices for self-reports are always smaller than are those for the reports about others. This pattern indicates that self-reports are consistently more accurate than reports about others, and this difference is significant across behaviors (Wilk's Lambda = .455, $F_{9,32} = 4.26$, p < .001). For both self reports and reports about roommates there is also a tendency for the absolute error index to be smaller in the count condition relative to the estimate condition--this effect of question wording is marginally significant across behaviors (Wilk's Lambda .654, $F_{9,32} = 1.88$, p < .091). Thus, consistent with the findings of Experiment 1, self-reports of behaviors are more accurate than reports about others. In addition, both self-reports and reports about others are somewhat more accurate when respondents are asked to count (versus estimate).

Insert Table 5 Here

Effects of Question Wording and Regularity on the Accuracy of Reports about Others

To examine the relationship between reporting accuracy, regularity, and question wording we used regression analysis. Estimating the following regression equation with OLS for the raw error and absolute error indices associated with reports about others tested the effects of these relationships.

$$Y = \alpha + \beta_1 D_1 + \beta_2 X_1 + \beta_3 X_1 D_1 + \Sigma \tau_i BEHAVE_i + \varepsilon$$

where:

Y = raw error or absolute error index, $D_1 = 1$ if question wording is count or 0 if question wording is estimate $X_1 =$ reported regularity of the target person's behavior, BEHAVE = eight dummy variables representing the nine behaviors.

We first examine regression results at the aggregate level—that is, across behaviors. We then look at the regression results within each behavior. Because the findings are similar for both error indices, we will discuss the results for the absolute error index only.

As shown in the first two columns of Table 6, there are three significant effects of interest in the regression model. First, there is a significant effect of question wording (B = -.893, t = 2.91, p < .004). The negative coefficient indicates that absolute error is lower (accuracy is greater) when respondents were asked to count (versus estimate). This finding is consistent with the findings of Experiment 1. Second, there is a significant main effect of regularity (B = -.12, t = 2.42, p < .016). Here, the negative coefficient indicates that absolute error is lower (accuracy is greater) for reports about more regular behaviors. This main effect however is qualified by a significant question wording by regularity interaction (B = .125, t = 2.17, p < .03). The relationship between regularity and absolute error is stronger when respondents are asked to estimate (slope = -.115, t=-1.81, p < .072) versus when respondents are asked to count (slope = .002, t= .056, ns). Thus, the regularity of the behavior does not appear to be related to absolute error when respondents are asked to count, but is negatively related to error when respondents are asked to estimate.

Insert Table 6 Here

We also ran the same regression for each individual behavior. These results are shown for the absolute error index only in Table 7. (Results were similar for the signed error index.) For the individual behaviors, only "snacking" and "checking email" show the same pattern of effects as observed for the complete model. These two behaviors were rated as moderate on regularity and had more variance on regularity than the other behaviors (see Table 4). So the inconsistent results across behaviors may be due to the restriction in range on the regularity measure.

Insert Table 7 Here

Effects of Knowledge and Question Wording on Response Accuracy

To examine the effects of knowledge and question wording on response accuracy, we

estimated the following regression equation (across behaviors):

$$\mathsf{Y} = \alpha + \beta_1 \mathsf{D}_1 + \beta_2 \mathsf{X}_1 + \beta_3 \mathsf{X}_2 + \beta_4 \mathsf{X}_3 + \beta_5 \mathsf{X}_2 \mathsf{D}_1 + \beta_6 \mathsf{X}_3 \mathsf{D}_1 + \Sigma \tau_i \mathsf{BEHAVE}_i + \varepsilon$$

where:

Y = raw error or absolute error index, D₁ = 1 if question wording is count or 0 if question wording is estimate X₁ = reported regularity of the target person's behavior, X₂ = proportion of times partner observed behavior, X₃ = closeness of relationship BEHAVE = eight dummy variables representing the nine behaviors.

We retained regularity of the behavior in the model as a covariate. The results are shown in the last two columns of Table 6. Again, the results are similar for both error indices, so we will focus on the results for the absolute error index.

The results of this analysis suggest that closeness of the relationship does affect the accuracy of partner reports, and further, that question wording moderates this effect. As found in the previous analyses, there is a significant effect of question wording (B = -.784, t = -2.69, p < .007), suggesting that absolute error is lower (accuracy is greater) when respondents were asked to count (versus estimate). In addition, there is a marginally significant main effect of

regularity (B = -.07, t = -1.69, p < .091), suggesting that absolute error is lower (accuracy is greater) for reports about more regular behaviors. More central to the current issue, there is a significant main effect of relationship closeness (B = -.04, t = -2.60, p < .01), suggesting that reports are more accurate (error is lower) for closer relationships. As expected, this effect is moderated by question wording (b = 4.68, t = 2.05, p < .04). When respondents are asked to estimate, the accuracy of their responses increases (error gets smaller) as the closeness of the relationship increases (slope = -.037, t = -2.11, p < .036). In contrast, when respondents are asked to count, relationship closeness is not related to response accuracy (slope = .008, t = .61, ns).

We predicted that the second knowledge measure, the proportion of times the respondent observed his/her partner engaging in the behavior, would improve reporting accuracy when respondents were ask to count, but not when they were asked to estimate. This knowledge measure is not significantly related to response accuracy, nor is the interaction with question wording significant. The correlation between the proportion of times the behavior had been observed and the closeness scale is relatively small but significant (r = .143, p < .000). Possibly, this shared variance is a problem in accounting for the separate effects of these two constructs.

Finally we ran the same regression equations for each individual behavior. For all behaviors except "visits to the library", the effects of the knowledge measures on response accuracy are not significant.

Discussion

The findings of this second experiment provide corroborative evidence for the results of the previous study. First, we found that self-reports of behaviors were consistently more accurate than were reports provided by others. This finding seems to hold across infrequent behaviors (Experiment 1) as well as a more diverse set of behaviors (Experiment 2). We also found that question wording is related to the accuracy of reports about others, with a tendency

for reports to be more accurate when a counting strategy is cued. Note that we increase the generalizability of our results in Experiment 2 by expanding our investigation to a larger set of behaviors.

Additionally, in Experiment 2, we measured the regularity of the behaviors and we find that regular behaviors are reported more accurately than irregular behaviors. Further, question wording moderates the effect of regularity on the accuracy of frequency reports of behaviors such that regularity only affects reporting accuracy when respondents are asked to estimate. These results are consistent with Menon (1993, 1997), but while her work only examined selfreports, our investigation extends these results to the realm of reports about others.

Finally, we show that the closeness of a relationship is related to accuracy of reports about others. Respondents in a close relationship with their partner provided more accurate reports of their partners' behavior, but only when an estimation strategy was cued. This finding is consistent with the idea that over a period of time, people acquire general knowledge about their others' behavior, which helps them construct an accurate estimate. Closeness of the relationship had no effect on reporting accuracy when a counting strategy was elicited. Thus, the closeness of the relationship does not help respondents retrieve specific occurrences of another person's behavior. We did not find any significant effects of observation on reporting accuracy.

General Discussion

Consumers make judgments about other peoples' behavioral frequencies, both explicitly during surveys and implicitly (e.g., when making household purchase decisions). In two experiments, we show that the process used to construct judgments about others' behavioral frequencies, the characteristics of the target behavior, and knowledge can affect the accuracy of such judgments. First, the findings suggest that like self-reports, the ability to accurately *estimate* another person's behavioral frequency is positively related to the regularity of the

behavior. In contrast, the regularity of the behavior does not affect reporting accuracy when people are encouraged to recall specific episodes of the behavior. Possibly, a rate-ofoccurrence is more likely to be available for regular behaviors. Alternatively, it may be easier to infer or construct an accurate rate when the behavior is regular. For example, for regular behaviors one's own rate of engaging in the behavior may be a good surrogate (or anchor) for inferring another person's behavioral frequency.

Second, we show how different forms of knowledge affect the ability of people to accurately predict others' behavioral frequency. Experiment 1 shows that discussion appears to improve access to specific episodes of behavior and increases reporting levels, but does not necessarily result in more accurate reports. Experiment 2 shows that people in closer relationships are better able to estimate (but not count) the frequency of their partner's behavior. Presumably, over time people form semantic representations of their partners' behavior that allow them to accurately construct a rate-of-occurrence. Observation of behavior, however, did not affect predictive accuracy. These findings support the importance of distinguishing between different ways of acquiring information about others and the implications of acquisition for how information about others is stored and retrieved.

Many consumer surveys, conducted both by the private sector and by government, rely on reports about others to obtain information about household members' behavior. Our findings hold several important implications for practitioners interested in increasing the accuracy of these proxy reports. First, we found that the wording of behavioral frequency questions affects the level of reporting for others as well as the accuracy of such reports. Asking people to provide a count of their partner's behavior (e.g. "How many times?") resulted in under-reporting relative to questions asking respondents to estimate their partner's behavior (e.g. "How often?"). When asked to estimate, people appear to use a variety of strategies, resulting in less accurate (and more variable) behavioral frequency reports. The results of Experiment 2 suggest that estimates of another person's behavior are more accurate when the behavior occurs regularly.

Thus, as with self reports (Menon 1997), when asking questions about others' behavior, encouraging the use of estimation is most appropriate when the behavior is regular.

Second, our findings suggest that screening potential respondents on their level of discussion on a topic with their partner or family member or the closeness of the relationship may help to reduce reporting error. For example, respondents are less likely to under-report their partner's behavior when discussion has been extensive. This finding parallels the results of Menon et al. (1995) with regard to the effects of discussion on measures of the agreement between self and reports about others. Further, respondents in a close relationship provide more accurate estimates of their partners' behavior. A closer relationship appears to affect the accessibility of semantic representations about other people, thus enhancing reporting accuracy.

Our research suggests several interesting directions for future research. First, while our results provide insight into how discussion and question wording facilitate episodic recall, less is learned about how these factors might affect inference making. It is clear that subjects were using a number of different strategies for estimating their own and their partner's behavior (see Conrad et al., 1998). As evidence of this, for both self and reports about others, the variance of absolute error measures is greater when respondents were asked to estimate than when they were asked to count (see Table 1). Thus, our experiment was not well equipped to see how conversation might affect inference making about another person's behavior. Second, future research could examine the role of different ways of acquiring information about others' behavior and how acquisition affects judgment processes and accuracy. The short-term nature of our experiments may not have facilitated the formation of a rate-of-occurrence for some behaviors. Using a more realistic, long-run context might help us to understand how discussion and observation ultimately affect the development of semantic memory about others' behaviors. Finally, the results of Experiment 1 suggest that in some situations, increasing discussion may actually distort memory for the original behavior (for both self and others). The creation of a

shared memory representation and how such a representation affects recall accuracy is an interesting direction for future research.

Experiment 1: Aggregate Actual Behavior versus Self and Reports about others by

Question Wording and Discussion

	Mean Actual Behavior Versus Reports				Correlations		
Condition:	Actual	Self- report	Reports about Partner	Self Relative Error ^a	Partner Relative Error ^b	Self with Actual	Partner with Actual
Question Wording:							
Count (n = 86)	26.47 (13.94)	28.05 (17.71)	22.42 (14.99)	.06	15	.64	.49
Estimate (n = 94)	27.73 (11.63)	28.98 (17.10)	28.73 (17.68)	.05	.04	.64	.17
Discussion:							
Low (n = 58)	25.81 (13.45)	28.91 (18.58)	22.48 (11.75)	.12	13	.69	.22
High (n = 61)	27.87 (11.16)	29.07 (17.81)	28.33 (18.48)	.04	.02	.65	.38
Control (n = 61)	27.64 (13.68)	27.64 (15.87)	26.18 (18.53)	0	05	.60	.35

Note: Standard deviations in parentheses.

^a Self Relative Error = (Self-report - Actual Behavior)/Actual Behavior

^b Other Relative Error = (Other-report - Actual Behavior)/Actual Behavior

	Accuracy Measures: Reports About Others			
Condition	Mean Raw Error (Report – Actual)	Mean Absolute Error Report – Actual	Correlation between Report about Partner and Actual Behavior	
Question Wording: Count	-1.33	4.99	.49 ^a	
(n=43) Estimate	(3.64) .89	(3.03) 6.23 (4.44)	(n=86) .17 ^b	
(n=47)	(5.15)	(4.11)	(n=94)	
Low	-1.11	5.12	.22 ^b	
(n=29) High	(3.27) .57	(3.35) 5.93	(n=58) .38ª	
(n=30) Control	(4.73) 0.00	(3.51) 5.84	(n=61) .35 ^a	
(n=31)	(5.47)	(4.14)	(n=61)	

Experiment 1: Accuracy Measures by Question Wording and Discussion

Note: Standard deviations in parentheses.

^{*}The cell sizes for the error indices are based on the number of dyads, while the cell sizes for

the correlations are based on the number of individuals.

^a p < .01; ^b p < .10

Retrospective Protocols: Percentage of Respondents Using Judgment Strategies for

Judgment Strategy:	% Self-reports (n=180)	% Reports about Partners (n = 180)
Recall specific activity: Specific activity mentioned in protocol.	32.2	7.2
<i>Estimate by location</i> : Recalls the number of locations examined and then estimates the number of activities examined per location.	38.3	20.6
<i>Estimate by relative importance of</i> <i>activities</i> : Uses the relative importance of activities versus accommodations and restaurants as a judgment cue.	15.6	10.6
<i>Estimate by characteristics of activities</i> : Uses characteristics of activities as a judgment cue (e.g. "I only looked at activities that were inexpensive.").	25.0	7.8
<i>Estimate by preference for activities</i> : Uses own or partner's likely preference for activities as a judgment cue (e.g. "Only a few activities sounded like fun.").	30.6	21.7
Guess: Mentions that they guessed.	2.2	16.7
Other	8.5	5.1
Strategies for reports about others only:		
Anchor on self and adjust . Uses own report as a basis for estimating partner's behavior (e.g. "I know she looked at more activities than I did.")		16.1
Based on conversation : Uses information learned during the conversation as a judgment cue.		28.9
Based on general knowledge about partner : Uses general knowledge about partner as a judgment cue (e.g. "She is detail-oriented.").		6.7

Self and Reports about Others for Experiment 1

Note: Because multiple strategies could be used, percentages do not sum to 100%.

Experiment 2: Average Actual Frequency and Regularity of Behaviors

		Actual Frequency		Reg	ularity
Behavior	n	Mean	Std. Dev.	Mean	Std. Dev.
Brushing Teeth	80	12.11	3.94	6.86	.46
Watching Television (hours)	78	10.30	6.89	5.83	1.18
Attending Class	78	7.56	3.35	6.04	1.20
Purchasing Food	77	6.55	3.77	4.97	1.42
Talk to Friends on Campus	73	6.25	4.80	4.80	1.71
Snacking	71	5.77	3.77	5.01	1.25
Checking Email	75	5.67	3.86	4.99	1.34
Drink from a Water Fountain	52	3.65	3.70	3.16	1.69
Visit the Library	57	2.68	1.93	3.90	1.52

Note: n = number of participants that engaged in the behavior during the diary period.

Experiment 2: Absolute Error Indices (|Report-Actual| / Actual) for Self Reports and

Reports about Others by Question Wording

	Cou	nt	Esti	Estimate		
-		Reports about		Reports about		
Behavior:	Self Reports	Others	Self Reports	Others		
Brushing	.077	.313	.097	.202		
Teeth	(0.10)	(0.31)	(0.11)	(0.29)		
Watching	.434	.627	.555	.995		
Television	(0.60)	(0.65)	(0.64)	(1.37)		
Attending	.171	.554	.318	.507		
Class	(.0.25)	(0.56)	(0.32)	(0.43)		
Purchasing	.257	.445	.364	.530		
Food	(0.20)	(0.29)	(0.31)	(0.62)		
Talk to Friends on Campus Checking Email	.425	.692	.905	1.99		
	(0.42)	(0.88)	(0.77)	(4.05)		
	.364	.602	.861	.999		
	(0.41)	(0.66)	(0.81)	(1.13)		
Snacking	.421	.700	.689	.989		
	(.361)	(0.91)	(1.08)	(0.99)		
Drink from a Water	.150	.908	.691	1.17		
	(0.32)	(1.39)	(1.11)	(1.88)		
Visit the	.164	.382	.378	.741		
Library	(0.32)	(0.49)	(0.43)	(1.01)		

Note: Standard deviations are in parentheses.

Experiment 2: Regression Results Across Behaviors

	Dependent Measure				
	Effects of Regularity and		Effects of Knowledge:		
	Question	r wording.	LIEGIS OF I	diowiedge.	
	Reports Reports about about Others: Others:		Reports about Others:	Reports about Others:	
	Raw Error	Absolute Error	Raw Error	Absolute Error	
Effect:	Index:	Index	Index:	Index	
Question Wording (QW)	-1.16 ^a	-0.89 ^a	-1.24 ^a	-0.78 ^a	
Regularity (REG)	-0.14 ^a	-0.12 ^a	-0.08 ^c	07 ^c	
Proportion Observed (OBS)			-0.07	09	
Closeness Scale (CLOSE)			-0.04 ^b	04 ^a	
QW x REG	.13 ^b	.13 ^b			
QW x OBS			.08	.05	
QW X CLOSE			.06 ^b	.05 ^b	
F	5.84 ^a	4.64 ^a	4.56 ^a	4.12 ^a	
R ²	.09	.07	.09	.09	

Note: Coefficients for dummy variables representing behaviors are not included.

^a p < .01; ^b p < .05; ^c p < .10

Experiment 2: Regression Results for Each Behavior, Dependent Variable is Absolute Error Index

	Independent Variables:			
Behavior:	Question Wording (QW)	Regularity (REG)	QW x REG	
Brushing Teeth	.73	.08	-0.08	
F = 1.95 Watching TV	3.03	.24	-0.49	
Attending Class	.63	.03	-0.14	
Purchasing Food	-0.40	-0.02	.07	
Talk to Friends on Campus $F = 1.07$	-1.95	-0.14	.31	
Snacking $F = 3.47^{b}$	-5.54 ^a	-0.82 ^a	.90 ^b	
Checking Email F = 3.01 ^b	-1.54 ^b	-0.25 ^ª	.25 ^c	
Drink from a Water Fountain $E = 54$.25	-0.04	17	
Visit the Library F = 1.90	-1.09 ^b	-0.11	.15	
Note: ${}^{a} p < .01; {}^{b} p < .05; {}^{c} p < .05;$	< .10			

References

Berschfield, E., Snyder, M. & Omoto, A. M. (1989). The relationship closeness inventory: Assessing the closeness of interpersonal relationships. <u>Journal of</u> Personality and Social Psychology, 57. 792-807.

Bickart, B. A., Blair, J., Menon, G. & Sudman, S. (1990). Cognitive aspects of proxyreporting of behavior. In M. Goldberg, G. Gorn, & R. W. Pollay (Eds.), <u>Advances in Consumer</u> <u>Research, 17</u> (pp. 198-206). Provo, UT: Association for Consumer Research.

Bickart, B. & Felcher, E. M. (1996). Expanding and enhancing the use of verbal protocols in survey research. In N. Schwarz & S. Sudman (Eds.), <u>Answering Questions: Methodology for</u> <u>Determining Cognitive and Communicative Processes in Survey Research</u> (pp. 115-142). San Francisco: Jossey-Bass.

Bickart, B, Menon, G., Schwarz, N. & Blair, J. (1994). The use of anchoring strategies in constructing proxy reports of attitudes. <u>International Journal of Public Opinion Research</u>, 6. 375-379.

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

Brewer, W. F. (1986). What is autobiographical memory? In Rubin, D. C. (Ed.), <u>Autobiographical memory</u> (pp. 25-49).

Brucks, M. (1988). Search Monitor: An approach for computer-controlled experiments involving consumer information search. <u>Journal of Consumer Research, 15</u>. 117-121.

Brucks, M. & Neebe, E. (1988). Search Monitor user's guide (Version 4.0).

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

Cohen, J. & Cohen, P. (1983). <u>Applied multiple regression/correlation analysis for the</u> behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. Conrad, F. G., Brown, N.R. & Cashman, E. R. (1998). Strategies for estimating behavioral frequency in survey interviews. <u>Memory: Special Issue on Survey Research, 6</u>. 339-366.

Dashen, M. (2000). The effects of retention intervals on self- and proxy reports of purchases. <u>Memory, 8</u>. 129-143.

Davis, H. L., Hoch, S.J. & Ragsdale E.K. E. (1986). An anchoring and adjustment model of spousal predictions. Journal of Consumer Research, 13. 25-37.

Fussell, S. R. & Krauss, R. M. (1992). Coordination of knowledge and communication: effects of speakers assumptions about what others know. <u>Journal of Personality and Social</u> <u>Psychology, 62</u>. 378-391.

Ickes, W., Stinson, L., Bissonnette, V. & Garcia, S. (1990). Naturalistic social cognition: empathic accuracy in mixed-sex dyads. <u>Journal of Personality and Social Psychology</u>, 59. 730-742.

Kasprzyk, D. & McMillen, D. (1987). SIPP: Characteristics of the 1984 panel. <u>Proceedings</u> of the Social Statistics Section, American Statistical Association. 181-186.

Kleppe, I. (1990). <u>Teoretisk versus subjektiv beslutningsinnfl ytelse: En eksperimentell</u> <u>studie</u>. Unpublished doctoral dissertation, Department of Marketing Economics, University of Bergin.

Kojetin, B. A. & Miller, L. A. (1993). <u>The intrahousehold communications study: estimating</u> <u>the accuracy of proxy responses at the dyadic level</u>. Paper presented at the annual conference of the American Association for Public Opinion Research, St. Charles, Illinois.

Kojetin, B. A. & Mullin, P. (1995). <u>The quality of proxy-reports on the Current Population</u> <u>Survey (CPS)</u>. Paper presented at the annual meeting of the American Association for Public Opinion Research, Fort Lauderdale, Florida.

Kojetin, B. A. & Jerstad, S. (1997). <u>The quality of proxy-reports on the consumer</u> <u>expenditure survey</u>. Paper presented at the Society for Consumer Psychology Conference, St. Petersburg, Florida. Means, B. & Loftus, E. (1991). When personal history repeats itself: Decomposing memories for recurring events. <u>Applied Cognitive Psychology, 5</u>. 297-318.

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

Menon, G. (1997). Are the parts better than the whole? The effects of decompositional questions on judgments of frequent behaviors. <u>Journal of Marketing Research, 32</u>. 335-346.

Menon, G., Bickart, B., Sudman, S. & Blair, J. (1995). How well do you know your partner? Strategies for formulating proxy-reports and their effects on the convergence to self-reports.

Journal of Marketing Research, 32. 75-84.

Moore, J. C. (1988). Self/proxy response status and survey response quality: A review of the literature. Journal of Official Statistics, 4. 155-172.

Schober, M. F. & Clark, H. H. (1989). Understanding by addressees and overhearers. <u>Cognitive Psychology, 21</u>. 211-232.

Schwarz, N. & Wellens, T. (1997). Cognitive dynamics of proxy responding: The diverging perspectives of actors and observers. <u>Journal of Official Statistics, 13</u>. 159-179.

Strube, G. (1987). Answering survey questions: The role of memory. In H. J.

Hippler, N. Schwarz, & S. Sudman (Eds.), Social Information Processing and Survey

<u>Methodology</u> (pp. 86-101). New York: Springer-Verlag.

Sudman, S., Bickart, B., Blair, B. & Menon, G. (1994). A comparison of self and proxyreporting. In N. Schwarz & S. Sudman (Eds.), <u>Autobiographical Memory and the Validity of</u> Retrospective Reports, (pp. 251-266). New York: Springer-Verlag.

Tomkins, S. (1978). Script theory: Differential magnification of affects. H. E.

Howe, Jr. & M. M. Page (Eds.), Nebraska Symposium on Motivation, 26, (pp. 201-236).

Lincoln, NE: University of Nebraska Press.

APPENDIX

WORDING OF THE BEHAVIORAL FREQUENCY QUESTIONS

Instructions:

We are interested in finding out how extensively people searched for various types of information. Therefore, in the following questions you will be asked to <u>count (estimate)</u> the number of specific kinds of options you or your partner examined. For example, you might be asked to <u>count (estimate)</u> the number of activities you obtained information about. Your answer should include any relevant options examined, even if you or your partner did not seriously consider the option when choosing a vacation.

Please try your best to answer these questions. For each question, just try to **<u>count (estimate)</u>** how many options you or your partner examined.

Finally, it is important that you answer each question in order. Do not skip ahead in the questionnaire, and do not return to earlier answers.

Question:

<u>Count (Estimate)</u> the **TOTAL** number of **ACCOMMODATIONS (RESTAURANTS, ACTIVITIES)** that **YOU (YOUR PARTNER)** looked at information about during the vacation search task.

_____ Accommodations (Restaurants, Activities)