Chapter 1

Introduction

1.1 Recent Developments in Telephone Survey Research

In 2008, the number of cell-phone-only households in the U.S. surpassed the number of landline-only households (Blumberg and Luke 2009). That moment was but one milepost in the ongoing transition from fixed, residential wirelines to mobile, personal communication devices. The implications of this trend for survey researchers are alarming. Landline random digit dialing (RDD), a popular sample design since the 1970's, now covers less than 80 percent of the general population (ibid.).[1] Evidence that this relatively low coverage rate leads to biased survey estimates is somewhat mixed. Many health and lifestyle estimates show clear coverage bias (Blumberg and Luke 2007), while some political attitudes and behavior estimates do not (Brick et al. 2006; Keeter et al. 2007; Link et al. 2007; Mokrzycki, Keeter, and Kennedy 2010). Estimates for certain subgroups, particularly low-income and young adults, are at very high risk of coverage error (Blumberg and Luke 2007; Keeter et al. 2007). Post-stratification techniques have been shown to reduce coverage error in landline RDD surveys but not eliminate it

[1] Some versions of list-assisted landline RDD (Casady and Lepkowski 1993) yield coverage rates well below 80 percent of the U.S. adult population. The coverage degradation is due to intentional exclusion of numbers in 100-banks with few or no directory-listed numbers, a decrease in residential landline assignment density, and an increase in alternative dial-tone providers, such as cable, that have lower listed rates (Fahimi, Kulp, and Brick 2008).

completely (Brick et al. 2006; Keeter 2006; Kennedy 2007b; Mokrzycki, Keeter, and Kennedy 2010).

To address the growing coverage gap, some survey organizations have begun supplementing landline RDD samples with samples of randomly-selected cell phone numbers. This approach has proven feasible in that the public's willingness to be interviewed on cell phones is comparable to their willingness to be interviewed on landlines (Brick et al. 2006; Keeter et al. 2007; Link et al. 2007). Nonresponse is still pervasive, but participation has been sufficient to support cell phone sampling as a promising methodology. Critically, cell phone samples have proven effective for capturing demographic groups (e.g., young adults) that had dwindled in landline samples since the early 1990's due to nonresponse and cell phone substitution (Keeter et al. 2007).

While these findings are encouraging, research on cell phone surveys is still in its infancy and much remains unknown. In particular, only a small number of studies have provided insight into the levels of nonresponse bias or measurement bias that researchers should expect to encounter. Fewer studies still have taken an experimental approach to investigating these error sources. This chapter reviews research that has been conducted on cell phone surveys and discusses several issues worthy of additional attention.

1.2 Overview of Cell Phone Survey Research

A major thrust of cell phone research has been describing respondent characteristics, particularly those who are cell-phone-only. A review of relevant professional conference presentations illustrates the point. Table 1.1 shows the number of cell phone survey papers presented from 2000 through 2009 at the Joint Statistical

Meetings or the Annual Meeting of the American Association for Public Opinion Research. These 158 papers were identified in conference programs using word searches for “cell”, “mobile”, and “wireless”. The most popular topic by far is description of cell phone respondents. One likely explanation for the ubiquity of descriptive studies is their simplicity. After the data are collected, it is fairly straightforward to compute the survey frequencies separately for each sampling frame (landline RDD and cell RDD).

Table 1.1. Number of Conference Presentations Addressing Cell Phone Surveys, by Year and Topic

	Operations	Descriptive/ Coverage	Sample Design	Nonresponse	Measurement	Weighting/ Estimation	Total
2000		2					2
2001							0
2002		2					2
2003	1	1	1				3
2004	2	13					15
2005	3	7	2	1	2		15
2006	2	7		1			10
2007	8	10	1	1		3	23
2008	8	17	11	3	2	5	46
2009	4	20	6	6	3	3	42
Total	28	79	21	12	7	11	158

Figures based on papers presented at the Annual Meeting of the American Association for Public Opinion Research and the Joint Statistical Meetings between 2000 and 2009. Posters and roundtable discussions are not included. Topic was coded based on the title and, in some cases, the abstract.

Descriptive studies have documented the demographics (Kim and Lepkowski 2002; Tucker, Brick, and Meekins 2007), health (Blumberg and Luke 2007), political attitudes (Traugott and Joo 2003; Keeter et al. 2007), voting behavior (Albaghal 2005; Keeter 2006; Mokrzycki, Keeter, and Kennedy 2010), media consumption (Fleeman 2006; Keeter et al. 2007), and a number of other attributes of the cell-only population. Research conducted by Stephen Blumberg and Julian Luke (2007) is especially important because their data come from the National Health Interview Survey, which uses area-probability sampling and in-person interviewing. Unlike cell phone RDD data, the NHIS data are not threatened by high levels of nonresponse, particularly nonresponse related to telephone usage. Information from descriptive studies has been critical to understanding the risk from coverage error to certain estimates under a landline RDD design. While some future descriptive studies will be necessary as Americans' telephone behaviors continue to evolve, the marginal benefit to the field from each additional descriptive report is arguably converging to zero. The survey research community has a fairly detailed understanding of cell phone user characteristics. The critical questions now center on how best to contact them, interview them, and combine the results with data from other samples (and potentially other modes) to produce reliable estimates.

Several of the first operational papers on cell phones focused on basic steps for fielding a cell phone sample (Fleeman 2007; Howes 2008; Brick et al. 2007). While most data collection procedures are similar to landline interviewing, there are several important differences. An important initial topic was adhering to federal laws governing the dialing of cell phone numbers (Kulp 2004). Other operations topics include calling times (Best and Hugick 2009; Fleeman 2007; Schroeder and Meekins 2009), caller ID (Dayton et al. 2009), and the precision of geographic codes (Christian and Dimock 2009; Fleeman 2007). Research on calling protocols is important because of the implications for nonresponse error, sample sizes, and survey costs. Fortunately, experimental designs for improving protocols in cell phone surveys have already been established. The large set of landline RDD studies that used randomized experiments to evaluate alternative design options (e.g., incentive amounts) provides a useful model for designing cell phone studies.

Another burgeoning topic in cell phone research is sample design. A fundamental question is whether cell phone users who have a residential landline (dual users) should be screened out of cell RDD samples. Screening out dual users from the cell phone sample avoids an overlap in the coverage of landline and cell phone samples, simplifying estimation. One downside of screening, however, is the cost associated with dropping willing respondents. The consequences for nonresponse bias from screening are still coming into view. Depending on the relationship between response propensity and the survey variable of interest, screening for cell-phone-only adults theoretically could increase or decrease bias. Several researchers have addressed this topic (Brick et al. 2006; Dimock, Christian, and Keeter 2009; Kennedy 2007b; Lohr, Grandjean, and Taylor 2009), but there is currently no consensus on screening for cell-only adults.

Other sample design studies have addressed efficiency in the RDD frame, as well as the use of alternative sampling frames. Salvanto (2008) and Brown and ZuWallack (2009) evaluated Mitofsky-Waksberg approaches to sampling cell phone numbers. Unfortunately, structural properties of the Telcordia database make this technique less efficient for cell phone numbers than for landlines. With respect to alternative sampling frames, researchers have considered registered voter files (DiCamillo 2008), lists of residential addresses (Fleeman and Wasikowski 2008; Han and Cantor 2008; Link et al. 2008), and Web panels (Turakhia, Schulman, and Bohinsky 2008). Non-telephone data collection is outside the scope of this dissertation, but studies indicate that address-based sampling in particular is a promising approach for surveys with relatively long field periods (Fleeman 2009; Link et al. 2008, 2009; Norman and Sigman 2009; Skalland, Barron, and Wooten 2009).

Other methodological topics are critical to understanding the error properties of cell phone data but have received relatively little attention. As shown in Table 1.1, nonresponse, measurement, and weighting/estimation are woefully understudied in the cell phone setting. Over the past decade, these topics have been the focus of twelve, seven, and eleven conference papers, respectively. The lack of attention to these nonsampling topics is troubling given the potential magnitude of the associated errors. The next two sections discuss nonresponse and measurement error in the context of cell phone surveys. Weighting and estimation is also a critical area for future research but will not be discussed in depth here. Brick, Cervantes, Lee, and Norman (in press) is recommended for a discussion of current practices regarding weighting and estimation for cell and landline RDD surveys.

1.3 Nonresponse in Cell Phone Surveys

Nonresponse is a serious threat to data quality in cell phone samples, just as it is in landline samples. Unfortunately, the research literature on this topic is quite thin, and studies to date have been observational rather than experimental in nature. A major obstacle to nonresponse research is the inherent difficulty in studying a behavior that yields little if any information. That said, experimental designs for studying nonresponse exist (for a review, see Groves 2006). This section reviews what is known about nonresponse in cell phone surveys and suggests areas for future investigation.

Nonresponse rates in cell phone samples are typically equivalent to or slightly lower than rates for parallel landline samples (Brick et al. 2006; Link et al. 2007; Keeter et al. 2008), which generally range between 60% and 90% in the United States (Holbrook, Krosnick, and Pfent 2008). A high level of nonresponse signals an elevated risk of bias but does not guarantee it. For estimates to be biased, there must be a relationship between the likelihood (or propensity) of responding and the variable of interest, and the relationship would need to persist after weighting.

Our understanding of when such relationships are likely to exist is incomplete. Several factors influencing response decisions have been identified, but they explain only a portion of the overall variance in response behavior. Perhaps the most robust finding is that people who use their cell phone frequently are more likely to participate than infrequent users (Brick et al. 2006). Using data from a dual frame RDD survey, Brick and his colleagues also found that people interested in technology are more likely to participate in cell phone surveys than those who are not (Brick et al. 2006). Vehovar and Callegaro (2007) suggest several more causes of nonresponse in cell phone surveys based on data from an in-person study in Slovenia. Respondents in a face-to-face survey were probed about their willingness to participate in a hypothetical cell phone survey. Expressed willingness to participate was negatively associated with age and positively associated with computer literacy, participation in previous surveys, size of social network, and extroverted personality type.

These factors, which can be thought of as nonresponse causes or mechanisms, explain a relatively small proportion of the overall variance in response behavior. This suggests that other mechanisms have yet to be identified. In other words, the model for response decisions in cell phone surveys is incomplete. The mechanisms that have already been identified and additional ones posited in this dissertation (see Chapter 2) can be classified into two main classes: mechanisms that transcend device and mechanisms that are specific to the cell phone and how it is used. The former class represents mechanisms that influence response decisions in both landline and cell phone surveys. For example, people who strongly distrust the government would likely refuse a federally-sponsored survey regardless of whether they were contacted on a landline or on a cell phone. Other nonresponse mechanisms that are likely to transcend device include social integration (Abraham, Maitland, and Bianchi 2006; Groves and Couper 1998), civic engagement (Brehm 1993; Cialdini, Braver, and Wolf 1992; Couper, Singer, and Kulka 1998; Groves, Presser, and Dipko 2004; Putnam 1995; Voogt and van Kempen 2002), economic conditions (Harris-Kojetin and Tucker 1999), and respondent valuation of salient survey features (Groves, Singer, and Corning 2000) among others. These mechanisms are well-documented in the literature. It remains to be seen whether or not they are as central to cell phone survey participation decisions as they are to participation decisions in other modes.

Another potential class of nonresponse mechanisms has to do with properties of cell phones and how users interact with them. This class may be described as device-related mechanisms. The finding of higher response propensities among heavy cell phone users (Brick et al. 2006) may reflect greater interest in and adeptness at using mobile technology, which may also explain the Vehovar and Callegaro (2007) finding about computer literacy. If so, then we may also expect to see a relationship between usage of non-telephony functions, such as text messaging, and response propensity.

Response propensity may also be affected by attitudes about cell phones. Anecdotal evidence indicates that a fair number of refusals occur because the person answering was surprised and upset by the fact that they were being called on their cell phone to participate in a survey. Other people, particularly those without a landline, may view their cell phone as a broad access point for anyone needing to reach them – acquaintances and strangers alike. Variability on this dimension of viewing the phone as a broad access point for the public or a narrow access point reserved for close acquaintances, seems promising as a predictor of response propensity.

It is important to note that any empirical search for the dominant mechanisms influencing participation decisions has limits. The propensity to respond varies across people, across surveys, and within surveys. For example, an individual may have a high propensity to respond to health surveys and a low propensity to respond to financial surveys. Response propensity can also vary *within* a survey, across different recruitment protocols (Groves, Singer, and Corning 2000; Olson 2007). For example, switching from no incentive to a $20 incentive would affect some people's response propensity even though nothing substantive about the survey would have changed. The main point is that the reasons why people decline to respond will differ from survey to survey. Consequently, the most that empirical studies can realistically achieve is to identify mechanisms of nonresponse that are robust across a broad range of surveys.

1.4 Measurement Error in Cell Phone Surveys

Another aspect of cell phone surveys in need of more empirical attention is measurement. Even if people cooperate with the survey request, there is still an opportunity for error if the responses that they give are inaccurate or incomplete.

Researchers have speculated that people are more likely to misreport when interviewed on a cell phone as compared to a landline (Steeh 2004; Lavrakas et al. 2007). This hypothesis is based on several observations: (*i*) People often multitask while talking on their cell phone (*ii*) The sound quality of cell phone calls may be inferior to that of landline calls (*iii*) People on their cell phones may be exposed to more distracting environmental stimuli (*iv*) People reached on a cell phone may be out in public and feel as though they have less privacy than people interviewed at home on a landline, which may lead them to censor responses to sensitive questions.

Fortunately for survey researchers, studies testing for an effect from device (landline/cell) on measurement error have not detected any meaningful differences. Several teams of researchers have documented equivalent rates of item nonresponse in cell phone and landline RDD interviews, after controlling for demographics (Brick et al. 2007; Steeh 2004; Witt, ZuWallack, and Conrey 2009). Witt and her colleagues (2009) report that item nonresponse was slightly more prevalent in cell phone interviews than landline interviews after controlling for age, gender, and race, but the effect was not statistically significant. Among dual users, Witt and her colleagues found no difference in item nonresponse propensity between people who predominantly use their landline and those who predominantly use their cell phone.

Researchers have also documented parity between landline and cell phone interviews with respect to the length of responses to open-ended questions (Brick et al. 2007; Witt, ZuWallack, and Conrey 2009). Brick and his colleagues analyzed responses to four sensitive questions in multilevel models controlling for age, sex, and home ownership, and they found no significant effect from device. Kennedy (2007a) investigated straight-lining and response order effects and found no significant differences between landline and cell phone data.

Each of these studies, however, was observational in nature. Consequently, any effect from the device is likely to have been confounded with differences in the composition of the samples. This is a potentially important limitation because some of the variables known to be different in landline and cell phone samples (e.g., age) are also known correlates of cognitive shortcuts such as item nonresponse and non-differentiation of battery items (Bell 1984; Slymen et al. 1984; Frisina and Thomas 2008). Some researchers attempt to adjust for this by introducing demographic controls, but this approach rests upon the assumption that all of the factors confounding the sample comparison are included in the set of controls.

A second limitation of previous work is the nature of the dependent variables. Item nonresponse rates and the lengths of open-ended questions are important, but they are only indictors of measurement error, not the thing itself. Ideally, we would compare the deviation between the survey responses and the corresponding "true values" for landline versus cell phone interviews. In RDD surveys, however, records containing true values are generally not available. For attitudinal questions, the existence of records is particularly unlikely, and the mere notion of a "true value" is ambiguous. Absent direct computation of measurement error, examining a wider range of measurement error indicators would be useful.

1.5 Conclusions

A tremendous amount of research has been conducted on cell phone surveys in a relatively short period of time. Some 158 papers on the topic have been presented at national professional conferences in the past decade. This proliferation is encouraging because it demonstrates researchers' desire to advance the method and share lessons learned with each other. Perhaps the best understood aspect of cell phone surveys is the demographic and lifestyle characteristics of respondents that they bring into dual frame RDD surveys. Descriptive papers constitute half of the research that has been presented.

While this research has been illuminating, there is a pressing need to better understand other aspects of cell phone surveys. At or near the top of the list is learning more about the risk to estimates from nonresponse. In this chapter, it was argued that assessing the risk of nonresponse bias begins with identifying the main reasons why people do not respond. Brick and his colleagues (2006) showed that people who use their cell phone infrequently and those less interested in technology are less likely to respond to cell phone surveys than others. These factors explain some of the variance in response decision, but other important factors almost certainly exist. Two classes of potential nonresponse mechanisms seem likely. One class consists of mechanisms familiar from landline RDD surveys and other modes. These include social integration, civic engagement, interest in the survey topic, affinity for the sponsor, and the like. Another class consists of mechanisms related to the cell phone itself. These mechanisms may include concern about the cost of the call, using the device for non-telephony purposes (e.g., text messaging), and attitudes about whether the cell phone is a broad access point for the public or a narrow access point intended only for close acquaintances.

Another topic that has received little empirical attention is measurement. Researchers have speculated that data quality from cell phone interviews will be inferior to that from landline interviews. The few studies to address this issue did not detect any meaningful differences. That said, those studies have been based on comparisons that may be confounded by differences in sample composition. Furthermore, the measurement error outcomes that have been considered are fairly limited. Evaluation of more error indicators or better yet – direct error computations using record-checks – would provide additional insight on possible differences in measurement between landline and cell phone interviews.

These are not the only topics that merit additional research. Techniques for weighting cell phone data are in their infancy. Many statistical challenges remain as survey statisticians work on accounting for multiple frames, differential probabilities of selection within frames, and differential response propensities. Hopefully, future insights about nonresponse will inform the development of nonresponse weighting procedures.

Continued research on cell phone surveys is essential as society moves toward mobile, personal devices that are integral to how people communicate with the outside world. Mobile technologies are evolving at a prodigious pace, with numerous potential consequences for survey data quality. Identifying causal mechanisms behind new error sources should help survey researchers to adapt to the changing technological landscape.

Chapter 2

Mechanisms Generating Nonresponse in Cell Phone Surveys

2.1 Introduction

Nonresponse rates, particularly in telephone surveys, have been increasing for several decades in developed countries (Curtin, Presser, and Singer 2005, de Leeuw and de Heer 2002; Tortora 2004). Studies featuring both landline and cell phone random digit dial (RDD) samples have shown that nonresponse rates for cell phone samples tend to be at least as high as those for landline samples (Brick et al. 2006; Link et al. 2007; Keeter et al. 2008). These rates typically range between 60% and 90% in the United States (Holbrook, Krosnick, and Pfent 2008). High nonresponse rates threaten survey data quality in two primary ways – elevating the risk of nonresponse bias and increasing data collection costs and variances. Nonresponse bias has the potential to render estimates inaccurate and dangerously mislead survey data consumers. Data collection costs increase because more resources are required to achieve a fixed number of completed interviews. As a consequence, resources may be diverted from other design elements that influence data quality, such as sample size or interviewer training. For both of these reasons, nonresponse is one of the most pernicious problems in telephone survey research.[2]

[2] The focus of this chapter is unit nonresponse, as opposed to item nonresponse. Unit nonresponse occurs when the sampled person does not respond to the request to be surveyed. Item nonresponse occurs when a person cooperating with the interview does not answer a particular question.

Relatively little is known about the nature of nonresponse in cell phone surveys and whether the error properties differ from those observed in landline surveys. As mentioned above, the rates are similar, but it is unclear whether the relationships between nonresponse mechanisms and survey measures are also similar. There is good reason to believe that some dynamics would be the same. For example, people who strongly distrust the government would likely refuse a federally-sponsored survey regardless of whether they were contacted on a landline or on a cell phone. Other mechanisms, however, may work differently and have divergent consequences for nonresponse error. One mechanism that may work differently for landline and cell phone surveys is concern about the financial cost of the call. People may be less likely to answer a cell phone survey that could involve a charge (under U.S. business models) than a landline survey that involves no charge.

The focus of this chapter is identifying mechanisms, or causes, of nonresponse in cell phone surveys. Mechanisms of nonresponse refer to respondent attributes (lifestyle, attitudes, behaviors, and the like) that determine people's propensity to respond to a survey. Prior research suggests that infrequent cell phone usage may be one mechanism of nonresponse (Brick et al. 2006). In addition, response decisions may be influenced by attitudes, such as whether an individual likes or dislikes the accessibility that cell phones afford. It may also be influenced by usage styles, such as using the phone for text messaging, listing to music, and Web browsing. It is an open question as to whether these hypothesized mechanisms are more or less influential in participation decisions than mechanisms already documented in the literature.

Identifying factors influencing response decisions has three practical goals. The first is to assist cell phone survey researchers in preemptively identifying survey statistics that are at serious risk of nonresponse bias. Research has shown that nonresponse may bias estimates of telephone usage and interest in technology (Brick et al. 2006) and possibly other topics at risk that researchers have yet to identify. The second goal is to inform survey recruitment protocols. Some potential reasons for nonresponse could be made more or less salient depending on how interviewers describe the survey to the prospective respondents during recruitment. The third goal is to provide survey statisticians with information about nonresponse that will be useful in developing post-survey adjustments. Identifying the dominant mechanisms of nonresponse would assist survey statisticians in selecting the optimal set of variables to include in post-survey weighting adjustment. The optimal adjustment in the dual frame context may be different from that used for landline RDD surveys, but any such difference has not been well established in the cell phone literature to date.

The next two sections provide a theoretical framework for thinking about nonresponse mechanisms and review studies that used data from traditional modes to evaluate various mechanisms. In Section 2.2 there is a description of the research design used in this study and an explanation of how nonresponse mechanisms were operationalized. The analysis is presented in Section 2.3, followed by a summary and conclusions in Section 2.4.

2.1.1 Distinguishing Nonresponse Mechanisms from Response Propensities

Survey methodologists have been quite successful in explaining how nonresponse may affect estimates (Bethlehem 2002; Groves 2006; Lessler and Kalsbeek 1992) but less successful in explaining why nonresponse occurs in the first place. With respect to the former, Groves (2006) has emphasized that nonresponse is a statistic-specific phenomenon. One estimate may be subject to egregious nonresponse bias while another estimate in the same survey may not be biased at all. It is not enough for survey designers to simply make a judgment that an entire survey is or is not at risk of nonresponse bias. Rather, each key survey measure should be evaluated in terms of what is known about the covariance between that measure and the likely nonresponse mechanisms.

This may be complicated because the mechanisms influencing participation vary across sample members (e.g. Couper 1997; Groves, Singer, and Corning 2000). Response propensity theoretically takes on a number between 0 percent and 100 percent that summarizes the likelihood that a sampled person will respond to a given survey. Values near 100 reflect a very high propensity to respond, and values near zero reflect a very low propensity to respond.

Several properties of response propensities are critical to nonresponse research. Response propensity varies across people, across surveys, and within surveys. For example, Barbara may have a high propensity to respond to health surveys and a low propensity to respond to financial surveys, while Kevin's propensities may be the reverse. Response propensity can also vary *within* a survey, across different recruitment protocols (Groves, Singer, and Corning 2000; Olson 2007). For example, changing from no incentive to a $20 incentive would affect response propensities even though nothing substantive about the survey would have changed. The focus of this study is on determinants of participation that are robust across a broad range of surveys, particularly cell phone RDD surveys of the general population.

Prior research examining nonresponse mechanisms is limited in two important ways. First, nearly all research on nonresponse in telephone surveys has been in the context of landline samples. Other fundamental nonresponse research is based on mail surveys (e.g., Dillman 1978; Heberlein and Baumgartner 1978) or in-person interviewing (e.g., Dunkelberg and Day 1973; Groves and Couper 1998). It is unclear how well findings from these studies generalize to cell phone surveys. The second limitation is that researchers seldom are able to evaluate numerous nonresponse mechanisms in head-to-head testing. For example, Brick and his colleagues (2006) found that topic salience and household inaccessibility influence response propensities in cell phone surveys, but they could not gauge whether these mechanisms are more or less powerful than civic engagement and social integration, which are both known correlates of survey participation. The study described below addresses both of these limitations of previous work on nonresponse in cell phone surveys.

2.1.2 Mechanisms Evaluated in the Literature

Sociologists and social psychologists have posited numerous theories as to why some people may have higher or lower response propensities, across a broad range of surveys. These mechanisms can be grouped into three main categories: societal-level, survey-level, and sample person-level. Figure 2.1 illustrates how these three tiers of mechanisms factor into response decisions.

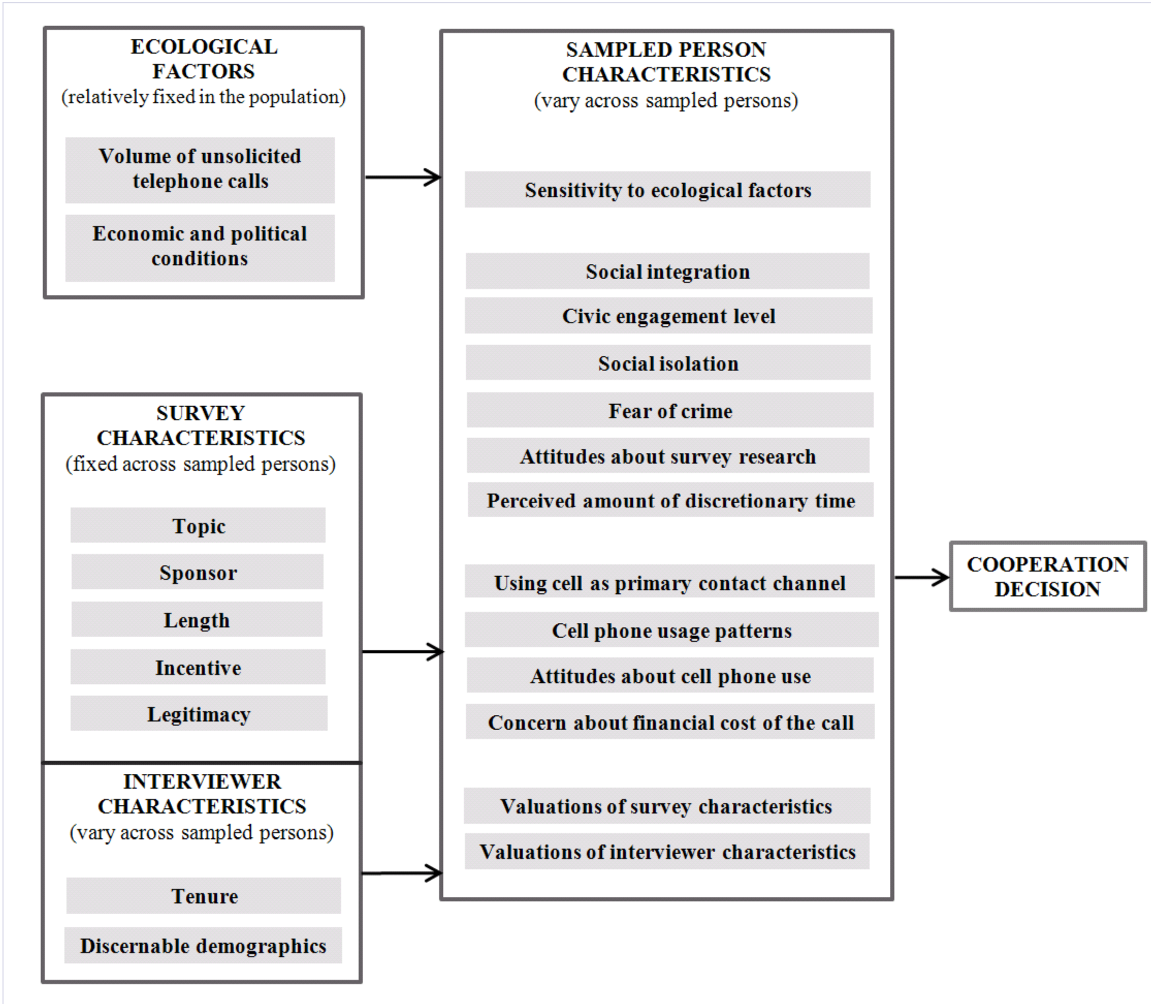

Figure 2.1. A Conceptual Framework for Cell Phone Survey Cooperation

The highest-level mechanisms are ecological. The most prominent mechanism of nonresponse in this tier is the growth in the number of surveys and telemarketing efforts over the past several decades. Since the late 1970's, the number of legitimate surveys, fraudulent surveys, and telemarketing efforts conducted in the U.S. grew rapidly as participation rates declined (Curtin, Presser, and Singer 2005; Remington 1992; Schleifer 1986). As Remington (1992) notes, "Americans have a limited reservoir of goodwill to expend on intrusive and unsolicited telephone contacts of any kind… The sheer volume and occasional dishonesty of telemarketing calls depletes the scarce stock of time individuals are willing to allocate to telephone-related contact." There is little if any causal evidence linking the volume of survey research to reduced survey response propensities, but the correlation is convincing enough for many survey methodologists (Tourangeau 2004). The research that perhaps comes closest to an experimental design is by Link and his colleagues (2006) on the effect of the National Do Not Call Registry on response rates to the Behavioral Risk Factor Surveillance Survey. They do not find any clear indication that the registry helped stem the decline in response rates.

It is extremely difficult to quantify the size of the over-surveying effect because the people who are most sensitive to this phenomenon are perhaps the least likely to participate in research attempting to study it. As a consequence, studies measuring attitudes about the volume of survey and telemarketing calls almost certainly underestimate negative viewpoints. Other ecological nonresponse mechanisms that researchers have documented are national economic and political conditions (Harris-Kojetin and Tucker 1999), though empirical support for these mechanisms is also tenuous.

Ecological mechanisms are a fixture of society. People differ in their sensitivity to these factors and the extent to which they let them shape their participation decisions. The ecological factors themselves, however, present a fairly uniform stimulus across the population. People in virtually all areas of the continental U.S. have experienced an abundance of unsolicited telephone calls, just as they tend to experience economic recessions together. That said, some regional and person-level variation exists. For example, battleground states are targeted for more surveys during election years, and the unemployment rate fluctuates from state-to-state. It is also the case that cell phones have yet to be targeted by vast numbers of survey and telemarketing efforts. The trend of declining cooperation rates across virtually all surveys, however, indicates that societal factors play a critical role in participation decisions.

The next class of mechanisms is specific to the survey. Methodological research has shown that survey features can have at least a modest effect on response propensities. These features include topic interest (Goyder 1987; Groves, Presser, and Dipko 2004; Groves et al. 2006; Heberlein and Baumgartner 1978), sponsor legitimacy (Everett and Everett 1989; Heberlein and Baumgartner 1978; Jobber and O'Reilly 1998; National Academy of Sciences 1979), perceived burden (Dillman, Sinclair, and Clark 1993; Galesic and Bosnjak 2009; Hansen 2006; Tourangeau et al. 2009), incentives (Church 1993; Kulka 1994; Curtin, Singer, and Presser 2007), and privacy and/or confidentiality concerns (Jobber and O'Reilly 1998; Jones 1979; Singer, Van Hoewyk, and Neugebauer 2003; Singer, Von Thurn, and Miller 1995). Critically, these features influence response propensity if and only if they are communicated to and internalized by the sampled person. Groves, Singer, and Corning (2000) detail how people's valuation of salient survey features can influence their propensity to cooperate with a survey request.

Within the class of mechanisms associated with a given survey, are interviewer characteristics that may influence participation decisions. Differences in participation rates across interviewers may be substantial, but identifying the interviewer characteristics that explain such differences is difficult. Researchers have documented a positive correlation between interviewer tenure and response propensity (Couper 1991; Durbin and Stuart 1951; Groves and Fultz 1985; Hansen 2006), but the effect is not always consistent (Singer, Frankel, and Glassman 1983; Schyberger 1967). Studies examining main effects from interviewer gender (Baruffol, Verger, and Rotily 2001; Campanelli and O'Muircheartaigh 1999; Hansen 2006; Pickery and Loosveldt 2002) and age (Singer, Frankel, and Glassman 1983; Norris and Hatcher 1994) have yielded mixed results. Researchers have also investigated the influence of vocal characteristics (Oksenberg and Cannell 1988). Even if such effects are present, the implications for nonresponse bias are minimal barring a connection between key survey measures and the interviewer trait (see Schuman and Converse 1971 on race and Kane and Macaulay 1993 on gender). Meaningful dependencies between interviewer characteristics and response decisions are not especially likely in CATI surveys given the limited information potential respondents have about the interviewer.

The third class of nonresponse mechanisms is specific to the individual sampled person. This class traditionally has included factors such as the sampled person's level of social integration (Abraham, Maitland, and Bianchi 2006; Groves and Couper 1998), social isolation (Goyder 1987), and civic engagement (Brehm 1993; Cialdini, Braver, and Wolf 1992; Couper, Singer, and Kulka 1998; Groves, Presser, and Dipko 2004; Putnam 1995; Voogt and van Kempen 2002). People with stronger ties to their community, stronger identification with mainstream culture, and greater interest in civics tend to participate at higher rates across all surveys. Fear of crime has been shown to suppress response propensities for in-person surveys (Groves and Couper 1998; House and Wolf 1978), though this effect may be greatly reduced in telephone surveys. Even over the phone, however, crime could be salient to participation decisions, given the existence of telemarketing scams, particularly targeting the elderly (for example, Brambila 2009 and King 2009).

Two other factors thought to influence response decisions are attitudes about the value of survey research (Rogelberg et al. 2001) and perceived amounts of personal discretionary time or "busyness" (Groves and Couper 1998). There is empirical support for the idea that people who view survey research positively are more likely to respond than those with negative views (Rogelberg et al. 2001). Recent studies have not, however, found support for the idea that people with less discretionary time have lower response propensities (Abraham, Maitland, and Bianchi 2006; Fricker 2007).

Virtually all of these mechanisms share two important properties: they are difficult to operationalize and have limited explanatory power. Concepts such as social isolation and perceived amount of discretionary time are somewhat ambiguous and do not lend themselves to clean quantification. Out of necessity, researchers have often used demographic and household composition measures as proxy variables for the mechanisms that they actually believe to be influencing response propensity. Groves and Couper (1998) note that demographic correlates of nonresponse, such as education and race, are not causal to response decisions, but rather they are indirect measures of the social psychological constructs that actually influence response. Differences between how the constructs are operationalized in research and the construct themselves raise questions about the internal validity of the results. The use of proxy variables for constructs of interest is relevant to the research described below because some of the same conventions used in the literature are employed.

In their study of nonresponse in the American Time Use Survey (ATUS), Abraham, Maitland, and Bianchi (2006) tested both social integration and busyness as hypothesized mechanisms. Following Groves and Couper (1998), they used the number of adults in the household and the presence of children in the household as indicators of social integration. They found no main effects for these variables but did detect an interaction effect with marital status, whereby the presence of children had no significant effect on survey response for married adults, but the presence children did raise the probability of response for unmarried adults. Other indicators of social integration that have been proposed include living in a multi-unit structure (Groves and Couper), living in rental housing (Lepkowski and Couper 2002; Rizzo, Kalton, and Brick 1996; Zabel 1998), and living in an urban area (Groves 2006). Abraham and her colleagues found no support for the hypothesis that busy people were less likely to respond, when using hours worked per week and the presence of children in the household as proxies. That said, their ability to measure the effect of perceived busyness may have been undermined by the fact that their sample had already proven to be willing survey participants through their earlier participation in the Current Population Survey (CPS).

Tests of alternative reasons for nonresponse have relied on other demographics as proxies. Social isolation is one theory for why racial and ethnic minorities generally have lower response propensities than non-Hispanic whites (Groves and Couper 1998). According to the theory, social isolates feel out-of-touch with the mainstream culture of society. They may explicitly reject some of the norms of the dominant society, including participation in a variety of social and political activities – including surveys. Alienated groups are often defined in terms of race and ethnicity, and so indicators for African-American race and Hispanic ethnicity are used as proxies for this mechanism.

Fear of crime is another construct that is often measured with crude indicators. Groves and Couper (1998) used indicators such as residence in urbanized areas, gender, and seniors living alone. The tendency for telemarketing scams to target the elderly suggests that there may also be an interactive effect between age and fear of crime on response propensities.[3] Among the elderly, concerns about victimization, however, may be outweighed by their greater civic participation. Political engagement has proven to be predictive of high response propensities in a number of studies (Brehm 1993; Cialdini, Braver, and Wolf 1992; Couper, Singer, and Kulka 1998; Groves, Presser, and Dipko 2004; Putnam 1995; Voogt and van Kempen 2002). Older and wealthier Americans are generally more likely to be registered and to vote, and so age and income are sometimes used as proxies for political engagement.

Finally, the belief that survey research is useful to society may also influence response propensities across surveys. Given that Americans are generally not exposed to statistics and survey research until college (if ever), a high level of education is perhaps the demographic indicator most strongly correlated with positive valuation of survey research.

These classes of mechanisms – ecological, survey-specific, individual social-psychological – tell only part of the story as to why some people respond and others do not. In the nonresponse literature, much of the variance in response decisions remains unexplained after having accounted for these factors. This is true for all modes of data collection. In the telephone survey context, evidence is mounting for another important class: device-related mechanisms. In particular, how people use their telephones appears to influence their propensity to respond. Perhaps the most robust finding is that people who use their cell phone more often are more likely to participate in cell phone surveys

[3] The potential for interaction effects is not apparent in the schema presented in Figure 2.1, which points to a limitation of the model. The model notes that fear of crime varies across people, but it does not suggest that the variation may depend on other known factors.

than those who use it less often (Brick et al. 2006). It is an open question as to whether this result stems from a greater likelihood of contact among frequent users, a greater likelihood of cooperation, or both.

A study by Vehovar and Callegaro (2007) using Eurostat data from Slovenia suggests several more causes of nonresponse in cell phone surveys. Respondents in a face-to-face survey were probed about their willingness to participate in a cell phone survey. Expressed willingness to participate was greater among those with high computer literacy scores than low scores, among younger respondents than older ones, among those with large social networks than those with small ones, among those who had participated in surveys in the past than those who had not, and among extroverted personality types than introverted types. It is not clear from the report whether these effects remain significant in multivariate analysis. Furthermore, these findings are limited because they are based on self reports of hypothetical behavior rather than observed behavior in an actual cell phone survey.

This set of empirical results leaves us with an imperfect understanding as to what mechanisms determine response propensity in cell phone surveys – and how those mechanisms compare to considerations that shape response decisions in other modes, especially landline surveys. Presumably, frequent cell phone users tend to have higher response propensities to cell phone surveys because their usage patterns and attitudes about their phones differ from those who use their cell phone less often. Perhaps heavy usage reflects greater interest in and adeptness at using mobile technology, which may also explain the Vehovar and Callegaro (2007) finding about computer literacy. If so, then there may also be a relationship between response propensity and usage of non-telephony functions, such as text messaging.

2.1.3 Extending the Literature – Nonresponse Mechanisms Related to Device

Several distinct cell phone usage patterns seem particularly relevant to survey response propensity. Some cell phone owners only use their device to make outgoing calls, as in the case of an emergency. Presumably, many of these people keep their cell phone turned off when it is not in use. It is expected that "outgoing only" users have lower response propensities in cell phone surveys than others. The major barrier would be establishing contact, but this behavior could very well be associated with a lower propensity to cooperate as well.

At the other extreme are people who use their cell phone all the time, even when there is no pressing need to communicate with someone. For example, a 2006 study conducted by the Pew Research Center, AOL, and the Associated Press found that 41% of cell phone users say that they fill free time when they are traveling or waiting for someone by making cell phone calls. This behavior implies frequent usage, but it also seems to reflect a deeper affinity or psychic reward from using the cell phone.

Another usage behavior that may influence response propensity is sharing the cell phone with another person. According to a 2007 study in three states (Georgia, New Mexico, and Pennsylvania), roughly 10% to 15% of cell phone owners share their phone at least "one-third of the time" with another adult in the household (Link et al. 2007). It seems likely that cell phone sharing would increase the likelihood of establishing contact with an eligible adult, though it is not clear that this behavior would be related to cooperation.

In addition to these behavioral characteristics, several attitudes may influence response decisions. Cell phone users vary in how they conceptualize their phone. Some view it as a private line through which only close friends and family are supposed to reach them. Anecdotal evidence indicates that a fair number of refusals occur because the person answering was surprised and upset by the fact that they were being called on their cell phone to participate in a survey. Others, particularly cell-only adults, view it as a broad access point for anyone needing to reach them – acquaintances and strangers alike.

A separate attitudinal dimension is affinity/dislike for the notion that carrying a cell phone makes the user more accessible to others. In the Pew/AOL/AP survey, about one-fifth (22%) of cell phone users agreed with the statement that "too many people try to get in touch with me because they know I have a cell phone." Others clearly like the increased accessibility. It is expected that those liking the increased accessibility have higher contact and cooperation propensities in general, compared to other cell phone users.

One final attitude that may be causal to participation decisions is concern about the financial cost of a call. In the US, the recipient pays, and some cell users have limited resources in terms of the ability to pay for nonessential calls, or they may be limited in the number of minutes on their service plan. According to the Pew/AOL/AP survey, some 44% say that they wait to make most of their calls for the hours when they do not count against their "anytime" minutes in their basic calling plan. Many survey organizations sampling cell phone numbers offer an incentive to defray this expense, as suggested by American Association for Public Opinion Research guidelines (AAPOR

2008), but this practice is not universal (Jones 2008). It seems plausible that for cell phone surveys not offering an incentive, cooperation (and possibly even contact) propensities may be lower among those who worry about the cost of the call. Even in surveys offering an incentive, the effect may not disappear altogether if the cell users are not confident that they will receive the post-paid incentive.

2.1.4 Noncontact versus Noncooperation

The previous discussion focuses primarily on overall response propensity. There is theoretical and practical value, however, in decomposing this concept into contact propensity and cooperation propensity. Some mechanisms may influence one component of nonresponse but not the other, while other mechanisms may influence both. The distinction is important because techniques for addressing noncontact (call schedule adjustment, pre-notification, voicemail messages) may differ from those targeting noncooperation (incentives, recruitment appeals, interviewer training).

The mechanisms reviewed in section 2.1.2, such as social integration and civic engagement, have been discussed most often in the context of cooperation, but certainly some may be relevant to contact as well. The discussion of the mechanisms related to device (section 2.1.3), was more explicit as to whether a given factor is expected to affect contact or cooperation. In the analysis below, nonresponse is assessed in the aggregate, as well as, decomposed into contact and cooperation. The goal is to understand better how traditional versus device-related mechanisms work to determine response outcomes, so making a distinction between these components will provide greater insight.

2.1.5 Experimental Designs for Identifying Causes of Nonresponse

The ideal design to identify mechanisms of nonresponse in cell phone surveys would feature a random sample of the general population and variables on the frame indicating the relevant attitudinal and behavioral characteristics for all cases (both respondents and nonrespondents). If, for example, the age and phone usage patterns of everyone in the sample were known, it would be possible to measure how well usage predicts response to a cell phone survey when controlling for age. An even richer design would randomly assign half of the sample to be called on a landline and the half to be called on a cell phone. This would facilitate comparison of the characteristics predicting nonresponse to surveys on each device.

Unfortunately, no general population frame contains micro-level data for attitudinal and behavioral nonresponse mechanisms. Researchers have used three primary tools to circumvent this problem: experimental manipulations, special population frames, and repeated measures designs. The first two approaches, which are sometimes combined, tend to have strong internal validity but are limited in scope. Using a general population sample, the researcher can experimentally manipulate one or more features of the recruitment protocol and evaluate the effects on response outcomes (as in Singer, Van Hoewyk, and Maher 2000 or Brick et al. 2005). Using specialized samples, such as membership lists, researchers can implement the same type of manipulation but with greater leverage on the theoretical mechanisms. Groves and his colleagues (2004, 2006, 2009) compare the response behaviors of persons sharing a common attitude about the manipulated feature (as assumed by their membership on the list) to the response behaviors of other sample members who are not on the list. This approach has the advantage of isolating the nonresponse mechanism of interest, but it generally does not permit simultaneous testing of alternative mechanisms. In addition, use of membership lists raises the possibility that the results may not apply to general population samples.

In studies where numerous attributes of a population are known, researchers often emphasize comparisons between respondents' and nonrespondents' values. Prominent examples of this approach include Assael and Keon (1982), Bolstein (1991), Kennickell and McManus (1993), Lin and Schaeffer (1995), and Olson (2006). These studies are better suited for assessing nonresponse bias, however, than identifying nonresponse mechanisms. Variables available on the frame may be associated with key survey measures, but they are unlikely to be causal to response decisions. This approach is not suitable for the research objectives in this investigation because I seek to test a large set of hypothesized nonresponse mechanisms simultaneously. The variables used to measure the mechanisms are not available on any pre-existing frame.

An alternative technique, and the one employed in this study, is to conduct a repeated measures experiment. A preliminary survey (Wave 1) is used to record the attributes of interest on a diverse set of sample members. The researcher can capture numerous attitudinal and behavioral variables in this Wave 1 survey. The participants are later re-contacted for a subsequent survey (Wave 2), and their response behaviors are recorded. Response (or not) to the Wave 2 survey is the outcome of interest. The researcher tests competing mechanisms by modeling response to Wave 2 from the attitudinal and behavioral predictors captured in Wave 1 and observing which measures are most effective in explaining response. When the Wave 2 survey is a subsequent panel study interview, then the analysis is measuring panel attrition (as in Lepkowski and Couper 2002). The study at hand, however, is intended to study nonresponse in one-off surveys, and to facilitate this the Wave 2 survey was made to appear unrelated to the Wave 1 survey (this is discussed in Section 2.2.1).

Several teams of survey methodologists have used this repeated measures design to generate new knowledge about nonresponse. For example, Abraham, Maitland, and Bianchi (2006) leveraged the fact that the ATUS sample is drawn from the outgoing rotation groups of the CPS. They modeled response to the ATUS (their "Wave 2") using data collection in the CPS (collectively, their "Wave 1"). Similarly, Groves, Singer, and Corning (2000) examined nonresponse to a mail survey follow-up of Detroit Area Study respondents.

The primary limitation of this research design is that it assumes that respondents to the first survey are representative of the entire population of inference, with respect to response behavior. Even if the second survey (Wave 2) is designed to appear independent, there is still a sense in which the experiment is measuring attrition rather than nonresponse to a one-off survey. Respondents to Wave 1 are, by definition, more cooperative than those who did not respond to Wave 1. Thus, conditioning the experiment on Wave 1 respondents has the potential to introduce bias. The critical unanswerable question is whether the Wave 1 respondents differ from the nonrespondents with respect to the considerations that determine their participation decisions. Encouragingly, a study conducted by the National Agricultural Statistical Service (McCarthy et al. 2006) indicates that cumulative burden from prior surveys does not uniformly have a negative effect on response propensity, at least in their establishment surveys. In this study, I attempt to minimize any bias from Wave 1 nonresponse in two ways. Large, national samples were drawn from both the landline and cell RDD frames so that the starting sample has desirable representative qualities. Second, the analyses were conducted with and without weighting to adjust for Wave 1 nonresponse, and both sets of results are discussed. An investigation of the effects from panel attrition on the results in this Chapter is presented in section A.3 of the Appendix.

2.2 Research Methods

2.2.1 Sample Design

A repeated measures experiment with a unique randomization component was conducted to examine mechanisms of nonresponse in cell phone RDD and landline RDD surveys. The experiment, shown in Figure 2.2, consisted of two surveys. In the first survey (Wave 1), some 1,072 landline RDD interviews and 1,053 cell RDD interviews were conducted in a national survey of the US adult population. Interviewing was conducted December 9 to 23, 2008. The Wave 1 questionnaire included a special module measuring the attitudinal and behavioral characteristics that are hypothesized to be causal to response decisions across a broad range of telephone surveys.

Two months later, a subset of the Wave 1 respondents were re-interviewed for a second survey (Wave 2), which was designed to appear independent and unrelated to Wave 1. The key dependent variable in the experiment is response to Wave 2. Response outcomes to Wave 2 are modeled using the predictors captured in Wave 1, because the independent variables are known for both the Wave 2 respondents and nonrespondents. To fully leverage the design, subjects were randomly assigned to be contacted for Wave 2 on either their landline or their cell phone. The randomization step is useful because it helps to break the confound between sampling frame and respondent characteristics. Past

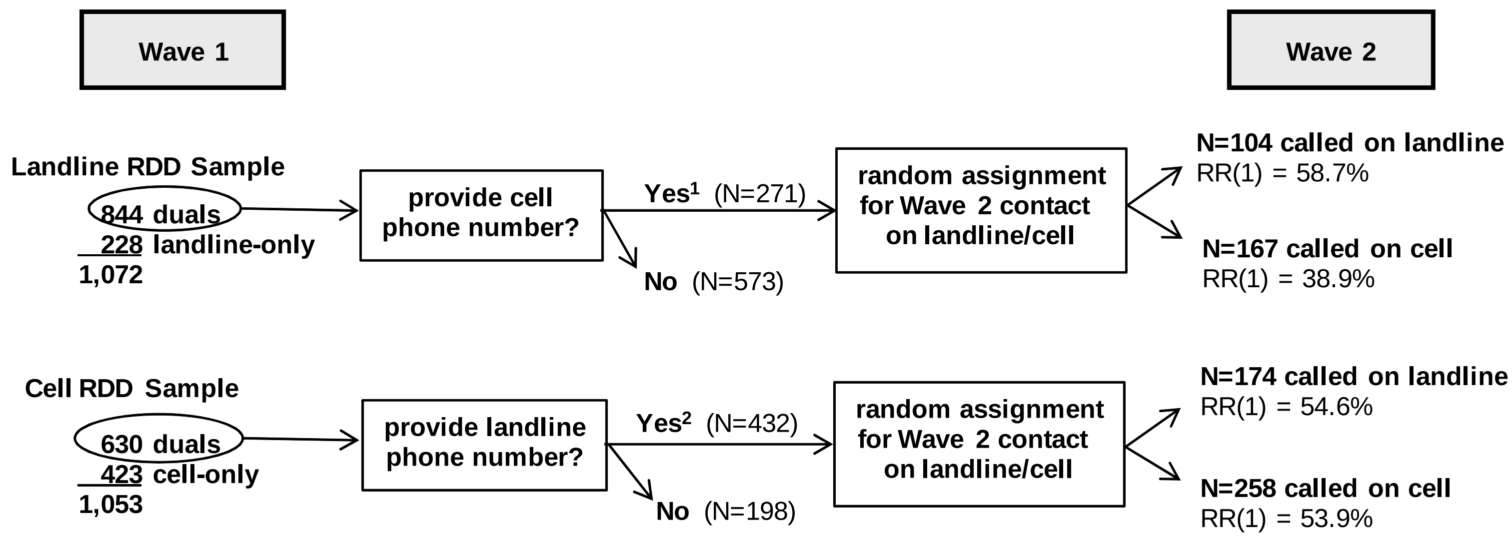

Figure 2.2. Experimental Design and Response Outcomes

[1]Some 79 of the 271 dual users from the Wave 1 landline RDD sample provided their cell number only after a follow-up call was made.

[2]Some 93 of the 432 dual users from the Wave 1 cell RDD sample provided their landline number only after a follow-up call was made.

studies show, for instance, that dual users sampled from the landline frame differ on several dimensions from dual users sampled from the cell frame (Keeter et al. 2008; Kennedy 2007b).

Two steps were required to create the randomized sample for Wave 2. During the Wave 1 interview, respondents were asked about their telephone service status: cell-only, landline-only, or both (henceforth "dual user"). Having both types of phones was a requisite for randomizing the device on which they were called for Wave 2. Therefore, only the dual users identified in Wave 1 were eligible for the Wave 2 survey, while landline-only and cell-only respondents were dropped from the experiment. Re-interviews with all Wave 1 respondents was not possible given the study budget. The dual users were of greatest analytic value because of the potential for randomization, so they were the focus of the study.

The second step required for randomization of Wave 2 device was obtaining the respondent's other telephone number. At the end of the Wave 1 interview, dual users sampled from the landline RDD frame were asked to provide their cell phone number, and dual users sampled from the cell RDD frame were asked for their landline number. Unsatisfactory levels of cooperation with this initial effort to obtain phone numbers prompted the conduct of a brief callback. Dual users who initially declined to provide the number for their other phone were re-contacted about one week later and simply asked for it again for "quality control purposes" associated with the Wave 1 survey. At no time were respondents informed that they would be re-contacted two months later for a second, ostensibly unrelated survey.[4]

After Wave 1 data collection, all respondents who had only a landline, had only a cell phone, or failed to provide sufficient contact information were dropped from the experiment and all analysis discussed here. The remaining cases were re-interviewed two months later for Wave 2. Wave 2 interviewing was conducted from February 3 to 22, 2009, using the same calling protocol as in Wave 1. Landline numbers were called a maximum of six times, and cell phone numbers were called a maximum of four times. Anyone called on a cell phone was offered $10, post-paid, to defray any charges. Respondents were not remunerated for landline interviews. These procedures were implemented so that both surveys reflected current practices of organizations conducting dual frame surveys (Keeter et al. 2007; Link et al. 2007). Steps were also taken to ensure that the Wave 1 and Wave 2 respondent was the same person and not someone else in the household. In the Wave 1 landline sample, the last birthday method was used to select the adult to be interviewed, and in the cell sample the adult answering the phone was interviewed. At the end of the Wave 1 survey, all dual users were asked for their age, gender, and first name. In Wave 2, interviewers asked to speak with the Wave 1 respondent. If the first name was not given, then interviewers used the age and gender data to identify the correct adult.

The Wave 2 survey request was made to appear completely independent so that response decisions would be as representative as possible of a one-off survey. The topic and sponsor mentioned during the recruitment protocol were changed for Wave 2. In

[4] This use of low-level deception was deemed justified by the institutional review boards involved on the grounds that respondents were not harmed in any way and the manipulation was critical to the research objectives.

Wave 1, interviewers stated that they were calling on behalf of the Everett Group, an opinion research firm in Washington, D.C., and that they were conducting a survey about "some issues that are important to the country." In Wave 2, interviewers stated that they were calling for Princeton Survey Research Associates (PSRA), and that they were conducting a "survey for policymakers interested in how people feel about some local and national issues." Both statements were essentially true because the Everett Group was the client funding the research and PSRA was the data collection company responsible for the interviewing and remuneration mailings.

2.2.2 Variables Used to Capture Nonresponse Mechanisms

As mentioned above, the Wave 1 instrument was used to measure respondents' attributes on the factors asserted to influence response decisions. This section details how those factors (or mechanisms) were operationalized in the questionnaire. Table 2.1 reports the mechanisms tested, the corresponding question wording, and expectation as to whether the mechanism should influence contact propensity, cooperation propensity, or both.

Table 2.1. Hypothesized Nonresponse Mechanisms Measured in Wave 1

Hypothesize Nonresponse Mechanism	Question Wording	Expected to affect…
Sensitivity to the volume of unsolicited telephone calls	- People get a lot of unwanted contacts in the form of telemarketing, junk mail, and e-mail spam. How much of a nuisance are these things for you? A very serious nuisance, somewhat serious nuisance, mild nuisance, or not a nuisance at all?	contact propensity and cooperation propensity
Social integration	- How many children under the age of 18, if any, do you have living at home? - How many people 18 or older live in your home, including yourself? - Are you married, single and never married or in some other category? - Do you own or rent your home?	contact propensity and cooperation propensity
Social isolation	- So we may represent all people fairly, do you consider yourself to be Latino, or of Hispanic or Spanish descent? - What race or races would you use to describe yourself?	cooperation propensity
Fear of crime	- I'd like to ask you about some different aspects of your life. For each one, please tell me whether you are very satisfied, somewhat satisfied, not too satisfied, or not at all satisfied… Your personal safety from crime where you live?	cooperation propensity
Civic engagement	- Which of the following best describes your age group? - What is the highest level of education you have completed?	cooperation propensity
Attitudes about survey research	- The terms "poll" or "research survey" are often used to disguise a sales pitch* - Answering questions in polls or research surveys is an interesting experience* - Answering questions in polls or research surveys is a waste of time*	cooperation propensity
Reliance on cell as primary contact channel	- I give my cell phone number when providing contact information on forms and applications*	contact propensity
Frequency of usage	- When you get a call on your cell phone, do you answer almost always, most of the time, some of the time, rarely, or never?	contact propensity and cooperation propensity
Use cell to fill free time	- I often make cell phone calls to fill up my free time while I'm traveling or waiting for someone*	contact propensity and cooperation propensity
Sharing cell phone	- Are you the only person who uses your cell phone or do you share it with someone else?	contact propensity
Use cell for outgoing calls but not incoming calls	- I only use my cell phone to MAKE calls - not to receive calls*	contact propensity
Non-telephony usage	- Do you ever use your cell phone to (a) send and receive text messages (b) take pictures or video (c) connect to the internet (d) watch video or TV shows?	contact propensity and cooperation propensity
Mastery of cell phone technology	- When I get a new electronic device, I often need someone else to show me how to use it* - I enjoy reading news and industry articles about the latest in cell phone technology*	contact propensity and cooperation propensity
Desire for cell phone to make the users more accessible to others	- I like that my cell phone allows me to be more available to others*	contact propensity and cooperation propensity
Cost concerns	- Some people try to keep costs down by not using too many minutes on their phone plan. How about you? When you use your cell phone, do you think about the cost of the call most of the time, some of the time, rarely, or not at all?	contact propensity and cooperation propensity

*The response scale (not shown to conserve space) is "strongly agree, agree, disagree, or strongly disagree."

The first set of mechanisms comes from the nonresponse literature. Sensitivity or frustration with the volume of unsolicited contacts was measured by asking, “People get a lot of unwanted contacts in the form of telemarketing, junk mail, and e-mail spam. How much of a nuisance are these things for you? A very serious nuisance, somewhat serious nuisance, mild nuisance, or not a nuisance at all?” Fear of crime is assessed using an item from a broader battery asking about satisfaction with various aspects of the respondent’s life. The question stem and item read, “I'd like to ask you about some different aspects of your life. For each one, please tell me whether you are very satisfied, somewhat satisfied, not too satisfied, or not at all satisfied… your personal safety from crime where you live?” To measure general attitudes about surveys, interviewers administered a subset of items from a study by Schleifer (1986). Respondents were asked to report whether they strongly agreed, agreed, disagreed, or strongly disagreed with each of three statements (i) The terms “poll” or “research survey” are often used to disguise a sales pitch (ii) Answering questions in polls or research surveys is an interesting experience (iii) Answering questions in polls or research surveys is a waste of time.[5] Finally, following previous studies (Abraham, Maitland, and Bianchi 2006; Groves and Couper 1998; Lepkowski and Couper 2002) household and demographic variables were used as proxies for other theoretical constructs. Presence of children, number of adults in the household, home ownership, and marital status serve as indicators of social integration. Race and Hispanic ethnicity serve as indicators of social isolation.

The second set of mechanisms concerns how respondents use their cell phones. The frequency of usage question read, “When you get a call on your cell phone, do you

[5] The Cronbach coefficient alpha for these items was highest (alpha=0.58) when the sales pitch item was deleted. In the analysis section, the “interesting experience” and “waste of time” items are combined using equal weighting and used as an independent variable called the “hostility to surveys scale.”

answer almost always, most of the time, some of the time, rarely, or never?" Three other items were used to identify specific types of usage: "I often make cell phone calls to fill up my free time while I'm traveling or waiting for someone," "I only use my cell phone to MAKE calls - not to receive calls," and "Are you the only person who uses your cell phone or do you share it with someone else?" These items featured a four-point agree/disagree response scale. A separate question was used to measure non-telephony usage of cell phones: "Do you ever use your cell phone to (a) send and receive text messages (b) take pictures or video (c) connect to the internet (d) watch video or TV shows?" Respondent interest and mastery of mobile technologies are measured with the questions, "When I get a new electronic device, I often need someone else to show me how to use it," and "I enjoy reading news and industry articles about the latest in cell phone technology", respectively. Two additional hypothesized mechanisms related to device were concern about the cost of the call (measured using an ordinal four point scale) and (dis)agreement with the statement, "I like that my cell phone allows me to be more available to others."

In most cases, the ordinal independent variables are expected to have a linear effect on response propensity, but not always. Consideration of how the variables relate substantively suggests that there may be several non-linear effects, interactive effects, and potentially some collinear variables. For example, respondent age may have a non-linear effect given that cell phone usage is less common among seniors than it is among adults under age 40. Separately, there may be an interactive effect with age and fear of crime, given that telemarketing scams often target senior citizens. There is also evidence that attitudes about cell phones and their usage vary by race, age, and education (Traugott et al. 2006). With respect to collinearity, several of the cell phone usage variables may be measuring highly similar constructs. For example, it may be the case that everyone who says they "like that my cell phone allows me to be more available to others" disagrees with the statement "I only use my cell phone to MAKE calls - not to receive calls." Each of these issues is addressed in the analysis, and the results are discussed below.

2.3 Findings

2.3.1 Response in Wave 1

The first survey (Wave 1) was designed to resemble other national, dual frame RDD surveys being conducted by major U.S. survey organizations. A common finding is that the response rate is similar for the landline and cell RDD samples (Keeter et al. 2007). As expected, this result replicated in Wave 1. The response rate (AAPOR(3)) was 20% for cases sampled from the landline frame and 19% for cases sampled from the cell phone frame, typical for studies using a four or six callback protocol. The incidence of cell and landline users in the Wave 1 samples was also in line with recent dual frame surveys (Link et al. 2007). Some 79% of the landline frame respondents reported being dual users, as did 60% of the cell phone frame respondents.

In order to be eligible for Wave 2, respondents needed to be a dual user and provide the number to their landline or cell phone (depending on which number was not already known from the sampling frame). The willingness of cell phone respondents to provide their landline number was substantially greater than the willingness of landline respondents to provide their cell phone number. This result mirrors differences in how landline and cell sample respondents incorporate cell phones into their daily life. Only half of the landline sample dual users say that they provide their cell phone number on forms and applications. This compares with 71% of cell sample dual users. At the end of Wave 1, 23% of the landline sample dual users had provided their cell number and 54% of the cell sample dual users had provided their landline number. These item-cooperation rates fell short of the thresholds needed to have enough cases eligible for Wave 2, if reasonable levels of power were going to be achieved. In response, The Everett Group organized a brief follow-up calling effort targeting Wave 1 dual users who declined to provide their phone number. These cases were called at most two times, and the interviewer simply asked for the phone number, explaining that it was needed as part of the survey in which they had recently participated. This follow-up calling effort yielded 79 missing cell phone numbers and 93 missing landline numbers. At the end of the follow-up effort, a total of 271 landline frame dual users and 432 cell frame dual users were combined to create the sample for Wave 2. These cases were randomly assigned to be called for Wave 2 on either their landline or cell phone.

2.3.2 Response in Wave 2

The Wave 2 sample was released in two replicates of equal size. In the first replicate, 50% of the cases were assigned to be called on their cell phone and 50% were assigned to be called on their landline. After a week of interviewing, it was clear that the response rate was much higher in the landline condition. Other methodological analyses planned for this study required that the responding samples in the two treatment groups be similar in size. The differential response rate threatened that goal. In response, the treatment allocation rates were revised for the second Wave 2 replicate. Assignment was still random, but this time about 70% of the cases were designated for calling on the cell number and 30% on the landline number.

The Wave 2 response rate for all four treatment groups are reported in Figure 2.2 above. Propensity to respond in Wave 2 was strongly related to the frame through which the person was originally sampled. Among those originally sampled through the landline RDD frame, the Wave 2 response rate was significantly higher for those called on their landline (AAPOR RR(1)=58.7%) than for those called on their cell phone (RR(1)=38.9%; $X^2(1)=10.03$, $p<.01$). By contrast, people originally sampled through the cell RDD frame were about equally likely to respond to Wave 2, regardless of whether they were called on their landline (RR(1)=54.6%) or their cell phone (RR(1)=53.9%; $X^2(1)=0.02$, *n.s.*).[6]

2.3.3 Response to the Cell Phone Survey: A Baseline Model Using Variables from the Literature

The central aim of this study is to identify factors influencing response propensities in cell phone surveys. The nonresponse literature suggests one set of mechanisms, and early research on cell phone surveys suggests another set of mechanisms related to how the device is used. Both sets are tested empirically in this study. First, logistic regression is used to model response to Wave 2 using the mechanisms in the literature. The results are presented in the "baseline model" column of Table 2.2. The model is estimated using the 415 dual users randomly assigned to be contacted for Wave 2 on their cell phone who provided a working, valid number.

[6] The productivity rates (contact, cooperation, and eligibility) for both waves are reported in Appendix Table A.1.1. The final disposition codes are reported in Appendix Table A.1.2.

In the baseline model, participation is regressed on indicators of social integration, civic engagement, social isolation, fear of crime, and general attitudes about surveys. Also included is an indicator for the frame from which the case was sampled in Wave 1. If people were equally accessible through the two frames, this coefficient would be zero. Previous studies, however, suggest a relationship between RDD frame and response propensity (Brick et al. 2006; Kennedy 2007b), and so the decision was made to retain the frame indicator in the model. The strength of this relationship was also clear from the variation in response rates across the experimental treatments, shown in Figure 2.2.

Table 2.2. Logistic Regression Models Predicting Response to the Wave 2 Cell Phone Survey

		Baseline Model		**Final Model**	
Parameter	*df*	Estimate	s.e.	Estimate	s.e.
Intercept	1	-0.50	(0.71)	-3.98**	(1.28)
Sampled from landline RDD frame	1	-0.38***	(0.11)	-0.39***	(0.11)
Hostility to surveys scale	1	-0.23*	(0.10)	-0.23*	(0.10)
Age	1	0.19*	(0.09)	0.26**	(0.09)
Black	1	-0.41*	(0.20)	-0.52*	(0.20)
Single adult household	1	-0.24	(0.15)		
Hispanic	1	-0.15	(0.23)		
Fear of crime	1	-0.15	(0.16)		
Married	1	-0.12	(0.13)		
Child in the household	1	-0.06	(0.13)		
Renter	1	0.01	(0.15)		
Sensitivity to telemarketing, junk mail, spam	1	0.01	(0.11)		
Education	1	-0.01	(0.06)		
Give cell number as contact info	1			1.03**	(0.35)
Non-telephony usage scale (text, picts, web)	1			0.52*	(0.21)
Outbound calling only	1			-0.45*	(0.19)
Answer cell most/all of the time	1			0.32	(0.17)
Share cell with someone else	1			0.24	(0.18)
Non-telephony scale x Give cell number as contact info	1			-0.17**	(0.06)
Area under ROC curve (c)		0.64		0.69	
-2 Log Likelihood		547.4		523.1	
Re-scaled r-square		0.09		0.16	

****p* <.001 ***p* <.01 **p* <.05

Note - Both models are based on the 415 cases randomly assigned to be contacted for Wave 2 on their cell phone. The likelihood ratio test comparing the models is highly significant (P[X^2(2)>24.3] < .0001), indicating that the final model explains more variance than the baseline model.

In fact, the sampling frame is the most powerful predictor of response to the cell phone survey among those tested in the baseline model. People sampled through the landline RDD frame in Wave 1 were 1.5 times less likely to participate in the cell phone component of Wave 2 compared to people sampled through the cell RDD frame, when controlling for the other factors in the model. Respondent age, race, and the hostility to surveys scale also were strong predictors of response to the cell phone.

Another notable finding from the baseline model is the lack of an effect associated with several ubiquitous correlates of nonresponse. People with more education tend to be more likely to participate in surveys than those with less. No such effect was found in the cell phone survey, however. One potential explanation was that the null result was a consequence of including a measure of the individual's attitudes about survey research, but the effect from education remained non-significant even when the hostility to surveys scale was dropped. Being married, owning a home, and having children under age 18 at home also tend to be positive predictors of response in surveys, a pattern that has been attributed to greater social integration (Groves and Couper 1998). None of these variables, however, is a significant predictor of response to the cell survey. The measure of frustration with unsolicited contacts, such as telemarketing, junk mail, and e-mail spam also showed no effect. This mechanism is particularly difficult to study in a survey, however, given that people most sensitive to unwanted contacts were probably unlikely to participate in Wave 1, and thus unlikely to be represented in the analysis.

One potential concern about the validity of the results in Table 2.2 is the fact that the analysis is conditioned on people who were cooperative with the first survey. If nonrespondents to Wave 1 base their response decisions on different factors than those who were cooperative, the parameter estimates will be off in terms of measuring effects in a one-off cell phone survey. I addressed this issue by re-running the model using a weight that adjusts for probabilities of selection (number of adults in the household) and differential response propensity across demographic groups. Raking was used to align the sample to U.S. adult population benchmarks for gender, age, education, race/ethnicity, region, and telephone service. The results are not dramatically different from the unweighted results. When the weight is used, the frame indicator remains significant (p<.01), the hostility to surveys scale becomes marginally non-significant (p=.09), age becomes non-significant (p=.33), and African-American race becomes non-significant (p=.92). The other predictors remain non-significant with the weighting, and the fit of the model does not change appreciably. These differences relative to the unweighted model are logical given the demographic variables used in the weighting. The weight corrects for age and race, so it is not surprising that these predictors lose their statistical significance when the weight is applied. The fact that the predictors in the baseline model work in much the same way with or without the weight suggests that nonresponse in Wave 1 does not severely undermine efforts to identify factors that influence response decisions in cell phone surveys. That said, the weight manipulation is an indirect test, and so the possibility that the results would be different if there was no nonresponse in Wave 1 cannot be ruled out completely.

A separate question concerning the baseline model is whether the results are different if the frame indicator is dropped. As mentioned above, support for that variable being in the model was solely empirical, not theoretical. When the baseline model is re-run excluding the frame indicator, the model fit declines substantially (-2 Log L= 559.4, c=.60, re-scaled r^2=.05). This result is unsurprising given the strong effect captured by the frame indicator. The coefficients, however, change little. The hostility to surveys scale remains significant, age and race become non-significant, and the other non-significant variables in the baseline model remain non-significant.

In the earlier versions of the baseline model, interaction terms between sampling frame and the other predictors were tested. Presumably, some of the variables, like education, could behave differently for those coming from one frame versus the other. No significant interaction terms, however, were found. The interactive effect of age and fear of crime mentioned in section 2.1.2 was also not significant.

Several other model specification decisions deserve mention. As indicated by the degrees of freedom in Table 2.2, all non-binary predictors were treated as linear variables. This is relevant for age, education, sensitivity to telemarketing, junk mail and spam, and the hostility to surveys scale. The model was re-estimated with these variables coded as categorical, but the results were essentially unchanged. The figures presented are based on the linear effects coding because the factors were expected to be related to response propensity in a linear fashion. For age and education, I suspected that there might also be a curvilinear effect. Quadratic terms for these variables were tested in the presence of the main effects, but both quadratic terms failed to reach statistical significance.

2.3.4 Response to the Cell Phone Survey: A Final Model Using Device-Related Variables and Variables from the Literature

In addition to the predictors from the literature tested in the baseline model, I sought to measure the influence of device-related factors. The "final model" shown in Table 2.2 was developed in two stages. The significant predictors from the baseline model ($p<.05$) were retained and the non-significant predictors were dropped. The second step was to add in the device-related mechanisms. Several of these variables did not have a perceptible effect on response to the Wave 2 cell survey. In the interest of parsimony, most of the non-significant device-related variables were excluded from the final model.

The fit of the final model is substantially better than that of the baseline model, reflecting the ability of the device-related mechanisms to explain additional variance in response behavior. Summary measures of model fit are reported at the bottom of Table 2.2. The likelihood ratio test comparing the models is highly significant ($p<.0001$), indicating that the fit of the final model is superior to that of the baseline model. The distributions of predicted probabilities also provide information on the relative performance of the models. In general, greater variance in predicted probabilities indicates better fit of the data. The predicted probabilities of response for the final model range from .04 to .82. The range of predicted probabilities from the baseline model is somewhat narrower (0.15 to 0.78). Histograms of both distributions are presented in Figure 2.3.[7]

[7] Another approach for comparing the performance of different models is cross-validation, which facilitates objective comparison of multiple models in terms of their respective fractions of misclassified cases. Unfortunately, cross-validation is not part of SAS PROC LOGISTIC or any other SAS regression procedure. Akaike's Information Criteria (AIC), however, simulates the cross-validation situation and has been recommended as a convenient substitute (Shtatland, Kleinman, and Cain 2004). Lower AIC values are preferable to higher values. The AIC for the final model is 545.1, which compares to 573.4 for the baseline model. Information criteria are not tests of significance, so they do not indicate that the better of two models is "significantly better." They are, however, appropriate for a non-nested case such as this and they do not depend on asymptotic Chi-Square approximations.

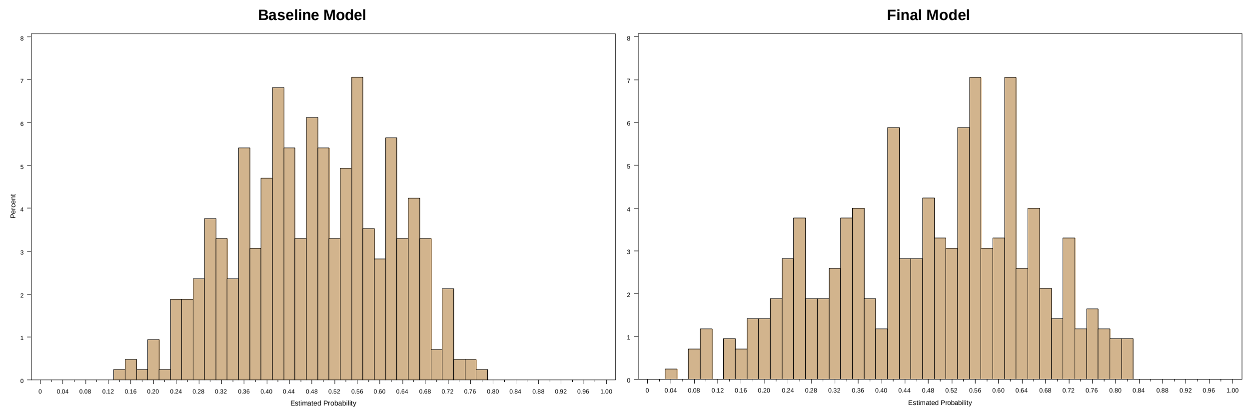

Figure 2.3. Histograms of Response Propensities from the Baseline and Final Models of Response to the Wave 2 Cell Phone Survey

One of the most influential device-related mechanisms related to nonresponse is whether or not the individual uses the cell phone as a primary channel of contact with the outside world. Specifically, those who say that they give their cell phone number when providing contact information on forms and applications were significantly more likely to respond to the cell phone survey as compared to those who do not give their cell number. This result appears to reflect a conceptualization of the cell phone as a broad access point, which promotes frequent answering of unfamiliar callers. People who give out their number probably expect that they will be called on their cell phone by strangers with whom they might wish to speak. Accordingly, they are probably more inclined than others to answer when the incoming number is unfamiliar or unavailable. Telephone survey researchers will clearly benefit as more and more people conceptualize their cell phone as a "public channel" as opposed to a private one. This positive effect may turn negative, however, if cell phones are eventually targeted for thousands of telemarketing efforts just as landlines are.

Another strong predictor of response is usage of non-telephony functions on cell phones, particularly text messaging, taking pictures or video, and connecting to the Web.

As expected, greater usage of non-telephony features is associated with greater propensity to respond, even when controlling for telephony usage. The items were scaled using equal weighting (Cronbach's alpha=.71). Watching video or television shows on the cell phone was also captured in the Wave 1 questionnaire, but this activity was quite rare; only 3% of the cell users do this weekly. It did not improve the reliability of the scale and was excluded.

There is evidence that non-telephony usage interacts in a fairly nuanced way with giving the number as contact information. Among those using their cell phone as a broad access point, the non-telephony behaviors are associated with lower response propensity. For those keeping their number private, by contrast, texting, taking pictures, and connecting to the Web are associated with higher response propensities to the cell phone survey. In other words, the response propensities of people keeping their cell number private are buoyed by texting, video capturing, and Web browsing. For people making their cell number public, by contrast, their response propensity is generally independent of non-telephony activities with their cell phone. One interpretation of this finding is that the critical factor is "total time spent per day using the cell phone." This quantity is probably already high for people using the cell as their primary telephone, so the other activities may have a relatively small marginal effect. Among those trying to keep their cell number private, however, the non-telephony activities may represent a substantial relative boost in the total amount of time spent using the phone.

The only other device-related mechanism to show a significant effect is using the cell phone only for making calls, not receiving calls. Based on all the cell phone users interviewed in Wave 1, 15% reported using their cell only for outgoing calls. This behavior clearly has a negative effect on response propensity, though this group was still somewhat accessible. Of the 52 dual users in the Wave 2 cell condition who reported that they use their cell only for outgoing calls, some 25% responded to Wave 2. This compares to a Wave 2 cell condition response rate of 51% for those saying that they use their cell for incoming and outgoing calls. This result echoes the finding by Keeter and his colleagues (2008) that a sizable number (47%) of the "cell-mostly" respondents in a cell RDD sample could have been reached if the interviewer had called them "just now" on their landline phone.

While several device-related mechanisms proved to be predictive of response to the cell phone survey, quite a few were not. When Wave 2 response is regressed on the usage variable alone, it is highly significant ($p<.01$), replicating the relationship between usage and response propensity documented by Brick and his colleagues (2006). In the final model, however, the effect from usage is explained better by the other device-related mechanisms in the model, and the usage variable becomes marginally non-significant ($p=.06$). Using the cell phone as a primary channel of contact with the outside world and using your cell for a range of activities appears to influence response propensities more than the frequency with which one answers calls. I was also interested in the effect that sharing a phone has on response propensity to a cell phone survey, but that too did not reach statistical significance in the final model. This null result suggests that cell phone samples probably do not need a non-response adjustment for shared/unshared phones, though they may need an adjustment for the differential probabilities of selection (Brick, Edwards, and Lee 2007).

Several other hypothesized mechanisms proved to have essentially no influence on response and were dropped from the final model presented in Table 2.2. These variables are mastery of cell phone technology, using the cell phone to fill free time, and concern about the cost of the call. Just being adept at using cell phones or using them for unessential calls do not appear to have much bearing on response behavior to surveys. Cost concern did not show any effect in this study, but it is important to note that this test was flawed because an incentive was given to those participating on their cell phone. Offering the incentive may have suppressed concern that may otherwise be present in a cell phone surveys. Given this, one cannot conclude that cost concern is always independent of response decisions.

As with the baseline model, several variations of the final model were tested to determine how the predictors perform when the sampling frame indicator is dropped and, separately, when weighting is used to adjust for Wave 1 nonresponse. In both cases, the results are highly similar to those presented in Table 2.2. The only difference is that race is not statistically significant when the weight is applied. This reflects the fact that race is one of the demographic dimensions that is adjusted for in the weighting .[8]

One concern with testing numerous independent variables is the potential for more than one indicator to be measuring the same construct. For example, answering incoming calls frequently and using the cell phone for outgoing calls only are closely-related behaviors. If two or more independent variables are highly correlated with each

[8] As discussed in the methods section, it was necessary to follow-up with some Wave 1 respondents in order to persuade them to report their cell phone or landline number. Tests were conducted to determine whether needing the follow-up effort was related to Wave 2 response propensity. In short, no relationship was found. Among those not requiring the follow-up, the Wave 2 response rate was 51%, which compares to 52% for those requiring the follow-up. This null result holds when the comparison is done separately for each sampling frame. When the final model (Table 2.2) is re-estimated to include an indicator for the follow-up effort (yes/no), the effect associated with follow-up indicator is not significant (p=.45).

other, the reliability and predictive power of the model as a whole is not reduced, but estimates of individual coefficients may be erroneous. Three checks were performed to detect multicollinearity. The first check was to look for unusually large standard errors associated with the estimated coefficients. As shown in Table 2.2, no particularly large standard errors are present. The second check was an examination of the correlation matrix for the independent variables. Again, no compelling evidence of multicollinearity was found. The third check was to review the variance inflation factors (VIF). This was performed because it is possible to have sets of variables that are highly interdependent, even when no pair of variables has a strong correlation. The VIF values were computed using weighted least squares regression. In the logistic regression setting, a general rule is that VIF values greater than 2.5 are thought to signal multicollinearity, while lower values signal that multicollinearity is probably not present (Allison 1999). The VIF values for the coefficients in the final cell phone response model range from 1.06 to 1.64, indicating that the model is not compromised by this issue.

2.3.5 Modeling Nonresponse to the Landline Survey

Based on the literature, it is not clear whether the factors driving response propensities in cell phone surveys differ from those in landline surveys, although the analysis above suggests some differences. To further explore this issue, I examined the mechanisms predicting response to the landline component of Wave 2. The models above focused on the 415 dual users randomly assigned to be contacted for Wave 2 on their cell phone. In this section, analysis is based on the 270 dual users called for Wave 2 on their landline.

Rather than imposing a model from the cell phone data, I constructed a new model that best fits the landline data (Table 2.3). Not surprisingly, most of the strong predictors of response to the landline component are measures of social psychological characteristics, as opposed to measures of cell phone usage. Unfortunately, there was not enough space in the questionnaire to collect measures of landline usage analogous to those collected for cell phone usage. For example, studies suggests that had interviewers asked about caller ID and answering machines, the models would indicate that usage of those technologies predicts response propensity (Link and Oldendick 1999; Tuckel and O'Neill 1996).

A major difference between the final cell phone model and the landline model is the absence of an effect from the sampling frame in the landline model. The frame indicator was not significant (p=0.95) and was dropped from the model. This result corroborates other research suggesting that many "cell-mostly" adults are still accessible through their landline phones. The large effect from frame in the cell phone model suggests that the complement is not supported; "landline-mostly" adults do not appear to be as accessible through cell phone surveys. This makes sense considering that many landline-mostly people probably turn their cell phone off at times, but rarely does anyone (even cell-mostly adults) turn off their landline.

Table 2.3. Logistic Regression Model Predicting Response to the Wave 2 Landline Survey

Parameter	Estimate	s.e.
Intercept	0.44	(0.93)
Hostility to surveys scale	-0.30*	(0.13)
Age	0.28**	(0.09)
Black	-0.52*	(0.23)
Education	0.27***	(0.08)
Give cell number as contact info	-0.36*	(0.15)
Outbound calling only	0.40	(0.23)
Area under ROC curve (c)	0.74	
-2 Log Likelihood	322.9	
Re-scaled r-square	0.21	

***p <.001 **p <.01 *p <.05

Note.- Figures are based on the 270 cases randomly assigned to be contacted for Wave 2 on their landline.

Another notable difference from the final cell phone model is the powerful effect of education. Education was not a predictor of response to the cell phone survey, but it is the strongest predictor of response to the landline survey, based on standardized coefficients. The exact explanation for this finding is unclear. Due to randomization, the overall distribution of education level for the Wave 2 cell phone sample was statistically equivalent to the education distribution for the Wave 2 landline sample (Chi-square=6.94, df=6, p=.33). Bivariate analyses show that the differential effect from education can be traced to the most highly educated. People with graduate degrees were extremely likely to participate in the Wave 2 landline survey (80%), but they were about average in their propensity to participate in the Wave 2 cell phone survey (49%). One post-hoc explanation is that well-educated Americans are aware that many legitimate scientific surveys are fielded using landline surveys, but they are unaware that such research is now being conducted with cell phones. Thus, the tendency for the highly-educated to support survey research may be dampened in cell phone surveys, due to lack of familiarity with the method and corresponding disbelief that cell phone survey calls are legitimate. This explanation is only speculative and would benefit from empirical testing.

Two device-related mechanisms also appear to contribute to response decisions in landline surveys. Providing a cell phone number as contact information was positively associated with response to the cell phone component, and it is negatively associated with response to the landline component – a logical reversal. Similarly, only using the cell phone for outgoing calls is negatively associated with cell phone survey response, and it is positively associated with landline survey response, though the effect falls short of statistical significance (p=.08). These two results indicate that telephone usage and survey response propensity is something of a zero-sum game for many people. They may have both devices, but they are more likely to participate in a survey when reached on the device that they use most often.

Overall, there is evidence that the factors influencing participation decisions in cell phone and landline surveys are similar but not identical. Major social psychological forces tend to work the same way with either device; older people and Caucasians are more likely to participate in both landline and cell phone surveys. The way that people use their phone(s) also matters. The device that people most commonly use to communicate with the outside world has a sizable positive effect on response propensity, as does proclivity to use the device for non-telephony activities.

2.3.6 Decomposing Nonresponse into Noncontact and Noncooperation

In this section, I return to the Wave 2 cell phone cases. Some mechanisms tested in this study were expected to affect contact propensity while others were expected to affect cooperation decisions (see Table 2.1). The distinction is unimportant for adjustment protocols, but it may be important from an operational or response rate improvement standpoint. I estimated separate logistic regression models for contact and cooperation using the independent variables in the final model discussed above.[9] The results are presented in Table 2.4.

There is no overlap between the set of factors predicting contact and those predicting cooperation. The likelihood of contact is lower among people who only use their cell for outgoing calls. Cooperation, by contrast, is predicted by age, race, attitudes about surveys, using the number as a primary contact channel, and using the phone for non-telephony activities. Interestingly, the sample frame indicator is a significant predictor of cooperation (p<.001) but not contact (p=.19).

[9] Non-contact cases had a final disposition of "No answer/Busy", "Voicemail", or "Other non-contact." Non-cooperation cases had a final disposition of "Callback", "Hung up during introduction/Refused", or "Interrupted."

Table 2.4. Logistic Regression Models for Nonresponse, Noncontact, and Noncooperation in the Wave 2 Cell Phone Survey

Parameter	Response model estimate	Contact model estimate	Cooperation model estimate
Intercept	-3.98**	-3.05	-2.93*
Sampled from landline RDD frame	-0.39***	-0.23	-0.40***
Hostility to surveys scale	-0.23*	0.25	-0.33**
Age	0.26**	0.25	0.24**
Black	-0.52*	0.22	-0.60**
Give cell number as contact info	1.03**	0.95	0.99**
Non-telephony usage scale (text, picts, web)	0.52*	0.36	0.51*
Outbound calling only	-0.45*	-0.65**	-0.31
Answer cell most/all of the time	0.32	0.30	0.26
Share cell with someone else	0.24	-0.07	0.29
Non-telephony scale x Give cell as contact info	-0.17**	-0.12	-0.17*
Sample size	415	415	373^
Area under ROC curve (c)	0.69	0.73	0.68
-2 Log Likelihood	523.1	242.2	470.4
Re-scaled r-square	0.16	0.14	0.15

****p* <.001 ***p* <.01 **p* <.05

Note - All models are based on the cases assigned to be contacted for Wave 2 on their cell phone.

^Based on cases contacted for Wave 2.

One takeaway from Table 2.4 is that the traditional nonresponse literature offers little guidance for understanding what factors influence contact in cell phone surveys. The sole significant predictor observed here relates to the data collection device itself and how it is used. For survey designers, establishing contact with people who only use their cell phone for outgoing calls may require supplementing cell phone surveys with other sampling frames (e.g., landline RDD or USPS Delivery Sequence File). More research is needed to determine, however, whether the failure to reach these "outgoing only" people yields enough nonresponse bias to justify the expense of recruiting them through another frame.

Cooperation, by contrast, appears to be influenced by several of the same social psychological factors observed in other modes, as well as device-related behaviors.

There is no obvious design strategy specific to cell phone interviewing that researchers can use to overcome the social psychological influences on response decisions. Device-related factors, by contrast, may become less important influencers on response over time. Cell phones are still a relatively new technology. As their usage become more common, some of the current differences between heavy users and light users may dissipate.[10,11]

2.3.7 *Effect of Reweighting on the Estimates*

From a practical standpoint, it is useful to know whether adjusting for the nonresponse mechanisms actually reduces nonresponse bias. No data on all sample members, such as those employed in record check surveys, are available for this study, so direct computation of bias was not possible. It is possible, however, to compare unadjusted estimates to estimates adjusted for the differential response documented above. If there is a material difference between the adjusted and unadjusted estimates, this would indicate an association between the nonresponse mechanisms and outcomes measured in Wave 2. If there are no differences, this would indicate that the survey outcomes are independent of the factors influencing response propensity.

[10] Another way to conceptualize this analysis is to study the three outcomes (noncontact, noncooperation, and nonresponse) simultaneously using multinomial regression. I specified noncontact as the reference group and estimated the effects from the independent variables in the final model. The substantive results do not change dramatically. When the probability of cooperation is considered relative to the probability of noncontact, most of the variables from the cooperation model in Table 2.4 remain significant at the $p<.10$ level (frame, age, giving cell number as contact information, non-telephony usage, and the interaction term). In addition, outbound-only calling and the frequency of answering show significant effects. In the model estimating the probability of contact versus noncontact, the significant predictor from the contact model in Table 2.4 (outbound-only calling) retains its effect, and the hostility to surveys scale shows a significant positive effect.

[11] Separate models predicting cooperation and contact were also estimated using the Wave 2 landline cases. In the contact model, only the intercept was significantly different from zero. In the cooperation model, age and education were statistically significant positive predictors, while African American race, providing the cell number as contact information, and hostility toward surveys were significant negative predictors.

Table 2.5 presents three sets of estimates for questions administered in Wave 2. The first column reports the unweighted estimates. The results in the middle column are weighted and intended to represent the current practice of several survey organizations conducting dual frame RDD surveys.[12] The weight for the middle column was created by raking the Wave 2 sample to population distributions for gender, age, race, ethnicity, and region. The Wave 2 sample consists entirely of dual users, and so I used demographic benchmarks for dual users in the U.S., which describes about two-thirds of the adult population (Blumberg and Luke 2009). The dual user benchmarks were computed from the most recent, publically available National Health Interview Survey (NHIS) data set (2008). The far-right column of Table 2.5 reports estimates adjusted not just for common demographic variation, but also for the mechanisms modeled above. This weight was created in stages. For cases in the Wave 2 cell phone condition, I estimated their response propensity from the final model in Table 2.2. For the cases in the landline condition, response propensity was estimated using the landline response model in Table 2.4. Thus, for each Wave 2 sample cases I computed both an estimated response propensity and a weight factor from the demographic raking. The weight used in the far-right column is the product of the demographic weight factor and the inverse of the propensity score.[13]

[12] There is substantial variation in how survey organizations weight data from dual frame RDD surveys. Treatment of within-household selection probabilities differ, as do the set of variables used for post-stratification. The weighting approach employed here treats cell phones as a person-level device and landlines as a household level-device.

[13] Researchers differ in their procedures for developing nonresponse adjustments. Brick and his colleagues (2006) created a weighting class adjustment for nonresponse based on region and number of attempts and then raked the adjusted data to demographic controls. Abraham, Maitland, and Bianchi (2006) multiplied a pre-existing demographic adjustment by the inverse of the estimated response propensity. The types of data available for modeling nonresponse in this study are akin to those available to Abraham and her colleagues, and the procedure used here resembles theirs.

The outcomes reported in the table represent every closed-ended Wave 2 question that was not part of an experiment. Experimental items were excluded because they featured special question wording manipulations designed to test measurement error hypotheses (see Chapters 3 and 4). Fortunately, the set of non-experimental items is quite diverse. Questions in Table 2.5 probe attitudes about the economy, the space program, confidence in institutions, and affinity for common recreational activities. Certainly not all survey topics are represented, but this set covers several common survey themes. This diversity is useful because of the statistic-specific nature of nonresponse bias.

Table 2.5. Effects of Alternative Weights on Wave 2 Survey Estimates

Question	Unweighted	Weight Used for Estimates: Demographic Raking	Weight Used for Estimates: Demographic Raking and Propensity Adjustment
Are you satisfied or dissatisfied with the way things are going in this country today?			
(Dissatisfied)	73.6% (2.3)	74.2% (2.4)	73.1% (2.8)
Would you say there are plenty of jobs available in your community or are jobs difficult to find?			
(Difficult to find)	80.8% (2.1)	82.2% (2.0)	83.7% (2.2)
How would you rate economic conditions in this country today?			
(Poor)	63.9% (2.5)	63.3% (2.7)	63.8% (3.0)
…Please tell me how much confidence you have in that institution using a scale from zero to 10, where 10 means "great confidence" and zero means "no confidence at all." (Mean)			
a. The public school system	5.55 (0.12)	5.64 (0.12)	5.63 (0.13)
b. The criminal justice system	5.27 (0.12)	5.39 (0.13)	5.46 (0.14)
c. The space program	6.31 (0.12)	6.39 (0.13)	6.38 (0.16)
d. The military	7.86 (0.10)	7.87 (0.11)	7.87 (0.12)
e. Congress	4.17 (0.13)	4.30 (0.14)	4.41 (0.16)
f. The presidency	6.48 (0.17)	6.58 (0.18)	6.62 (0.21)
Please tell me whether you like doing this a great deal, like doing it somewhat, dislike doing it somewhat or dislike doing it a great deal. (Like a great deal)			
a. Reading a newspaper	36.4% (2.5)	34.7% (2.6)	32.5% (2.9)
b. Watching a movie	46.4% (2.6)	46.5% (2.8)	48.5% (3.2)
c. Cooking a meal at home	42.5% (2.6)	42.2% (2.8)	44.9% (3.2)
d. Reading a book	55.3% (2.6)	55.6% (2.8)	53.8% (3.2)
I'm going to read some statements about the U.S. space program… (Agree)			
a. The program is important to national security	70.8% (2.4)	69.6% (2.7)	68.0% (3.1)
b. " " does a good job keeping astronauts safe	77.8% (2.2)	77.4% (2.4)	76.5% (2.8)
c. " " needs more funding to improve its technology	50.3% (2.6)	50.7% (2.8)	49.7% (3.2)
d. " " is a waste of taxpayers' money	13.9% (1.8)	13.5% (1.9)	14.3% (2.2)
e. " " contributes to American pride and patriotism	79.7% (2.1)	80.3% (2.2)	78.9% (2.6)
f. " " inspires young people to study science and math	82.2% (2.0)	82.4% (2.1)	83.4% (2.3)

Figures are based on Wave 2 respondents (N=360). The percentages shown are for the modal category. The standard errors, computed using SURVEY procedures in SAS, are shown in parentheses.

Several of the estimates change noticeably when an adjustment is made for the nonresponse mechanisms that go beyond demographic correlates in the literature. That is, several estimates in the last column (with the propensity adjustment) differ not only from the unweighted estimates but also from the estimates in the middle column, which reflect one current practice of weighting to several demographic dimensions. The propensity adjusted estimate for the proportion who like reading newspapers a great deal is roughly four percentage points lower (32.5%) than the unweighted estimate (36.4%) and two percentage points lower than the weighted estimate based on demographic raking (34.7%). The propensity adjusted estimates for affinity for watching movies and cooking at home, however, are at least two percentage points higher than the estimates based on no weighting or demographic weighting only. Finally, the propensity adjustment yields estimates reflecting more negative evaluations of the job market and less agreement that the space program is important to U.S. national security. Benchmarked against the unweighted results, estimates change by an average of 0.6 percentage points when the demographic raking weighted result is used, and they change by an average of 1.6 percentage points when the demographic raking and propensity adjusted weighted result is used. These differences are generally larger than those reported by Abraham, Maitland, and Bianchi (2006) in their analysis of nonresponse adjustments for the ATUS. This difference across studies is likely related to the fact that the nonresponse rate in the ATUS is lower than in the present study. By extension, the potential for bias was smaller in the ATUS.[14]

The propensity adjustment appears to work differently than the demographic weighting for a number of estimates. To quantify this, consider the difference between a given weighted estimate and the corresponding unweighted estimate. For seven of the 19 items (37%), the direction of the difference was different for the propensity-adjusted estimate and the estimate based on just the demographic raking. For example, the demographic weighting pushes the estimated proportion of people dissatisfied with the

[14] Neither independent samples nor paired samples difference in proportions tests are appropriate for the comparisons in Table 2.5 because estimates are based on a single set of observations. Regardless, if independence is assumed and the tests are performed, then none of the differences in estimates is statistically significant. This is not surprising given that the differences are relatively small in comparison with the sampling errors shown in parentheses.

way things are going in the country slightly higher, while the combined demographic and propensity weighting pushes it slightly lower. The magnitude of these differences is not large, but the inconsistency in the direction of the two different weights suggests that the device-related factors relate differently to the outcome variables than the set of demographic factors.

Overall, the results from the weighting analysis are mixed. There is some indication that adjusting for the device-related mechanisms (in addition to adjusting for demographics) could reduce bias. Some propensity weighted figures were as much as four percentage points different from the unweighted figures. It is also the case that the propensity adjustment moved estimates in the opposite direction from the demographic adjustment, for several outcomes. That said, for some research endeavors the magnitudes of the differences between the weighted estimates may not be of great consequence. It is also true that benchmarks are not available for this testing, and so it is inappropriate to draw definitive conclusions as to whether one of the weighting protocols outperformed the other in terms of reducing bias.

2.4 Summary and Conclusions

Factors that influence survey response propensity in cell phone surveys may differ systematically from those in landline surveys, due to operational and sociological differences between the two devices. Identifying device-related mechanisms of response is important for anticipating (and avoiding) nonresponse bias in survey estimates. Prior studies documented a link between phone usage levels and response propensity (Brick et al. 2006), but do not elaborate on the nature of this dependency and whether it operates at the cooperation or contact stage. I designed a unique experiment to extend this line of research by testing a wide range of mechanisms – those suggested by the literature as well as newly conceptualized mechanisms related to the device.

In order to study influences on response behavior, I first needed to collect the relevant characteristics from the subjects. This was accomplished through a preliminary survey featuring samples from both the landline and cell RDD frames. The use of national RDD samples (as opposed to listed samples) is one of the strengths of this design. The dual users identified in the preliminary survey were randomly assigned to be contacted for a second, ostensibly unrelated survey on their landline or their cell phone. This randomization step is new to cell phone research and is critical to fleshing out the connection between sampling frame and response propensity. Several conclusions can be drawn from the results.

1. *The sets of factors influencing response to landline and cell phone surveys appear to be different, though overlapping.*

The best-fitting model for response to the landline component of the study differed somewhat from the best-fitting model for the cell phone component. In particular, education has a strong positive effect on landline response but no effect on cell phone response. Age and race work essentially the same for both devices, but the effects from device-related mechanisms change direction. Using the cell phone as a primary channel of contact with the outside world is negatively associated with landline survey response and positively associated with cell phone survey response.

2. *Cell-mostly adults can be reached effectively through the landline frame. By comparison, landline-mostly adults are less accessible through the cell frame.*

One of the strongest effects documented in this study is that dual users sampled through the landline RDD frame are *less* likely to respond to a cell phone survey than a landline survey. By contrast, device does not appear to matter to the dual users sampled through the cell RDD frame. This effect persists even in the presence of powerful demographic and device-related controls. I suspect that the sampling frame indicator captures the effect of unobserved factors related to differential device usage, but this is only a post hoc explanation.

3. *How people use their cell phone (device-related mechanisms) has a strong influence on response propensity – perhaps comparable to the strength of effects from social psychological characteristics.*

The addition of device-related variables significantly improved the explanatory power of models predicting response to the cell phone component of the study. A baseline model tested mechanisms discussed in the literature. Demographic variables served as proxies for social psychological factors. Positive effects from age and Caucasian race documented in surveys using all major modes, replicated here for cell phones. When a set of device-related measures was added to the model, several proved to be highly significant. The strength of the standardized effects from the best device-related mechanisms is comparable to that from race and larger than those from other social psychological indicators (e.g., education, household size, marital status).

4. *Device related mechanisms and social psychological mechanisms tend to influence cooperation rather than contact.*

All of the social psychological mechanisms and most of the device mechanisms showed effects for cooperation but not contact. The only factors associated with contact were behaviors related to device. People who only used their cell phone for outbound calling and do not typically provide the cell number as contact information were significantly less likely to be contacted than other cell phone users.

5. *Weighting procedures that adjust for the observed device-related mechanisms may yield bias reduction beyond that achieved through normal demographic raking.*

I constructed two sets of weights to adjust for nonresponse: (a) a demographic raking weight and (b) a weight that incorporates the demographic raking and the differential response propensities modeled using all of the significant mechanisms observed. Estimates in the second survey change by an average of 0.6 percentage points when the demographic raking weighted result is used (compared to the unweighted figures). Estimate change by an average of 1.6 percentage points when the demographic raking and propensity adjusted weighted result is used (compared to the unweighted figures). In addition, the two weights shift the unweighted figure in different directions for more than one-third of the estimates. I cannot measure bias directly to determine which weight is better, but the results suggest that some measures may still contain nonresponse bias even after demographic raking. This analysis gives some insight to how nonresponse mechanism may be related to bias, but unfortunately the propensity adjustment used here is not available in one-off cross-sectional surveys. This means that researchers must either develop additional population benchmarks or other adjustment techniques to reduce bias above and beyond the reduction accomplished through post-stratification to demographics and telephone service.

I posit a three-tiered model (Figure 2.1) for conceptualizing factors that influence cell phone survey participation. Ecological factors such as the volume of surveys and unsolicited requests in society reside at the highest level. These factors may influence norms for how people react to unsolicited requests and the potential concerns that come to mind. Features of the survey made salient to the respondent comprise the middle tier. Myriad studies have documented the effects on participation from incentives, topic, sponsor, calling protocols, and interviewers. The lowest tier represents characteristics of the sample person. These characteristics include valuation of the salient survey features and social psychological factors, such as social integration and civic engagement. This study was designed to verify whether or not device-related behaviors and attitudes also belong in the sample person-level of the model for cell phone surveys. The results suggest rather strongly that they do. Brick and his colleagues (2006) provided the first empirical evidence for this, and the present study extends that work.

Central to this endeavor was testing as many of the known nonresponse mechanisms as possible in a multivariate setting. One limitation of previous nonresponse studies is that they tend to focus on just one mechanism (e.g., incentives or topic interest), making it difficult to gauge the power of the effect relative to other factors. The study at hand was fairly successful in exploring simultaneous effects, but there were several limitations. In particular, it was not feasible to measure respondent valuations of the surveys features, such as affinity for the sponsor and interest in the topic. I had planned to measure topic interest, but the topic of the second survey changed after the Wave 1 data collection for client-related reasons. I intentionally avoided measuring sponsor affinity because it was more important to have the second survey appear unrelated to the first, so as to create conditions similar to a one-off survey. Other mechanisms were excluded only because there was not enough space in the questionnaire. These include population density (ZIP code), personality type, and size of social circle (Vehovar and Callegaro 2007). As discussed in section 2.1.5, the study is also limited because it measures nonresponse to a follow-up survey rather than to a true one-off survey. I acknowledge that this may influence the results somewhat, but this is an unavoidable consequence of the experimental design. It is encouraging to know that adjusting for nonresponse to the first survey does not change the results in any meaningful way.

This study has several implications for the design of recruitment protocols and nonresponse adjustments, though additional research would be useful. One intriguing result is that device-related behaviors appear to influence cooperation, not just contact. It may be fruitful to experiment with new recruitment language addressing this. Having interviewers mention the difficulty of reaching some people on their landline, for instance, may help increase cooperation in cell phone interviews (and vice versa in landline interviews). This idea could be pursued in future research. With respect to nonresponse adjustment, the ideal information would be the population parameter for thinking about the cell number as public or private information. This attitude appears to explain more variance in response behavior than a sheer usage measure, and could theoretically be used as a post-stratification dimension. Currently, however, the population distribution for this variable is unknown. No studies have shown a benefit in terms of bias reduction from raking to benchmarks defined by usage measures (as in the NHIS), but this approach should be considered in future studies as telephone behaviors continue to evolve. Results from this study confirm that how the cell phone is used strongly predicts response propensity, so it seems likely that future studies may find that incorporating usage into weighting adjustments helps to reduce bias.

Chapter 3

Use of Cognitive Shortcuts in Landline and Cell Phone Surveys

3.1 Introduction

In 2003, survey researchers in the U.S. began supplementing landline RDD surveys with samples of cell phone numbers. This practice improved coverage of the general public, but it also raised concerns about the consequences for measurement error. Researchers have speculated that people are more likely to respond inaccurately when interviewed on a cell phone as compared to a landline (Steeh 2004; Lavrakas et al. 2007), but empirical studies have found essentially no meaningful evidence to support this hypothesis (Steeh 2004; Brick et al. 2007; Kennedy 2007a; Lavrakas, Tompson, and Benford 2009; Witt, ZuWallack, and Conrey 2009). These studies have been limited, however, with respect to experimental design and the outcome variables assessed. This chapter reports the results of a unique experiment designed to address these challenges and generate new knowledge about the possible effects of telephone device (cell/landline) on measurement error.

The hypothesis that respondents may respond less accurately on a cell phone than on a landline is based on several observations. First, people often multitask while talking on their cell phone. The same is true for landlines, but the cognitive demand of the competing activities (e.g., traveling, shopping) may be greater with cell phones due to the freedom of movement that the technology affords. People conversing on a cell phone can move about anywhere that they have coverage, unlike landline users who generally must remain within a certain radius from the phone's main unit. Landlines that are not wireless feature a cord, tethering the user even closer to a fixed point in the household. Also, many cell phone users use hands-free headsets, such as Bluetooth, freeing their hands for use in other tasks. Anecdotal evidence suggests that use of a hands-free headset is more prevalent for cell phones than landlines, but I was not able to find authoritative data to verify this. Headset or not, intensive multitasking may distract people from the task of responding. Second, during the early adoption of cell phones, the sound quality of calls was often inferior to that of landline calls. While cell phone audio fidelity has improved over time, there may still be a difference for some people between the clarity they have on one device versus the other. Degraded sound clarity could make it more difficult for respondents to hear the question wording and for interviewers to record their responses accurately. Third, people on their cell phones may be exposed to more distracting environmental stimuli, such as in-store announcements or obstacles along one's path in transit. Distractions certainly exist in the home as well, but the assertion is that they may be more severe away from home. Similar to multitasking, environmental stimuli may distract respondents from the task of responding and increase the likelihood of misreporting. Finally, people reached on a cell phone may be out in public and feel as though they have less privacy than people interviewed at home on a landline (though other household members may be present in landline interviews as well). A lower sense of privacy in cell phone interviews may lead to more censoring of responses to sensitive questions.

Together, these considerations represent four possible mechanisms of differential measurement error between landline and cell phone interviews. One observation about the mechanisms is that they may impair the response process in different ways. For example, distraction from multitasking may lead respondents to one type of error (e.g., incomplete memory search), while poor sound quality may lead to an entirely different error (e.g., answering a different question than what was asked). The next section discusses connections between the hypothesized mechanisms and specific stages of the response process.

3.1.1 Implications for the Cognitive Response Process

One piece that is absent from the cell phone survey literature is a discussion of how the hypothesized mechanisms (more intensive multitasking, lower audio fidelity, greater environmental distractions, and more censoring) might interfere with specific stages of the response process (Tourangeau 1984). Making this connection would be helpful in identifying approaches for reducing error associated with any observed device effect.

The effects of multitasking and ecological distractions would probably be similar. Both mechanisms imply that while the respondent is answering questions, he or she is also reacting to something else. The cognitive demand from processing the competing thoughts could interfere with each of the stages in the response process. Respondents may fail to comprehend a question as intended if the distraction prevents them from processing the entire question. This may be particularly problematic for questions with important inclusions or exclusions that dictate how one should respond. Distractions may undermine the retrieval stage by leading the respondent to employ a less taxing, less precise strategy for pulling considerations for memory. For example, respondents may decide simply to estimate the number of alcoholic drinks that they consumed in the past week, as opposed to recalling each drink and summing. Alternatively, distractions could lead respondents to end their retrieval process sooner than had there not been other demands on their mental processing at that moment. At the judgment stage, distractions may hasten the respondent's effort to integrate all of the considerations that were retrieved from memory. Partial integration, rather than full integration, could lead to an inaccurate response. At the last stage, distracted respondents may fail to optimally map their judgment onto the response scale. Imagine trying to articulate approval of the U.S. president using a 101-point thermometer scale while searching for a city on a map or maneuvering a cart through a crowded grocery store. Respondent distraction, whether it be from multitasking or the surrounding environment, poses a potentially serious threat to the accuracy and reliability of survey data.

I expect that poor audio quality and lack of privacy would work differently. At the comprehension stage, poor sound may impair the respondent's ability to hear the question. If respondents do not request that the question be repeated, they may answer based on a false understanding of the information requested. Also, low audio fidelity may impair the interviewer's ability hear the respondent's answer. The respondent may answer accurately, but the response recorded may be erroneous. Censoring responses for fear of being overheard is specific to the final stage of the response process. Respondents know what the correct answer is, but they edit that answer and ultimately report something else. Figure 3.1 illustrates my expectations for how the hypothesized measurement error mechanisms interfere with various stages of the response process.

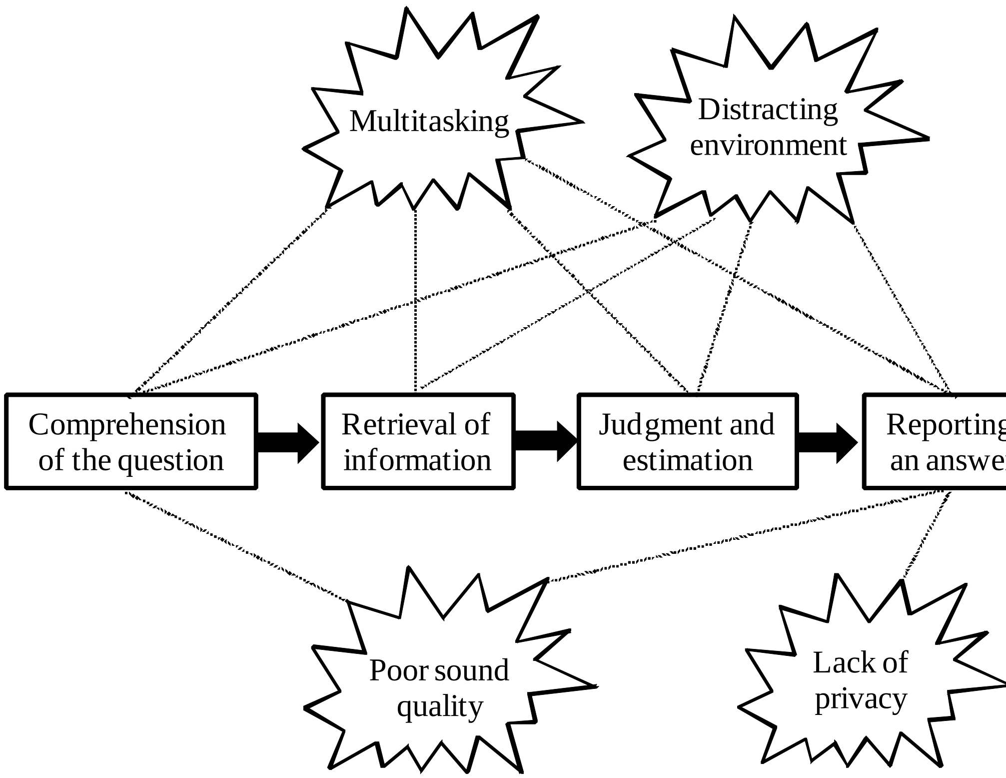

Figure 3.1. Hypothesized Cell Phone Device Effects Interfering with the Response Process
Note – The model of the response process is borrowed from Tourangeau, Rips, and Rasinski (2000).

3.1.2. Multitasking as a Threat to Performance

Of the four mechanisms, multitasking may be the most pernicious. The sound quality of cell phone calls will most likely continue to improve and serious environmental distractions to respondents are potentially rare, but multitasking during survey interviews is already common. In a 2004 survey, over one-quarter of cell phone respondents reported that they were either driving or at work (Brick et al. 2007). Driving while talking on a cell phone is dangerous with or without a hands-free phone because the conversation is at least as distracting as manipulating a handheld phone (Strayer, Drews, and Crouch 2006).

Psychologists and human factors engineers have been studying multiple task performance for several decades, yielding a substantial body of knowledge. These literatures are relevant because they detail how multitasking impairs performance on the task of interest. Four findings are particularly relevant: People process multiple tasks sequentially through rapid toggling, rather than simultaneously (Jersild 1927); tasks differ in their difficulty and, thus, their level of interference (Norman and Bobrow 1975); low-level multitasking can actually improve performance (Rauscher, Shaw, and Ky 1993; Thorne and Debener 2008); and the very young and the very old are less capable of multitasking than others due to physiological changes to the brain during the life cycle (Kray and Lindenberger 2000; Reimers and Maylor 2005).

People can perform multiple familiar activities, such as walking and dialing a phone, simultaneously with relative ease. These highly practiced skills can be done while thinking about other things. Action planning, by contrast, is usually performed sequentially. Unfortunately, formulating a response to a cognitively demanding question is more akin to action planning than to executing a highly practiced skill. Studies have shown that when people attempt to perform multiple complex tasks at the same time, performance deteriorates with concomitant increases in the error rate and the time needed to complete the tasks (Jersild 1927; Spector and Biederman 1976). The implication is that if respondents are engaged in another complex task while answering a difficult question, both their efficiency and accuracy will suffer. Presumably, the magnitude of the error depends on whether the respondent considers the survey interview to be the primary task or the secondary task.

Psychologists have not converged on a particular model to describe how people perform concurrent tasks (for a review, see Meyer and Kieras 1997), but there is broad support for the importance of interference (Broadbent 1957; Kahneman 1973; Wickens 1980; Friedman et al. 1982). Tasks interfere with each other if they compete for a limited resource, namely activation of the same part of the brain (Klingberg 1998). It is not clear from the literature whether interference jeopardizes some stages of the response process (comprehension, retrieval, judgment, mapping/reporting) more than others. Presumably, respondents may switch back and forth between formulating a response and a competing task (e.g., composing an e-mail message) during *any* of the four steps in the Tourangeau model.

If this is true, then the greatest potential for error could well be during the comprehension stage, as the respondent listens to the question and interprets its meaning. The comprehension stage is particularly vulnerable because the temporal dynamic is unforgiving. If the respondent switches tasks while the question is being read, then a word or phrase may be missed and the entire question misinterpreted.[15] The other steps, by contrast, take place entirely in the mind of the respondent, allowing him or her to adjust for interference. Respondents can put the other steps (retrieval, judgment, reporting) "on hold" as they toggle back and forth to the competing task. The difference is that little, if any, progress in the formulation of an accurate response need be lost due to interference in the latter three stages. The respondent can begin a search of long term

[15] If the respondent asks the interviewer to repeat the question, then the problem would be repaired. Such requests are fairly rare in survey interviews, however (Schaeffer and Maynard 2002).

memory, switch to the competing task, and later resume the memory search. To be sure, multitasking during any part of the response process may increase the probability of misreporting. Distraction may be particularly harmful, however, to the comprehension stage.

Encouragingly, one set of empirical results on multitasking has positive implications for survey research. Researchers have shown that low-level multitasking can actually improve performance on tasks like answering questions (Rauscher, Shaw, and Ky 1993). For example, Thorne and Debener (2008) found that subjects performed an auditory frequency discrimination task more quickly when presented with uninformative visual stimuli, as compared to a control condition in which no visuals were presented. This suggests that browsing a magazine or flipping through television channels during a telephone interview may actually improve respondent performance. There is a threshold, however, beyond which the stimulus can be so engrossing that it impairs attention to the target task and undermines performance (for an example, see Furnham and Bradley 1997). The relationship between stimulus intensity and task performance is, thus, believed to resemble a bell curve with low to moderate stimulation yielding better performance than zero or high levels of stimulation (Wallis 2006). In explaining why low-level stimulation can enhance performance, researchers cite the priming of neurophysiological activities associated with the target task. [16]

In addition to the main effect from multitasking, researchers have documented a powerful interactive effect from age. The very young and the very old are less capable of

[16] A related finding is that listening to certain types of music primes abstract reasoning abilities (Rauscher, Shaw, and Ky 1993). Subsequent studies have found mixed support for the so-called “Mozart effect” (for a review see Chabris 1999), but this line of research does illuminate the potential for background sounds to improve performance on cognitive tasks.

multitasking than others due to physiological changes to the brain during the life cycle (Kray and Lindenberger 2000; Reimers and Maylor 2005). Studies using functional magnetic resonance imaging have shown that the switching of attention from one task to another occurs in the brain's prefrontal cortex (Koechlin et al. 1999; Dove et al. 2000). The prefrontal cortex allows humans to abandon a task momentarily and later resume the task at the same place, which facilitates multitasking (Dempster 1991,1992; Fuster 1980; Gilbert et al. 2006; Goldman-Rakic 1987). The age interaction occurs because the prefrontal cortex is one of the last regions of the brain to mature (for a review see Dumontheil, Burgess, and Blakemore 2008) and one of the first to decline with aging (Albert and Kaplin 1980; Daigneault, Braun, and Whitaker 1992; Dempster 1992). As a result, young children do not multitask well, and neither do most adults age 60 and older. A key implication for survey researchers is that elderly respondents may be more likely to misreport when multitasking than younger ones, assuming a constant level of multitasking intensity. If multitasking is a greater problem among cell phone respondents than landline respondents, there may be serious consequences for the comparability of data across devices.

3.1.3 Prior Studies on Cell/Landline Measurement Error Differences

Studies testing for an effect from device (landline/cell) on measurement error have not detected any meaningful differences, but they have also been limited in important ways. In this section I first summarize their results and then comment on their design.

Several teams of researchers have documented equivalent rates of item nonresponse in cellular and landline RDD interviews, after controlling for demographics (Brick et al. 2007; Steeh 2004; Witt, ZuWallack, and Conrey 2009). Witt and her colleagues (2009) report that item nonresponse was slightly more prevalent in cell phone interviews than landline interviews after controlling for age, gender, and race, but the effect was not statistically significant. Among dual users, Witt and her colleagues found no difference in item nonresponse propensity between people who predominantly use their landline and those who predominantly use their cell phone.[17]

Researchers have also documented parity between landline and cell interviews with respect to the length and richness of responses to open-ended questions (Brick et al. 2007; Witt, ZuWallack, and Conrey 2009). Brick and his colleagues analyzed responses to four sensitive questions (difficulty meeting monthly living expenses, voting in the 2000 and 2004 presidential elections, and frequency of reading the editorial page of a newspaper) in multilevel models controlling for age, sex, and home ownership, and they found no significant effect from device. Kennedy (2007a) investigated straight-lining and response order effects and found no significant differences between landline and cell phone data.

Each of these studies, however, has been observational in nature. Consequently, any effect (or lack of effect) from the device is likely to have been confounded with differences in the composition of the samples. This is a critical limitation because some of the variables known to be different in landline and cell phone samples (*e.g.*, age) are

[17] Lavarakas, Tompson, and Benford (2009) compare error indicators for cell phone respondents reached away from home versus at home. This is different from the studies discussed in the text, which compare landline and cell phone respondents. Lavrakas and his colleagues found no significant differences between cell phone respondents reached at home versus away with respect to item nonresponse, the strength of correlations between attitudes and demographics, or the amount of differentiation to battery questions.

also known correlates of cognitive shortcuts such as item nonresponse and non-differentiation of battery items (Bell 1984; Slymen et al. 1984; Frisina and Thomas 2008). Some researchers attempt to adjust for this by introducing demographic controls, but this approach rests upon the tenuous assumption that all factors confounding the sample comparison are included in the set of controls. A more rigorous approach to isolating device effects would be to randomize respondents to device in an experimental design. That way the comparison groups would be essentially identical (subject to random variation), except for the device on which they are interviewed. The study described below features just such a randomization component.

A second limitation of previous work is the nature of the dependent variables. Item nonresponse rates and the lengths of open-ended questions are important, but they are only indicators of measurement error, not the thing itself. Ideally, the researcher would compare the deviation between the survey responses and the corresponding "true values" for landline versus cell phone interviews. In RDD surveys, however, records containing true values are generally not available. For attitudinal questions, the existence of records is particularly unlikely, and the mere notion of a "true value" is ambiguous. Absent direct computation of measurement error, examining wider range of measurement error indicators would be useful. The more error indicators tested, the more confidence researchers can have in their assessment of the difference in data quality between landline and cell phone sample.

3.1.4 Framework for Investigating Measurement Error

Most of the proxy variables for measurement error in the cell phone literature are indicators of cognitive shortcutting or satisficing (Simon 1957; Krosnick 1991). Krosnick and others have shown that respondents vary in how much cognitive effort they are willing or able to expend in answering survey questions. Highly motivated or highly able respondents may "optimize" by carefully formulating an accurate response. Others with less motivation or less cognitive capacity may "satisfice," taking various shortcuts to complete the interview. If respondents are indeed more distracted when interviewed on a cell phone than on a landline, then presumably they make greater use of cognitive shortcuts when interviewed on a cell phone. Some of these indicators may also detect error from poor sound quality. I would not expect, however, indicators of cognitive shortcutting to capture the censoring of responses.[18]

The present study seeks to test a model of why respondents may be more likely to respond inaccurately on a cell phone as compared to a landline. The model begins by positing that there is a set of mechanisms intrinsic to the device that make the experience of answering questions on a cell phone different from answering questions on a landline. The mechanisms are (*i*) distraction from multitasking, (*ii*) poor audio quality, (*iii*) distraction from environmental stimuli, and (*iv*) pressure to censor socially undesirable responses. The next step in the model posits that these mechanisms are causal to respondents using cognitive shortcuts. For example, a respondent in a noisy store may be less likely to internalize an important caveat to a question, as compared to someone in a quiet setting. The final aspect of the model is that cognitive short cuts are indicators of

[18] A study by Holtgraves (2004) suggests that response times may be useful for evaluating respondents' social desirability concerns.

data quality, specifically measurement error. Using this framework, I investigate the effect from device on measurement error by testing for differential use of cognitive shortcut among those interviewed on a landline versus those interviewed on a cell phone. By also measuring the mechanisms themselves, it should be possible to establish the link between the device properties and the error, if it exists.

3.2 Research Methods

3.2.1 Experimental Conditions and Sampling Frames

The data for this study come from the repeated-measures experiment described in section 2.2.1. The purpose of the second survey, Wave 2, was to re-interview respondents after randomly assigning them to be interviewed for Wave 2 on their landline or their cell phone. The randomization step prior to Wave 2 was necessary to minimize the confound between sample composition and device, which limits the analytic power of previous cell phone studies. In prior studies, the cell and landline comparison groups differed with respect to more than just device; they also differed on age, gender, race, and possibly other characteristics that affect responses but are not readily captured in statistical controls.

As shown in Figure 2.2, there was a substantial amount of attrition in the experiment. Most of the attrition was by design (landline-only and cell-only cases were dropped prior to Wave 2 in order to keep costs down), but some attrition also occurred from nonresponse and noncooperation with reporting telephone numbers. I conducted a panel mortality analysis to investigate the potential for these sources of attrition to bias the tests reported below (see section A.4 in the Appendix). Three data quality indicators were computed from the Wave 1 data: item nonresponse rates, non-differentiation to a

Wave 1 battery, and estimating responses to behavioral frequency questions. The respondents who completed both waves of this study were slightly less prone to taking cognitive shortcuts, but there is no indication that this pattern undermines the device comparisons reported below.

3.2.2 Survey Instruments

The questionnaires[19] were designed to measure two types of phenomena: (1) mechanisms that may impair the response process in cell phone surveys and (2) the result of that impairment, specifically evidence of respondents using cognitive short cuts. The four mechanisms evaluated in this study are: being away from home in a distracting environment, engagement in other activities during the interview (multitasking), use of a hands-free headset, and the audio fidelity of the call.

At the end of both the Wave 1 and Wave 2 interviews, respondents were asked "Would you mind telling me if I reached you today AWAY from home or AT home?" This was followed by a pair of questions designed to measure multitasking. Interviewers recorded verbatim answers to the questions "What did you happen to be doing when I called?" and "Some people multitask when on the phone while others do not. What, if anything else, did you happen to do while we were talking?" By contrast, the sound clarity of the call and use of hands-free headsets were only measured in Wave 2. The sound question read "For data quality purposes, how would you rate the clarity of the phone connection for this call? Would you say the connection was perfect – like we were talking face to face, very good, good, fair, or poor – like you could barely hear me at

[19] The wording for all of the questions used in this study is provided in Appendix sections A.7 and A.8.

times?" The headset question read "Right now, are you using a hands-free headset, or are you just holding the phone to your ear?"[20]

The key dependent variables in the study are indicators of cognitive shortcutting (Table 3.1). All seven indicators were measured in Wave 2 – after the randomization of respondents to device. Several of the indicators are being used for the first time in a cell phone measurement error study. These indicators are failing to account for important in/exclusions, correlations between attitudes and related behaviors, and estimating answers to behavioral frequency questions. Other indicators have been used in prior studies, albeit with a more limited experimental design. These indicators are recency effects, non-differentiation to battery-format questions, length of open-ended responses, and responding "don't know" or declining to answer. The hypothesis associated with each cognitive shortcutting indicator (Table 3.1) derives from the survey measurement error literature. Unfortunately, there was not sufficient space in the Wave 2 questionnaire to administer questions with socially desirable responses. Differential censoring in cell phone versus landline interviews is, therefore, beyond the scope of this study.

[20] Transportation safety researchers have found that conversation task performance does not differ as a function of phone interference when comparing hand-held phones versus headsets (Mazzae et al. 2004). Despite this, phone interface is still of interest in present study because the possible effects have not been explored in the survey literature. Furthermore, the tasks evaluated in transportation studies (e.g., recalling words) may not generalize to the task of responding to cognitively challenging questions.

Table 3.1. Indicators of Cognitive Shortcutting and Associated Hypotheses

Indicator	Stage of response process implicated	Device effects hypothesis
Important inclusions/exclusions	Comprehension	Smaller treatment vs. control difference among cell respondents, relative to landline respondents
Estimation as a response strategy for behavioral frequency questions	Retrieval, Judgment	Cell respondents more likely to estimate and less likely to enumerate, relative to landline respondents
Tendency to select response options presented recently	Comprehension, Response	Cell respondents more likely to select response options read recently, relative to landline respondents
Non-differentiation when rating items in a battery	Retrieval	Lower within-battery variance for cell respondents, relative to landline respondents
Correlations between attitudes and behaviors	Multiple stages	Weaker correlations for cell respondents, relative to landline respondents
Length of responses to open-ended questions	Retrieval	Shorter answers from cell respondents, relative to landline respondents
Item nonresponse	Multiple stages	Higher item nonresponse rates for cell respondents, relative to landline respondents

Indicator 1. Important Inclusions/Exclusions. Researchers have used experiments featuring important inclusions or exclusions in questions stems to test hypotheses about respondent attention to question wording (for an example, see Fuchs 2008). Generally speaking, a random half of the subjects (treatment group) are presented the question with the important in/exclusion, while the rest (control group) are presented the same question without the in/exclusion language. A significant difference in the expected direction between the group means is evidence that respondents were paying close attention to the question wording. In the present study, the statistic of interest is the difference in the differences. The hypothesis is that the difference between the treatment and control group means is smaller for the cell phone cases than for the landline cases. This indicator was tested using two different questions in Wave 2, so as to provide internal variability

on the construct. One question was borrowed from the Behavioral Risk Factor Surveillance Survey (BRFSS). The control group was asked "Now I would like to ask you about moderate physical activities, including brisk walking, bicycling, vacuuming, gardening, or anything else that causes a small increase in breathing or heart rate. On days when you do moderate activities for at least 10 minutes at a time, how much TOTAL time in MINUTES per day do you spend doing these activities?" The treatment group was asked the same question but also told "Please do not include time spent at work or commuting to work." The other question asked about internet usage. The control group was asked "During a typical week, about how many hours do you spend using the internet? Please include internet time at work, at home or in other locations, and visiting any kind of web site." The treatment group was asked the same question but told "Do not include time spent doing e-mail."

Indicator 2. Estimating Answers to Behavioral Frequency Questions. Researchers have demonstrated that there are multiple strategies that respondents use to answer behavioral frequency questions (Brown 1995; Conrad, Brown, and Cashman 1998). These strategies include enumerating individual episodes, retrieving a rate from memory, and answering based on a general impression. When the respondent has a high number of episodes to report, studies have shown that the response time is longer when enumeration is used (due to the counting of individual episodes) than when a different strategy is used (Conrad, Brown, and Cashman 1998). The longer response times are evidence that enumeration requires more cognitive effort than alternative strategies, such as estimating. This link between cognitive effort and response strategy represents another opportunity to test for an effect from device on cognitive shortcutting. The hypothesis in the present study is that cell phone respondents are less likely to use enumeration than landline respondents when asked behavioral frequency questions.

It should be noted that several factors have been shown to influence respondents' choice of response strategy – not just willingness to expend cognitive effort. These other factors include the actual frequency of the behavior, how the question is worded, the time allocated to respond (Blair and Burton 1987; Burton and Blair 1991), and the regularity and distinctiveness of the behavior (Conrad, Brown, and Cashman; Menon 1993). That said, under the current experimental design, it is reasonable to assume that these factors are balanced across the comparison groups. A set of dual users were randomly assigned to be interviewed on their cell or their landline. This randomization is expected to balance the groups with respect to their true scores, and the questions were administered in the same way for both groups. Consequently, any differences observed in the distribution of response strategies should be attributable to the device difference.

The behavioral frequency question used in the study read, "How many round trips have you taken by airplane in the past two years?" After answering, respondents were asked "Which of these best describes how you came up with your answer? Did you know the number off the top of your head; did you think about EACH individual trip and add them up; did you think about TYPES of trips and use that to estimate; or did you estimate based on a GENERAL IMPRESSION?" Using a retrospective protocol with a closed-ended question is an imperfect technique for measuring response strategy (Ericsson and Simon 1993). Respondents may have difficulty reconstructing what strategy they used, and studies using open-ended protocols reveal more strategies than the four listed in this study (Brown 1995; Conrad, Brown, and Cashman 1998). The retrospective, closed-ended protocol was selected for the current study, however, because questionnaire space was extremely limited, and the only distinction of interest was essentially between estimation versus non-estimation. The analytic objective was simply to test for a device effect on the choice of a high versus low effort strategy; it was not to map the strategies onto a precise, detailed classification.

Indicator 3. Recency (Response Order) Effects. Limitations of working memory capacity dictate that response options read last are the easiest to remember, resulting in a cognitive bias (Krosnick and Alwin 1987). Furthermore, satisficing theory argues that some respondents answer survey questions by selecting the first acceptable response option offered, rather than taking time to select an optimal answer (Krosnick 1991). When response alternatives are read aloud, the final options are more cognitively accessible than those read first, resulting in a recency effect.

The hypothesis in the present study is that recency effects will be greater for cell phone respondents than for landline respondents. This indicator was tested using two different questions in Wave 2, so as to provide internal variability on the construct. One question asked the respondent to select the main reason why America continues to explore space from a list of five options, including “to inspire and motivate our children” and “to keep our nation safe.” The options were read in a certain order for half the respondents and in the reverse order for the other half. The other recency effect question asked “Which of the following do you think will be America’s greatest rival in space exploration over the next 20 years? Russia, China, Japan, North Korea, the European Space Agency, or Iran?” A random half of the respondent were administered the same question with the list of countries reversed.

Indicator 4. Non-differentiation. Battery questions require the respondent to use a common set of response options in evaluating multiple items, such as rating confidence in various institutions. Respondents seeking to minimize their cognitive effort may elect to give the same response for each item in the battery (straight-line). By this logic, researchers may consider low variance in an individual's responses to reflect a lack of effort in formulating answers (Krosnick 1991). Some responses may erroneously be classified as evidence of error, but low variance on batteries is a widely used indicator of measurement error (e.g., Chang and Krosnick 2009; Fricker et al. 2005). In the present study, the hypothesis is that the respondent-level variances on battery items will tend to be lower for cell phone respondents than for landline respondents. The Wave 2 questionnaire contained two batteries. The first battery read, "I'd like to ask you about some institutions in American society. As I read each one, please tell me how much confidence you have in that institution using a scale from zero to 10, where 10 means 'great confidence' and zero means 'no confidence at all.'" The institutions evaluated were Congress, the military, and the public school system, the criminal justice system, the space program, and the Presidency. The other battery featured a five-point Likert scale ranging from strongly agree to strongly disagree. Respondents were asked to react to six statements about the U.S. space program, such as "The space program does a good job keeping astronauts safe."

Indicator 5. Correlation between Attitudinal and Related Behavioral Reports. If a respondent expresses a strong opinion about a given issue, then presumably, their behavior should align with that position. For example, people who voice strong support for the merits of recycling, are probably more likely to recycle regularly than those who express skepticism about recycling. In the present study, the hypothesis is that the correlation between related attitudes and behaviors is weaker for the cell phone cases than for the landline cases. Four attitude/behavior pairs were tested in the study. Attitudes were measured first using a battery question that read, “Now I’m going to name a few activities. For each one, please tell me whether you like doing this a great deal, like doing it somewhat, dislike doing it somewhat or dislike doing it a great deal.” The items were reading a newspaper, watching a movie, cooking a meal at home, and reading a book. Then respondents were asked whether or not they happened to do each of those activities yesterday.

Indicator 6. Length of Responses to Open-ended Questions. Several teams of researchers have used the length of responses to open-ended questions as a measure of data quality when comparing landline and cell phone interviews (Brick et al. 2007; Witt, ZuWallack, and Conrey 2009). This practice is based on the assertion that, in general, shorter answers (or fewer mentions) reflect less cognitive effort than longer answers. In the present study, the hypothesis is the cell phone respondents will tend to give shorter answers than landline respondents. Wave 2 contained two open-ended questions. One asked respondents for their thoughts on what should be the next major mission for the U.S. space program. Interviewers were not instructed to probe responses. The other open-ended question asked respondents to list the ways in which their lives have been directly improved by the space program. Interviewers were instructed to probe once for “any other ways” in which the respondent’s life was improved by the program.

Indicator 7. Item Nonresponse. When a respondent fails to answer a survey question, it is usually due to one of three reasons: the respondent feels that the question is inappropriate (e.g., too personal); the respondent does not know or cannot remember the answer; or the respondent declines to put forth the mental effort necessary to give a substantive response (Krosnick 1991). Only the third reason represents a cognitive shortcut, but item nonresponse is still a popular measure of data quality. In the present study, the hypothesis is that cell phone respondents have higher item nonresponse rates than landline respondents.

In interpreting the results below, it is important to bear in mind that the seven indicators highlight potential problems at *different stages* in the response process. As noted in Table 3.1, recency effects and inattention to important in/exclusions suggest error at the comprehension stage. Estimation, non-differentiation, and short open end responses may reflect errors in retrieving considerations from long-term memory. Estimation may also reflect a failure to accurately form a judgment based on the material retrieved. Other indicators may reflect error in multiple stages of the response process. For example, a "don't know" response may indicate that the respondent decided to avoid the response process altogether. Given the diversity in the dependent variables, it is perhaps unlikely that the experiment will yield significant devices differences across the board. Even if only one of the seven indicators shows evidence for a device effect, this may be enough to warrant researcher attention.

The results section below contains three parts. The first part compares landline and cell phone respondents on the seven indicators of cognitive shortcutting. Most device differences are small, in line with previous studies, but several significant differences suggest there may be a relationship between data quality and device in some instances. Theoretically, I expect any observed device effects to be explained by the proposed mechanisms (multitasking, poor audio quality, being away from home, use of a hands-free headset). The middle section reports the incidence of the mechanisms among the landline and cell phone respondents. In the last section, I test whether the device effects on measurement error go away when I account for the mechanisms posited to explain the device difference.

3.3 Results: Are Cell Phone Respondents More Likely to Use Cognitive Shortcuts Than Landline Respondents?

3.3.1 Indicators of Cognitive Shortcuts

This section reports the results of statistical tests to determine whether cell phone respondents were more likely to use cognitive shortcuts than landline respondents in a national dual frame RDD experiment. All subjects had both a landline and a cell phone, and they were randomly assigned to be interviewed on one device or the other. The use of landline RDD and cell phone RDD samples for this study has the advantage of generalizability to other surveys using those frames, but it has the disadvantage of providing no record data that can be used to investigate measurement error directly. Absent such data, I evaluate the risk of measurement error using seven different approaches: failing to account for important in/exclusions, estimating behavioral frequency answers, recency effects, non-differentiation to battery format questions, correlations between attitudes and related behaviors, length of open-ended responses, and item nonresponse. Where possible, I present multiple tests of a single approach in order to achieve internal variation and enhance external validity. For dual frame survey designers, null findings are a desirable outcome. If randomized trials show no evidence that cell phone respondents are more likely to use cognitive shortcuts than landline respondents, this would support the idea that data from these two methods can be combined with minimal concern about device differences in response accuracy.

3.3.2 Attention to Question Wording (Important Inclusions/Exclusions)

The first opportunity to use a cognitive short cut is at the beginning of the response process when the interviewer reads the question and the respondent interprets its meaning. A pair of split-ballot experiments was used to assess how closely respondents were paying attention to question wording, and whether attention level differed by device. Presumably, respondents who are listening carefully account for all of the considerations mentioned in the question when they formulate their response. Those who are not paying close attention, by contrast, may to fail to account for considerations in formulating their answer.

In Wave 2 half of the respondents were asked the BRFSS question about the number of minutes per day they spend doing moderate physical activity. The other half were asked the same question with the statement “Do not include time spent at work or commuting to work.” I operationalize paying attention by assessing the effect of this exclusion on the response distributions. A significant difference in response distributions would suggest that those administered the exclusion statement paid attention to the exclusion and adjusted their answer accordingly. In this device effects comparison, if cell phone respondents are taking cognitive shortcuts at the interpretation stage, then I would expect the effect from the exclusion to be less dramatic in the Wave 2 cell phone condition than in the landline condition.

The results, displayed in Figure 3.2, support the hypothesis that respondents were less likely to be paying attention to the question when interviewed on a cell versus a landline. When respondents in the cell condition were told to exclude exercise at work, the mean minutes reported changed from 84.7 to 88.6, which was a non-significant shift in the opposite direction from what one would expect. Among those randomly assigned to be interviewed on Wave 2 on their landline phone, however, the effect of the exclusion was dramatic. The mean for those receiving the exclusion was 66.3 minutes, which compares to a mean of 102.8 minutes for the group not given the exclusion statement. In all four comparison groups, there were a handful of respondents reporting exceptionally high values, which would have unduly influenced the results. To prevent this, the response distributions were capped at the 95th percentile.

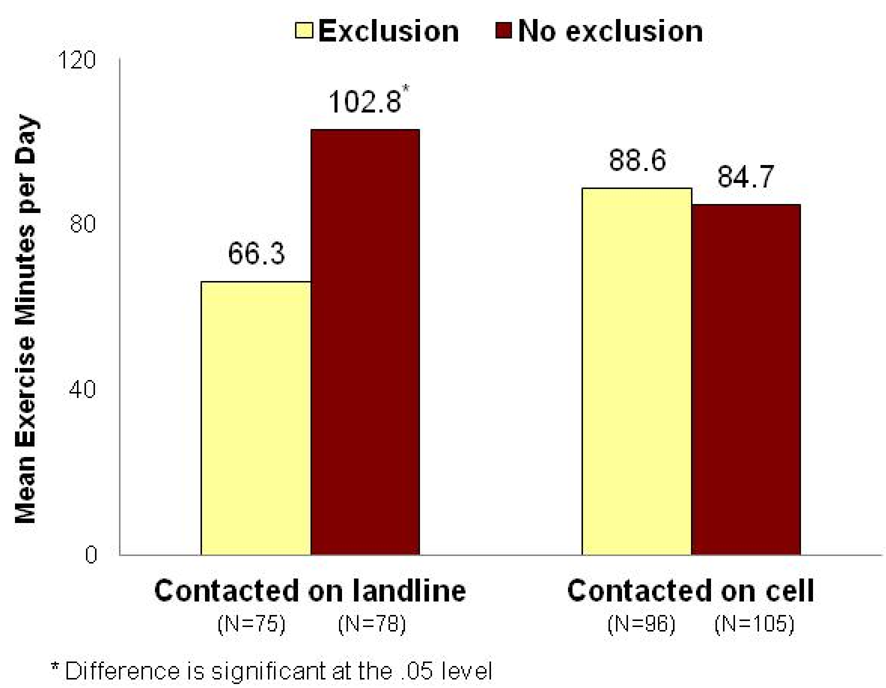

Figure 3.2. Mean Exercise Minutes per Day by Question Wording and by Device

A second test of this hypothesis was conducted using an item that asked respondents to report the number of hours per week that they spend using the internet. A random half were instructed to exclude time spent doing e-mail. The results for each device are displayed in Figure 3.3. For this item, both the landline and cell groups show the expected effect. The mean number of hours using the internet is lower for those instructed to exclude e-mail time than those not given that instruction. Consistent with the result of the exercise experiment, the effect of the exclusion appears to be more dramatic for the landline group than the cell group, but the effect is not statistically significant. In the landline group, the exclusion resulted in a drop in the mean from 13.2 hours to 10.2 hours, for a difference of 3.0 hours and a relative change of 25% (t=1.36, df=148, p=0.176). In the cell group, the exclusion resulted in a drop in the mean from 16.2 to 14.5, for a difference of 1.7 hours and a relative change of 9% (t=0.82, df=200, p=.415). A generalized linear model was estimated to determine if the interaction of device (landline/cell) and questionnaire form (exclusion yes/no) had a significant effect on the reported number of internet hours. The interaction term was not significant, indicating that the appearance of a larger effect (closer attention to the question) in the landline condition is not statistically reliable.

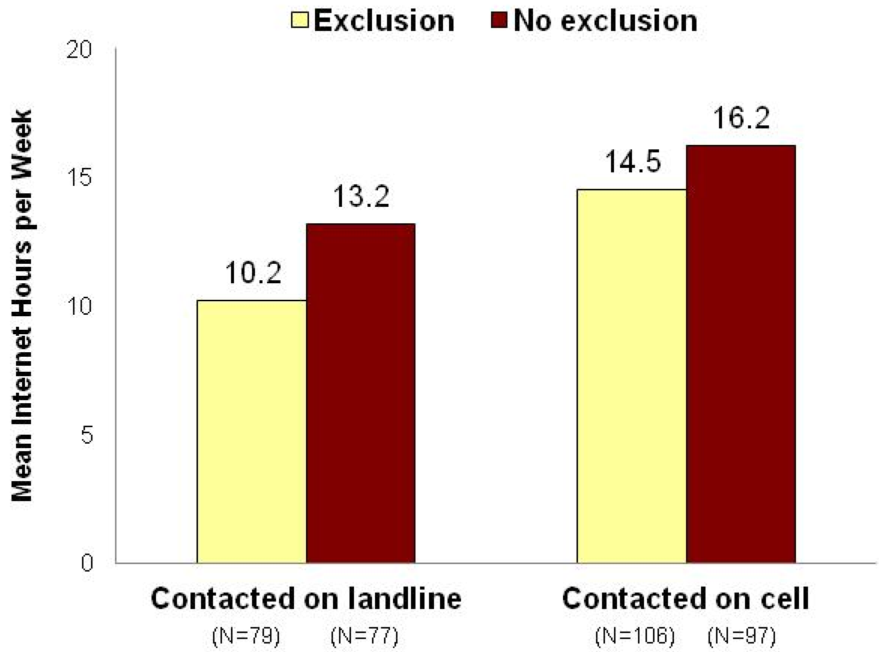

Figure 3.3. Mean Internet Hours per Week by Question Wording and by Device

Two factors may explain why the difference between the landline and cell groups was less dramatic in the internet experiment than the exercise experiment. First, the exercise question was notably more verbose than the internet question. The exercise question contained nearly twice as many words (60) as the internet question (33). This longer question presented a stronger stimulus because people are generally less likely to pay attention when they are read a long-winded statement than a short one. So the internet question may have simply been too short to discriminate between those paying careful attention and those giving less attention. Another possible explanation is that the randomization to Wave 2 device was not sufficient to balance the landline and cell groups with respect to the interest and usage of technology. The Wave 2 respondents in the cell condition may have been more interested in technology than those in the landline condition. This is evidenced by the difference in the mean hours per week using the internet. Collapsing across in/exclusion groups, the Wave 2 cell phone respondents reported an average of 15.7 hours per week on the internet, compared to 11.9 hours for the landline respondents. If greater usage indicates greater interest, then the cell phone group may have perked up and paid more attention to this particular item simply because they tend to be heavier internet users than the landline group. The two factors suggest that the device effect, should it be robust, was not significant in the Figure 3.3 comparison because the target question was too short and asked about the wrong subject (one that was of differential interest to the two experimental groups).

The results from this pair of experiments suggest that people are somewhat less likely to be paying careful attention to questions when they are interviewed on a cell phone as opposed to a landline. This effect is likely to be more pronounced for lengthier questions. Given that this is the first test of this particular hypothesis, this effect would need to be replicated in other telephone surveys using other questions before concluding with greater confidence that attention level is device-dependent.

3.3.3 Response Strategies for Behavioral Frequency Questions

A common approach for differentiating between respondents answering carefully and those answering less carefully is to administer a cognitively demanding question. Behavioral frequency questions require the respondent to report a numeric summary of the number of times that they did a certain behavior, such as visiting a doctor or eating out at a restaurant. On balance, enumerating individual episodes tends to be more cognitively demanding than answering based a general impression or a extrapolating from a known rate of doing the behavior (Conrad, Brown, and Cashman 1998). If cell phone respondents are indeed more distracted than landline respondents, then presumably they would be less likely to enumerate than landline respondents.

Two metrics were used to test this. The first is a self-report of the strategy used when given four options: direct retrieval from memory, recall and count, recall types and count, and use of general impression. The other metric is an indicator for whether the respondent reported a prototypical value. Respondents who are estimating are more likely to report a round value such as 50 than a similar number such as 49. The hypothesis being tested is that people are more likely to provide a prototypical response when interviewed on a cell phone as compared to a landline.

In Wave 1, respondents were asked how many overnight trips they have taken in the past two years. Presumably, each person can provide the correct answer (or one very close) if they are willing and able to recall each individual trip and sum them. Immediately after reporting how many trips they had taken, respondents were asked which response strategy they used. This same pair of questions was also asked in Wave 2, with the exception that trips were defined as round trips taken by airplane. This change was made to make the questions fit better substantively with the other content of the Wave 2 survey.

The results, which are presented in Table 3.2, indicate that respondents are no more likely to estimate when they are interviewed on a cell phone as compared to a landline phone. Nearly one-third (32%) of those interviewed on their cell phone for Wave 2 estimated based on a general impression or by thinking about types of trips. The proportion of those interviewed for Wave 2 on their landline who estimated is highly similar (35%). This lack of difference supports the null hypothesis that respondents are equally likely to estimate when interviewed on a landline and on a cell phone.[21]

Looking at the device comparison from Wave 1,[22] which is based on two probability samples with no randomized assignment, two differences are clear. First, the overall incidence of estimating is much higher. This is likely attributable to two factors. One is the slight variation in question wording. In general, people reported many more "overnight trips" in Wave 1 (median = 7, mean = 54.0) than "round trips taken by airplane" in Wave 2 (median = 2, mean = 5.1). The response distributions to the Wave 1 and Wave 2 questions are presented in Figures 3.4 and 3.5, respectively.

Another explanation for the change in the overall rate of estimation could be the change in the pool of subjects from Wave 1 to Wave 2. Those responding to the Wave 2 survey were demonstrably more cooperative than the Wave 2 nonresponders, and this characteristics may be related to a general willingness to participate and perhaps even participate more thoughtfully than their peers.

[21] While the likelihood of enumerating was about equal across the Wave 2 device conditions (34% for landline cases versus 37% for cell phone cases), the thoroughness of the enumeration could have differed. For example, respondents with many trips to report may have terminated the enumeration process sooner if they were interviewed on their cell phone as opposed to their landline. The difference in the mean number of trips reported was not significantly different, however, for the self-reported enumerators in the Wave 2 landline and cell phone conditions (t=0.02, df=65.7, p=0.99). This result suggests that respondents who chose to enumerate devoted similar levels of effort to the task, regardless of device.

[22] The analysis of Wave 1 data were re-run excluding the 34 landline RDD respondents who reported being interviewed away from home. Presumably, these respondents were actually interviewed on a cell phone due to porting or call forwarding. This would have contaminated the device comparisons presented in Table 3.2. The Wave 1 results do not, however, change at all when these cases are excluded.

Table 3.2. Response Strategy for Behavioral Frequency Question, by Device[A]

	Wave 1		Wave 2	
	Landline RDD Sample	Cell RDD Sample	Called on Landline	Called on Cell
Response Strategy[B,C]				
Knew number off top of head	21%	18%	30%	29%
Thought about each individual trip and added them up	20%	19%	34%	37%
Thought about types of trips and used that to estimate	24%	24%	19%	15%
Estimated based on a general impression	30%	34%	16%	17%
Don't know/Refused	5%	5%	1%	1%
	100%	100%	100%	100%
Percent giving a prototypical value[D]	25%	32%	5%	8%
Sample size	(763)	(933)	(102)	(138)

[A]Respondents who reported zero trips are excluded from this analysis.

[B]The proportion of respondents estimating (thinking about types or using a general impression) was not significantly different for the Wave 1 landline RDD and cell RDD samples ($X^2(1)=3.27$ $p=.071$) when cases with "don't know" or "refused" values are excluded.

[C]The proportion of respondents estimating (thinking about types or using a general impression) was not significantly different for the Wave 2 landline and cell conditions ($X^2(1)=0.06$ $p=.801$) when cases with "don't know" or "refused" values are excluded. This result is the same when controlling for Wave 1 frame.

[D]Prototypical values are considered here to be multiples of 10 greater or equal to 10.

The second difference to notice comparing the Wave 1 and Wave 2 results is that in Wave 1 there is no strong evidence that cell phone respondents are more prone to estimating (58%) than the landline respondents (54%). The difference is marginally non-significant ($X^2(1)=3.27$, $p=.071$). Given that this result did not replicate in the randomized Wave 2 test, there is no compelling evidence that people are more likely to estimate their response to a behavioral frequency question when interviewed on a cell phone as compared to a landline.

The categories of estimating based on a "general impression" and estimating by thinking about "types of trips" are purposefully combined in this testing. Examination of the types of responses given by respondents using the various strategies revealed that people using these two strategies in fact did estimate more than people using the other strategies. I computed an indicator for whether the respondent provided a prototypical response (10, 20,…) or not. In both waves the proportion of respondents utilizing “general impression” or “types” estimation who gave a prototypical response was approximately 20% greater than the same proportion among those utilizing the other strategies. In Wave 1 some 43% of the respondents who estimated gave a prototypical value, while 18% of other respondents did so ($X^2(1)=110.1$, $p<.0001$). In Wave 2 some 24% of the respondents who estimated gave a prototypical value, while 3% of other respondents did so ($X^2(1)=24.6$, $p<.0001$). Put simply, response distributions provided evidence that those who said that they estimated (using types or general impression) did, in fact estimate, while those who said that they did not estimate appear to have been truthful.

Respondents who reported a round or prototypical number are, presumably, more likely to have given an inaccurate answer than respondents reporting a non-prototypical number. To be sure, some responses that are prototypical may also be entirely accurate, but the rate of this type I error is not likely to be related to device. The proportion of respondents reporting a prototypical value for the number of trips taken in the past 2 years in shown in the middle of Table 3.2.

Here again, looking only at the non-randomized Wave 1 comparison, the reader might conclude that estimating is more prevalent among cell phone than landline respondents (32% versus 25%, respectively, $X^2(1)=10.51$, $p=.0012$). In the randomized Wave 2 comparison, the effect is in the same direction but the magnitude of the difference between the groups does not reach statistical significance ($X^2(1)=1.05$, $p=.31$).

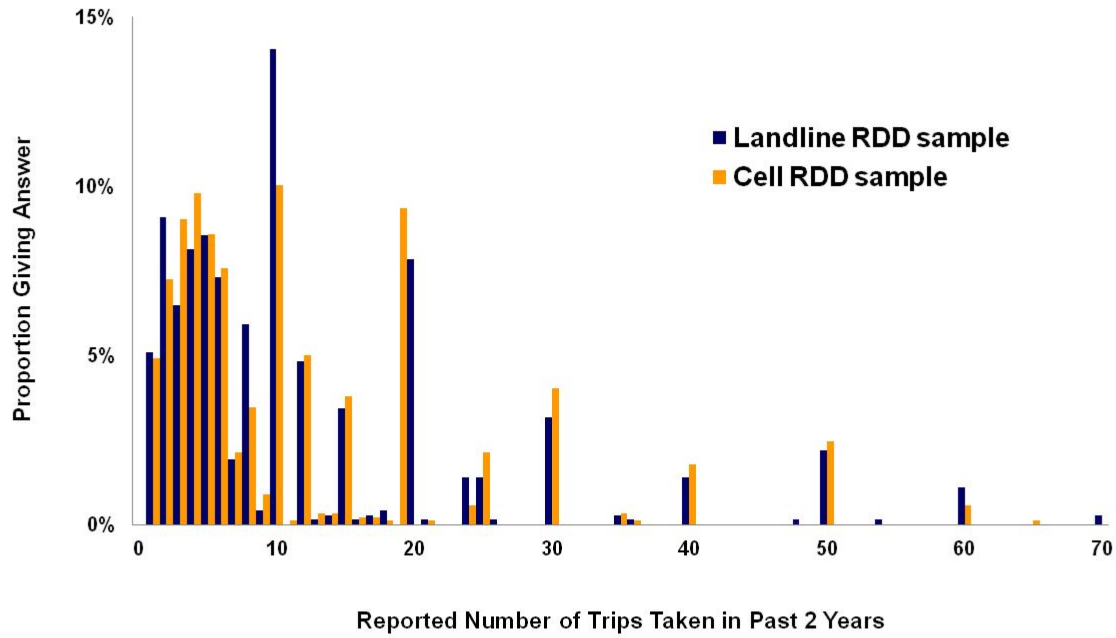

Figure 3.4 Histogram of Responses to the Wave 1 Behavioral Frequency Question, by Device

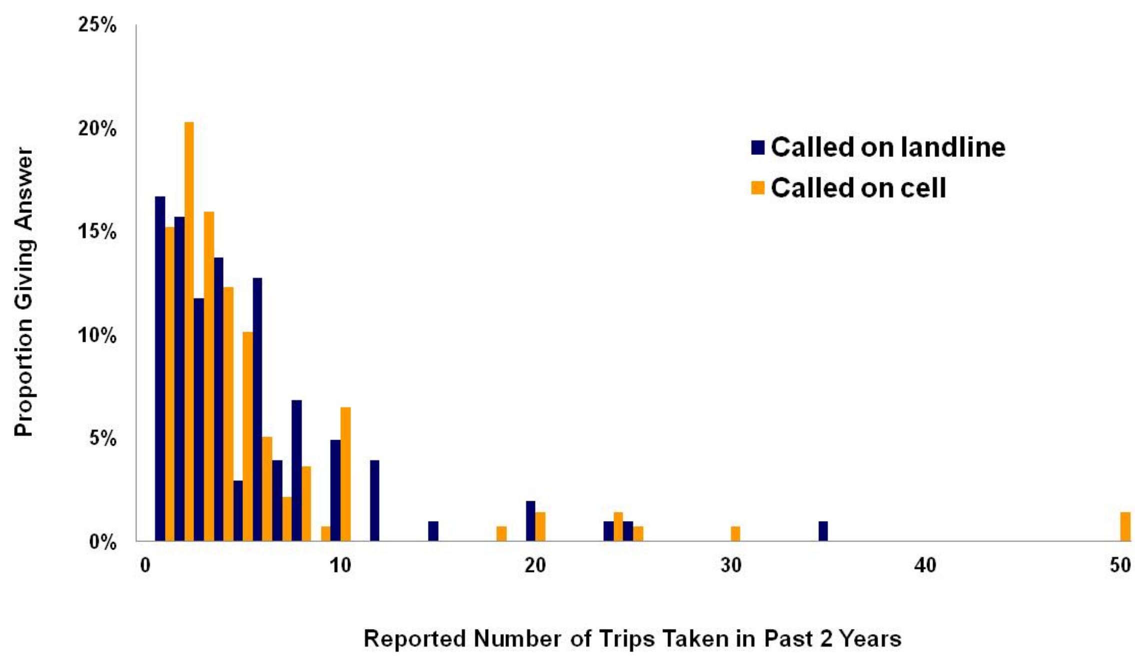

Figure 3.5 Histogram of Responses to the Wave 2 Behavioral Frequency Question, by Device

Considering both the self-reported response strategies and the incidence of providing prototypical values, it does not appear that the type of telephone device is related to estimating. This conclusion is based primarily on the Wave 2 comparisons, which are designed to isolate the effect of device through randomization. The Wave 1 comparisons, by contrast, are confounded with differences in the samples with respect to age, education, and other factors that may be related to the use of cognitive shortcuts.

The Wave 2 result suggests that a large effect from device is highly unlikely. That said, it is still possible that a more modest effect exists, but that this particular test was not sensitive enough to detect it. Two factors support this latter line of reasoning – the relatively small Wave 2 sample sizes and the nature of the behavioral frequency question. As reported in Table 3.2, there were only 102 and 138 cases in the Wave 2 comparison groups. The true population distribution of values for the frequency question asked in Wave 2 could also have suppressed a device effect. Examination of the response distributions (Figure 3.5) shows that only a small proportion of the Wave 2 respondents took more than a handful of airplane trips in the past few years, reducing the need to consider many instances of the construct in question. The fewer episodes of the behavior there are to recall, the less burdensome it is to enumerate them. A question asking about a more common behavior, such as trips to the grocery store, might yield a more sensitive test because true values would tend to be higher and the question would be more difficult to answer accurately.

3.3.4 Recency Effects

Some cognitive shortcuts may be unintentional but, nonetheless, have consequences for measurement error in survey estimates. If respondents are presented with a long list of response options, they may be unable to retain all of those options in their working memory as they try to formulate a response. In telephone surveys, the options are presented orally, so respondents may be more likely to select one of the last few options read. Knäuper (1999) demonstrated that such recency effects tend to be stronger among older adults than younger ones because working memory capacity decreases with age. Based on the recency effects literature and speculation about cell phones, there are two potential reasons for less working memory capacity – advanced age and distraction. If people responding on cell phones are more distracted by their multi-tasking activities or environmental stimuli, then they may have less cognitive capacity available for answering the survey questions. If this pattern is real, then it should manifest itself as more frequent endorsement among recently-presented response options. In particular, the recency effect would be more pronounced in the Wave 2 cell treatment group than the landline group.

The Wave 2 instrument featured three pairs of questions designed to test for recency effects. The question stem for the first pair was “Which one of the following do you think is the MAIN reason why America continues to explore space?” Then five potential reasons were read: to keep our nation safe; to inspire and motivate our children; to maintain our status as an international leader in space, to provide benefits on earth, or because it’s human nature to explore. Half the respondents received the options in that order and the others received the same options read in the reverse order. The question stems for the next pair was “Which of the following do you think will be America’s greatest rival in space exploration over the next 20 years?” Six countries or entities were then read in one order for half the respondents and in reverse order for the other half. The question stem for the final pair was “If the U.S. space program had more money, which of the following would be the best way to spend it?” This was followed by a listing of six potential NASA missions, which were administered in the same split-ballot fashion.

The results of the recency effect experiments are presented in Figures 3.6 through 3.8. The bars represent the difference in the proportion of respondents endorsing the option when the list was read as shown versus the proportion endorsing it when the list was read in the reverse order. If the recency effect was strong, then the options listed at the bottom would have large positive values, reflecting a higher rate of endorsement when the option is read last (as shown) and a lower rate of endorsement when the option is read first (when the list is reversed). By the same token, if the effect is strong, the response options at the top of the list would have negative values, reflecting a lower endorsement rate when the options are read as shown with the item presented first, compared to when the option is read last.

Across the device conditions, there is modest evidence that respondents tended to favor options read most recently. Critically, there is no consistent evidence that recency effects were more pronounced for respondents interviewed on a cell phone than those interviewed on a landline. None of the values depicted by the bars in the three figures is reliably different from zero, after adjusting the significance threshold for multiple comparisons using the Benjamini-Hochberg procedure[23] (1995). This failure to detect differences may be related to the modest sample sizes, though overall the tendency to select recent response options does not look to be strong with either device.

I also used an alternate test that considered the proportion of respondents who endorsed one of the options presented in the latter half of the list, as it was administered to them. The results of this test also indicated that cell phone and landline respondents were equally likely to exhibit a recency effect. For example, in the first experiment, some 39% of the cell phone respondents endorsed the 4th or 5th option read. The proportion selecting one of these options among the landline respondents was highly similar (38%). This null finding suggests that cell phones do not place greater demands on working memory capacity than landlines.

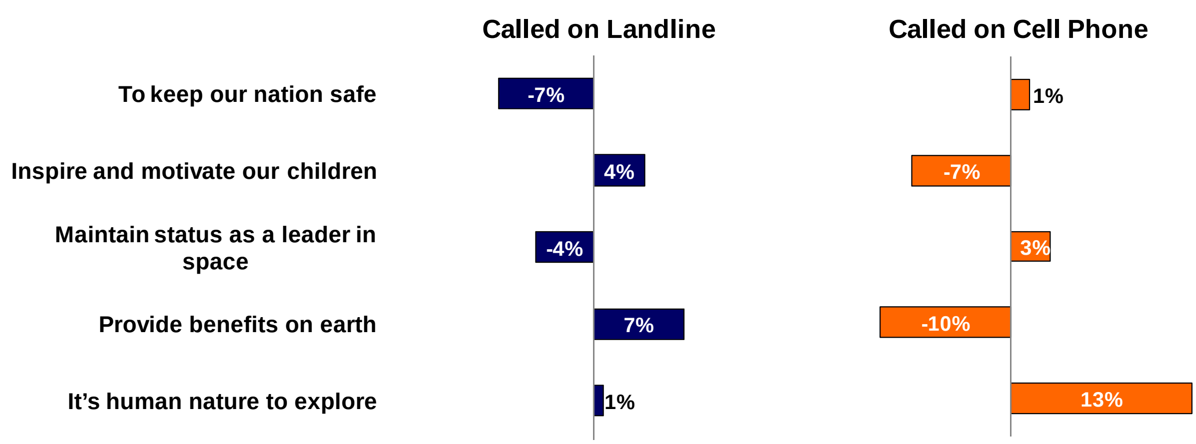

Figure 3.6. Difference in Response Frequencies When List Is Read as Shown Versus Read in the Reverse Order, by Device.
Note.– Sample sizes: landline N = 156, cell phone N = 204. Question wording: "Which of the following do you think is the MAIN reason why America continues to explore space?"

[23] The Benjamini-Hochberg procedure uses a sequential approach to control for the false discovery rate in multiple comparisons. It yields much greater power than the Bonferroni technique that limits the family-wise Type I error.

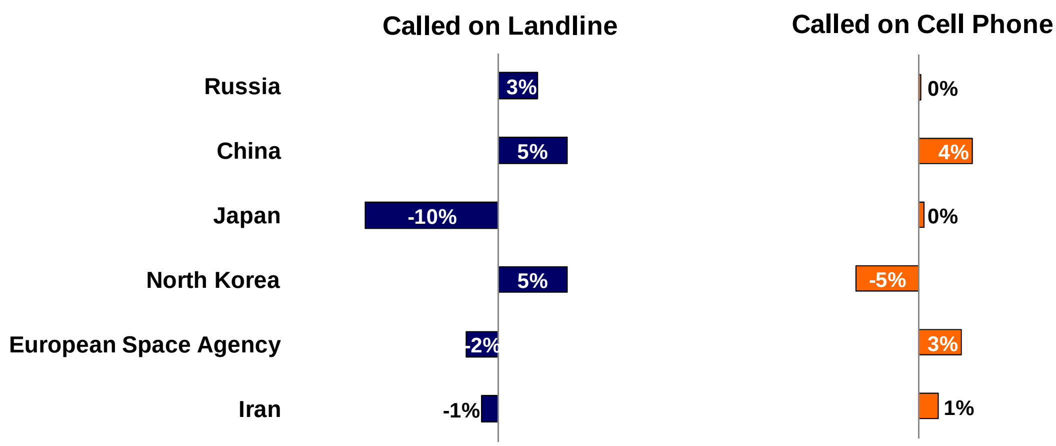

Figure 3.7. Difference in Response Frequencies When List Is Read as Shown Versus Read in the Reverse Order, by Device.
Note.– Sample sizes: landline N = 156, cell phone N = 204. Question wording: "Which of the following do you think will be America's greatest rival in space exploration over the next 20 years?"

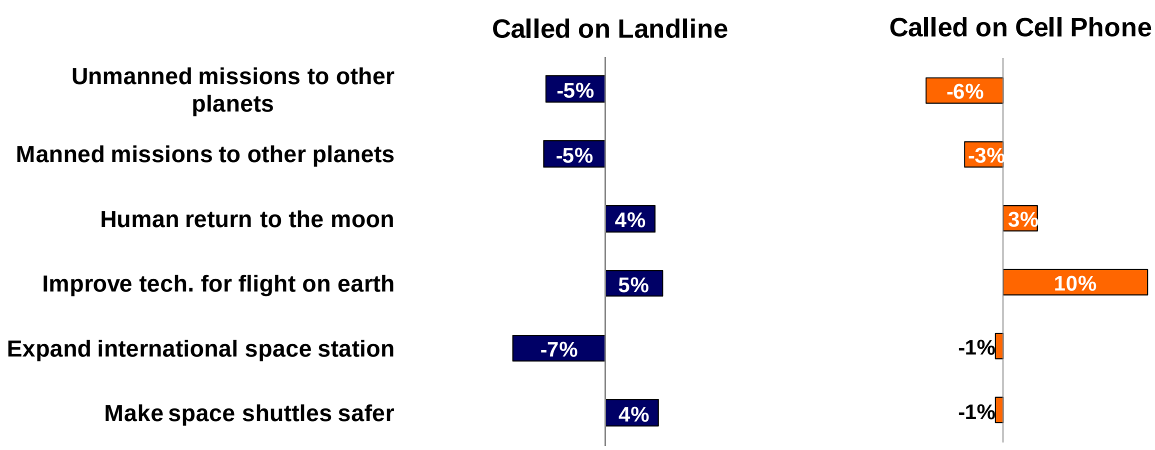

Figure 3.8. Difference in Response Frequencies When List Is Read as Shown Versus Read in the Reverse Order, by Device.
Note.– Sample sizes: landline N = 156, cell phone N = 204. Question wording: "If the U.S. space program had more money, which of the following would be the best way to spend it?"

One possibility is that the effect was masked by a confound from the composition of the comparison groups. This possibility was ruled out using multivariate logistic regression models. For each of the three experiments, I modeled the probability that the respondent selected one of the options presented in the latter half of the list. Respondent age and education were included as covariates in order to control for cognitive capacity. Indicators for Wave 2 device and the ordering of the response options were also included in the model. An effect of device was not detected in any of the three models. Even if an effect does exist, these results suggest that it is too small to be of concern in most instances.

One limitation of these recency effect experiments is that the stimulus was relatively weak. The General Social Survey (GSS) question on desirable qualities that a child might have (Krosnick and Alwin 1987) features 13 response options and has been used to great effect in measurement error studies. Technically, the GSS question is used to demonstrate a primacy effect (tendency to select options presented first) because it is administered in printed form on a show card, as opposed to a telephone survey question received aurally. The basic point, though, is that the cognitive demand of selecting among 13 options may be substantially greater than the demand of selecting among five or six options, as in the questions at hand. The GSS item was too time consuming (and off topic) to use in this study, but it might be a good candidate for future studies investigating measurement error in cell phone interviews.

3.3.5 Non-differentiation among Items in a Battery

Another indicator for cognitive shortcutting is not differentiating among items presented in a battery (Krosnick 1991). If the respondent is indifferent to the accuracy of his/her responses, it is easier to give the same rating of each item in a robotic fashion, than to assess each item individually and formulate the most appropriate response for that particular item. The hypothesis that I wish to test is whether respondents tend to differentiate among items less if they are interviewed on a cell phone as opposed to a landline.

Two batteries in the Wave 2 survey are appropriate for this test. In the first battery, respondents were asked to rate their confidence in six U.S. institutions (*e.g.*, the military, the public school system) using a scale from zero to ten. The other battery consisted of six statements about the U.S. space program. Respondents registered their strength of agreement or disagreement with each statement using a five-point Likert scale. In each battery, the order in which the items were presented was randomized across respondents. The metric used to capture "differentiation" is the treatment group mean of the respondent-level variances. First, I calculated the variance of each respondent's answers to the battery. Then, I computed the mean of those variances separately for each treatment group. Higher group means are more desirable because they reflect greater differentiation between the battery items. The results are presented in Figure 3.9.

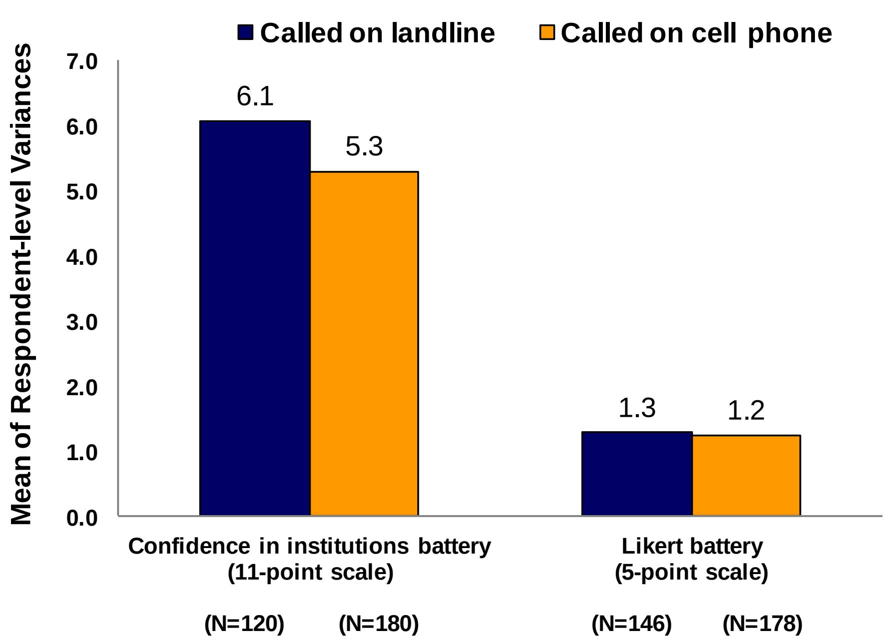

Figure 3.9. Differentiation in Rating Items in Two Batteries, by Device
Note. – Figures shown are the group mean of the respondent-level variances for the battery.

There is some support for the hypothesis that respondents differentiate more in a landline interview than a cell interview, but this is true for only one of the two batteries tested. The mean of the respondent-level variances in responses to the confidence in institutions battery was 6.1 for the landline group and 5.3 for the cell phone group. The one-sided t-test for this difference falls just short of the .05 significance threshold (t=1.39, df=218.7, p=.084). In the Likert scale battery, however, the degree to which landline respondents and cell phone respondents differentiated was virtually the same. The mean of respondent variances was 1.3 for the landline group and 1.2 for the cell phone group. One possible explanation for the null result in this second battery is that the response scale was shorter by half, meaning that respondents had less opportunity for differentiation. It follows that any effect from device would probably be smaller in the second battery than the first. That said, the null result in the second battery and the lack of significance in the first raises questions as to how robust a device effect on battery differentiation may be.

3.3.6 Correlations between Attitudinal and Behavioral Reports

For many survey topics, it is reasonable to expect that people's attitudes are correlated with their behavior. For example, if someone says that they strongly oppose environmentalist policies, then I would expect that they probably do not recycle plastics, glass, and other materials. Large discrepancies between how people say they feel and what they do with respect to common activities, would raise questions as to whether the reports are accurate. In this study I leverage this logic to determine whether device (landline/cell) is related to the strength of association between attitudinal reports and behavioral reports. If attention levels are indeed lower in cell phone interviews, then I would expect cell phone respondents to answer less accurately, attenuating the strength of the correlations between attitudes and behaviors. To be sure, perfect correlations are not realistic. Life is complicated, and people often do to not act on many of their feelings. That said, the baseline disagreement should be the same for those randomly assigned to the landline group and those assigned to the cell phone group. This suggests that any observed difference in correlations should be attributable to device.

A set of four attitudinal items and a matching set of behavioral items were specifically administered in Wave 2 to test this. The attitudinal question asked respondents to rate how much they like or dislike four activities (reading a newspaper, watching a movie, cooking a meal at home, and reading a book) using a four point scale ranging from “like a great deal” to “dislike a great deal.” The follow-up behavioral question asked respondents to report whether or not they did each of those activities yesterday. Asking the behavioral question immediately after the attitudinal question could have induced greater correlation in responses to these two items, relative to spacing them out in the interview. Unfortunately, there is no way to test this empirically in this study. Any such context effect arguably would not jeopardize the internal validity of the test because the questions were administered in the same fashion for both landline and cell phone respondents. That said, the close sequencing of the questions may have diminished any device difference by priming the behavioral reports in both conditions.

Two metrics designed for ordinal data are used to summarize the strength of the association between the attitudes and behaviors. The second and third columns of Table 3.3 report the Goodman-Kruskal gammas for the landline group and cell group, respectively. The fourth and fifth columns display the Stuart's tau-c values for these groups. Both gamma and tau-c can range from -1 (perfect negative association) to +1 (perfect positive association), with the value zero representing statistical independence. Gamma makes no adjustment for either table size or ties, but tau-c does include an adjustment for the size of the table (the number of levels of the variables). The cross-classification tables for this analysis were 4 x 2, as the attitudinal scale had four points and the behavioral question was dichotomous.

Table 3.3. Association between Attitudinal and Behavioral Reports for Four Activities, by Device

	Goodman-Kruskal Gamma		Stuart's tau-c	
	Landline	Cell	Landline	Cell
Reading a newspaper	0.84 (0.71, 0.96)	0.78 (0.66, 0.90)	0.54 (0.41, 0.67)	0.54 (0.42, 0.65)
Watching a movie	0.25 (-0.03, 0.53)	0.51 (0.30, 0.72)	0.14 (-0.02, 0.29)	0.28 (0.15, 0.42)
Cooking a meal at home	0.60 (0.40, 0.81)	0.54 (0.36, 0.72)	0.35 (0.20, 0.50)	0.35 (0.21, 0.48)
Reading a book	0.73 (0.57, 0.89)	0.70 (0.55, 0.84)	0.47 (0.33, 0.61)	0.43 (0.31, 0.55)

Note.- Confidence intervals (95%) are presented in parentheses.

The results in Table 3.3 indicate that there is no perceptible effect of device on the strength of association between related attitudinal and behavior reports. The tau-c values are identical for two of the activities (reading a newspaper and cooking). The difference in tau-c values for the other two items run in opposite directions, though the 95% confidence intervals for the point estimates are overlapping. The four pairs of contingency tables (not shown) indicate that cell phone respondents were somewhat more likely to say they liked an activity but did not do it yesterday. Collapsing across the four items, the cell respondents did this 37% of the time compared to landline respondents reporting this way 34% of the time. The result is intriguing, but the difference is not large enough to reach statistical significance. Based on the data at hand, I find no compelling evidence that attitudes are more predictive of behavior in either landline or cell phone interviews.

3.3.7 Length of Open-Ended Responses

Open-ended questions tend to be more cognitively demanding than closed-ended questions because they force respondents to formulate their own answer as opposed to providing them with a set of alternatives. Providing a terse answer to an open-ended question may be another shortcut that respondents can use to answer the question by expending a minimal amount of cognitive effort. Short answers are not necessarily less accurate than longer ones, but the length of a response presumably is correlated with the level of cognitive effort used to produce the answer. If cell phone respondents are prone to taking cognitive shortcuts, then their answers may be shorter than those of landline respondents.

Wave 2 contained two open-ended questions. One asked respondents for their thoughts on what should be the next major mission for the U.S. space program. Interviewers were not instructed to probe responses. The other open-ended question asked respondents to list ways in which their lives have been directly improved by the space program. Interviewers were instructed to probe once for “any other ways” in which the respondent’s life was improved by the program.

Item nonresponse rates to these two questions did not vary significantly by device. Approximately one-third of those interviewed on a cell phone and on a landline did not answer the “next major mission” question (32% versus 33%, respectively). For the “improved your life” question, 55% of cell phone respondents did not provide an answer or said their life had not been improved by the space program, compared to 53% of landline respondents.

I first tested for differences in the distribution of response length as measured by the number of words. For both open ends, the mean and median response lengths were smaller for the cell phone respondents than the landline respondents, but not significantly so. The mean number of words to the "next major mission" question was 6.64 words for the cell phone respondents as compared to 7.16 words for the landline respondents (t=0.95, df=221, p=0.34). On the "improved your life" question, the cell phone respondent mean was 8.33 words and landline respondent mean was 8.71 words (t=0.70, df=157, p=0.69). These results are based only on those answering the questions. Values were capped at the 95^{th} percentile in order to avoid undue influence from a small number of respondents who gave extremely long responses. The distributions of the response lengths are displayed in Figures 3.10 and 3.11.

In a separate analysis, I regressed response length on device using a negative binomial model. Respondents not answering the questions (number of words = 0) were retained in the analysis. A negative binomial model was selected instead of Poisson model because the outcome variable is count data and the variance is greater than the mean, indicating overdispersion.[24] Overdispersed variables are better modeled by a negative binomial distribution than a Poisson distribution. The results were unambiguous. In the negative binomial model for both questions, device had no effect on response length, and including controls for demographics related to satisficing (age and education) made no difference.

[24] Overdispersion is the presence of greater variability in the observed data than would be expected based on a given simple theoretical model.

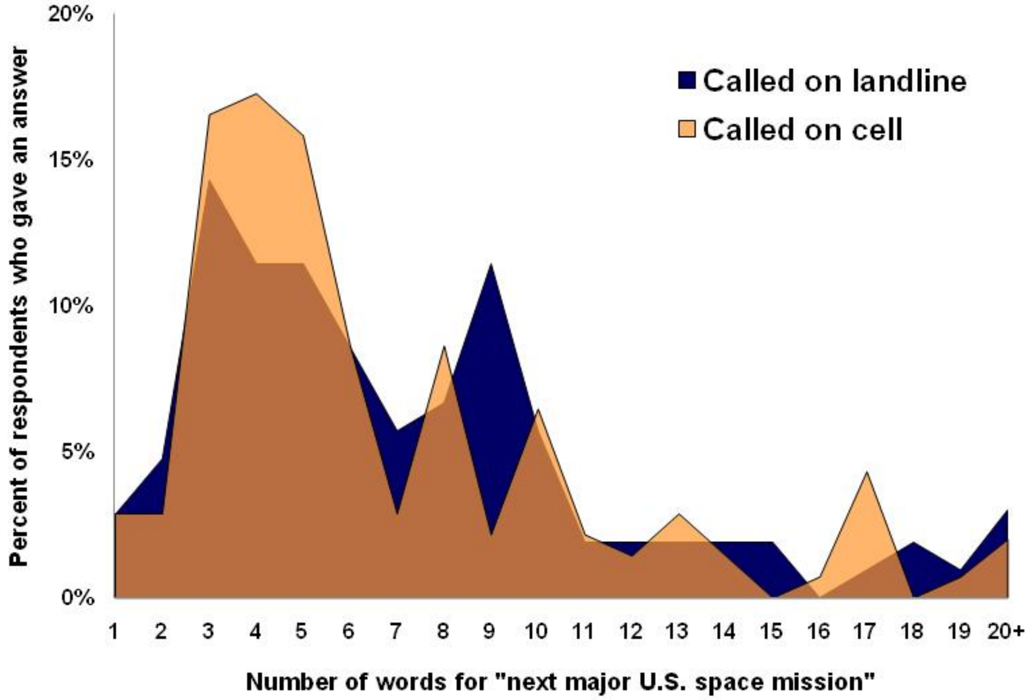

Figure 3.10. Length of Open-ended Response To "Next Major Space Mission" Item
Note. – The figure excludes respondents not giving an answer.

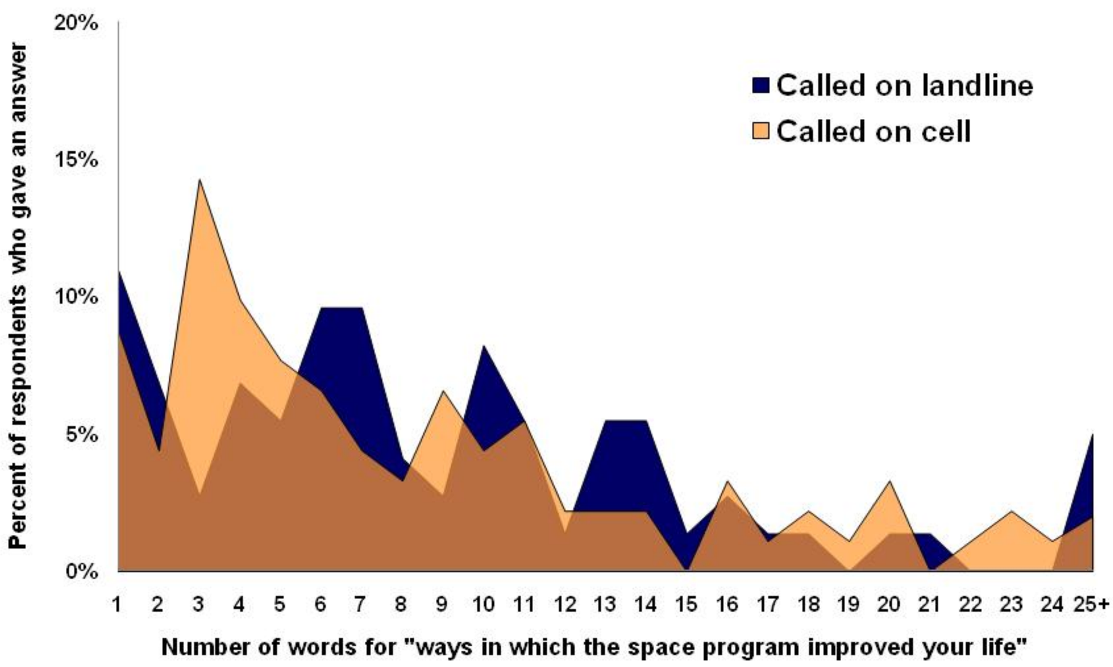

Figure 3.11. Length of Open-ended Response To "Ways the Space Program Improved Your Life" Item.
Note. – The figure excludes respondents not giving an answer.

While there was no evidence that overall response length was affected by device, an examination of the distributions in Figures 3.10 and 3.11 raises another possibility. There appears to be a spike in the distribution for cell phone respondents corresponding to answers three to five words long. This spike appears for both questions and does not look to be as prominent in the distributions for landline respondents. This suggests that there may be an effect from device limited to one part of the distribution – short responses. To test this, I created an indicator for responses between one and five words long. The chi-square tests for both open-ended questions fall just short of the more liberal .10 threshold. Some 55% of the cell respondents answered the "next mission" question in five or fewer words, compared to 45% of the landline respondents ($X^2(1)=2.71$, $p=.100$). For the other open-end, 45% of the cell respondents gave short responses, compared to 33% of the landline respondents ($X^2(1)=2.51$, $p=.113$). The sample sizes for each comparison group ranged from 73 to 139 cases and were smaller for the "ways the space program improved your life" question than the "next mission" question.

Another metric of the cognitive effort used in responding is the number of items mentioned, rather than sheer length in words. The "ways the space program improved your life" question was well-suited to this analysis, whereas the "next mission" open-end was not. Again, there is some suggestion that cell respondents are giving less thoughtful responses, but the difference is not significant. The distribution for number of mentions by device is presented in Table 3.4. Some 22% of landline respondents mentioned three or more ways in which they felt that the space program had improved their life. This compares with 16% of the cell phone respondents. The difference in these proportions and the overall Chi-square test reported in the table are not, however, statistically significant.

Table 3.4. Number of Mentions to Open-Ended Question, by Device

Number of mentions	Called on landline	Called on cell
One	52%	54%
Two	26%	30%
Three	11%	7%
Four	5%	8%
Five or more	6%	1%
	100%	100%
Sample size	(73)	(91)

Note.- The difference in the distributions is not significant (X^2=3.53, *df* =4, *p* =.47). Figures exclude respondents not providing any mentions.

In summary, I find some indications that respondents may give shorter answers to open-ended questions when interviewed on a cell phone as compared to a landline, but the results are not statistically significant. More research with larger sample sizes is needed.

3.3.8 Item Nonresponse

The final cognitive shortcut considered in this study is item nonresponse. Respondents may decline to answer questions for a number of reasons. They may object to the question, not understand the question, not know the answer, or want to avoid embarrassment from answering the question. Another possibility, posited by Krosnick (1991), is that some respondents decline to answer because they are satisficing (Simon 1957). Whatever the cause, item nonresponse is problematic for the researcher because it represents a loss of information. In this section I test whether the rate of item nonresponse in Wave 2 differed between those randomly assigned to be interviewed on their landline versus those interviewed on their cell phone.

In total 49 questions in Wave 2 were suitable for this analysis. The distribution for the number of items not answered for each treatment group is presented in Figure 3.12. The direction of the effect is contrary to expectations. Some 39% of the Wave 2 cell respondents answered every item, as compared to 30% of the landline respondents (X^2=3.20, *df*=1, *p*=.074). A logistic regression model was estimated in order to verify that the device effect on item nonresponding once or more versus not at all persists when controlling for proxies of cognitive capacity and interest in the survey topic. The parameter estimates for the logistic regression are reported in Table 3.5. The effect of device remains significant at .10 level (*p*=.056) when controlling for education, age, and topic interest.

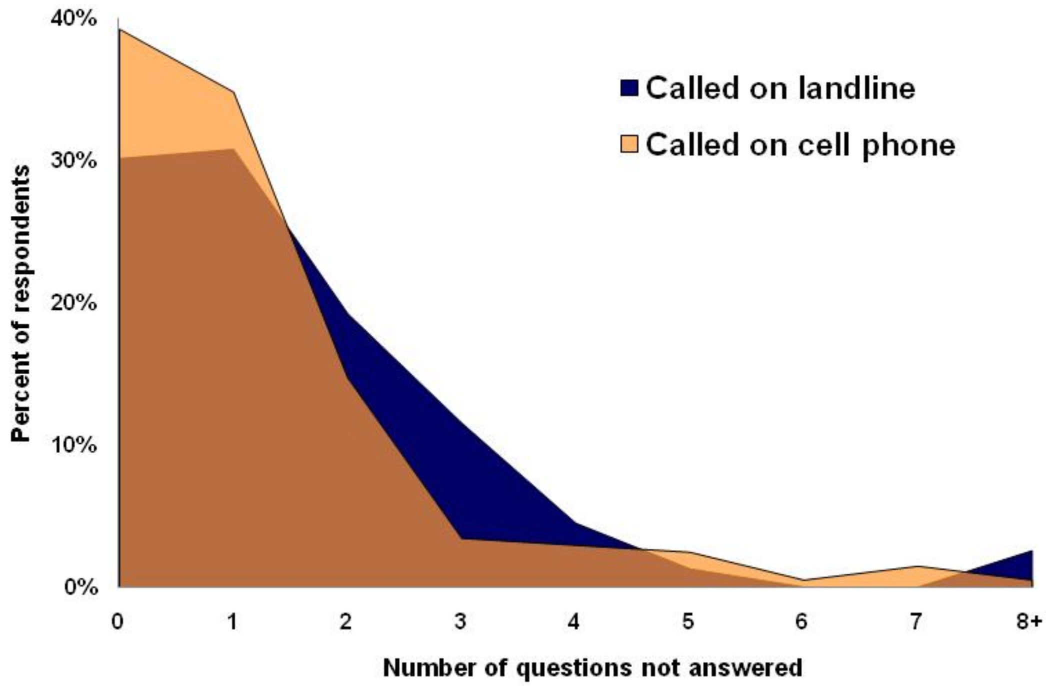

Figure 3.12. Distribution of Number of Wave 2 Questions Not Answered, by Device

Among those who did not answer one or more items, however, there appears to be no effect from device on the number of items not answered. When a negative binomial model is fitted to the distributions for the number of items missed, the effect of device is not significant.

Table 3.5. Logistic Regression Predicting Nonresponse to One or More Items in Wave 2

	Estimate	s.e.	Wald X^2	p-value
Intercept	-0.33	0.33	1.00	0.317
Device				
Landline (reference)				
Cell phone	-0.23	0.12	3.64	0.056
Education				
High school or less (reference)				
Trade school or some college	0.02	0.20	0.01	0.925
Bachelor's degree	-0.33	0.19	3.04	0.081
Graduate school degree	0.00	0.21	0.00	0.986
Age				
18 to 24 (reference)				
25 to 34	-0.46	0.35	1.65	0.200
35 to 44	-0.19	0.27	0.50	0.481
45 to 54	0.02	0.24	0.01	0.926
55 to 65	-0.45*	0.23	3.98	0.046
65 or older	0.59*	0.28	4.34	0.037
Interest in the survey topic	-0.46**	0.13	11.69	0.001

***p<.001 **p<.01 *p<.05 Model based on N=356 cases.
Area under ROC curve c = .683, -2 Log L = 428.6

These results suggest that cell phone respondents are at least as likely as landline respondents to answer survey questions. When item nonresponse is viewed dichotomously (once or more versus never) being on a cell phone appears to be a negative predictor, but when it is modeled as a continuous variable (number of items unanswered) then being on a cell phone does not appear to be a predictor. More evidence would be needed before concluding that item nonresponse rates will tend to be lower for cell phone interviews than landline interviews as suggested by these data.

3.3.9 Summary of Experiments Testing for a Device Effect on Cognitive Shortcutting

Table 3.6 presents a summary of the results from the seven approaches used in this study to test for a device effect on the respondents' use of cognitive shortcuts in telephone surveys. The evidence is generally quite weak, but some of the experiments suggest that respondents may be slightly less attentive to the task of responding when interviewed on a cell phone as opposed to a landline. Cell phone respondents were somewhat less likely to be listening carefully to question wording, differentiating between items in a battery, and giving lengthy responses to open-ended questions. When asked a BRFSS question about the number of minutes per day spent doing moderate physical activity, the mean for landline respondents dropped significantly when they were told to exclude exercise at work, while the mean for cell phone respondents did not drop at all. A second exclusion experiment showed more similarity between landline and cell phone responses.

A comparison of responses to a multi-item battery reveals some evidence that people may differentiate more between items when interviewed on a landline as compared to a cell phone. Responses to a second battery showed no device difference, but there was much less variance overall in the second battery. With respect to open-ended items, there was no device difference in overall response length. There is a slight indication, though, that cell phone respondents were more likely to give very short responses (five words or fewer) than landline respondents. It is also the case that several other intriguing results could represent a real device effect but they did not reach statistical significance in this study.

While these findings warrant caution and further investigation, they are not evidence that data from cell phone surveys are substantially inferior to data from landline surveys. The magnitude of the device differences found in this study were generally small, in keeping with prior research. The largest device effect observed (in the exclusion experiment) could likely be avoided by keeping questions short and easily understood. One practical conclusion is that cell phone respondent may have slightly less cognitive capacity to answer surveys questions than landline respondents, so questionnaire designers may want to err on the side of striking or revising questions that are cognitively demanding.

Table 3.6. Summary of Experiments Testing for Greater Usage of Cognitive Shortcuts in Cell Phone versus Landline Interviews

Indicator of cognitive shortcutting	Finding	Considerations	Evidence for device effect
Important inclusions/exclusions	The effect of an important exclusion is more dramatic in the landline condition than the cell condition. This appears to have replicated within Wave 2 but only one of two experiments was significant.	The non-significant test was based on a relatively short question, which may have supplied insufficient stimulus.	Evidence for device effect
Estimation as a response strategy for behavioral frequency questions	The incidences of estimating and giving prototypical responses were similar for the Wave 2 landline and cell groups.	Many Rs appear to have had low true scores and, therefore, little motivation to estimate. This reduced the sensitivity of the test.	No evidence of an effect
Tendency to select response options presented recently	Recency effects were no stronger in the cell group than the landline group	The lists of response options may not have been long enough to detect the effect.	No evidence of an effect
Non-differentiation when rating items in a battery	The average variance of responses to items in a battery was lower for cell Rs than for landline Rs. But the result was only marginally significant (.10 level) and did not replicate in another battery.	In the non-significant battery, the response scale may have been too small to detect the effect.	Weak evidence on an effect
Correlations between attitudes and behaviors	The strength of associations between attitude and predictor items for four activities did not vary systematically by device.	Only four activities were tested, and the correlations tended to be low.	No evidence of an effect
Length of responses to open-ended questions	Results suggest that responses may be shorter when Rs are interviewed on a cell phone, but the finding is not statistically significant.	Only two open-ends were considered, and insufficient power appears to have seriously damaged this test.	Weak evidence on an effect
Item nonresponse	Cell phone respondents were slightly more likely to answer every question than landline respondents, but the result was only marginally significant (.10 level).	Conditioning on Wave 1 responders yielded relatively low item-nonresponse levels.	Weak evidence for an effect in the opposite direction

3.4 Explaining the Device Effect

3.4.1 Factors That May Impair the Response Process, Particularly in Cell Phone Surveys

The previous section presents evidence that, for some questions, people are more prone to misreport if they are interviewed on a cell phone as opposed to a landline. This section explores why. Four mechanisms may lead respondents to answer less accurately on a cell phone than on a landline: the respondent's location (at home or away), engagement in other activities during the interview (multitasking), use of a hands-free headset, and the audio fidelity of the call. Several of these factors can also impair the response process for landline interviews, but presumably the problem is more severe in the cell phone setting because the ways that users interact with their cell phones are often dramatically different from the way they use a landline phone. For each mechanism, I first report the incidence level among the cell phone versus landline respondents, and then I report whether the mechanism does or does not help to explain the device differences detailed above.

3.4.2. Respondent Location

The mobility permitted by cell phones implies that respondents can be interviewed in virtually any location allowed by their service provider. This is quite different from landline technology, which usually requires the phone user to be in or near the residence where the phone is connected. The physical location of the respondent was measured in both the Wave 1 and Wave 2 surveys, and the results are presented in bold face in Table 3.7. Two main observations can be made. As expected, respondent location is closely related to device. In Wave 1, the device on which respondents were interviewed also indicated the sample frame from which they were selected. About one third (35%) of the cell RDD sample respondents were interviewed away from home. This compares to 4% of the landline RDD respondents. Similarly in Wave 2, about one-quarter of dual users interviewed on their cell phone were reached away from home, compared to 1% of those called on their landline. If the sampling frames were perfect (with respect to the correspondence between the type of phone to which a number is *believed* to service and the type which a number *actually* services), then all landline respondents would be reached at home. The two most likely explanations for the deviation are porting and call forwarding.

Table 3.7. Location and Activities of Respondents When the Interview Began, by Device

	Wave 1		Wave 2	
	Landline RDD Sample	Cell RDD Sample	Called on Landline	Called on Cell
Respondent reached away from home	**4%**	**35%**	**1%**	**24%**
In car	0%	8%	0%	5%
Shopping/Running errands	0%	5%	0%	4%
Watching TV/Video gaming/On computer	1%	3%	0%	2%
Working	0%	6%	1%	5%
Other activity/Refused	3%	12%	0%	6%
Reached at home	**96%**	**65%**	**99%**	**76%**
Watching TV/Video gaming/On computer	33%	25%	43%	34%
Eating/Drinking	17%	11%	15%	11%
Household chores	9%	3%	6%	6%
Reading	6%	4%	8%	6%
Working	4%	1%	4%	1%
Other activity/Refused	27%	21%	23%	17%
	100%	**100%**	**100%**	**100%**
Sample size	(1,072)	(1,053)	(156)	(204)

* Figures may not add up to 100% due to rounding

The other key observation about the respondent location findings in Table 3.7 is that the incidence of cell phone respondents who are interviewed away from home is relatively high. Even though taking a cell phone survey away from home is possible, we would not necessarily expect it to be common. On this issue, the Wave 1 results are most appropriate for generalization to other surveys because it was a cross-sectional survey with no screen for telephone service. As noted above, about one-third of the cell phone interviews were conducted with people away from their home. If being away from home is shown to generate problematic interviewing conditions, then this issue of cell phone mobility may have serious implications for measurement error in future telephone surveys.[25]

3.4.3 Multitasking During the Interview

In this study of cognitive short cuts, respondent location is of interest not because location per se matters (whether the respondent is sitting on their sofa or on a park bench), rather because location can be an indicator for interviewing condition outside the home that may undermine the respondent's ability to provide accurate answers. Examples of such potentially distracting conditions include being at work, in a restaurant, or shopping. To this point, respondents in both waves were asked what they happened to be doing when the interviewer called. The results are summarized and nested in Table 3.7. Among those reached away from home, many were working, shopping or traveling in a car. Some 8% of the Wave 1 cell RDD respondents reported being in a car when the

[25] One piece of useful information that is missing is the location of nonrespondents. During the design of this study, consideration was given to having interviewers attempt to get this information from refusers who were about to hang up. This feature was ultimately not implemented due to historically low rates of compliance with such last-ditch efforts (Kulka, McNeill, and Bonito 1982; Lynn 2003), but this could be a subject for future research.

call came. This was also the case for 5% of the Wave 2 respondents contacted on their cell phone. Alarmingly, there is reason to believe that not all of these respondents were passengers. When asked what they were doing when the call came, most of those categorized in Table 3.7 as "in car" said that they were driving. It could be the case that by "driving," some of these respondents actually meant that they were riding in a car being operated by someone else. It seems doubtful, however, that this explanation applies to everyone in that category. This finding is troubling given the potential threat to respondent safety. It is also somewhat surprising that the incidence of this activity was so high considering explicit interviewer instructions to avoid this scenario. During the survey introduction, interviewers calling cell phone numbers were instructed to say, "If you are now driving a car or doing any activity requiring your full attention, I need to call you back later." This low but non-trivial incidence of respondents who say (at the end of the interview) that they are driving deserves greater scrutiny in research specifically designed to measure this behavior.

It is important to keep in mind that respondents who report that they were doing an activity (*e.g.*, household chores) when their phone rang did not necessarily continue to do that activity during the interview. This distinction unfortunately cannot be made using the measures from this study. Given the nature of the survey (a cold call from an unfamiliar sponsor), many respondents probably continued to do low-exertion activities during the interview.

A follow-up question was designed to capture other activities that the respondent was doing during the interview. The question was, "Some people multitask when on the phone while others do not. What, if anything else, did you happen to do while we were talking?" Responses to this item were combined with responses to the "What did you happen to be doing when I called?" item, and the results are summarized in Table 3.8 The percentages in Table 3.8 represent the proportion of respondents who reported being engaged in various activities either when they were contacted or during the interview. Combining responses from these two questions seemed to produce the best possible representation of the ways in which respondents multitasked during the interview. Percentages sum to more than 100% (column-wise) because many respondents said that they were doing one activity when the call came (*e.g.*, watching TV) and then they engaged in another activity during the course of the interview (*e.g.*, folded laundry). To be clear, Table 3.7 provides the best possible summary of what respondents were doing at the start of the interview, and Table 3.8 provides that most comprehensive report of what they did during the interview. Activities in which respondents were originally engaged but were abandoned before the administration of the first question, will be over-reported in Table 3.8, but this error is not likely to be related to the device.

Table 3.8. Proportion of Respondents Engaged in Other Activities During the Interview, by Device

	Wave 1		Wave 2	
Activity	Landline RDD Sample	Cell RDD Sample	Called on Landline	Called on Cell
Watching TV/Video gaming/On computer	34%	35%	49%	45%
Eating/Drinking/Preparing meal	18%	17%	24%	16%
In car	1%	11%	0%	8%
Working	3%	8%	5%	7%
Household chores	10%	7%	10%	7%
Shopping/Running errands	0%	6%	0%	5%
Reading	6%	5%	10%	8%
Other activity	38%	20%	16%	14%
Sample size	(1,072)	(1,053)	(156)	(204)

* Column figures sum to more than 100% because some respondents reported multiple activities.

A central observation from Table 3.8 is that while the distribution of multitasking is generally similar for landline and cell interviews, there are some potentially important exceptions. In Wave 1, one quarter of the cell phone RDD sample respondents were engaged in activities such as shopping, working, or riding in a car. Each of these activities has the potential to be very distracting, and they are generally unique to cell phone respondents.

3.4.4 Hands-free Headsets

The desire to multitask while on the phone is one reason people use hands-free headsets. In Wave 2, the incidence of hands-free headsets was 6% for both the landline condition respondents and the cell condition respondents (see Table 3.9). Headset usage was also independent of respondent location (at home or away), but it appears to be related to multitasking activities. The proportion of headset users who were shopping or in a car was double the proportion of non-headset users doing these activities (14% versus 7%, respectively), though the sample sizes are too small for this difference to be significant.

Table 3.9. Audio Fidelity Ratings and Use of Hands-free Headsets, by Device

	Wave 2	
	Called on Landline	Called on Cell
Are you using a hands-free headset, or just holding the phone to your ear?		
Using hands-free headset	6%	6%
Holding phone to ear	94%	93%
Don't know/Refused (VOL)	1%	0%
	100%	100%
How would you rate the clarity of the phone connection for this call?*		
Perfect, like we were talking face to face	37%	33%
Very good	46%	51%
Good	15%	10%
Fair	3%	5%
Poor, like you could barely hear me at times	0%	0%
Don't know/Refused (VOL)	0%	0%
	100%	100%
Sample size	(156)	(204)

* The difference in distributions is not statistically significant (X^2=4.24, df=3, p=0.24)

3.4.5 Audio Fidelity

Another reason that researchers have posited as to why measurement error may be greater in cell phone interviews than landline interviews is differential audio fidelity. In Wave 2, respondents rated the clarity of the phone call on a scale that ranged from "perfect, like we were talking face to face" to "poor, like you could barely hear me at times." The results are shown in Table 3.9. Those interviewed on their landline were slightly more likely to describe the connection as perfect (37% versus 33%), but the difference in response distributions was not statistically significant. Given this parity with respect to headset usage and audio fidelity, these do not appear to be likely causes of differential measurement error in landline versus cell phones interviews

3.4.6 Do the Hypothesized Mechanisms Explain Usage of Cognitive Shortcuts?

If the mechanisms discussed above are, in fact, the reasons why measurement error may differ for landline and cell phone interviews, then an important prediction can be made. The device differences in cognitive shortcutting observed in Section 3 should disappear when controlling for the mechanisms. In theory that is what I would expect to happen, but the empirical reality is more complicated. Of the seven shortcutting indicators, only a few showed any sign of a device difference, and some of the mechanisms, such as use of a hands-free headset, are about equally popular among landline and cell phone users. Furthermore, even the most promising mechanisms were measured somewhat imprecisely. For example, some respondents may have been eating or drinking but failed to report that to the interviewer. Any such error in the mechanism variables would undermine my ability to explain device differences. In addition, the incidence level for most of the multitasking indicators (e.g., shopping, working) was too low to use them as independent variables. In the analysis below, only the effects of eating/drinking and watching TV/using a computer could be measured reliably.

For this section I re-ran the cognitive shortcutting analysis in Section 3, but this time the mechanisms were included as independent variables. The reader may recall that the strongest evidence for differential cognitive shortcutting came from the question wording exclusion experiment. Cell phone respondents were less likely than landline respondents to account for an important exclusion mentioned in the question. I reanalyzed the data comparing the effect of the exclusion among those interviewed away from home versus those interviewed at home, irrespective of device. The data from respondents interviewed at home showed the expected drop in reported exercise time when respondents were told to exclude time spent at or commuting to work (95.5 min. versus 75.2 min., t=-1.83 p=.07). The data from respondents interviewed away from home did not show the expect drop (62.8 min. versus 94.5 min, t=1.27, p=.210). Unfortunately, only 49 respondents were reached away from home, and they were split between the two question versions (exclusion mentioned yes/no), so the sampling error associated with that test is quite large. The other exclusions experiment, which asked about internet usage, showed no effect from being at home or away from home.

Failure to attend to the exclusion was also weakly associated with eating or drinking during the interview and poor sound quality. In both of the exclusion experiments, respondents who reported eating or drinking were less likely to attend to the exclusion than others. The exclusion lowered the means for the non-eaters and non-drinkers by 17 min. and 3 min. in the exercise and internet experiments, respectively (t=-1.48, p=.14 and t=-1.65, p=.10). It did not lower means for those eating or drinking (t=0.00, p=1.00 and t=0.17, p=.86). Respondents reporting relatively poor sound quality were less likely to attend to the exercise exclusion than those reporting better sound quality. This effect does not replicate, however, for the internet exclusion, and sample sizes were small for both tests. Some 59 respondents reported only “good” or “fair” sound quality (as opposed to “very good” or “perfect” sound quality), and they were split between the two questionnaire forms. There was no relationship between attention to question wording and use of a hands-free headset or watching television/using a computer.

Unfortunately, because of the nature of the test, it was not possible to measure the effect of device (cell/landline) while controlling for these mechanisms. The important exclusions, recency effects, and attitude/behavior correlation experiments all rely on aggregate tests. The dependent variables (e.g., differences in group means) have no meaning, except in the aggregate, and the questionnaire form splitting required for some of the experiments yielded small sample sizes in many instances.

One cognitive shortcut experiment that did allow for simultaneous testing of device and the mechanisms was non-differentiation among items in a battery. The reader may recall that there was some indication that respondents may differentiate among item less when interviewed on a cell phone as opposed to a landline, but the results fell short of statistical significance. Table 3.10 shows the effect of device on variation in responses to a battery when controlling for the hypothesized mechanisms of reporting error.

Table 3.10. Linear Models Predicting Differentiation Among Items in Two Batteries

	Model estimating differentiation in confidence in institutions battery		Model estimating differentiation in Likert battery	
	Estimate	s.e.	Estimate	s.e.
Intercept	3.55*	(1.57)	1.09***	(0.25)
Called on cell phone	-0.69	(0.57)	-0.09	(0.09)
Away from home	-0.27	(0.81)	0.16	(0.14)
Multitasking indicators				
Eating/Drinking	0.01	(0.67)	0.01	(0.11)
Watching TV/On the computer	-0.08	(0.56)	-0.03	(0.09)
Using hands-free headset	1.65	(1.04)	-0.05	(0.17)
Poor sound quality for the call	0.84	(0.71)	-0.15	(0.11)
Education	0.12	(0.25)	0.09*	(0.04)
Age	1.35**	(0.46)	0.05	(0.07)
Sample size	(297)		(330)	
Adjusted r-square	0.023		0.002	

***p <.001 **p <.01 *p <.05

Note.- Figures are based on those who answered every item in the battery. Negative coefficients indicate that the predictor is associated with less differentiation among items in the battery.

The behavior being modeled is differentiation, so positive predictors are associated with desirable response behavior and negative predictors are associated with undesirable behavior (lack of differentiation). Unfortunately, the mechanisms provide little insight into this behavior. None is a significant predictor of variance among items in either of the batteries, and they do not appear to explain what effect, if any, is associated with device. Not surprising, both models fit the data poorly, and neither overall F-test is significant. The results do not change in any meaningful way when the device indicator is dropped from the models. The only statistically significant effects are associated with the demographic controls. Age and education were included in the models because both have been shown in the literature to be associated with satisficing. Age and education are both positively associated with greater differentiation and, presumably, less reporting error. This analysis suggests that device may be weakly associated with non-differentiation, but the effect is not statistically significant and the factors that I expected to underlie this phenomenon are not supported.

Another cognitive shortcut that cell phone respondents seemed to be using slightly more than landline respondents is giving short answers to open-ended questions. Two such questions were asked in the Wave 2 questionnaire. I modeled the length of response using the same set of predictors shown in Table 3.10. Specifically, I estimated generalized linear regression models predicting the log of the number of words. The log transformation was used because the distribution for the number of words was highly skewed, and the transformation helped to normalize the data. The results (Table 3.11) were quite similar to those from the non-differentiation analysis. For both open-ended questions, the model fit is poor and there is no evidence that the mechanisms lead to shorter responses. In one of the models, watching television/using a computer is significantly and positively associated with the length of response, but this effect did not replicate for the other open-ended question.

Table 3.11. Linear Models Predicting the Length of Responses to Two Open-ended Questions

	Model for length of response to "ways the space program has improved your life"		Model for length of response to "next major mission for the space program"	
	Estimate	s.e.	Estimate	s.e.
Intercept	1.45**	(0.47)	1.74***	(0.25)
Called on cell phone	-0.09	(0.16)	-0.08	(0.09)
Away from home	0.28	(0.23)	0.01	(0.14)
Multitasking indicators				
Eating/Drinking	0.18	(0.20)	0.09	(0.11)
Watching TV/On the computer	0.37*	(0.16)	-0.10	(0.09)
Using hands-free headset	0.01	(0.25)	-0.28	(0.17)
Poor sound quality for the call	0.27	(0.20)	-0.05	(0.11)
Education	-0.01	(0.07)	0.03	(0.04)
Age	0.12	(0.14)	0.05	(0.08)
Sample size	(164)		(244)	
Adjusted r-square	0.003		0.000	

***p <.001 **p <.01 *p <.05

Note.- The outcome being modeled is the log of the number of words recorded. Respondents not answering the question are excluded from the model.

A similar counter-intuitive result was found when I re-analyzed the behavioral frequency response strategy experiment. Respondents who said that they were eating or drinking during the interview were less likely to estimate than those who did not. The results are presented in Tables 3.12. Two logistic regression models were used to determine whether or not the hypothesized mechanisms were predictive of estimating a response to the behavioral frequency items (number of roundtrips by airplane in the past two years). The first model shown in Table 3.12 predicts whether the respondent said that they estimated their response. The second model in Table 3.12 predicts whether or not the respondent reported a prototypical number (e.g., 10, 20,…). Presumably, prototypical numbers are less likely to be accurate than non-prototypical numbers, all else equal. Eating or drinking was not significantly associated with prototypical responses, but the direction of the effect suggests better data from respondents doing this.

One post hoc explanation is that, under some circumstances, low-level multitasking may actually improve respondent behavior. People who are eating or drinking are probably stationary and less likely to be seriously distracted by another activity, such as driving or shopping. That is, the fact that they are engaged in a low-level distraction keeps them from doing something more intensive that could actually impair their ability to respond carefully. Alternatively, snacking or watching television may put respondents at ease, possibly even elevate their mood, making them more amenable to exerting cognitive effort in responding. These ideas are purely speculative, though, and would benefit from empirical testing.

The other notable finding in Table 3.12 is that respondents interviewed away from home were much more likely to give a rounded, prototypical number than those interviewed at home. Nearly one-quarter (23%) of respondents interviewed away from home gave a prototypical response, compared to 5% of those reached at home ($X^2(1)=19.0$, $p=<.0001$). According to the model on the left, however, they were no more likely to say that they estimated than respondents reached at home. Unfortunately, not enough data were collected to determine exactly why respondents reached away from home might be more prone to using this shortcut. Presumably, they could be more distracted than those reached at home, but more evidence is needed.

Table 3.12. Logistic Regressions Predicting Use of Estimation as a Response Strategy to a Behavioral Frequency Question

	Model predicting estimation (self-report)		Model predicting the report of a prototypical number	
	Estimate	s.e.	Estimate	s.e.
Intercept	-0.41	(0.41)	-2.25	(0.67)
Called on cell phone^	-0.08	(0.16)	-0.08	(0.28)
Away from home^	0.12	(0.22)	0.86**	(0.29)
Multitasking indicators				
Eating/Drinking	-0.89*	(0.42)	-0.83	(0.80)
Watching TV/On the computer	-0.25	(0.31)	-0.03	(0.49)
Using hands-free headset^	-0.20	(0.31)	0.03	(0.43)
Poor sound quality for the call	0.20	(0.19)	-0.29	(0.35)
Education^				
Trade school or some college	-0.57*	(0.28)	-0.47	(0.46)
Bachelor's degree	0.06	(0.24)	-0.17	(0.40)
Graduate school degree	0.40	(0.23)	0.29	(0.37)
Age^				
35 to 64	0.01	(0.23)	0.67	(0.44)
65 or older	-0.30	(0.28)	0.20	(0.55)
Sample size	(236)		(236)	
Re-scaled r-square	0.09		0.16	

***p <.001 **p <.01 *p <.05
^Reference categories: called on landline, respondent at home, not using headset, high school education or less, age 18 to 34.

I also found that responding away from home was related to use of another cognitive shortcut, but in the opposite direction from what as expected. Table 3.5 showed that cell phone respondents appeared to be somewhat less likely than landline respondents to item nonrespond in Wave 2 (B=-0.23, s.e.=0.12, p=0.056), but it was not clear why. The results in Table 3.13 help to answer this question. The model on the left predicts item nonresponse in Wave 2 (once or more) from device, the hypothesized mechanisms, and common demographic correlates. As in Table 3.5, a measure of respondent interest in the survey topic is also included, given that people with greater interest in the subject were much more likely to answer every question (as evidenced by the significant negative coefficient).

Table 3.13. Two Regression Models Predicting Item Nonresponse

	Logistic model		Negative binomial model	
	1=Item nonresponded at least once 0=Answered every item		Number of items not answered	
	Estimate	s.e.	Estimate	s.e.
Intercept	-0.62	(0.46)	-0.16	(0.27)
Called on cell phone^	-0.15	(0.13)	-0.23	(0.12)
Away from home^	-0.38*	(0.18)	-0.25	(0.20)
Multitasking indicators				
Eating/Drinking	-0.18	(0.31)	-0.27	(0.16)
Watching TV/On the computer	0.04	(0.25)	0.07	(0.12)
Using hands-free headset^	0.12	(0.24)	0.03	(0.25)
Poor sound quality for the call	-0.23	(0.16)	-0.07	(0.16)
Education^				
Trade school or some college	0.08	(0.20)	0.03	(0.15)
Bachelor's degree	-0.34	(0.20)	-0.43*	(0.17)
Graduate school degree	-0.09	(0.21)	-0.32	(0.17)
Age^				
35 to 64	-0.29	(0.17)	-0.03	(0.16)
65 or older	0.50*	(0.23)	0.47*	(0.18)
Interest in the survey topic	-0.48***	(0.14)	-0.26***	(0.06)
Sample size	(356)		(356)	
Re-scaled r-square	0.14		n/a	
Lagrange multiplier test	n/a		X^2=13.1, p <.001	

***p <.001 **p <.01 *p <.05

^Reference categories: called on landline, respondent at home, not using headset, high education school or less, age 18 to 34.

This more comprehensive model suggests that the weak device effect shown in Table 3.5 may be attributable to many cell phone respondents being away from home. Over two-thirds (68%) of those reached at home failed to answer at least one question, as compared to less than half (47%) of those reached away from home. It is not clear why being away from home is associated with better response behavior on this metric. One possibility is that people who respond away from home are actually more committed to the task of responding than those reached at home. People with low commitment or motivation may consent to the interview when reached at home (e.g., out of boredom) but refuse when reached away from home. This logic runs counter to what researchers have speculated about location effects and would benefit from further empirical testing. A competing explanation is that respondents reached away from home tend to answer every question because their threshold for response accuracy is lower. That is, respondents reached away from home may not care as much that their answer is correct as those reached at home. If this is true, then a negative effect from device would be expected to manifest is other tests, such as the recency effect experiments, but generally it did not.

The model on the right side of Table 3.13 evaluates item nonresponse somewhat differently. It models the number of items that the respondent failed to answer, as opposed to the logistic regression model on the left which treats item nonresponse as a binary outcome (once or more versus never). The negative binomial model[26] on the right is useful because it leverages information collected about the number of times respondents failed to answer. The results of the negative binomial model are similar to those of the logistic regression. The elderly and the less educated are more likely to item nonrespond than young adults and those with a college education. The sign on the coefficients suggests that item nonresponse is less prevalent for respondents reached away from home or interviewed on a cell phone, but those two effects are not statistically significant. When the model is re-run without the cell phone indicator, the coefficient for being away from home and the coefficient for eating/drinking become statistically significant.

Overall, I find only scattered, inconsistent evidence that the hypothesized mechanisms of cell/landline device effects actually explain variation in use of cognitive

[26] A negative binomial model was used instead of a Poisson model for these count data because the variance was substantially greater than the mean. The Lagrange multiplier test is highly significant, indicating that the data are overdispersed and follow a negative binomial distribution rather than a Poisson distribution

shortcuts. Results of these tests are summarized in Table 3.14.[27] Being away from home was weakly associated with giving rounded answers and less attention to question wording, but it was positively associated with answering all of the questions in the survey. Similarly, eating or drinking during the interview was associated with less attention to question wording but positively associated with more demanding response strategies to a behavioral frequency question. Relatively poor sound quality seemed to impair responses slightly, but too few respondents reported problems with sound quality to measure this effect well. There was no evidence that using a hands-free headset affected the accuracy of responses.

The multivariate models indicate the danger of the hypothesized mechanism to response accuracy are minimal, especially compared to other factors. For example, age and education explain some of the variation in item nonresponse and differentiating among battery items. These associations were expected based on the literature, and they were statistically significantly in this study. The hypothesize mechanisms, by contrast, appear to be largely unrelated to the error indicators. People interviewed away from home, using a hands-free headset, or multitasking generally did not appear to be more likely to use the cognitive shortcuts. In other words, the successful replication of effects associated with age and education indicate that the experiment went as intended, and so the general absence of effects associated with the mechanisms should be taken seriously.

These findings reinforce the idea that the accuracy of cell phone response data is generally comparable to that of landline data. None of the tangible differences between cell phone and landline interviews showed a consistent relationship with the seven error

[27] Two of the cognitive shortcuts, selecting response options presented recently and attitude/behavior report correlations, are not discussed in detail in this section because there was no evidence that people were more prone to using these shortcuts when interviewed on a cell phone.

indicators. In fact, there is some intriguing evidence that some properties of cell phones may actually improve data quality. Additional research is needed, however, before this conclusion can be made definitively.

Table 3.14. Summary of Evidence That Four Hypothesized Mechanisms Explain Differential Usage of Cognitive Shortcuts

	Indicator of Cognitive Shortcutting						
Hypothesized Mechanism	**Attention to Important Exclusions**	**Estimating Responses**	**Recency Effect**	**Low Variance on Battery Qs**	**Attitude/Behavior Correlations**	**Length of Open-ends**	**Item Nonresponse**
Respondent interviewed away from home	**Mixed.** Rs interviewed away from home were less likely to attend to an exercise exclusion than Rs at home. The effect did not replicate for an internet question exclusion.	**Evidence.** Rs interviewed away from home were more likely to give a prototypical response than those at home.	**No evidence**	**No evidence**	**No evidence**	**No evidence**	**Evidence contrary to expectation.** Rs reached away from home were more likely to answer all the questions in the survey than Rs reached at home.
Mutlitasking during the interview	**Evidence.** Rs who were eating or drinking were less likely to attend to both the exercise and internet exclsuions.	**Evidence.** Rs who ate or drank during interview were less likely to estimate than others.	**No evidence**	**No evidence**	**No evidence**	**No evidence**	**No evidence**
Use of hands-free headset*	**No evidence**	**No evidence**	**No evidence**	**No evidence**	**No evidence**	**No evidence**	**No evidence**
Audio fidelity*	**Mixed.** Rs reporting poorer sound quality were less likely to attend to the exercise exclusion than Rs reporting better sound quality. The effect did not replicate for the internet exclusion, and sample sizes were small.	**No evidence**	**Mixed.** Poor audio quality is positively associated with a recency effect for one question, but shows no association on two other questions.	**No evidence**	**No evidence**	**No evidence**	**No evidence**

*The number of respondent who were using a hands-free headset or reported poor sound quality was quite small, limiting the power of these tests.

3.4 Conclusions

Several differences between landline and cell phones have led survey researchers to question whether the quality of data from cell phone interviews is as high as that from landline interviews. Cell phone respondents may be reached away from home in a distracting environment, they may be wearing a hands-free headset and doing other activities while talking, or the audio fidelity of the call may be poor. Previous studies, however, have found virtually no evidence that cell phone respondents were providing less accurate answers than their landline counterparts. The results from these studies were limited in that the effect of the device (landline versus cell) was confounded with differences in the composition of the comparison groups. This study was designed to break the confound by randomly assigning respondents to be re-interviewed on either their landline or cell phone. A special instrument was developed to capture two sets of factors: (1) seven indicators of cognitive shortcutting were used to determine whether respondents are, in fact, less likely to respond carefully on a cell phone as compared to a landline and (2) hypothesized mechanisms that could explain why such a device effect may exist.

The results from this study are generally consistent with the literature, with a few intriguing exceptions. As in previous studies, there is no strong or consistent evidence that the answers respondents give when interviewed on a cell phone are inferior to those given when interviewed on a landline. Of the seven cognitive short cut indicators tested, there was no evidence of a device effect for four indicators, weak evidence for two, and strong evidence of a device effect for only one. This is good news for dual frame survey designers who combine data from landline and cell samples to produce estimates.

Results from this study suggest that differences in response accuracy between such samples are likely to be small, especially for questions that are not cognitively demanding. I remind the reader that censoring of responses was beyond the scope of this study, though it is addressed by Lavrakas and colleagues (2007) and Brick and colleagues (2007).

The places where there was evidence of a possible device effect, however, merit attention. A split-ballot experiment indicates that respondents were less likely to pay close attention to an important exclusion in a question when interviewed on a cell phone versus a landline. This implies that cell phone respondents may have been distracted or putting forth less effort in interpreting the question. There was also weak evidence that cell phone respondents were less likely to differentiate among items in a battery and more likely to give short open-ended responses. These indicators suggest that respondents' propensity to do the minimal amount of cognitive work needed to answer the question (satisfice) may be somewhat greater when interviewed on a cell phone than a landline.

The external validity of these results extends to most of the cell phone population in the U.S. but not all of it. The randomization component of the experiment necessitated focusing on adults with both a landline and a cell phone (dual users). The experiment did not include adults who are cell-phone-only. Fortunately, data from Wave 1 were available to assess whether cell-only respondents differ from dual users on several of the cognitive shortcutting indicators evaluated in the study. The comparisons from the first survey show no meaningful differences. Cell-only and dual user respondents did not differ with respect to item nonresponse rates, degree of differentiation among items in a battery, or providing rounded, prototypical answers to a behavioral frequency question.

This supplementary analysis suggests that the exclusion of the cell-only population had minimal impact on the external validity of the study.

Unfortunately, our understanding as to why measurement error may be greater in cell phone interviews remains fairly limited. Researchers have speculated that being away from home, multitasking, or subject to other distractions may inhibit cell respondents' ability to respond carefully. This research provides little empirical support, however, for these mechanisms. Low-level multitasking and being interviewed away from home are associated with more error on some indicators but less error on others. Two of the mechanisms, poor audio quality and use of a hands-free headset, showed little if any bearing on response quality.

The results from this study highlight the need for more research on this topic. There were numerous effects in the expected direction, but many did not reach statistical significance. The device comparisons featured fairly low statistical power due to essentially unavoidable attrition at multiple stages in the repeated measures study. Future studies with larger comparison groups may provide more clarity on some of the marginal findings reported here. Another limitation of this study is that the risk of measurement error is assessed using proxies or indicators of cognitive short cuts. By their nature these indicators are imperfect as they fail to capture some respondents who give inaccurate answers but are not detected (false negatives) and incorrectly implicate others who gave accurate responses, such as rounded numbers, that just happened to appear incorrect (false positives). The direction of these errors in the indicators are not likely to be related to the device used in the interview, but the errors would most certainly dilute the effect that we are seeking to measure. A record-check study would facilitate direct computation of measurement error and may be a useful follow-up to this measurement error research.

Chapter 4

Influences of Telephone Device on Survey Response Distributions

4.1 Introduction

Interviewing respondents on cell phones raises a set of measurement concerns that researchers are just beginning to address. The emphasis of empirical work to date has been on determining whether the quality of data collected in cell phone interviews is lower than that collected with other modes/devices, particularly landlines (Brick et al. 2007; Kennedy 2007a; Lavarakas et al. 2007; Steeh 2004, Witt, ZuWallack, and Conrey 2009; Chapter 3). Fortunately, results suggest that cell phone data quality is generally comparable to that from landline surveys. Other measurement concerns, however, have yet to receive serious attention. One such issue is whether the mobility of cell phone respondents can influence response distributions. If respondents lack a crystallized, pre-existing attitude on a survey topic, they may use cues from their environment in formulating their response. To the extent that cues encountered away from home differ from those at home, people may answer some items differently depending on their physical location. That is, respondent location could affect the *mean* of certain response distributions. A related issue is whether the *variance* of survey estimates is greater when data are collected over cell phones versus landlines. The mechanisms leading to greater variance in cell phone interviews could be either a location effect or differential cognitive effort. This study addresses these gaps in the literature regarding the influence of device (cell phone/landline) on response distributions.

4.1.1 The Effect of Environmental Cues on the Mean of Survey Measures

The majority of questions commonly posed in surveys are almost certainly unaffected by whether the respondent is interviewed at home versus another location. People generally report the same educational history, financial circumstances, political ideology, and so forth when interviewed one place or another. There are, however, a set of question domains that could conceivably be sensitive to respondent location. These include health status, evaluations of one's community, and level of satisfaction with one's social life, to name a few. For example, respondents may be more likely to report having smoked 100 cigarettes in their lifetime (as in the Behavioral Risk Factor Surveillance Survey) if they are interviewed while having an occasional smoke at a social outing than if they are reached at home where, perhaps, they never smoke. That cell phone respondents sometimes participate *away* from home whereas landline respondents almost always participate *at* home may introduce a systematic influence for certain measures.

The possibility that location can affect how a respondent answers a survey question is suggested by psychological research on survey measurement. A foundational piece is the four-step model of the cognitive processes involved in answering a survey question (Tourangeau 1984). Tourangeau notes that answering entails comprehension of the item, retrieval of relevant information, use of that information to make required judgments, and finally the selection and reporting of an answer. He and other researchers have demonstrated that the second and third steps – retrieving relevant information and making a judgment – can be influenced by the cues used for the task. The most extensively researched cues are those provided by earlier questions in the interview, often immediately preceding the target item (Bishop, Oldendick, and Tuchfarber 1982; Hyman and Sheatsley 1950; Shuman and Presser 1981; Tourangeau et al. 1989). Numerous split-ballot experiments have shown that such contextualizing questions can change the distribution of responses to the target item, although the effect is sometimes elusive (Schwarz and Sudman 1992). Respondents also may derive information from the response scale to formulate their answer (Schwarz et al. 1985). Respondents assume that the options presented reflect the researcher's knowledge of the distribution of opinions or behaviors in the population, and they proceed to use the scale as a frame of reference in estimating their own behavior (Schwarz and Bienias 1990).

Critically, the cues that respondents use to formulate responses can come from a variety of sources, not just the questionnaire. Characteristics of the interviewer, the climate of opinion at the time of the survey, the stated purpose of the survey (Galesic and Tourangeau 2007), the respondent's mood (Schwarz and Clore 1983), and even the weather at the time of the survey (Schwarz and Clore 1993) can serve as response cues. In this study, I investigate whether respondent location belongs on this list and, if so, the implications for dual frame RDD surveys. The hypothesized model is that the set of stimuli (potential retrieval and judgment cues) that respondents encounter when interviewed away from home may differ in systematic ways from the stimuli that they encounter when interviewed at home, and some respondents use these cues in formulating their response.

An effect from location would be most likely to manifest itself on attitudinal questions for which respondents do not have a pre-existing, crystallized opinion. Without a predetermined attitude to report, respondents construct answers “on the fly” based on considerations that come to mind (Zaller and Feldman 1992; Tourangeau, Rips, and Rasinski 2000). Which considerations come to mind is determined by their momentary accessibility. Tourangeau and his colleagues (2000) argue that accessibility is influenced by both chronic and temporary factors. Factors documented in the literature include the wording of the questions, instructions given to the respondent (Ottati et al. 1989; Wilson, Kraft, and Dunn 1989), and the content of earlier questions (Tourangeau and Raskinksi 1989; Tourangeau et al. 1989). It seems plausible that factors external to the interview (e.g., location) could also be influential, if they are somehow related to the question at hand. Considering that location is to some extent a function of device, the responses of landline and cell phone respondents may differ systematically on items that are potentially sensitive to the respondent’s location.

Recent psychological and marketing research supports the hypothesis that environmental cues can influence real-world judgment and decision-making. Early studies investigated the influence of direct exposure to cues (e.g., Zajonc 1968). More recent work has shown that even indirect, incidental cues can have an effect, so long as they are perceptually or conceptually related to the target construct (Berger and Fitzsimons 2008; Lee and Labroo 2004; Monahan, Murphy and Zajonc 2000; Whittlesea 1993). For example, Berger and Fitzsimons (2008) found that respondents were more likely to recall orange-associated products, such as Sunkist and Orange Crush, when the color orange was more prevalent in the environment (the day before Halloween compared to several days after). Thus, cue exposure can not only affect attitudes towards the exposed object but also towards objects that share a conceptual link.

Importantly, researchers have documented directionality to priming effects. Objects are evaluated more favorably and chosen more frequently if people have been exposed to perceptually- (Moreland and Beach 1992; Nedungadi 1990) or conceptually-related (Berger and Fitzsimons 2008) cues. For example, mere exposure to people can increase ratings of their physical attractiveness (Moreland and Beach 1992). Psychologists have demonstrated that the mechanism underlying priming effects is the fluency of processing. Cues or primes can automatically activate representations in memory, leading them to become more accessible and to spread to related constructs through an associative network (Anderson 1983; Collins and Loftus 1975; Neely 1977). Priming a construct in memory leads to the spontaneous priming of related constructs in memory. Because ease of processing is often positively valenced (Harmon-Jones and Allen 2001), this process can lead to more positive evaluations and selection likelihood (Lee and Labroo 2004; Nedungadi 1990). These effects require that respondents attribute the ease of processing to positive qualities of the target rather than a spurious source. This attribution condition is similar to that examined in Schwarz and Clore's (1983) "feelings-as-information" study, which demonstrated that survey respondents reject information that they attribute to an inappropriate, transient source.

While attitudinal responses are perhaps the most susceptible to environmental cues, several studies suggest that responses to factual questions can also be affected by context (Strack, Schwarz, and Wankë 1991; Tourangeau and Rasinski 1988; Shuman and Presser 1981). Research to date has focused primarily on cues from the questionnaire, but it is possible that external, environmental cues could be used as well. If the meaning of the question is ambiguous, then respondents' interpretation may be informed by cues in their environment. For example, if someone without a computer at home is asked if they regularly use the Internet, they may respond in the negative at home but may respond in the positive if they are interviewed at work or some other place with internet access.

While there is theoretical and empirical support for environmental priming, this does not necessarily mean that survey estimates will be noticeably affected. An important moderating factor is that the circumstances of respondents reached away from home are generally quite diverse. In the study described below, cell phone respondents were interviewed boarding planes, carving deer meat, attending a baptism, unloading bricks from a truck, feeding horses, riding a bicycle, pumping gas, waiting in a hospital, shopping at Lowe's, getting a haircut, and watching soccer practice. This diversity reduces the chance that they will use a common response cue related to their surroundings. Furthermore, the experiences of respondents interviewed at home can be quite varied, which should also dampen any effect from location.

4.1.2 The Effect of Environmental Cues on the Variance of Survey Measures

The diversity of respondent activities may reduce the likelihood of a directional effect on survey estimates, but it may increase variances. This increase would be relative to interviews conducted in a more controlled environment. This hypothesis can be expressed using a measurement error model that assumes each respondent (j) has a true value (x_j) for a given survey item (Hansen, Hurwitz, and Bershad 1961). Respondents do not always report that value, however, because of random error (e_j) as well as other error sources. The variable response error (e_j) is often assumed to be normally distributed with a mean of zero. The situation is further specified to note the occasion of the interview or trial (t), which is relevant because response error varies from interview to interview.[28]

$$y_{jt} = x_{jt} + l_{jt} + e_{jt}$$

To this model, I have added the term l_{jt}, signifying the random error introduced by the location of respondent j during interview t. The hypothesis that I seek to test is that the random error associated with location (l_{jt}) is larger in cell phone interviews than in landline interviews of the same population. The measurement error model is presented because the error is best understood in a longitudinal framework where respondents are interviewed multiple times (t=1,2,…) under the same (or similar) essential survey conditions.

Two factors could explain this effect – location and cognitive effort. As discussed above, cell phone respondents are reached in myriad locations while landline respondents are almost always reached at home. To the extent that the respondent's environment provides cues for retrieval and judgment in answering questions, I expect that responses to certain items may vary as respondent location varies. In this chapter, I set aside the issue of whether or not the change in responses represents a change in the true value.[29] I seek only to document the association between location and variation in responses.

[28]More detailed models decompose the survey response further to account for effects from interviewers, correlations between various error sources, and other factors.
[29] For example, some people's perceptions about how satisfied they are with their job may depend on whether they are asked when they are working or when they are not working. In theory, the answers could differ yet both be accurate. This study does not attempt to determine whether one answer is more accurate than the other. Instead, the aim is just to determine if the answers differ.

The other mechanism that may explain an effect of location on response variance is differential cognitive effort. There is little evidence to suggest that respondents are less willing to work hard (expend cognitive effort) when they are interviewed on a cell phone rather than a landline (Steeh 2004; Brick et al. 2007; Kennedy 2007a; Witt, ZuWallack, and Conrey 2009), but this research question is far from closed (see Chapter 3). If we assume for the moment that people tend to respond less carefully on cell phones, then the implication for the variance of estimates arguably could go in either direction (higher variance in cell phone surveys or lower). Seam effect studies suggest that a lack of cognitive effort leads to lower variance in responses because reporting the same answer over and over is easier than carefully formulating the correct answer each time (Rips, Conrad, and Fricker 2003). This logic may apply even in studies that do not use retrospective reports. All that is required is for the path of least resistance to point to the same response each time. The response order literature suggests that in telephone surveys, respondents putting forth minimal effort may favor options read more recently. If the order of the options is not randomized and respondents use this cognitive shortcut, then the variance of estimates may be artificially low.

It could also be argued, however, that lower cognitive effort in cell phone interviews would yield higher variances, relative to landline interviews. Researchers have long speculated some respondents select their responses randomly from the list of available options (Converse 1964, 1970, 1974). Random responding is more likely when the respondent does not know the answer or is not motivated to identify the correct response. If random responding is more prevalent in cell phone interviews than landline interviews, then the variance of estimates based on cell phone data are probably higher.

To study the effect of location – and, by extension, device – on response distributions, I carried out a unique, repeated measures experiment. The experiment began with a national, dual frame random digit dial (RDD) survey, referred to as Wave 1. The questionnaire included a battery of six items expected to be sensitive to respondent location. Two months later, a subset of the Wave 1 respondents were re-interviewed in an ostensibly unrelated survey (Wave 2). The Wave 2 questionnaire also contained the six-item battery, but otherwise featured a different topic and sponsor. Critically, the participants in Wave 2 were randomly assigned to be contacted for that interview on either their landline phone or their cell phone. Eligibility for Wave 2 was contingent upon being a dual user (having both a cell phone and a working landline at home).

The randomization step and the re-interview make this study uniquely suited for addressing the change in means and change in variance hypotheses. The effect of device on the direction of responses is assessed using within-subject tests, because both landline and cell phone responses were collected from the same subset of participants. I address the issue of differential random error by comparing the consistency of answers from those interviewed on their landline twice to those interviewed on their cell phone twice, as well as those interviewed once on each device.

4.2 Research Design

4.2.1 Sampling Frames and Calling Protocols

The experimental design for this study is described in section 2.2.1. A critical feature of the experiment is that respondents were randomly assigned to be interviewed for Wave 2 on either their cell phone or on their landline. In total, 703 respondents for which both a landline and a cell phone number were known comprised the sample for Wave 2. The AAPOR(1) response rate was 56% in the Wave 2 landline condition and 48% in the cell phone condition. Some 61 of the Wave 1 landline RDD respondents completed Wave 2 on their landline phone, and 65 completed Wave 2 on their cell phone. Among the Wave 1 cell RDD respondents, 95 completed Wave 2 on their landline versus 139 on their cell phone. For the remained of the Chapter, I will employ a shorthand to refer to the four experimental groups. Cases sampled through the landline RDD frame and called for Wave 2 on their landline phone are the “landline-landline” cases, while those sampled through the landline RDD frame and called for Wave 2 on their cell phone are the “landline-cell” cases. The cases sampled through the cell RDD and called for Wave 2 on their landline or cell phone are the “cell-landline” and “cell-cell” cases, respectively.

Response outcomes to Wave 1 were similar to other national dual frame RDD surveys (Keeter et al. 2008) with a 20% response rate in the landline sample and a 19% response rate in the cell phone sample (both AAPOR RR(3)). Details about nonresponse in the study are reported in Chapter 2. While the attrition in the study is nontrivial, it is also unlikely to compromise the analysis presented in this chapter seriously. For nonresponse to be a problem, there would need to be a connection between the reasons for not responding (in Wave 1 or Wave 2) and the response behaviors examined. Whether respondents use different retrieval and judgment cues when being interviewed on a cell phone versus a landline has no obvious connection to nonresponse mechanisms (e.g., topic interest or phone usage level).

Instability in responses, however, could be related to nonresponse if response propensity is correlated with motivation to respond accurately (Biemer 2001; Cannell and Fowler 1963; Groves and Couper 1998; Olson 2006). People with lower response propensities would also be less likely to respond carefully. This would mean that the respondents studied here (who completed both interviews) were more interested in responding carefully and, thus, less likely to have large response errors. Critically, there is no obvious reason to suspect that the factors determining response propensity are related to the interaction between device and random response error. The nonresponse that occurred in this study possibly made the second dependent variable (random response error) harder to detect, but not in a way that favors one device over the other.

4.2.2 The Questionnaires

The survey instruments used for this study reflect a trade-off between capturing the construct of interest and making the results as generalizable as possible to other cell phone surveys. To measure the influences of device on response distributions, it was helpful to have as many items as possible administered in both waves to facilitate within-subject comparisons. Repeating much of the content, however, would have tipped off respondents to the fact that Wave 2 was, in fact, a re-interview. Only one battery of items was administered twice (along with age and gender in the demographics section). The battery was worded as follows.

Next, I'd like to ask you about some different aspects of your life. For each one, please tell me whether you are very satisfied, somewhat satisfied, not too satisfied, or not at all satisfied. First, how would you rate your satisfaction with…?

a. The condition of roads and highways where you live
b. Traffic conditions on roads where you live
c. Your personal safety from crime where you live
d. Your social life
e. Your physical health
f. Your family life

These particular items were designed to be sensitive to respondent location. Survey questions about other topics would not necessarily be expected to yield the same results as those recorded in this study. The first three items were selected because they address constructs that may be visible from a cell phone respondent's location or at least more salient if the respondent is away from home. For example, respondents may have the general impression that local roads are in good condition, but if they are interviewed while traveling down a particularly uneven street, then that experience may color their response. The last three items were used because they tap constructs about which respondents may not have a stable, crystallized opinion. For example, how people evaluate their level of satisfaction with their social life might change depending on whether they are reached while watching television at home or out visiting friends.

The expected direction of the device effect is clear for some items but unclear for others. For example, I expect that respondents will express greater concern about crime when interviewed on a cell phone than on a landline. The rationale is that people generally feel safer inside their home than outside it. This particular effect may interact with the time of day the interview is conducted, assuming that concern about being victimized outside the home is elevated at night. A location effect on reports of satisfaction with family life, however, could conceivably work in either a positive or negative direction. Family dynamics may seem more satisfying when the respondent in reached outside the home if, as the adage goes, "absence makes the heart grow fonder" or, perhaps, these dynamics feel more satisfying when the respondent is reached at home in midst of enjoyable interactions with family members. Unfortunately, there is no prior research that I know of reporting the effects of location on the responses to these questions. I do not discount the possibility that competing directionality predictions may be equally (or even more) appropriate.

The six-item satisfaction battery is also used to address response variance. The hypothesis being tested is that response instability is greater in cell phone surveys than in landline surveys of the same population. Changes in true scores on these items are expected to be uncorrelated with device. That is, Wave 2 respondents interviewed on their cell phone are assumed to be no more likely to have a change in their opinion than those interviewed on a landline. It is somewhat unfortunate that the same battery must be used to study response instability and changes in the direction of responses, but these were the only substantive items administered in both waves. Fortunately, the experimental design provides protection against contamination of this random response error analysis. One group of study participants (landline-landline) was interviewed on their landline phone in both surveys, while another group was interviewed on their cell phone in both surveys (cell-cell). The cases are used to compare the stability of responses in landline versus cell phone interviews. Thus, even though a common set of items is used to investigate changes in the both the direction and variance of responses, the cases used for these two analyses are mutually exclusive.

The other key construct measured in both questionnaires is the respondent's location. At the end of each interview, respondents were asked whether they were reached away from home or at home. Responses to these questions are used to isolate the specific property of cell phones that I assert could influence response distributions, namely mobility.

4.2.3 Potential for Memory Effects

In the analysis presented below, responses to the first and second waves are treated as independent observations. While this may not be strictly true, precautions were taken to minimize any such dependency. The main threat to the independence of the responses is the potential for memory effects. During the Wave 2 interview, respondents may have remembered being asked the same question two months prior. Those who could recall their response in Wave 1 may have elected to repeat this response in Wave 2, to spare themselves the cognitive effort of formulating an answer for a second time.

Several measures were taken to make the Wave 2 survey appear independent of the Wave 1 survey. In the Wave 2 introduction, interviewers mentioned a different sponsor and topic. Even more important, perhaps, was the difference in content. Wave 1 would have been memorable for its focus on attitudes about the military. Wave 2, by contrast, focused almost exclusively on space exploration. Even if respondents suspected a connection between the two surveys, they were probably not certain that it was, indeed, a re-interview. Unfortunately, the manipulation check that was planned for the study – a question at the end of Wave 2 asking if they remembered taking a similar survey recently – was cut prior to data collection due to constraints on the length of the interview.

Critically, even if respondents had suspected that Wave 2 was a re-interview, the experimental design guards against a systematic effect on the analysis. The order of the interview was crossed with device, meaning that roughly half of the cell phone interviews come from Wave 1 and the other half come from Wave 2. Thus, any effect from device (landline/cell) is *not* confounded with interview order or, by extension, memory effects. Furthermore, if memory effects artificially increased the correlation between Wave 1 and Wave 2 responses, then this probably would have a conservative effect on the testing, making it harder to detect respondent-level change.

4.3 Results

4.3.1 Respondent Location

The general model being tested in this study is that the data collection device (landline/cell) dictates respondent location, and location determines the set of environmental cues that may prime certain considerations and ultimately influence responses. In this model, location is a conditional variable that determines the stimulus, which is the set of environmental cues. In order for differential environmental cuing to be a serious concern, a substantial proportion of the cell phone respondents would need to be away from home during the interview. Across the two waves, nearly one-third of respondents interviewed on a cell phone reported being away from home. In Wave 1, which is most generalizable to one-off dual frame telephone surveys, some 35% of the cell RDD sample respondents were interviewed away from home. The results are displayed in Figure 4.1.

If the sampling frames were perfect (with respect to the correspondence between the type of phone to which a number is *believed* to service and the type which a number *actually* services), then all landline respondents would be reached at home. The two most likely explanations for the deviation are porting and call forwarding. Another explanation could be that the respondent took the survey in a home, just not the place they consider their own home. This scenario implies an error in the within-household adult selection. Another possibility is reporting error to the location question. Some landline respondents may have in fact been at home, but for some reason reported otherwise. For Wave 2, about half of the landline numbers dialed were obtained through self-reports obtained near the end of the Wave 1 interview. Wave 1 respondents may have misreported (intentionally or unintentionally) their landline number and provided a number that actually belonged to a cell phone. Unfortunately, I lack the necessary information to determine what exactly led to some landline respondents reporting that they were reached away from home.

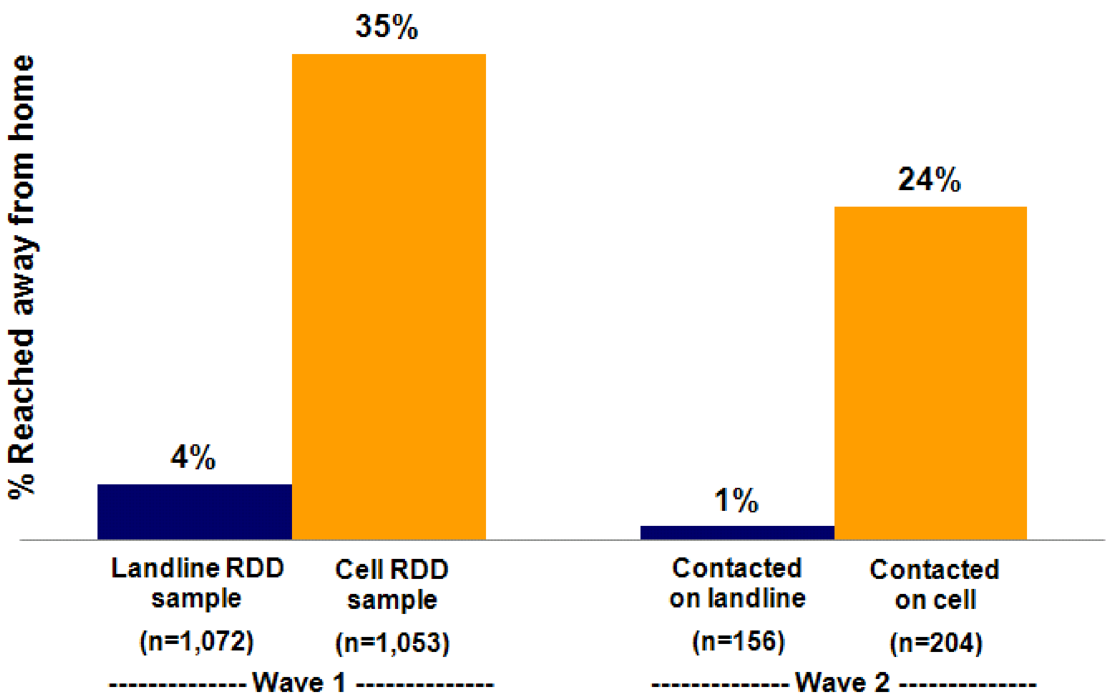

Figure 4.1. Proportion of Respondents Reached Away from Home, by Device

The key finding, though, is that a sizable proportion of cell phone respondents are interviewed away from home. Based on the current study and prior work by Brick and his colleagues (2007), a reasonable estimate is that about 35% to 45% of cell respondents in dual frame surveys are interviewed away from home. Had only a trivial fraction of the cell phone respondents been in a location different from the landline respondents, then there would be no serious risk to response distributions, even if the effect from location was strong.

4.3.2 Influence of Location on Responses to Six Experimental Questions

Having established device differences in respondent location, I next investigate whether location influences responses to some questions. In this section of the analysis, I consider only respondents who were interviewed on both devices, irrespective of order (*i.e.*, the landline-cell and cell-landline cases). The goal is to isolate the effect of the device (specifically the mobility of cell phones) by comparing responses given when respondents were interviewed on their landline versus the responses given when the *same respondents* were interviewed on their cell phone. These answers are combined to compute a single set of landline interview results. The answers to the cell phone interviews are combined across waves in an analogous fashion. In total, 146 study participants were interviewed on both devices. Of them, about one-third (n=50) were interviewed away from home in at least one of the interviews. Nearly all (92%) of the interviews conducted away from home were on a cell phone, as suggested by Figure 4.1.

There is modest evidence that respondents answer several of the items differently when interviewed at home on a landline versus away from home on a cell phone. Figure 4.2 shows the response distributions for the six experimental items when respondents were interviewed on their landline (gray bars) and on their cell phone (black bars), among those who were ever reached away from home. When interviewed on their cell phone, more than two-thirds (68%) said that they were "very satisfied" with their social life. When these same people were interviewed on their landline phone, only half (52%) said they were very satisfied with their social life. I tested the statistical significance of this difference using the Wilcoxon signed-rank test because the data are paired and the difference on these four-point measures is ordinal. The device difference for the social life item is significant (p<.05) despite a case base of only 50 respondents.

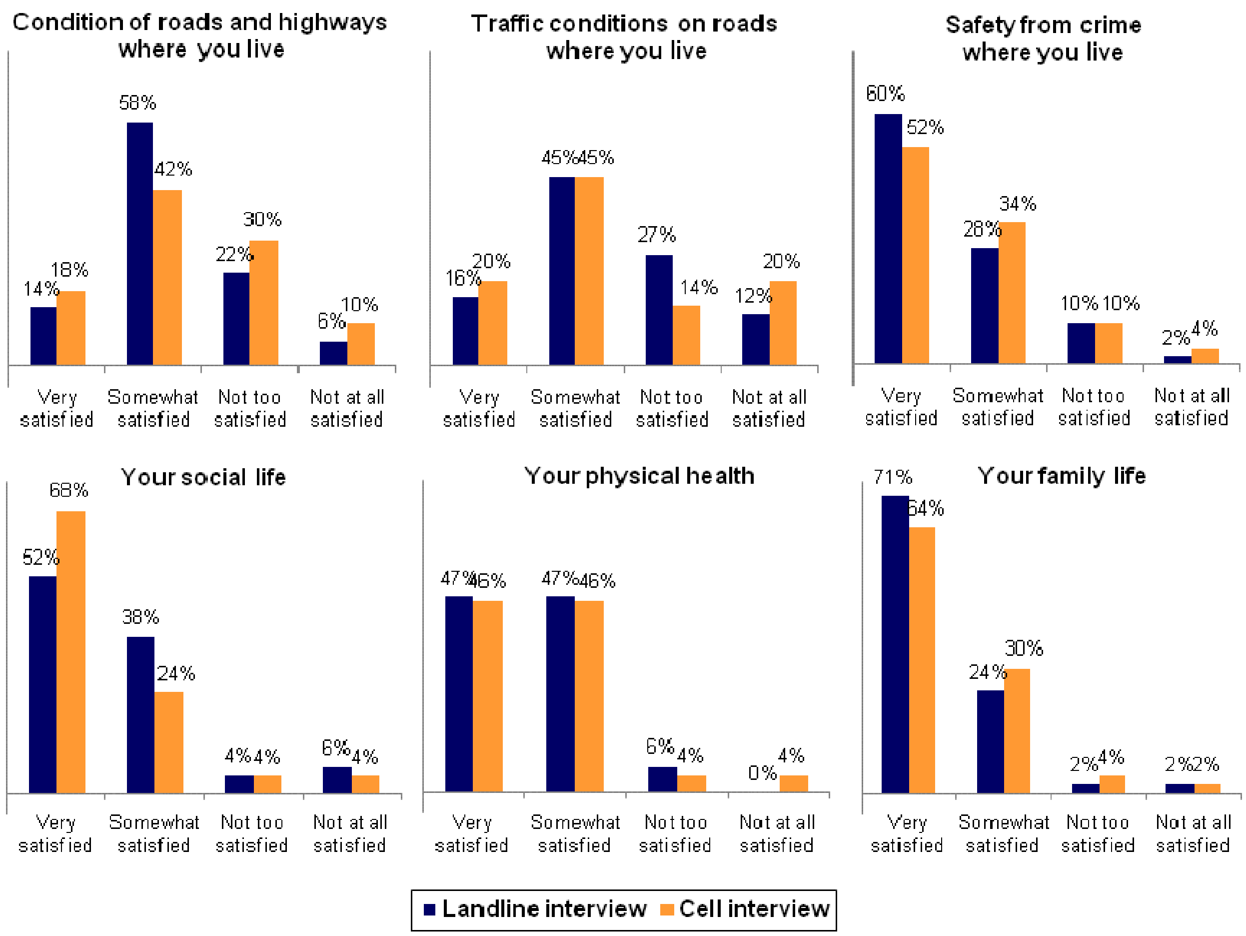

Figure 4.2. Repeated-Measures Results for Six Items Potentially Sensitive to Respondent Location, by Device
Note. - Respondents are represented twice in each of the six graphs (their landline response and cell phone response). Each marginal summing to 100% is based on 50 respondents, each of whom was interviewed away from home in Wave 1 or Wave 2. The significance level from the Wilcoxon signed-rank test is *p*=.21 for the condition of roads and highways, *p*=.88 for traffic conditions, *p*=.28 for safety from crime, *p*=.03 for social life, *p*=1.00 for health, and *p*=.79 for family life.

The ratings of satisfaction with road conditions and safety from crime also seem to be affected by the respondents' location, but the effects are weaker and cannot reliably be detected with the available case sizes. When interviewed on their cell phone, some 40% said they were not too satisfied or not at all satisfied with the condition of local roads and highways, but fewer (28%) expressed dissatisfaction when these same respondents were interviewed on their landline (p=.21). In addition, people appear to be somewhat more likely to report being very satisfied with their safety from crime when interviewed on a landline (60%) than when they are interviewed on their cell phone (52%), but again the difference does not reach statistical significance (p=.28).[30] As for the other three items in the battery, there is no clear indication of an effect from device on evaluations of local traffic conditions, personal physical health, or family life.

Another way to evaluate these data is to disregard the device comparison and, instead compare responses given when the respondents were at home with responses given when the same respondents were away from home, regardless of device. An advantage of this alternate approach is that it allows for the inclusion of cases in the cell-cell condition. When the analysis in Figure 4.2 is re-run based on all 94 cases interviewed once away from home and once at home, the effect of location looks to be similar if somewhat weaker. The *direction* of the differences for evaluations of one's social life (greater satisfaction reported when away from home), the condition of local roads (greater satisfaction reported at home), and safety from crime (greater satisfaction

[30] I sought to test for an interaction between device and time of day on responses to the safety from crime item. Presumably, cell respondents interviewed away from home after dark would express less satisfaction with their safety than if they were interviewed during the day. Unfortunately, only 23 of the subjects interviewed on both devices were ever reached away from home after dusk. I did not detect a significant device effect when looking at this small subset, but I do not rule out the possibility that an effect would be detected with a larger case base.

reported at home), all are present when the larger base of 94 cases is used, but the magnitude of the differences falls short of statistical significance (Wilcoxon signed-rank test p=.24, p=.22, p=.78, respectively). Interestingly, there is indication that respondents report greater satisfaction with their physical health when they are interviewed away from home (52% very satisfied) as opposed to at home (45% very satisfied) though this result also falls short of the traditional significance threshold (Wilcoxon signed-rank p=.11).

These results suggest that respondent location (at home versus away) may influence responses to some survey questions, but the scope of topics for which this issue arises is likely quite narrow, and low statistical power limits my ability to draw definitive conclusions. Importantly, the analysis above does not address whether the differences shown are large enough to influence survey estimates based on a full sample. The results reported above are based on the subset of respondents who were interviewed away from home during either Wave 1 or Wave 2. Focusing on those cases allows us to address the theoretical question of whether an effect from location ever exists at the level of the individual respondent. From a practical standpoint, however, researchers need only be concerned with this issue if it actually moves survey estimates. The likelihood that location differences could noticeably affect a mean, for example, is a function of the size of the effect and the proportion of respondents reached away from home on their cell phone.

To address this larger question, I compared the cell and landline responses for all 149 respondents interviewed on both devices. When this larger base is used, the differences in landline versus cell phone response distributions decrease (because the scope is no longer limited to those reached away from home in the cell interview), but the statistical power is greater. Looking at all participants interviewed on both devices, the tendency for people to express less satisfaction with the condition of local roads and highways on a cell phone versus a landline becomes marginally significant (p=.09). The tendency for people to report greater satisfaction with their social life in a cell phone interview, however, loses its significance. None of the other items showed a significant difference between cell and landline responses when the larger base is used. The fact that respondents interviewed away from home are the minority in cell phone samples, coupled with the modest (at best) effect from location, equates to little or no device effect on the distribution of responses for most items. That said, the differences observed for the social life and road condition items do suggest that location may matter for a small set of items, especially if cell phone samples become a larger portion of dual-frame designs and/or more people consent to taking surveys away from home.

4.3.3 Respondents in the Landline-Landline and Cell-Cell Conditions

The analysis of Figure 4.2 assumes that the differences observed are attributable to the change in device as opposed to other factors. To check this assumption, I analyzed responses to those interviewed on the same device in Wave 1 and Wave 2. That is, I replicated the analysis for those assigned to the landline-landline and cell-cell conditions. I expected that there would be no differences between the Wave 1 and Wave 2 means on the battery items for these respondents. This was borne out in the results, with the exception of one statistically significant difference, which may be attributable to random error.

In the landline-landline condition, 55 respondents answered the battery items in both waves. There were no significant differences between their answers in Wave 1 versus Wave 2, based on Wilcoxon signed-rank tests for the six items. In the cell-cell condition, 126 respondents answered the battery items in both waves. For one of the six items, satisfaction with your physical health, the difference in the Wave 1 and Wave 2 responses was significant for this group (p<.01). The largest change was in the "very satisfied" category. In Wave 1, some 45% of cell-cell respondents reported they were very satisfied with their physical health, compared to 53% reporting this in Wave 2. One likely explanation for this difference is type I error, or random noise. Between the landline-landline and cell-cell conditions I performed 12 significance tests. Roughly speaking, I would expect about one significant difference among these 12 tests based on chance alone. A significant difference in the cell-cell condition was made somewhat more likely by the fact that the sample size in that group was about double that in two other treatment groups and double the sample size for the results in Figure 4.2. Critically, there was no a priori reason to anticipate this difference between Wave 1 and Wave 2 evaluations of physical health for the cell-cell group, and there is no obvious post hoc explanation for the difference. Based on these considerations, I conclude that between Wave 1 and Wave 2, the centrality of the response distributions did not change systematically in the conditions where the device was kept constant. A comparison between treatment groups on the Wave 1 versus Wave 2 response distributions is presented in section A.6 of the Appendix.

4.3.4 The Effect of Device on the Reliability of Responses

The analysis above addresses whether the type of telephone used in a survey interview can have a systematic, directional effect on the responses to certain questions. Another possibility is that the type of telephone can affect the variance of responses. I evaluate the effect of device on the variance of responses by comparing the consistency of Wave 1 and Wave 2 responses for the different treatment groups. The six-item satisfaction battery discussed above is one of the few items included on both questionnaires, and so it is the focus of this analysis as well. If the variance of responses is greater in cell phone interviews than landline interviews, then I would expect to observe several patterns in the data. Among those sampled from the landline RDD frame, I would expect greater agreement in Wave 1 and Wave 2 responses among those re-interviewed on a landline than among those switched to a cell phone in Wave 2. Also, I would expect to see an overall effect, whereby there is greater disagreement in wave-to-wave responses among those interviewed on a cell phone in both waves than among those interviewed on a landline in both waves.

In this study I attempt only to measure the effect of device on the instability of responses over time. I do not attempt to measure response biases or make definitive statements about simple response variance because the necessary conditions are not in place. Measurement of simple response variance assumes that there are no changes in the true values between the two interviews, that the first and second measurements are independent (e.g., no memory effects), and that the essential aspects of the measurement protocol are constant across the interviews. The switching of some respondents from one device to another is a potential violation of the third assumption. In fact, whether or not the device switch represents a change in the essential measurement conditions is exactly what I seek to learn from the experiment.

In light of these considerations, two alternative measures were used to gauge response instability. The first is the average (absolute) difference between Wave 1 and Wave 2 responses. The absolute values of the differences between the Wave 1 and Wave 2 responses were computed at the respondent level for all items in the battery. This yielded six variables with values 0, 1, 2, or 3 because the battery featured a four-point satisfaction/dissatisfaction scale and cases with missing values were dropped. I then computed the respondent-level mean of those absolute values. The treatment group means of those values are presented in Table 4.1.

There is no evidence that responses were less consistent in cell phone interviews than landline interviews. The average wave-to-wave change in response was not significantly different for the landline-landline and the landline-cell groups (0.48 and 0.40, respectively, t=1.35, df=107, p=.18). There is also no indication that wave-to-wave differences were larger for those interviewed on their cell phone in both waves (cell-cell) than those interviewed on their landline in both waves (landline-landline) (t=1.01, df=84.5, p=.32). A check of the disaggregated results shows that this pattern was generally consistent across all of the items in the battery; no single item is driving the results of these tests.

Table 4.1. Measures of Reliability in Wave 1/Wave 2 Responses, by Treatment Group[A]

Treatment Group	*N*	Mean Difference in Wave 1-Wave 2 Responses to Satisfaction Battery[B]		Mean of Gross Difference Rates for Satisfaction Battery Items	
Landline-Landline	(54)	0.48	(0.30)	38.6%	(0.07)
Landline-Cell	(55)	0.40	(0.31)	35.9%	(0.06)
Cell-Cell	(124)	0.43	(0.24)	37.8%	(0.04)
Cell-Landline	(79)	0.44	(0.30)	38.7%	(0.05)

[A] Standard errors are shown in parentheses.

[B] For each respondent, the difference in his/her Wave 1 and Wave 2 responses was computed for all six items. To create a summary measure for the battery, I computed the respondent-level mean of the absolute values of those six differences. The values presented in the table are the group means of the average differences. Higher values denote greater disagreement in responses across waves.

I also used the gross difference rate (GDR) as a measure of (un)reliability in responses across waves. This statistic has the advantage of being more familiar from the response error literature (Bailar 1968; Biemer et al. 1991; Brick and West 1992; Hansen, Hurwitz, and Pritzker 1964; O'Muircheartaigh 1991). The GDR is the proportion of cases in which the reinterview response is different than the response in the original interview. Originally, the GDR was defined for continuous variables (Hansen, Hurwitz, and Bershad 1961), but in recent years it has often been applied to dichotomous outcomes. The data in the current study are ordinal, collected with a four-point scale. They were dichotomized by collapsing categories so that the recoded variable was 1 if the Wave 1 and Wave 2 responses were identical and 0 otherwise. The GDR was computed for each of the six items in the battery, and the treatment group means across the six items are presented in the far-right column of Table 4.1. The results for the GDR are similar to the absolute differences. The level of inconsistency in responses across the two waves was at least as high for landline interviews as for cell phone interview.

I also tested the effects of respondent location and multitasking directly. There was no relationship with consistency of responses across waves. Respondents who were interviewed away from home in either survey had the same mean difference in Wave 1 versus Wave 2 responses as respondents who were reached at home in both surveys. Similarly, respondents who were engaged in various activities during the interview, such as shopping, eating, watching TV, or riding in a car, had mean Wave 1 – Wave 2 differences to the battery items that were comparable to respondents not engaged in these activities. In sum, there is no indication from this study that researchers should expect greater response variances in cell phone interviews than in landline interviews. There is no overall effect from device on the consistency of responses and no effect from respondent location or multitasking.

4.3.5 Alternative Explanations for Micro-level Changes in Response

The comparisons between landline and cell phone data discussed above rely on the assumption that any differences between the groups can be attributed to the device (specifically, the mobility permitted by cell phones) or random variation. Other factors that could potentially explain the differences are changes in respondents' "true scores" between Wave 1 and Wave 2 and differences in the Wave 1 and Wave 2 questionnaires. For example, a respondent may have fallen ill just prior to the Wave 2 interview and consequently reported less satisfaction with his/her health in that survey than in Wave 1. There is no reason to believe, however, that such changes in true scores would be systematically related to the sequence of devices used in this study (e.g., whether the respondent happened to be interviewed on their landline in Wave 1 or in Wave 2). Thus, changes in true score almost certainly do not account for the differences observed.

The difference in the Wave 1 and Wave 2 questionnaires is a somewhat more credible concern, but that too is unlikely to have affected the comparison. The battery of the six experimental items discussed above was administered using the exact same wording in both waves. In Wave 1, the battery was located toward the end of the survey after about 15 minutes of questions probing attitudes about the U.S. military. This included questions asking respondents to rate how well they thought certain words and phrases described service branches such as the Navy or the Air Force. They were also asked about their level of knowledge about the services, recruiting goals, and other issues. In Wave 2, by contrast, the satisfaction battery was the third question in the survey, after a general question about the way things are going in the U.S. and an item about the availability of jobs in the respondent's community.

Clearly, the content preceding the target battery differed in these two surveys, but it is unlikely that this affected the device comparison for two reasons. Most critically, treatment assignment (landline/cell phone) was crossed with the questionnaire. That is, about half of the landline responses represented in Figure 4.2 were collected in Wave 1 and the remainder were collected in Wave 2. Likewise, the cell phone responses in Figure 4.2 were collected in both waves. This means that any difference in the context of the six-item battery in Wave 1 versus Wave 2 would be nearly eliminated by the crossing of the device and questionnaires in the experimental design.

Furthermore, the measurement error literature suggests that context effects tend to occur only when the preceding contextual questions are related to the target item in some meaningful way. For example, context effects have been shown to occur when the preceding question raises considerations that could be relevant to the target question (Tourangeau and Raskinksi 1989; Tourangeau et al. 1989) or when the preceding question influences how the target question is interpreted (Hyman and Sheatsley 1950; Shuman and Presser 1981). In both Wave 1 and Wave 2, the preceding measures, by design, addressed topics unrelated in any obvious way to the six-item battery. Given the crossing of survey waves with the treatment and the relatively innocuous context used in both questionnaires, the possibility of a meaningful confounding effect contaminating these results is very small. The more important limitation appears to be the random variation associated with the small sample sizes available for this analysis. This variance makes it difficult to detect a significant effect from location, but it does not bias the comparison in any way.

4.4 Conclusions

Dual frame RDD surveys collect data from respondents interviewed in their homes as well as respondents interviewed away from home. This dynamic raises questions as to whether respondent location has implications for the responses recorded. Previous studies have found that cell phone data are generally of comparable quality to landline data (see Chapter 3), but they have not rigorously addressed whether the device affects response distributions. This experiment investigated the influence of device (landline/cell phone) on the direction of responses and their stability over time. Samples were drawn from both RDD frames, and respondents identified as dual users were randomly assigned to complete a second survey with overlapping content on either their landline or cell phone.

The results indicate that responses to some survey questions may, indeed, be influenced by the respondent's location. Respondents reported greater satisfaction with their social life and less satisfaction with the condition of local roads when they were interviewed away from home on their cell phone, compared to what they reported when interviewed on their home landline phone. The items used in this study were designed to provide a sensitive test, and three of six items showed at least weak evidence of an influence from location. Evaluations of infrastructure in the respondent's community and evaluations of their social lives appear to be on the list of topics where this issue is a concern. Future studies might consider looking for this effect in questions on employment, commuting, and vacationing – as these topics relate to being away from home.

For most survey estimates, though, differential environmental cues are unlikely to introduce bias for two reasons. The majority of cell phone respondents are interviewed in the same place as landline respondents (at home), and usually there is no reason to expect a correlation between the respondent location and the answers reported. Furthermore, the power of environmental cues is likely to be weaker than cues provided by the questionnaire, and many respondents would probably reject environmental cues on the grounds that they are spurious (Schwarz and Clore 1983).

In this study, responses were only affected by location when the respondent was reached on their cell phone *and* happened to be away from home. Results from this study and others indicate that about 35% to 45% of cell phone respondents are interviewed away from home. This incidence rate appears to be low enough that full dual frame sample estimates are generally unaffected by a location effect. In the future, though, the proportion of cell phone respondents reached away from home may increase, elevating the possibility that differential environmental cues may noticeably influence responses. This problem could be avoided by having interviewers call back respondents and interview them only at home, but this approach would need to be weighed against a possible increase in nonresponse error, assuming many callbacks would be unsuccessful.

While systematic error from environmental cues may be rare, this issue warrants further attention. This line of research may be informed by studies on ecological momentary assessment (EMA) (Stone and Shiffman 1994), which allows subjects to report repeatedly on their experiences in real time, in real-world settings, over time and across contexts. Numerous teams of researchers have successfully used cell phones to conduct EMA studies (for example, Collins, Kashdan, and Gollnisch 2003; Freedman et al. 2006; Galloway et al. 2008), though the error properties of the data are not fully understood. As health and social sciences researchers increasingly seek to measure mobile subjects/respondents, it is critical that researchers understand how the real world may affect the data collected.

The results of this experiment raise a cautionary, though not distressing, flag concerning the effects on survey data from interviews being conducted away from the relatively controlled confines of respondents' homes. There is some evidence that attitudes reported outside the home may be somewhat different from those reported inside the home, although they are not necessarily less accurate. In this study, respondent location is used as a proxy for the actual mechanism that would causes a change in response – differential environmental cues used in retrieving considerations and integrating response considerations. Studies designed to measure those cues directly would enhance understanding about the nature of locations effects and possibly help us to systematically identify the most susceptible survey topics

I also examined whether the stability of responses over time is greater in landline interviews than in cell phone interviews of the same population. I find no support for this hypothesis. The consistency of Wave 1 and Wave 2 answers was at least as high for those interviewed on their cell phone in both surveys as for those interviewed on their landline in both surveys. Unfortunately, only a small number of items were administered on both the Wave 1 and Wave 2 questionnaires, and those items were designed primarily to detect changes in the direction of the responses. Future studies should consider testing a greater variety of items, particularly those for which the respondents' true scores are unlikely to change between interviews. There is no compelling evidence that the variances of respondents' reports will be greater if they are interviewed on a cell phone instead of a landline, but more testing with a broader set of items would be helpful in definitively answering this question.

Chapter 5

Conclusion

5.1 Overview

The growth of personal communication devices and the decline of landline telephones have far-reaching implications for survey researchers. Landline samples must now be supplemented with cell phone samples in order to cover the general population. Alternative designs, such as address-based sampling also show promise (Fleeman 2009; Link et al. 2008, 2009; Norman and Sigman 2009; Skalland, Barron, and Wooten 2009), but telephone data collection is almost certain to remain a fixture in survey research. Understanding the error properties of cell phone data is critical to maintaining the credibility of well-designed telephone studies and producing accurate, reliable estimates from them.

Two error sources that merit research attention in cell phone surveys are nonresponse and measurement. Researchers have a limited understanding of whether estimates from dual frame RDD surveys are likely to be biased by high levels of nonresponse. Brick and his colleagues (2006) have documented a link between telephone usage levels and response propensity, but it is unclear what other factors influence participation decisions. With respect to measurement, evidence suggests that data quality from landline and cell phone sample is comparable. That finding, however, is based on studies that were not designed to isolate the effect from device (landline/cell), raising the possibility that results are confounded with differences in sample characteristics. A separate issue that has not been explored is whether people respond differently to certain questions (e.g., Are you a smoker?) depending on whether they are interviewed at home or away from home. Such differential response patterns could have important implications for dual frame RDD estimates. This dissertation addresses these gaps in the literature. The next three sections summarize how the questions were addressed empirically, what was learned, and the theoretical and practical implications.

5.2 Conclusions about Nonresponse in Cell Phone Surveys

A central aim of the study was to identify causal factors influencing people's decisions about participating in cell phone surveys. Many correlates of response decisions have been identified in the survey literature (e.g., race, urbanicity, age), but actual mechanisms underlying participation decisions are much more elusive (Dillman 1978; Groves and Couper 1998; Groves, Singer, and Corning 2000). There is reason to believe that some mechanisms (e.g., incentives, topic interest) affect cell phone survey decisions in the same way that they affect decisions in other modes, while a separate set of mechanisms may be unique to cell phone surveys (e.g., frequency of usage, type of calling plan). A repeated measures experiment was conducted to assess the impact of these factors on participation decisions to a cell phone survey. Details of the experiment are discussed in Chapter 2.

A key finding from the experiment is that several device-related mechanisms have a significant influence on response decisions, and the magnitude of the effects are on par with other mechanisms that have been documented in the nonresponse literature. In particular, using the cell phone as a primary channel of contact with the outside world and using the cell phone for non-telephony activities like texting and Web browsing are significant positive predictors of survey response. Several familiar correlates of nonresponse from other modes are also supported (e.g., positive effects associated with age and Caucasian race). Overall, the set of mechanisms influencing response decisions in cell phone surveys appears to overlap partially the set of predictors that tend to influence decisions in landline surveys.

The fact that how people use their phone affects their response propensity suggests that the survey nonresponse literature needs to be re-thought in the context of cell phones. The usage and primary contact information effects are inconsistent with theories espousing the importance of social exchange dynamics (Dillman 1978) and the respondent's valuation of salient survey features (Groves, Singer and Corning 2000). It could be that a social exchange model or a Leverage Salience Theory model aptly describes some people's response decisions, but they would be incomplete for others. Future theoretical work would benefit from acknowledging that all of these factors can be important for participation decisions.

There may be thresholds beyond which any one of these factors trumps all others. If the respondent holds the sponsor in extremely high esteem, then sponsor affinity may be the only factor that matters. If, by contrast, the sponsor is unknown and the topic is of marginal interest, but the person is accustomed to fielding numerous cell phone calls per day, then that conditioned behavior may be enough to achieve cooperation. This logic needs further model specification, but the underlying idea is that current models of survey response need to be expanded to explain the effects of telephone usage styles documented here and in Brick et al. (2006).

An unexpected result in the nonresponse study highlights the complexity of modeling response decisions. The original sampling frame (landline or cell RDD) continued to be a very significant predictor of response to the Wave 2 cell phone survey, even after controlling for demographics and cell phone-related attitudes and behaviors. This finding reinforces the idea that the probability of selection in a given frame is very different from the probability of response. Unfortunately, it is not yet clear why the probability of response differs, given that it is not explained by differential usage. If not differential usage, then what construct is the frame variable capturing?

To begin to answer this, it may help to recall that the frame effect operated at the cooperation stage and that it was not a significant predictor of response to the Wave 2 landline survey. One post hoc explanation is that landline RDD respondents are less trusting when they are called on their cell phone. It could be that when some people are called on their landline, they are comforted by the many years of experience they have talking on the landline with family, friends, and strangers alike. But when these people are called on their cell phone, they have relatively little experience on which to draw, and they may be less trusting that the caller's intentions are honorable. This explanation is merely speculative; other accounts are certainly possible. For social scientists focused on telephone surveys, it may be useful to identify the variable(s) that could be added to the cell phone response model to make the significant effect from sampling frame disappear. Such a variable could potentially be incorporated into weighting protocols.

An immediate practical concern is the threat to estimates from nonresponse bias. Two factors are likely to mitigate this concern as it relates to device usage – increasing cell phone adoption and independence from most survey measures. Cell phone ownership and usage is increasing. If European cell phone adoption models are informative for the U.S., then cell ownership will eventually plateau around 90% within ten years or so (Kuusela, Callegaro, and Vehovar 2008). As more and more Americans rely on a cell phone (or some other type of mobile communication device), the variance in usage currently observed in survey samples may very well decrease. In time, these device-related mechanisms may cease to be an important correlate of participation decisions. A second mitigating factor is that cell phone usage is unlikely to be strongly related with many variables commonly measured in surveys. This is especially true when one considers the strength of association remaining after controlling for demographics such as age, gender, race, and education.

That said, some variables are potentially at high risk of nonresponse bias. Brick and his colleagues (2006) point out that estimates of technology interest and usage may be overestimated with cell phone surveys, because people with higher values on these dimensions are disproportionately more likely to respond. In addition, Vehovar and Callegaro's findings (2007) suggest that estimates of social connectedness and social network sizes may be overstated with cell phone samples. These social health variables may, in turn, be related to indicators of mental and physical health. Depending on the strength of the relationships, it seems plausible that such variables may also be at risk of nonresponse bias.

The nonresponse chapter provides preliminary evidence that at least some of the nonresponse bias may be reduced though post-survey adjustments. Unfortunately, many of the variables used in the propensity adjustment are not available in cross-sectional surveys. Two of the most important components were using the cell phone number as contact information (or not) and non-telephony activities, such as texting and Web browsing. Future studies could experiment with these variables either as raking dimensions (if benchmark data are eventually collected) or in a more complex model-based adjustment that "weights up" infrequent cell phone users who respond to represent similar people who do not respond.

In addition to nonresponse adjustment, future studies should seek to refine theoretical models of nonresponse. Limitations of the repeated measures study analyzed in Chapter 2 represent areas for improvement. For example, nonresponse in Wave 1 may have distorted some measurements, especially those related to attitudes about surveys and telemarketing. Weighting simulations were conducted in an attempt to gauge the severity of this problem, but this approach provides only partial insight. An alternative would be to conduct Wave 1 using an in-person area-probability sample. This could potentially yield a much lower nonresponse rate in Wave 1 and also reduce the possibility that Wave 1 nonresponse is related to the mechanisms that are posited to influence response decisions to Wave 2. Assuming that regional variation in cell phone survey response mechanisms is not particularly large, then the cost of an area-probability Wave 1 survey could be reduced by sampling only within particular area, such as a state.

Another limitation of the nonresponse study described in Chapter 2 is the exclusion of several variables that may be important to response decisions. In particular, sponsor affinity, topic interest, and urbanicity are among the missing constructs. Preliminary plans for the dissertation study included these variables, but interview length restrictions and late changes to the questionnaire precluded their measurement. Future work could include these variables, which would be useful for understanding how Leverage Salience Theory (Groves, Singer, and Corning 2000) stacks up against competing models of participation decisions.

5.3 Conclusions about Measurement Error in Surveys

Chapter 3 addresses differences in measurement error between landline and cell phone interviews. Researchers have speculated that cell phone survey responses may be less accurate because of greater distractions from multitasking or the respondents' surroundings, inferior sound quality, and the presence of other people who may overhear responses to sensitive questions. Prior investigations compared responses from landline RDD respondents to those from cell RDD respondents. These investigations did not detect any meaningful differences in item nonresponse rates, the length of open-ended questions, or responses to questions with socially desirable answers – even when controls were in place for basic demographics like gender, age, and race. These studies were not entirely conclusive, however, because landline and cell RDD respondent samples are known to differ systematically with respect to more than just the device. It is possible that differences in sample composition confounded the (lack of an) effect from device.

This study attempted to reduce that confound substantially by randomly assigning people to be interviewed on either their landline or their cell phone. In addition to the randomization step, this study also extended the literature by evaluating several measurement error indicators that were new to cell phone research. In total, seven indicators of cognitive shortcutting were captured. Questions asking about the hypothesized causes of measurement error differences (multitasking, sound quality, and the like) were also included so as to better understand any observed device differences in the use of cognitive shortcuts.

The results from the measurement error tests were generally consistent with the literature, with a few intriguing exceptions. As in previous studies, there is no strong, consistent evidence that the answers respondents give when interviewed on a cell phone are inferior to those given when interviewed on a landline. Of the seven cognitive short cut indicators tested, there was no evidence of a device effect for four indicators, weak evidence for two, and strong evidence of a device effect for one. The one strong device effect was observed in an experiment testing respondent attention to question wording. Respondents were less likely to pay close attention to an important exclusion in a question when interviewed on a cell phone versus a landline. This implies that cell phone respondents may have been distracted or put forth less effort in interpreting the question. I also found weak evidence that cell phone respondents were less likely to differentiate among items in a battery and to give longer open-ended responses. These results suggest that respondents' propensity to do the minimal amount of cognitive work needed to answer the question (satisfice) may be somewhat greater when interviewed on a cell phone than a landline.

Unfortunately, empirical attempts to unpack the device effect into specific, causal mechanisms were not very successful. In theory, the effect from the device should disappear if the causal mechanisms (e.g. differential sound quality) are included in the model. In most instances, this cancelation did not occur. There are three possible explanations for this: low variance on the mechanism indicators, mis-measurement of the mechanisms, or the list of mechanisms is wrong or incomplete.

With respect to low variance, landline and cell phone respondents reported roughly equivalent levels of sound quality, and no respondent said that the sound quality was “poor.” Only 6% of cell phone respondents and an equal proportion of landline respondents reported using a hands-free headset. Landline and cell cases also did not differ greatly in their likelihood to be multitasking, though the cell phone respondents’ activities often sounded more distracting (e.g., shopping or working versus household chores). Only a handful of activities were common enough for their effects to be measured in multivariate analysis.

An alternative explanation is that the mechanisms were not measured with enough accuracy in this study. Indeed, analysis was based on respondents’ self-reports about what they were doing during the interview. It seems plausible that the self-reports understate the actual amount of multitasking that occurred. Respondents may have considered it disrespectful or counterproductive to acknowledge to the interviewer that they had been attending to another task throughout the interview. They may also have unintentionally left certain activities (e.g., watching television) out of their answer, under the false assumption that it was not important enough to mention. Future methodological studies could do more to measure the respondents’ environment and activities systematically. New cell phone applications, such as global positioning and physiological measurements, could potentially be leveraged for this task. Researchers would need to supply respondents with the necessary device and gain their consent regarding the data collection.

A third possible explanation is that the list of reasons why cell phone data may be less accurate than landline data is incorrect. Researchers have focused on distraction from multitasking, lower audio fidelity, and being out in public (Lavrakas et al. 2007), but these may not be the right mechanisms. It is possible that measurement error and nonresponse error stem from the same problem – many people are unaware that serious scientific research is being conducted with cell phone samples. Even if people participate, they may not be convinced that it is important for them to answer as accurately as they can.

The measurement error analysis was also limited in terms of small sample sizes and the reliance on indicators of measurement error (as opposed to direct computation). Several device effects were in the expected direction, but many did not reach statistical significance because the study was underpowered due to attrition at numerous stages in the experiment. Future studies with larger comparison groups may provide more clarity on some of the marginal findings reported here. Another limitation of this study is that the risk of measurement error is assessed using proxies or indicators of cognitive short cuts. One way to assess measurement error directly is through a record check study. A record check study would allow the researcher to compute the actual amount of error in each respondent report, as well as, the consequences for the mean squared error of the survey estimate.

Despite its limitations, the measurement error analysis has several implications for practitioners. On the positive side, the majority of the results support the conclusion in the literature that cell phone responses and landline responses can be equally accurate – if questions are not overly demanding. No device differences were found with respect to item nonresponse rates, correlations between attitudes and behavior, selection of response strategies, and recency effects. This suggests that researchers can transition from landline RDD designs to dual frame RDD designs without having serious concerns that measurement error will increase.

Results of several other tests, however, indicate that people may be more likely to misreport responses to cognitively demanding questions when interviewed on a cell phone, as compared to a landline. Multiple-item batteries and long questions containing important in/exclusions appear to be somewhat more problematic for cell phone respondents. The lesson for practitioners may be that dual frame questionnaires should be reviewed with an eye toward shortening question stems as well as multiple-item batteries. Branching may be a prudent alternative for questions that have important inclusions or exclusions. Additional empirical evidence is needed to confirm the device differences documented in this study. In the meantime, researchers should consider avoiding demanding questions in order to reduce the risk of measurement error, particularly in cell phone samples.

5.4 Conclusions about Effects from Respondent Location

The third major research question addressed in this study concerns the potential effects of respondent location. If respondents lack a crystallized, pre-existing attitude on a survey topic, they may use cues from their environment in formulating their response. To the extent that cues encountered away from home differ from those at home, people may answer some items differently depending on their physical location. That is, respondent location could affect the mean of certain response distributions. A related issue is whether the variance of survey estimates is greater when data are collected in cell phones versus landline interviews.[31]

The repeated measures experiment was particularly useful for investigating this issue. For this analysis, the critical group was the dual user interviewed once on their landline and once on their cell phone. A set of six questions was administered in both interviews. This permitted within-subject comparisons, which are statistically more powerful than between-subject device comparisons. The six items tested were purposefully chosen to be sensitive to the respondent's location. The results indicate that responses to some survey questions may, indeed, be influenced by location. Respondents reported greater satisfaction with their social life and less satisfaction with the condition of local roads, when they were interviewed away from home on their cell phone, compared to what they reported when interviewed at home on a landline. Responses were only affected by location when the respondent was reached on a cell phone *and* happened to be away from home. Not surprisingly, responses to some questions were unaffected. On questions about local traffic conditions, the respondent's physical health, and family life there was no effect from location on the answers reported.

The finding that location can influence responses to some survey items has important implications. Respondent location appears to belong on the list of potential cues that respondents use to formulate their response. This list includes the response scale (Schwarz et al. 1985), previous questions in the interview (Hyman and Sheatsley

---

[31] See Brick et al. (in press) for a related discussion about how nonsampling errors may be linked directly, sometimes causally, to sampling frame.

1950; Shuman and Presser 1981; Tourangeau et al. 1989), and interviewer characteristics (Schuman and Converse 1979), among other factors. Unlike the response scale or previous questions in the interview, however, a respondent's location is beyond researcher control. One way to exert control is to schedule callbacks with respondents in certain locations away from home. A callback protocol would ensure that respondents' environments are relatively uniform during the interview, but a downside is that it would likely increase interviewing costs as well as unit nonresponse.

As yet, such drastic measures do not appear to be necessary. The risk of differential respondent locations influencing survey estimates is a function of two factors. One is the proportion of respondents interviewed away from home. In current dual cell phone RDD samples, this proportion is approximately 35 to 45% (Brick et al. 2006; Chapter 4). The higher the proportion of respondents reached away from home, the greater the likelihood of an effect from location. This incidence rate appears to be low enough that full dual frame sample estimates are generally unaffected by a location effect. As cell phone usage increases and more respondents are reached away from home, this issue may become more serious.

The second factor moderating a location effect is the degree of heterogeneity in the environmental cues for respondents reached away from home. The more heterogeneous the cues, the less likely it is that mean and medians will be affected. If the environmental cues are wildly dissimilar, it is unlikely that location would have a systematic effect on responses. Conversely, if many of the respondents are, for example, interviewed in restaurants, then reported frequencies of cooking meals at home may be lower relative to the frequencies that would be reported if respondents were only interviewed at home.

The items used in this study were designed to provide a sensitive test, and three of six items showed at least weak evidence of an influence from location. Evaluations of infrastructure in the respondent's community and evaluations of their social lives appear to be on the list of topics where effects from respondent location is a concern. For most questions, however, it seems unlikely that there would be a correlation between the respondent location and the answers reported.

A related issue tested in this study is whether or not respondent location affects the variance of survey estimates. When variance is the outcome of interest, it is possible that heterogeneity in environmental cues would increase the likelihood of error. No evidence was found to support this hypothesis, however. The consistency of Wave 1 and Wave 2 answers was at least as high for those interviewed on their cell phone in both surveys as for those interviewed on their landline in both surveys. Unfortunately, only a small number of items were administered on both the Wave 1 and Wave 2 questionnaires. Future studies should consider testing a greater variety of items, particularly those for which the respondents' true scores are unlikely to change between interviews. From this study, we conclude that there is no compelling evidence that the variances of respondents' reports will be greater if they are interviewed on a cell phone instead of a landline, but more testing with a broader set of items would be helpful in definitively answering this question.

Follow-up studies can build on this research in several ways. These include testing other questions that may be sensitive to respondent location, collecting more detailed information about the respondent's environment, and larger sample sizes. Survey estimates on employment, commuting, vacationing, and sociological topics may be affected by the decision to interview people away from home as well as at home. In this study, respondent location is used as a proxy for the actual mechanism that would causes a change in response – differential cues used in formulating responses. Studies designed to measure those cues directly would enhance understanding about the nature of locations effects and possibly help us to identify systematically the most susceptible survey topics. Larger sample sizes would also permit greater confidence that the differences observed are robust and are likely to replicate in other surveys.

5.5 An Agenda for Future Research on Cell Phone Surveys

A change in experimental design may be appropriate for studies following up on work presented in this dissertation. The dual frame RDD repeated measures experiment emphasized generalizability to national surveys of the general public, in some instances at the expense of construct validity. For example, analysis of nonresponse in Wave 2 is potentially undermined by nonresponse in Wave 1, and measurement error tests rely upon indicators of cognitive shortcutting rather than direct computation of error. These issues stem from the decision to sample from the landline and cell phone RDD frames, which contain few if any variables that can be used for nonresponse an measurement error analysis.

Follow-up studies could abandon an RDD design in favor of richer sampling frames. Frame variables could be used to assess both nonresponse and measurement error directly. This would enhance construct validity and help to answer lingering questions about possible effects from attrition (see sections A.2 through A.5 in the Appendix). The downside to using a listed sample is that the behaviors of the list members may not be reflective of behaviors of the general population. This tradeoff would need to be carefully evaluated. On balance, though, smaller scale studies that enhance understanding of the mechanisms underlying device difference are a promising next step for research on nonresponse and measurement error in cell phone samples.

5.6 Implications for Telephone Survey Practitioners

This dissertation has several implications for survey practitioners. Chapters 3 and 4 provide evidence that the quality of data from cell phone surveys is comparable to that from landline surveys in most instances. Consequently, cell phone surveys appear to be a viable data collection method for many research purposes. That said, major weaknesses of landline surveys also apply to cell phone surveys, such as low response rates and social desirability pressure due to the presence of the interviewer.

While cell phone interviewing is a promising data collection method, there are numerous details in need of research attention. At the sample design stage, work needs to be done on identifying the optimal ratio for allocating sample to landline versus cell phone interviewing in dual-frame designs. Groves and Lepkowski (1985) evaluated allocation ratios for studies featuring landline interviews and in-person interviews. They make several points that are just as relevant in the landline/cell context as they are in the landline/in-person context. In particular, they demonstrate that the optimal allocation is statistic specific and that it is a function of cost, variance, and bias parameters. Fortunately, the optimal ratio for one survey estimate may be highly similar to the optimal ratio for another, and this pattern may hold for the most important estimates. Methodologists from the Pew Research Center have discussed efficiencies from setting the allocation ratio in a dual frame survey to mirror the allocation ratio of landlines and cell phones in the general population (Dimock, Christian, and Keeter 2009), but additional work is needed to validate or improve upon this approach.

Questionnaire design is also a topic that may need to be re-visited in the context of cell phone surveys. As discussed above, there is some evidence that long or otherwise cognitively demanding questions are not appropriate for cell phone RDD surveys. They can be administered, of course, but there is reason to doubt that the responses will be as accurate as the research objectives require. Future studies on measurement error should seek to push the boundaries of respondent effort in cell phone surveys in order to provide researchers with better defined guidelines for what constitutes a question that is overly burdensome and what does not. Attempts to “pushing the boundaries” should, naturally, only be done within reason and with the approval of the relevant institutional review board.

Finally, the nonresponse findings in Chapter 2 highlight the need to develop more sophisticated weighting protocols in cell phone surveys. Some researchers calibrate their sample to population benchmarks for telephone service but not necessarily to usage patterns (e.g., the proportion of the public thinking of their cell phone number as private versus public). Evidence here and in the study by Brick and his colleagues (2006) suggests that this practice may result in residual nonresponse bias because it does not account for the link between differential usage and response propensity. More work is needed to identify whether and how survey statisticians should adjust for differential response propensities across usage types.

Appendix A

A.1 Dispositions and Performance Rates

Table A.1.1 Productivity Rates in Wave 1 and Wave 2

	Called on Landline	Called on Cell Phone
Wave 1		
Contact Rate	78.4%	78.0%
Cooperation Rate	27.7%	28.3%
Eligibility Rate	86.2%	54.7%
Completion Rate	92.9%	86.3%
Response Rate (3)	20.2%	19.1%
Wave 2		
Contact Rate	91.9%	90.0%
Cooperation Rate	63.2%	55.6%
Eligibility Rate	98.7%	97.6%
Completion Rate	100.0%	99.5%
Response Rate (1)	56.1%	48.0%

Table A.1.2 Final Disposition for Wave 2 Sample by Treatment

	Called for Wave 2 on Landline	Called for Wave 2 on Cell Phone
Business/Government	0.0%	0.2%
Not working	2.2%	0.9%
No answer/Busy	4.7%	1.4%
Voicemail	2.9%	8.5%
Other non-contact	0.4%	0.0%
Callback	5.8%	11.1%
Refused/Hung up during intro	27.3%	28.5%
Ineligible due to language	0.7%	0.7%
Child cell	0.0%	0.5%
Interrupted	0.0%	0.2%
Completed interview	56.1%	48.0%
	100%	100%
Sample Size	(278)	(425)

A.2 Overview of Panel Mortality

In this section I evaluate the potential for panel attrition occurring between Waves 1 and 2 to compromise the analysis presented in Chapters 2, 3, and 4. Any effect of attrition is likely to be small in most instances because of the lack of association between the reason(s) for attrition and the dependent variable of interest. Possible exceptions to this are discussed in detail.

Most of the attrition in this study was purposefully built into the experimental design. Only respondents with both a landline and cell phone were eligible for Wave 2. Respondents with only a landline at home or only a cell phone were dropped after Wave 1. This eligibility constraint was imposed because respondents with just one type of phone were less valuable analytically (their Wave 2 device could not be randomly assigned), and the budget would not permit interviewing all Wave 1 respondents again in Wave 2. As shown in Table A.2.1, about one-third (31%) of the Wave 1 respondents were dropped prior to Wave 2 because of their landline-only or cell phone-only status. The other attrition that occurred was unintended, though it was also expected given the difficulty of gaining cooperation in one-off telephone surveys with no advance notification. Some 36% of Wave 1 respondents reported having both types of telephones but declined to provide the number necessary to perform the randomization for Wave 2. The reminder of the attrition came from nonresponse to the Wave 2 request.

Table A.2.1 Attrition Bewteen Wave 1 and Wave 2

	Number of cases	Percent of Wave 1 respondents
Ineligible for Wave 2 because respondent did not have a cell phone	228	11%
Ineligible for Wave 2 because respondent did not have a landline	423	20%
Ineligible for Wave 2 because respondent did not provide phone number	771	36%
Eligible for Wave 2 but did not respond	343	16%
Eligible for Wave 2 and responded	360	17%
Total (all Wave 1 respondents)	2,125	100%

A.3 Effects of Panel Mortality on the Chapter 2 Nonresponse Analysis

We expect that these forms of attrition vary in how they relate to the key dependent variables in the analysis. In Chapter 2, the primary dependent variable is nonresponse to Wave 2 (examined separately for the landline and cell phone conditions). Exclusion of the landline-only and cell-only cases could influence the results because the factors influencing response propensity may be related to telephone status. For example, concern about the cost of the call could be more influential to the decisions of dual users than to the decisions of cell-only adults.

Given that cell-only and landline-only respondents were not re-contacted for Wave 2, it is impossible to know exactly how the results would differ had they been retained in the study. One way to approximate that effect is to identify respondents who were re-contacted for Wave 2 (dual users) and have characteristics similar to the cell-only or landline-only respondents who were dropped. That is, I can use proxies to simulate how the results of the study would change if all cases had been re-contacted for Wave 2. It is important to note that this simulation addresses the effect of attrition that was purposefully designed into the study, but it does not address attrition due to nonresponse at Wave 1.

To identify landline-only and cell-only proxies, I identified a set of variables strongly correlated with telephone status in the literature (Blumberg and Luke 2008; Keeter et al. 2007; Tucker, Brick, and Meekins 2007). These variables included marital status, age, home ownership, race, and ethnicity, among others. I measured these variables for all respondents during Wave 1. I then used these variables to identify dual users in the study who were similar to landline-only respondents or similar to cell-only respondents. One logistic regression model based on all 2,125 Wave 1 cases was used to predict the probability that the respondent was cell-only, and a separate model based on the same cases was used to predict the probability that the respondent was landline-only. The predictors used to estimate these outcomes were age, gender, education, marital status, home ownership, race, Hispanic ethnicity, and high-speed internet connection (yes/no), region, children in the household, number of adults in the household.

Dual users with a high predicted probability of being cell-only (top quartile) were flagged as cell-only proxies. Similarly, dual users with a high predicted probability of being landline-only (top quartile) were flagged as landline-only proxies. After identifying these proxy groups, I re-ran the key models in the nonresponse chapter.

Table A.3.1 suggests how the final model of participation in the Wave 2 cell phone condition would have been affected if the cell-only respondent had been included in Wave 2. The model on the left comes from Chapter 2 and shows the estimated coefficients based on all 425 dual users in the cell phone condition. The model on the right is based on 79 “cell-only proxies” who are also dual users in the Wave 2 cell phone condition, but they also were shown to be highly similar to the cell-only respondents interviewed in Wave 1.

The results suggest that the conclusions in Chapter 2 regarding device-related mechanisms of nonresponse in cell phone surveys were unaffected by the exclusion of cell-only cases in the experiment. The model on the right shows that cell-only proxies were more likely to participate in the Wave 2 cell phone survey if they give their cell phone number out as contact info and use their phone for non-telephony activities, such as texting and Web browsing. These effects and their interaction are significant in both the “dual user” model from Chapter 2 and the “cell-only proxy” model shown here.

Table A.3.1. Logistic Regression Models Predicting Response to the Wave 2 Cell Phone Survey: Comparison of Results Based on Dual Users (Chapter 2) with Results Based on Cell-only Proxies

	"Final Model" Based on All Dual Users in Wave 2 Cell Condition		**"Final Model" Based on Cell-only Proxies[A]**	
Parameter	Estimate	s.e.	Estimate	s.e.
Intercept	-3.94**	(1.27)	-13.07*	(5.32)
Sampled from landline RDD frame	-0.40***	(0.11)	-0.33	(0.32)
Hostility to surveys scale	-0.24*	(0.10)	0.20	(0.28)
Age	0.26**	(0.08)	0.11	(0.30)
Black	-0.49*	(0.20)	-0.36	(0.39)
Give cell number as contact info	1.02**	(0.35)	3.00*	(1.45)
Non-telephony usage scale (text, picts, web)	0.53*	(0.21)	1.64*	(0.71)
Outbound calling only	-0.44*	(0.19)	n/a[B]	
Answer cell most/all of the time	0.31	(0.17)	0.50	(0.45)
Share cell with someone else	0.19	(0.18)	0.25	(0.50)
Non-telephony scale x Give cell number as contact info	-0.17**	(0.06)	-0.45*	(0.21)
Sample size	N=425		N=79	
Area under ROC curve (c)	0.69		0.73	
-2 Log Likelihood	535.5		88.9	
Re-scaled r-square	0.16		0.22	

***p <.001 **p <.01 *p <.05

[A]The "cell-only proxies" are 79 dual users in the Wave 2 cell phone condition who resemble the cell-only respondents in Wave 1.

[B]This coefficient could not be estimated because none of the 79 cell-only proxies reported using their cell phone only for outbound calling. That is, all of the cases used to estimate this model said that they both make and receive calls on their cell phone.

This analysis does raise questions, however about the conclusions in Chapter 2 regarding the effects associated with demographic variables. The left side of Table A.3.1 shows that age, race, and attitudes about surveys were all significant predictors of cell phone survey response when looking at dual users. The model on the right side, however, shows that these effects are not significant when looking at cell-only proxies. The age and race effects are in the same direction as the left-side model, but the standard errors are too large for them to be significant. Hostility to surveys may work differently for cell-only adults and dual users. The results do not indicate that the analysis in Chapter 2 is deeply flawed, but it does suggest that conclusions about the social-psychological mechanisms are not as well-supported as conclusions about the device-related mechanisms when generalizing to all cell phone users. It is important to bear in mind that the model on the right-side is somewhat underpowered (79 cases were used to estimate a model with 11 parameters) and the relies on proxies of cell-only adults rather than actual cell-only adults. Consequently, this analysis cannot be considered definitive. Future studies should consider re-contacting all telephone groups, not just dual users to avoid this complication.

Table A.3.2 presents the same type of comparison, but now the focus is on modeling response to the Wave 2 landline survey. Here too the results generally support the conclusions from Chapter 2, though some of the predictors in the model are not applicable for landline-only adults. For example, using one's cell phone for outbound calling only has a positive non-significant effect for landline-only proxies, but this coefficient would be zero for actual landline-only adults (who do not have a cell phone). Encouragingly, most of the demographics appear to affect response propensity in the same way for dual users and landline-only proxies alike. None of the coefficients in the landline-only proxy model is reliably different from zero (perhaps due to low power), but the direction of the coefficients suggests that there may be a positive effect from education and a negative effect from African-American race on propensity to respond to the landline survey. The strong, positive effect from age observed for the dual users in the left-side model did not replicate for the landline-only proxies, on the right side. This failure to replicate is very likely due to low variance on age among the landline-only proxies. Over 75% of the landline-only proxies are age 55 or older, which is more than double the rate among the dual users used for the left-side model (36%) and reduces the likelihood of observing an association with the outcome of interest. Overall, Table A.3.2 shows no strong evidence that the landline survey nonresponse conclusions in Chapter 2 are off-base, but this analysis is limited by small sample sizes and reliance on proxy cases.

Table A.3.2. Logistic Regression Models Predicting Response to the Wave 2 Landline Survey: Comparison of Results Based on Dual Users (Chapter 2) with Results Based on Landline-only Proxies

	"Final Model" Based on All Dual Users in Wave 2 Landline Condition		**"Final Model" Based on Landline-only Proxies^**	
Parameter	Estimate	s.e.	Estimate	s.e.
Intercept	0.09	(0.90)	1.09	(2.93)
Hostility to surveys scale	-0.24	(0.12)	-0.25	(0.26)
Age	0.29**	(0.09)	-0.05	(0.37)
Black	-0.49*	(0.23)	-0.73	(0.67)
Education	0.25**	(0.08)	0.33	(0.20)
Give cell number as contact info	-0.36*	(0.15)	-0.24	(0.37)
Outbound calling only	0.36	(0.22)	0.34	(0.39)
Sample size	N=278		N=49	
Area under ROC curve (c)	0.73		0.68	
-2 Log Likelihood	336.6		57.4	
Re-scaled r-square	0.20		0.16	

***p <.001 **p <.01 *p <.05

^The "landline-only proxies" are 49 dual users in the Wave 2 landline condition who resemble the landline-only respondents in Wave 1.

Another source of attrition relevant to the nonresponse analysis in Chapter 2 is the exclusion of 771 dual users who declined to provide their telephone number. This group consists of 573 dual users from the landline RDD frame who would not provide their cell phone number and 198 dual users from the cell phone DD frame who would not provide their landline number. The central difference between these dual users and the 703 dual users who did provide their phone number appears to be greater concern about privacy. Presumably, the people who declined to provide their phone number were suspicious that they would have been contacted again for new surveys or telemarketing efforts. It is also possible that they simply did not want to be re-interviewed about the same questions that they just answered. Specific reasons why people refused to provide their number were not recorded. Attitudes about surveys and frustration with unsolicited requests for information were among the key predictors on the nonresponse analysis in Chapter 2. Therefore, the exclusion of these 771 leery dual users could have distorted the results because they may possess large, negative values on these dimensions – values that may be underrepresented among the dual users who did provide their number.

Here again, I evaluate the potential bias by using proxy cases. Logistic regression was used to predict cooperation with reporting the telephone number based on variables measured in Wave 1. The model is based only on dual users identified in Wave 1 and considers cooperation with giving the number during the Wave 1 interview or during the follow-up effort conducted on the heels of Wave 1 in order to increase response on this variable. I modeled this item-level refusal using demographic variables, attitudes about surveys, and behavioral measures about phone usage. The results are reported in Table A.3.3.

The sampling frame (landline or cell RDD) from which the respondent was selected proved to be a very strong predictor, though this effect almost certainly has more to do with the difference in the information being requested than the sampling frame. Most of the dual users from the cell phone RDD sample (69%) provided their landline number, but less than a third (32%) of the dual users from the landline RDD sample provided their cell phone. In the table this is represented by the highly significant positive coefficient associated with asking for a landline number (asking for a cell phone number is the reference category). To the extent that a cell phone is viewed more as a “personal” device rather than a public access point for strangers, people are less likely to disclose that number than a landline number regardless of the frame used to contact them for this study. The frame variable appears to be capturing the added difficulty of obtaining someone’s cell phone number, relative to obtaining their landline number.

Table A.3.3. Logistic Regression Model Predicting Respondent Cooperation with Reporting Phone Number^

Parameter	Estimate
Intercept	0.47
Asked to provide landline number	1.29***
Give cell number as contact info	0.36***
Asked to provide landline number x Give cell as contact info	-0.24***
Hostility to surveys scale	-0.35***
Age	0.08
Rent	0.09
Asked to provide landline x Rent	-0.16*
Sample size	N=1,474
Area under ROC curve (c)	0.76
-2 Log Likelihood	1732.7
Re-scaled r-square	0.25

***p <.001 **p <.01 *p <.05

^Model is based on Wave 1 respondents who had both a landline and a cell phone.

The logistic regression also identifies several other significant predictors of cooperating with the request to provide the phone number. Renters and people who provide their cell phone number on forms and applications were more likely to cooperate with the request, but only if they were asked to provide their cell number (not their landline). The regression model also indicates that people believing that surveys are uninteresting or unimportant were much less likely to provide their phone number. Age and other demographics (dropped from the model) did not show a significant association with cooperation in giving the number.

The main research question of interest concerning cases refusing to report their phone number is whether or not their attrition biased the findings reported in Chapter 2. To address this, I used the propensity scores from the model shown in Table A.3.3. For each dual user I computed the propensity score for refusing to report the phone number (the complement to the outcome modeled in Table A.3.3). These values ranged from .13 to .96, with the high values assigned to cases that were more likely to have refused to provide their phone number, based on the predictors in the model. I created an indicator variable for the dual users giving their number who, nevertheless, had a propensity score in the top tercile. These 95 cooperative dual users are used as proxies for the 771 dual users who declined to provide their number. Some 43 of the refuser proxies were randomly assigned to the Wave 2 landline condition, and the rest were assigned to the Wave 2 cell phone condition. I re-ran the key regression models, as just I did in the proxy analysis above.

The results from re-running the cell phone response model are presented in Table A.3.4, along with the model estimates reported in Chapter 2. The results of the refuser proxy model nicely parallel the results of the Chapter 2 model. That is, the direction and relative magnitude of the effects are similar when the model is based on proxies for those who refused to report their phone number. Most of the effects are not statistically significant in the refuser proxy model, but that can be attributed to the small case base for the model (N=52). Some of the effects in the original model could not be estimated for this same reason, as some cells were empty. The data that are available, though, generally indicate that refusal to report telephone numbers did not systematically bias the conclusions drawn in Chapter 2. I reiterate, though, that this analysis relies on the strong assumption that the response behavior of the proxies is representative of the response behavior of the refusers had I been able to include them in Wave 2.

Table A.3.4. Logistic Regression Models Predicting Response to the Wave 2 Cell Phone Survey: Comparison of Results Based on Those Reporting Their Number (Chapter 2) with Results Based on Proxies for Those Refusing to Report Their Number

	"Final Model" Based on All Dual Users in Wave 2 Cell Condition		**"Final Model" Based on Proxies for Cases Refusing to Report Their Number[A]**	
Parameter	Estimate	s.e.	Estimate	s.e.
Intercept	-3.94**	(1.27)	-13.53*	(5.73)
Sampled from landline RDD frame	-0.40***	(0.11)	n/a[B]	
Hostility to surveys scale	-0.24*	(0.10)	-0.48	(0.46)
Age	0.26**	(0.08)	0.91	(0.52)
Black	-0.49*	(0.20)	n/a[C]	
Give cell number as contact info	1.02**	(0.35)	3.41	(2.02)
Non-telephony usage scale (text, picts, web)	0.53*	(0.21)	2.12*	(1.00)
Outbound calling only	-0.44*	(0.19)	-0.74	(0.65)
Answer cell most/all of the time	0.31	(0.17)	0.48	(0.94)
Share cell with someone else	0.19	(0.18)	-0.33	(0.65)
Non-telephony scale x Give cell number as contact info	-0.17**	(0.06)	-0.78	(0.48)
Sample size	N=425		N=52	
Area under ROC curve (c)	0.69		0.84	
-2 Log Likelihood	535.5		41.9	
Re-scaled r-square	0.16		0.36	

***p <.001 **p <.01 *p <.05

[A]These refuser proxies are 52 dual users in the Wave 2 cell phone condition who resemble the dual users in Wave 1 who refused to report their phone number and, thus, were dropped from the experiment. They were identified using the logistic regression model presented in Table A.3.3.

[B]This coefficient could not be estimated because all of the refuser proxies came from the landline RDD frame. None of the cases from the cell RDD were very likely to have refused to give their landline number.

[C]This coefficient could not be estimated because there was only one African-American respondent idenitified as being a refuser proxy in the Wave 2 cell phone condition.

The analogous comparison with respect to the Wave 2 landline survey response model is presented in Table A.3.5. Here too the results suggest that the Chapter 2 conclusions are robust to attrition from people not reporting their phone numbers. In the

Chapter 2 analysis, I observed that age and education had significant, positive effects on responding to the landline survey, while being black and using your cell phone number as contact information had negative effects. When I re-ran the model based on the refusers proxies, nearly all of these effects appear to replicate, though the case base for the model is too small (N=43) for the coefficients to be significantly different from zero. The race effect did not replicate, but there were only three African-Americans in the model, so it was not possible to reliably estimate the effect.

Table A.3.5. Logistic Regression Models Predicting Response to the Wave 2 Landline Survey: Comparison of Results Based on Those Reporting Their Number (Chapter 2) with Results Based on Proxies for Those Refusing to Report Their Number

	"Final Model" Based on All Dual Users in Wave 2 Landline Condition		**"Final Model" Based on Proxies for Cases Refusing to Report Their Number[A]**	
Parameter	Estimate	s.e.	Estimate	s.e.
Intercept	0.09	(0.90)	-0.33	(2.49)
Hostility to surveys scale	-0.24	(0.12)	-0.19	(0.32)
Age	0.29**	(0.09)	0.45	(0.29)
Black	-0.49*	(0.23)	0.16	(0.68)
Education	0.25**	(0.08)	0.31	(0.22)
Give cell number as contact info	-0.36*	(0.15)	-0.51	(0.55)
Outbound calling only	0.36	(0.22)	0.43	(0.51)
Sample size	N=278		N=43	
Area under ROC curve (c)	0.73		0.76	
-2 Log Likelihood	336.6		49.9	
Re-scaled r-square	0.20		0.24	

***p <.001 **p <.01 *p <.05

[A]These refuser proxies are 43 dual users in the Wave 2 landline phone condition who resemble the dual users in Wave 1 who refused to report their phone number and, thus, were dropped from the experiment. They were identified using the logistic regression model presented in Table A.3.3.

The possibility that attrition distorted some of the results presented in Chapter 2 cannot be completely dismissed. The rate of overall attrition was quite high, which indicates that the risk of bias is non-trivial. Based upon the data at hand, there do not appear to be any strong relationships between the reasons for attrition and the outcomes investigated. That said, the techniques available to me for testing such relationships are imperfect and liberal, in the sense that false negative results were more likely than false positive ones. I attempted to gauge the impact of attribution on the main nonresponse analyses by identifying groups of proxies that had similar characteristics to cases that were dropped from the study. It appears that the Chapter 2 conclusions would not change much had the cell-only and landline-only respondents been re-contacted for Wave 2. It also appears that the conclusions would not change had there been 100% cooperation with reporting telephone numbers. While the results of this analysis are reassuring, they are far from definitive. The best way to reduce these biases would be to replicate the experiment but include all Wave 1 respondents in the Wave 2 interview. Noncooperation with the telephone number question might be reduced by an extra incentive and/or revised interviewer language explaining why the number is needed.

A.4 Effects of Panel Mortality on the Chapter 3 Cognitive Shortcutting Analysis

In Chapter 3, seven dependent variables are used to test whether or not respondents are more likely to use cognitive shortcuts when interviewed on a cell phone versus a landline. Several tests treat Wave 1 and Wave 2 responses interchangeably, on the assumption that the data from the surveys are of comparable validity. A threat to this assumption is that about half of the cases eligible for Wave 2 did not respond. If the propensity to respond is related to response behaviors of interest, then there is a potential for bias. Chapter 3 discusses this possibility and also notes that the potential for bias is reduced by the fact that these relationships would have to be systematically related to the device (landline or cell) in order for the conclusions to be undermined. Therefore, the attrition occurring because of nonresponse to Wave 2 would only undermine the comparison of landline versus cell phone response behaviors if the reasons for attrition are relevant to one device but not the other.

I investigate this nonresponse attrition by comparing Wave 2 responders and nonresponders on indicators of cognitive shortcutting measured in Wave 1. The study was designed so that most of the cognitive shortcutting experiments were administered in Wave 2 (for which device was randomized), so only a handful of indicators are available from Wave 1. I compared responders and nonresponders with respect to item nonresponse in Wave 1, differentiation among items in a Wave 1 battery, and self-reported response strategy on a Wave 1 behavioral frequency question.

The main concern with panel attrition in Chapter 3 is that the Wave 2 measurement error comparisons may be confounded because differential attrition could have resulted in the randomized cell condition cases being poorer respondents than the randomized landline condition cases. I verified that the Wave 2 comparison groups were not confounded by panel attrition by comparing their response behaviors in Wave 1. The top two rows of Table A.4.1 show that respondents in the Wave 2 landline and cell conditions were highly similar with respect to item nonresponse in Wave 1. Some 27 questions in Wave 1 were administered to the entire sample. From those data I computed two metrics: the proportion of cases item non-responding at least once and the mean number of items not answered.[32] The Wave 2 experimental conditions did not differ significantly on either measure.

[32] The raw range for the number of questions not answered out of a total of 27 was 0 to 19. In order to avoid undue influence from outliers, the top of the response distribution was capped at the 99th percentile, which was 9 questions not answered.

A related concern is that exclusion of other groups (landline-only adults, cell-only adults, dual users not reporting their phone number, and Wave 2 nonrespondents) distorted the Chapter 3 analysis. There is some evidence that landline-only respondents differed with respect to item nonresponse. As shown in Table A.4.1, landline-only respondents were the most likely to item nonrespond in Wave 1, out of all the experimental groups. This relationship was statistically significant even after controlling for respondent age and education in a logistic regression (not shown). Landline-only respondents also tended to not answer more questions than others. The number of questions left unanswered are count data and the variance was greater than the mean, so a negative binomial model (not shown) was used to measure the effect of the various experimental groups on the number of items not answered when controlling for age and education. Being landline-only and being a dual user who refused to report the phone number were both associated with significant positive effects on the number of items unanswered.

Table A.4.1. Item Nonresponse in Wave 1, by Experimental Group

Experimental Group	% Did not answer one or more questions		Mean number of items not answered		Sample size
W2 Respondent (Landline condition)	31.8%	(0.038)	0.75	(0.13)	154
W2 Respondent (Cell phone condition)	32.7%	(0.033)	0.69	(0.10)	202
W2 Respondents (Total)	32.3%	(0.025)	0.72	(0.08)	356
All W2 Nonrespondents	32.5%	(0.026)	0.74	(0.08)	338
W2 Ineligible (refused to report number)	39.7%	(0.018)	1.03[B]	(0.07)	761
W2 Ineligible (landline-only)	48.2%[A]	(0.033)	1.27[B]	(0.13)	226
W2 Ineligible (cell phone-only)	29.3%	(0.022)	0.75	(0.08)	423

Note.- Standard errors are shown in the parentheses. Differences between Wave 2 landline and cell phone conditions are not significant.

[A]Logistic regression was used to test differences in item nonresponse (1=once or more, 0=never) across the experimental groups controlling for respondent age and education. Wave 2 respondents were specified as the reference group. The effect of being landline-only relative to being a Wave 2 respondent on the likelihood of item nonresponding was significant at the p <.01 level.

[B]A negative binomial model was used to test differences in the means across the experimental groups controlling for respondent age and education. Wave 2 respondents were specified as the reference group. The effects of being landline-only and being a dual user who refused to report the phone number were both statistically significant at the p <.01 level.

These results indicate that panel attrition probably led to less item nonresponse in Wave 2 than would be expected in a one-off dual frame RDD study. The Wave 2 questionnaire was different, so I cannot say definitively that this would happen, but the analysis presented in Table A.4.1 shows that some of the most egregious item nonresponders captured in Wave 1 were not included in the Wave 2 measurement error experiments. This does not necessarily imply, however, that the Chapter 3 conclusions were distorted. The goal of the Chapter 3 experiments was to isolate the effect of device, which is not readily possible with landline-only adults who do not use cell phones. If landline-only adults were included in the Wave 2 sample and folded into the “landline condition” means, then it is possible that the landline condition respondents would have exhibited more item nonresponse than was actually observed. Doing this, however, muddies the theoretical value of data from groups that were randomly assigned to device.

Understanding the measurement error from landline-only adults may be an important research objective, but it was not the objective of the study at hand. The other attrition group prone to higher levels of item nonresponse is dual users who refused to report their phone number. The exclusion of this group is even more innocuous because had they been included in Wave 2, they would have been randomly assigned to device. They may have increased the overall rate of item nonresponse in Wave 2, but there is no indication that this would have compromised the device comparison.

Another metric that can be used to assess the effect of panel attrition is non-differentiation in responses to a long (14-item) Wave 1 battery. Following other measurement error studies, I assert that lower variance in respondent-level answers reflects lower effort, all else equal. Wave 1 contained a battery of 14 statements about the U.S. military and foreign policy, and it featured a five-point response scale measuring strength of dis/agreement. Some 75% of the Wave 1 respondents answered all 14 items. The other 25% are excluded from this analysis. For each respondent, I computed the variance of their answers. The experimental group means of these variances are reported in Table A.4.2. Larger means in the table reflect greater differentiation and, presumably, higher data quality.

Table A.4.2. Respodent-level Variance on Wave 1 Battery

Experimental Group	Mean variance		Sample size
W2 Respondent (Landline condition)	1.35	(0.06)	126
W2 Respondent (Cell phone condition)	1.25	(0.05)	153
W2 Respondents (Total)	1.30	(0.04)	279
All W2 Nonrespondents	1.23	(0.04)	264
W2 Ineligible (refused to report number)	1.31	(0.03)	545
W2 Ineligible (landline-only)	1.34	(0.05)	152
W2 Ineligible (cell phone-only)	1.24	(0.04)	345

Note.- Larger values reflect greater differentiation between the 14 items in the battey and, thus, are assumed to reflect higher data quality. Results are based only on respondents who answered each item in the battey. Standard errors are shown in the parentheses.

Critically, none of the group means is significantly different from another. The Wave 2 landline condition cases exhibited about the same amount of differentiation on the battery as the Wave 2 cell phone condition cases (means of 1.35 and 1.25, respectively, t=1.29, df=262, p=.20). This suggests that the randomization was basically successful. Using Scheffé tests, none of the other pair-wise comparisons in group means were significant. These results support the argument that panel attrition did not seriously bias the conclusions from Chapter 3, especially those concerning non-differentiation in battery responses.

The final test available for evaluating the potential for panel attrition bias is examination of response strategies used to answer a behavioral frequency question in Wave 1. Respondents were asked to report the number of overnight trips that they had taken in the past two years. This was followed by a question asking them to report how they formulated their answer: did they know the answer off the top of their head, think about each individual trip and added them up, think about the types of trips they had taken and use that to estimate, or simply estimate based on a general impression. Studies have shown that enumerating individual episodes and adding them up is a more cognitively demanding response strategy than estimating the answer (Conrad, Brown, and Cashman 1998). In Chapter 3, I present evidence that respondents who reported using one of the latter two strategies (thinking about types of trips or using a general impression) were significantly more likely to give a round, prototypical response (indicative of estimating) than respondents who used one of the other strategies. If the experimental groups not captured in Wave 2 are shown to be more likely to estimate than those who were in Wave 2, this would indicate that the Chapter 3 conclusions might be biased by attrition. If, however, there are no differences across the experimental groups, then there is reason to believe that the Chapter 3 conclusions are not undermined by attrition. This behavioral frequency question was not administered to landline-only cases because they were not an important analytic group in the methods analysis presented in Chapter 3. The results for the other experimental groups are presented in Table A.4.3.

Table A.4.3. Response Strategy for Wave 1 Behavioral Frequency Question, by Experimental Group

	W2 Respondents (Landline condition)	W2 Respondents (Cell condition)	W2 Respondents (Total)	All W2 Nonrespondents	W2 Ineligible (refused to report number)	W2 Ineligible (cell phone-only)
Response Strategy[A]						
Knew number off top of head	17%	19%	18%	20%	20%	19%
Thought about each individual trip and added them up	24%	23%	23%	23%	18%	17%
Thought about types of trips and used that to estimate	29%	20%	24%	27%	22%	24%
Estimated based on a general impression	27%	35%	32%	27%	34%	33%
Don't know/Refused	3%	3%	3%	4%	6%	7%
	100%	100%	100%	100%	100%	100%
Percent giving a prototypical value[B]	31%	29%	29%	32%	35%	31%
Sample size	(142)	(184)	(326)	(301)	(703)	(366)

Note.- Respondents who reported zero trips are excluded from this analysis. Landline cases were not administered the revelant questions and are also excluded.

[A]The proportion of respondents estimating (thinking about types or using a general impression) did not vary significantly across the experimental groups.

[B]The proportion of respondents reporting prototypical values (multiples of 10, greater or equal to 10) did not vary significantly across the experimental groups.

There is essentially no evidence that respondents who dropped out during the study were more likely to use estimation as a cognitive shortcut than those used for the Chapter 3 analyses. When collapsing the “types” and “general estimation” categories, the data show that the proportion of Wave 2 nonrespondents who estimated their answer to this Wave 1 question (54%) is highly similar to the proportion of all Wave 2 respondents who estimated (56%). The proportion of cases ineligible for Wave 2 was slightly higher (57%), but the difference is not statistically significant. Furthermore, the randomization to device appears to have worked correctly, as respondents in the Wave 2 landline and cell conditions were about equally likely to estimate for this Wave 1 response. Examination of the values reported also shows no sign of attrition bias. Among dual users refusing to report their phone number, some 35% gave a prototypical value (e.g., 10,20,…). This is slightly higher than the proportion of Wave 2 respondents reporting a prototypical value (29%), but the difference is not statistically significant. Looking at both metrics, there is some indication that the respondents ineligible for Wave 2 were slightly more likely to use estimation as a cognitive shortcut, but the magnitude of this effect is not statistically significant and is unlikely to seriously bias the conclusions in Chapter 3.

In sum, panel attrition is not expected to distort the measurement error conclusions in any meaningful way. The chapter looks at differences in response behavior between respondents interviewed on a landline versus those interviewed on a cell phone. Even though there was attrition, there is virtually no evidence indicating that it would confound comparisons between groups randomly assigned to device. The more realistic concern is that people most likely to use cognitive shortcuts (on either device) were not captured in the Chapter 3 analysis, due to attrition. That is, nonrespondents at Wave 1 and Wave 2 may have been poorer responders in general relative to those who participated in both waves. The analysis presented above provides some evidence for this, but the magnitude of the differences tends to be very small and often not statistically significant. The respondents who completed both waves of this study were slightly better than average with respect to answering carefully, but there is no indication that this pattern undermines the device comparisons that are the focus of Chapter 3.

A.5 Effects of Panel Mortality on the Chapter 4 Response Distribution Analysis

Chapter 4 addresses two research questions: (*i*) whether the respondent's location may present cues used in formulating responses to certain questions and (*ii*) whether response variance tends to be greater if the respondent is interviewed on a cell phone rather than a landline. Neither of these questions can be investigated rigorously based on Wave 1 data alone. There are also too few cases to identify proxies for respondents who dropped out by design or through nonresponse. Unfortunately, this rules out the possibility of conducting a meaningful assessment of panel attrition effects.

If anything, panel attrition may have made the Chapter 4 analysis more conservative, in the sense that the type II error rate was probably higher than it would have been without attrition. Dual users who refused to give their telephone number or gave the number but failed to respond to Wave 2 were excluded from the Chapter 4 analysis. These cases were, by definition, less cooperative than those who completed Wave 2. If their lower propensity to cooperate reflects less interest or commitment in the study, then it is possible that these respondents would have been more likely to use environmental cues in responding than the respondents who were measured. Respondents who are less interested in participating may be more likely to use top-of-mind-considerations when answering the six experimental items analyzed in Chapter 4. If anything, panel attrition may have suppressed the effect of interest in Chapter 4, making the conclusions conservative. There is no way to evaluate this speculation with the data at hand, but it could potentially be address in a follow-up study.

A.6 Wave 1 and Wave 2 Response Distributions for the Satisfaction Battery, by Treatment

In Chapter 4, section 4.3.3 addresses differences in Wave 1 versus Wave 2 responses for respondents in the landline-landline and cell-cell conditions. The differences between the means for each of the battery items are reported separately for each treatment group. The two groups in which respondents were interviewed on both devices are combined in order to illustrate the main experimental comparison – results for landline versus cell phone interviews. In the far right column (IV), figures are based only on respondents reached away from home in either Wave 1 or Wave 2. This is the same subset of respondents used to compute the values in Figure 4.2.

Table A.6. Differences Between Means for Satisfaction Battery Items, by Experimental Condition[A]

	I. Landline-Landline Condition	II. Cell-Cell Condition	III. Combined LL-Cell and Cell-LL Conditions	IV. Combined LL-Cell and Cell-LL Conditions[B]
	W1 mean - W2 mean	W1 mean - W2 mean	LL mean - Cell mean	LL mean - cell mean
Condition of local roads and highways	0.00	0.12	-0.10	-0.12
Local traffic conditions on roads	0.22	0.10	0.01	-0.02
Your personal safety from crime	-0.10	-0.03	-0.08	-0.12
Your social life	-0.09	-0.05	0.01	0.21*
Your physical health	-0.09	0.16**	-0.06	-0.02
Your family life	-0.07	0.01	0.01	-0.04
Minimum sample size	55	126	142	48

*p <.05 **p <.01 ***p <.001 based on paired difference in means *t* -test.

[A]Each mean is based on responses to a four-point scale where 1=very satisfied, 2=somewhat satisfied, 3=not too satisfied, 4=not at all satisfied. This table reports means as a way to summarize the central tendency of the response distribution in a concise way. The distributions, however, are not necessarily Gaussian. Thus, the paired t-test results are intended only as rough indicators of statistically significant differences.

[B]Figures are based only on respondents interviewed away from home in either wave. Column III does not have this filter.

A.7 Summary of Panel Mortality Analysis

Two main analytic approaches were used to evaluate the effect of panel mortality on the experimental results: (1) re-estimation of key model parameters using proxies for cases that attrited and (2) analysis of data from Wave 1 (prior to attrition). Regression models were used to identify proxies for various attrition groups (landline-only, cell-only, and dual users refusing to give their phone number). The key models in the nonresponse chapter were re-estimated based on the proxy cases, and then the differences from the models estimates presented in the chapter were discussed. The proxy analysis suggests that the conclusions in the chapter are generally robust to the attrition that occurred. In particular, most of the device-relate effects replicated with the proxy cases. Replication of the demographic effects was somewhat more spotty, but there is no suggestion that the findings presented in the chapter would be dramatically different had the attrition not occurred.

Assessing the effect of panel mortality on the cognitive shortcutting chapter was more straight-forward. The chapter presents analysis of seven indicators of cognitive shortcutting based on data from Wave 2. The panel mortality analysis looks at a subset of those indicators based on data from Wave 1. The Wave 1 data are critical because they are available for the respondents who dropped out of the study. There is some evidence that respondents who dropped out due to non-cooperation showed more signs of satisficing than others, but the non-cooperation is unrelated to the device comparisons that are the focus of the chapter. More importantly, the respondents dropped from the experiment because they were not dual users appear to be quite similar on the outcomes of interest to the respondents who were retained in the study. So while the overall incidence of cognitive shortcutting may have been reduced somewhat by the attrition, the device comparisons appear to be unaffected.

A.8 Wording of Experimental Questions in Wave 1

D2. Now I'm going to read some statements that some people agree with but others disagree with. For each one, please tell me how much you agree or disagree. (Randomize) [IF NECESSARY: Please tell me whether you Strongly Agree, Agree, are Neutral, Disagree or Strongly Disagree.]

1. America should begin moving some of its troops from Iraq into Afghanistan.
2. America needs to have more people in the active-duty military.
3. I have great confidence in America's military.
4. The U.S. military is stressed so much that it's near the breaking point.
5. I'm confident America will win the Global War On Terror.
6. It's going to take many years to fight and win the Global War On Terror.
7. The military does a poor job of treating the psychological wounds of its fighting men and women.

(NOTE: Halfway through the randomized list of 14 items, have interviewer say the following: "Just a few more statements to go now…")

8. The world is a more dangerous place now than it was 20 years ago.
9. Using America's military power to prevent war is every bit as important as using that power to win wars.
10. The U.S. military services help the global economy when they prevent acts of terrorism or other criminal acts that could disrupt world trade.
11. Our fighting men and women get world-class medical treatment.
12. Most of the problems with the military medical system have been resolved.
13. The U.S. should use threat of military action as a negotiating tool when dealing with unfriendly countries.
14. I think America's military should buy its equipment from U.S. manufacturers, even if better equipment is available from European companies.

F1. Next, I'd like to ask you about some different aspects of your life. For each one, please tell me whether you are very satisfied, somewhat satisfied, not too satisfied, or not at all satisfied. First, how would you rate your satisfaction with…? [IF NECESSARY: Are you very satisfied, somewhat satisfied, not too satisfied, or not at all satisfied with that aspect of your life?]

a. The condition of roads and highways where you live
b. Traffic conditions on roads where you live
c. Your personal safety from crime where you live
d. Your social life

e. Your physical health

f. Your family life

Now, I have some questions about your telephone use. Your answers will help improve the way research like this is conducted in the future, as people start using new technology.

[ASK IF CELL SAMPLE]

F 2. Is a cell phone your only phone, or do you also have regular landline telephone service in your home?

1 – Only phone

2 – Have regular phone at home

8 – DK (VOL)

9 – Refused (VOL)

[IF R ASKS WHAT IS MEANT BY “REGULAR TELEPHONE, SAY: “A regular landline telephone is a phone that is wired to a jack in the wall.]

[ASK IF LANDLINE SAMPLE]

F 3. Do you, yourself, have a working cell phone?

1 – Yes, have cell phone, 2 – No, do not, 8 – DK (VOL, 9 – Refused (VOL)

[ASK IF CELL USER]

F4. Are you the only person who uses your cell phone, or do you share it with someone else?

1 – Only respondent, 2 – Shared cell phone, 8 – DK (VOL), 9 – Refused (VOL)

[ASK IF CELL USER]

F5. When you get a call on your cell phone, do you answer... (READ 1-5)

1 – Almost always

2 – Most of the time

3 – Some of the time

4 – Rarely, or

5 – Never

8 – DK (VOL)

9 – Refused (VOL)

[ASK IF CELL USER]

F6. Next, I'm going to read some statements. Please tell me whether you Strongly Agree, Agree, Disagree or Strongly Disagree with each one. [INSERT ITEM] [IF NECESSARY: Do you Strongly Agree, Agree, Disagree or Strongly Disagree with this statement?]

a. I like that my cell phone allows me to be more available to others

b. I give my cell phone number when providing contact information on forms and applications

c. When I get a new electronic device, I usually need someone else to show me how to use it

d. I often make cell phone calls to fill up my free time while I'm traveling or waiting for someone

e. I enjoy reading news and industry articles about the latest in cell phone technology

f. I only use my cell phone to MAKE calls – not to receive calls

[ASK IF CELL USER]

F7. People get a lot of unwanted contacts in the form of telemarketing, junk mail, and e-mail spam. How much of a nuisance are these things for you? (READ 1-4)

1 – A very serious nuisance

2 – A somewhat serious nuisance

3 – A mild nuisance, or

4 – Not a nuisance at all

8 – DK (VOL)

9 – Refused (VOL)

[ASK IF CELL USER]

F8. Next, I'm going to name some things that you may or may not do with your cell phone. Do you ever use your cell phone to…

IF "YES": And do you do this at least once a week OR not that often?

a. Send and receive text messages

b. Take pictures or video

c. Connect to the Internet

d. Watch video or TV shows

[ASK IF CELL USER]

F9. Some people try to keep costs down by not using too many minutes on their phone plan. How about you? When you use your cell phone, do you think about the cost of the call…? (READ 1-4)

1 – Most of the time

2 – Some of the time

3 – Rarely, OR

4 – Not at all

8 – DK (VOL)

9 – Refused (VOL)

[ASK ALL]

F10. Now I'm going to read some statements that some people agree with but others disagree with. For each one, please tell me whether you Strongly Agree, Agree, Disagree or Strongly Disagree. [INSERT ITEM] [IF NECESSARY: Do you Strongly Agree, Agree, Disagree or Strongly Disagree with this statement?]

a. The terms "poll" or "research survey" are often used to disguise a sales pitch

b. Answering questions in polls or research surveys is an interesting experience

c. Answering questions in polls or research surveys is a waste of time

G6. Which of the following best describes your age group? Just stop me when I get to the right category… (READ 1-6)

1 – 18 to 24

2 – 25 to 34

3 – 35 to 44

4 – 45 to 54

5 – 55 to 64

6 – 65 or older

8 – DK (VOL)

9 – Refused (VOL)

G7. How many children under the age of 18, if any, do you have living at home? ______

[ENTER 98=DK (VOL.) and 99=REFUSED (VOL.)]

G8. And how many people 18 or older live in your home, including yourself? ______

[ENTER 98=DK (VOL.) and 99=REFUSED (VOL.)]

G9. What is the highest level of education you have completed?

1 – Did not graduate high school

2 – High school graduate or GED

3 – Technical/vocational school graduate

4 – Some college

5 – Bachelor's degree

6 – Some graduate school but no advanced degree

7 – Graduate or doctorate degree

8 – Other (Specify: ______)

98 – DK (VOL)

99 – Refused (VOL)

[ASK IF CELL USER]

G10. We find that people who travel a lot sometimes have different attitudes about issues in this study. Could you please tell me, overall, how many overnight trips have you taken in the past 2 years?

[ASK IF CELL USER]

G11. That's helpful to know. Which of these best describes how you came up with your answer? (READ 1-4)

1 Did you know the number off the top of your head,

2 Did you think about EACH individual trip and add them up,

3 Did you think about TYPES of trips and use that to estimate,

4 Or did you estimate based on a GENERAL IMPRESSION?

8 DK (VOL)

9 Refused (VOL)

[ASK ALL]

G12. Are you (READ 1-3):

1 Married

2 Single and never married

3 Or in some other category

8 DK (VOL)

9 Refused (VOL)

G13. Do you own or rent your home?

1 Own home

2 Rent

3 (VOL) Some other arrangement

8 DK (VOL)

9 Refused (VOL)

G.14. So we may represent all people fairly, do you consider yourself to be Latino, or of Hispanic or Spanish descent?

1 Yes

2 No

8 DK (VOL)

9 Refused (VOL)

G.15. What race or races would you use to describe yourself? (READ 1-6)

1. American Indian or Alaska Native

2. Asian

3. Black or African American

4. Native Hawaiian or other Pacific Islander

5. White, OR

6. Some other race (Specify: _____)

8. DK (VOL)

9. Refused (VOL)

G16. Do you have a computer at home? (IF YES: Does the computer you use at home connect to the Internet through a dial-up telephone line, or do you have a high-speed connection?)

1 No computer with internet at home

2 Dial-up connect

3 High-speed connection

4 Other (VOL)

8 DK (VOL)

9 Refused (VOL)

That's about it. Sometimes we need to get back in touch with those who have participated in our research to find out more about their opinions. It is very important that we have good contact information. All information is kept strictly confidential, as required by professional codes of conduct.

[ASK IF F2=2]

G17_L. Earlier you mentioned you have a landline phone at home. Would you mind telling me that number?

IF R REFUSES: We ask people for their home phone number only because it's hard to reach them on a cell phone. Your number will be kept strictly confidential, and it would be used for just one brief follow-up. Are you okay with sharing your home phone number, or should we move on? (IF R SAYS THEY HAVE MORE THAN ONE LANDLINE NUMBER, ASK ONLY FOR THE NUMBER THEY USE "THE MOST")

[ASK IF F3=1]

G18_C.   Earlier you mentioned you have a cell phone. Would you mind telling me that number?

IF R REFUSES: We ask people for their cell phone number only because it's getting harder and harder to reach them on a home phone. Your number will be kept strictly confidential, and it would be used for just one brief follow-up. Are you okay with sharing your cell phone number? (IF R SAYS THEY HAVE MORE THAN ONE CELL NUMBER, ASK ONLY FOR THE NUMBER THEY USE "THE MOST")

[ASK IF CELL USER: F3=1 OR CELL SAMPLE]

G19. And could you tell me just your first name?

[ASK IF CELL USER: F3=1 OR CELL SAMPLE]

G20. Would you mind telling me if I reached you today AWAY from home or AT home?

[INTERVIEWER NOTE: IF THE RESPONDENT WAS AWAY FROM HOME DURING ANY PART OF THE CALL, THEN ENTER 1]

1 Away from home

2 At home

3 Other Specify (____________)

8 DK (VOL)

9 Refused (VOL)

[ASK IF CELL USER: F3=1 OR CELL SAMPLE]

G21. What did you happen to be doing when I called? [key in verbatim response]

[ASK IF CELL USER: F3=1 OR CELL SAMPLE]

G22. Some people multi-task when on the phone while others do not. What, if anything else, did you happen to do while we were talking? [key in verbatim response]

[ASK IF CELL SAMPLE]

G23. Finally, when the survey finishes, we'd like to send you $10 for your time. Can I please have your last name and a mailing address where we can send you the money? [INTERVIEWER NOTE: If R does not want to give full name, explain we only need it so we can send the $10 out to them personally]

A.9 Wording of Experimental Questions in Wave 2

P3. Next, I'd like to ask you about some different aspects of your life. For each one, please tell me whether you are very satisfied, somewhat satisfied, not too satisfied, or not at all satisfied. First, how would you rate your satisfaction with…? [IF NECESSARY: Are you very satisfied, somewhat satisfied, not too satisfied, or not at all satisfied with that aspect of your life?]

a. The condition of roads and highways where you live

b. Traffic conditions on roads where you live

c. Your personal safety from crime where you live

d. Your social life

e. Your physical health

f. Your family life

P4. How would you rate economic conditions in this country today – as poor, only fair, good, or excellent?

P5. Next, I'd like to ask you about some institutions in American society. As I read each one, please tell me how much confidence you have in that institution using a scale from zero to 10, where 10 means "great confidence" and zero means "no confidence at all." (First,) how would you rate your confidence in [INSERT ITEM; RANDOMIZE.] [IF NECESSARY: You can use any number between zero and ten (where 10 means "great confidence" and zero means "no confidence at all.)"]

a. The public school system

b. The criminal justice system

c. The space program

d. The military

e. Congress

f. The presidency

P6. How would you describe your overall interest in the U.S. space program? Are you… (READ 1-4)

1 Very interested

2 Somewhat interested

3 Not too interested

4 Not at all interested

8 DK (VOL)

9 Refused (VOL)

P7. Can you think of any ways that your life has been improved directly by the U.S. space program? [PROBE ONCE: Can you think of any other ways your life has been improved directly by the US space program?]

1 [ENTER VERBATIM RESPONSE]

2 Life has not been improved by space program in any way

8 Don’t know

9 Refused

[ASK FORM 1 ONLY]

P8. Which of the following do you think is the MAIN reason why America continues to explore space? Is it …? (READ 1-5)

1 To keep our nation safe

2 To inspire and motivate our children

3 To maintain our status as an international leader in space

4 To provide benefits on earth, OR

5 Because it's human nature to explore

6 Other (VOL)

8 DK (VOL)

9 Refused (VOL)

[ASK FORM 2 ONLY]

P9. Which of the following do you think is the MAIN reason why America continues to explore space? Is it…? (READ 1-5)

1 Because it's human nature to explore

2 To provide benefits on earth

3 To maintain our status as an international leader in space

4 To inspire and motivate our children, OR

5 To keep our nation safe

6 Other (VOL)

8 DK (VOL)

9 Refused (VOL)

[ASK ALL]

P10. So far the United States has landed astronauts on the moon, landed rovers on Mars, and sent satellites around the solar system. What do you think should be the next major mission for the U.S. space program?

P11. Now I'm going to read some statements about the U.S. space program that some people agree with but others disagree with. For each one, please tell me how much you agree or disagree. [RANDOMIZE; READ FOR FIRST ITEM, THEN IF NECESSARY: Please tell me whether you Strongly Agree, Agree, are Neutral, Disagree or Strongly Disagree.]

a. [The U.S. space program] is important to national security.

b. [The U.S. space program] does a good job keeping astronauts safe.

c. [The U.S. space program] needs more funding to improve its technology.

d. [The U.S. space program] is a waste of taxpayers' money.

e. [The U.S. space program] contributes to American pride and patriotism.

f. [The U.S. space program] inspires young people to study science and math.

[ASK FORM 1 ONLY]

P12. Which of the following do you think will be America's greatest rival in space exploration over the next 20 years? (READ 1-6)

1 Russia

2 China

3 Japan

4 North Korea

5 The European Space Agency, OR

6 Iran

7 None of the above (VOL)

8 DK (VOL)

9 Refused (VOL)

[ASK FORM 2 ONLY]

P13. Which of the following do you think will be America's greatest rival in space exploration over the next 20 years? (READ 1-6)

1 Iran

2 The European Space Agency

3 North Korea

4 Japan

5 China, OR

6 Russia

7 None of the above (VOL)

8 DK (VOL)

9 Refused (VOL)

[ASK FORM 1 ONLY]

P14. If the U.S. space program had more money, which of the following would be the best way to spend it? (READ 1-6)

1 UNmanned missions to other planets

2 Manned missions to other planets

3 Human return to the moon

4 Improving technology for flight on earth

5 Expanding the international space station, OR

6 Making space shuttles safer

7 Other (VOL. – SPECIFY)

8 DK (VOL)

9 Refused (VOL)

[ASK FORM 2 ONLY]

P15. If the U.S. space program had more money, which of the following would be the best way to spend it? (READ 1-6)

1 Making space shuttles safer

2 Expanding the international space station

3 Improving technology for flight on earth

4 Human return to the moon

5 Manned missions to other planets, OR

6 UNmanned missions to other planets

7 Other (VOL. – SPECIFY)

8 DK (VOL)

9 Refused (VOL)

[ASK ALL]

P16. People who fly in airplanes a lot may have different views about space than those who fly less often. How many round trips have you taken by airplane in the past 2 years?

IF A RANGE IS GIVEN, CLARIFY: An estimate is fine, but I need a single number. Could you give me your best estimate?

INTERVIEWERS: If necessary, say, "Please count each round trip just once, even if there were multiple plane rides."

P17. That's helpful to know. Which of these best describes how you came up with your answer? (READ 1-4)

1 Did you know the number off the top of your head,

2 Did you think about EACH individual trip and add them up,

3 Did you think about TYPES of trips and use that to estimate

4 Or did you estimate based on a GENERAL IMPRESSION?

8 DK (VOL)

9 Refused (VOL)

Next, I'd like to ask you a few lifestyle questions.

[ASK FORM 1 ONLY]

P18. During a typical week, about how many hours do you spend using the internet? Please include internet time at work, at home or in other locations, and visiting any kind of web site.

[ASK FORM 2 ONLY]

P19. During a typical week, about how many hours do you spend using the internet? Please include internet time at work, at home or in other locations, and visiting any kind of web site. Do not include time spent doing e-mail.

[ASK FORM 1 ONLY]

P20. Now I would like to ask you about moderate physical activities, including brisk walking, bicycling, vacuuming, gardening, or anything else that causes a small increase in breathing or heart rate. On days when you do moderate physical activities for at least 10 minutes at a time, how much TOTAL time in MINUTES per day do you spend doing these activities? Please do not include time spent at work or commuting to work. [IF NECESSARY: Just your best guess is fine.] [INTERVIEWER REFERENCE: 1 HR=60 MIN, 2 HRS=120 MIN, 3 HRS=180 MIN, 4 HRS=240 MIN, 5 HRS=300 MIN]

INTERVIEWERS: If R says 1 to 9 minutes, code as 0.

[ASK FORM 2 ONLY]

P21. Now I would like to ask you about moderate physical activities, including brisk walking, bicycling, vacuuming, gardening, or anything else that causes a small increase in breathing or heart rate. On days when you do moderate activities for at least 10 minutes at a time, how much TOTAL time in MINUTES per day do you spend doing these activities? [IF NECESSARY: Just your best guess is fine.] [INTERVIEWER REFERENCE: 1 HR=60 MIN, 2 HRS=120 MIN, 3 HRS=180 MIN, 4 HRS=240 MIN, 5 HRS=300 MIN]

INTERVIEWERS: If R says 1 to 9 minutes, code as 0.

[ASK ALL]

P22. Now I'm going to name a few activities. For each one, please tell me whether you like doing this a great deal, like doing it somewhat, dislike doing it somewhat or dislike doing it a great deal. (The first one is…/The next one is…) [INSERT ITEM; RANDOMIZE]. [READ FOR FIRST ITEM, THEN AS NECESSARY: Do you like doing this a great deal, like doing this somewhat, dislike doing it somewhat or dislike doing it a great deal?]

a. Reading a newspaper
b. Watching a movie
c. Cooking a meal at home
d. Reading a book

P23. Now, please tell me if you happened to do any of these activities yesterday or if you did not happen to do them yesterday. (First,) yesterday, did you…?

a. Read a newspaper
b. Watch a movie

c. Cook a meal at home

d. Read a book

Finally, I have just a few more questions for statistical purposes only.

P25. Right now, are you using a hands-free headset, or are you just holding the phone to your ear?

1 Using hands-free headset

2 Holding the phone

8 DK (VOL)

9 Refused (VOL)

P26. Would you mind telling me if I reached you today AWAY from home or AT home?

[INTERVIEWER NOTE: IF THE RESPONDENT WAS AWAY FROM HOME DURING ANY PART OF THE CALL, THEN ENTER 1]

P27. What did you happen to be doing when I called? [key in verbatim response]

P28. Some people multi-task when on the phone while others do not. What, if anything else, did you happen to do while we were talking? [key in verbatim response]

P29. What is your age?

P30. For data quality purposes, how would you rate the clarity of the phone connection for this call? Would you say the connection was… [READ 1-5]

1 Perfect, like we were talking face to face

2 Very good

3 Good

4 Fair, or

5 Poor, like you could barely hear me at times

8 DK (VOL)

9 Refused (VOL)

References

Abraham, Katharine G., Aaron Maitland, and Suzanne M. Bianchi. 2006. "Nonresponse in the American Time Use Survey: Who Is Missing from the Data and How Much Does it Matter?" *Public Opinion Quarterly* 70:676-703.

Albaghal, Tarek. 2005. "Can You Hear Me Now? Differences in Vote Behavior in the Cell and Landline Populations." Paper presented at the Annual Conference of the American Association for Public Opinion Research, Miami Beach, FL.

Albert, Marilyn S. and Edith Kaplin. 1980. "Organic Implications of Neuropsychological Deficits in the Elderly." In L. W. Poon (Ed.) *New Directions in Memory and Aging: Proceedings of the George A. Talland Memorial Conference*. Hillside, NJ: Erlbaum.

Allison, Paul D. 1999. *Logistic Regression Using the SAS System: Theory and Application*. Cary, NC: SAS Institute Inc.

American Association for Public Opinion Research. 2008. "Guidelines and Considerations for Survey Researchers When Planning and Conducting RDD and Other Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers." Available at http://www.aapor.org/uploads/Final_AAPOR_Cell_ Phone_TF_ report_041208.pdf.

Anderson, John R. 1983. *The Architecture of Cognition*, Cambridge, MA: Harvard University Press.

Assael, Henry and John Keon. 1983. "Nonsampling versus Sampling Errors in Survey Research." *Journal of Marketing* 46:114-123.

Bailar, Barbara A. 1968. "Recent Research in Reinterview Procedures." *Journal of the American Statistical Association* 63:41-63.

Baruffol, Eric, Pierre Verger, and Michel Rotily. 2001. "Using the Telephone for Mental Health Surveys: An Analysis of the Impact of Call Rank, Non-response and Interviewer Effect." *Population* 56:987–1010.

Bell, Ralph. 1984. "Item Nonresponse in Telephone Surveys: An Analysis of Who Fails to Report Income." *Social Science Quarterly* 65:207-215.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B* 57:289-300.

Berger, Jonah and Gráinne M. Fitzsimons. 2008. "Dogs on the Street, Pumas on Your Feet: How Cues in the Environment Influence Product Evaluation and Choice," *Journal of Marketing Research* 45:1-14.

Best, Jonathan and Larry Hugick. 2009. "Some Practical Issues in Cell Phone Interviewing from a Phone Room Perspective." Paper presented at the Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

Bethlehem, Jelke. 2002. "Weighting Nonresponse Adjustments Based on Auxiliary Information." In Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. Little (Eds.) *Survey Nonresponse*, New York: John Wiley & Sons, Inc.

Biemer, Paul P. 2001. "Nonresponse Bias and Measurement Bias in a Comparison of Face to Face and Telephone Interviewing." *Journal of Official Statistics* 17:295–320.

Biemer, Paul P., Robert M. Groves, Lars Lyberg, Nancy Mathiowetz, and Seymour Sudman. 1991. *Measurement Error in Surveys*. New York: Wiley.

Bishop, George F., Robert W. Oldendick, and Alfred J. Tuchfarber. 1982. "Political Information Processing: Question Order and Context Effects." *Political Behavior* 4:177-200.

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

Blumberg, Stephen J. and Julian V. Luke. 2007. "Coverage Bias in Traditional Telephone Surveys of Low-Income and Young Adults." *Public Opinion Quarterly* 71:1-16.

____. 2009. "Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July-December 2008." National Center for Health Statistics. Available from: http://www.cdc.gov/nchs/nhis.htm.

Bolstein, Richard. 1991. "Comparison of the Likelihood to Vote among Preelection Poll Respondents and Nonrespondents." *Public Opinion Quarterly* 55:648–50.

Brambila, Nicole C. 2009. "Seniors Increasingly Targeted by Abusers" Posted on The Desert Sun July 18, 2009. Available at http://www.mydesert.com/article/20090718/NEWS01/907180323/1026. Accessed August 12, 2009.

Brehm, John. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation*. University of Michigan Press.

Brick, J. Michael, Pat D. Brick, Sarah Dipko, Stanley Presser, Clyde Tucker and YangYang Yuan. 2007. "Cell Phone Survey Feasibility in the U.S.: Sampling and Calling Cell Numbers Versus Landline Numbers." *Public Opinion Quarterly* 71:23-39.

Brick, J. Michael, Ismael Flores Cervantes, Sunghee Lee, and Greg Norman. "Nonsampling Errors in Dual Frame Telephone Surveys." Survey Methodology, *in press*.

Brick, J. Michael, Sarah Dipko, Stanley Presser, Clyde Tucker and Yangyang Yuan. 2006. "Nonresponse Bias in a Dual Frame Sample of Cell and Landline Numbers." *Public Opinion Quarterly* 70:780-793.

Brick, J. Michael, W. Sherman Edwards, and Sunghee Lee. 2007. "Sampling Telephone Numbers and Adults, Interview Length, and Weighting in the California Health Interview Survey Cell Phone Pilot Study." *Public Opinion Quarterly* 71:793-813.

Brick, J. Michael, Jill Montaquila, Mary Collins Hagedorn, Shelley Brock Roth and Christopher Chapman. 2005. "Implications for RDD Design from an Incentive Experiment." *Journal of Official Statistics* 21:571-589.

Brick, J. Michael and Jerry West. 1992. "Reinterview Program for the 1991 National Household Education Survey." *Proceedings of the Survey Research Methods Section of the American Statistical Association*. Alexandria, VA: American Statistical Association.

Broadbent, D.E. 1957. "A Mechanical Model for Human Attention and Immediate Memory." *Psychological Review* 64:205-214.

Brown, Joshua and Randal ZuWallack. 2009. "Mitofsky-Waksberg Sampling of Cell Phone Numbers." Paper presented at the Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

Brown, Norman R. 1995. "Estimation Strategies and the Judgment of Event Frequency." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21:1539-1553.

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

Campanelli, Pamela and Colm O'Muircheartaigh. 1999. "Interviewers, Interviewer Continuity, and Panel Survey Nonresponse." *Quality & Quantity* 33:59–76.

Cannell, Charles F. and Floyd J. Fowler. 1963. "Comparison of a Self-Enumerative Procedure and a Personal Interview: A Validity Study." *Public Opinion Quarterly* 27:250–264.

Casady, Robert and James Lepkowski. 1993. "Stratified Telephone Survey Designs." *Survey Methodology* 19:103–113.

Chabris, Christopher F. 1999. "Prelude or Requiem for the 'Mozart Effect'?" *Nature* 400:826-827.

Chang, LinChiat and Jon A. Krosnick. 2009. "Comparing Sample Representativeness and Response Quality." *Public Opinion Quarterly* 74:1-38.

Christian, Leah M. and Michael Dimock. 2009. "Where Do We Find You? How to Locate Cell Respondents." Paper presented at the Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

Church, Allan H. 1993. "Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis." *Public Opinion Quarterly* 57:62-79.

Cialdini, Robert B., Sanford L. Braver, and W. Shapard Wolf. 1992. "A New Paradigm for Experiments on the Causes of Nonresponse." Unpublished paper, Arizona State University.

Collins, Allan M. and Elizabeth F. Loftus. 1975. "A Spreading-Activation Theory of Semantic Processing." *Psychological Review* 82:407–428.

Collins, R. Lorraine, Todd B. Kashdan, and Gernot Gollnisch. 2003. "The Feasibility of Using Cellular Phones to Collect Ecological Momentary Assessment Data: Application to Alcohol Consumption." *Experimental and Clinical Psychology* 11:73-78.

Conrad, Fred, Norman R. Brown, and Erin R. Cashman. 1998. "Strategies for Estimating Behavioral Frequency in Survey Interviews." *Memory* 6:339-366.

Converse, Philip. 1964. "The Nature of Belief Systems in Mass Publics." In David Apter (Ed.) *Ideology and Discontent*, New York: Free Press.

____. 1970. "Attitudes and Non-Attitudes: Continuation of a Dialogue." In Edward Tufte (Ed.) *The Quantitative Analysis of Social Problems*, Reading, MA: Addison-Wesley.

____. 1974. “Comment.” *American Political Science Review* 68:650-660.

Couper, Mick P. 1991. “Modeling Survey Participation at the Interviewer Level.” *Proceedings of the Survey Research Methods Section of the American Statistical Association*. Alexandria, VA: American Statistical Association.

Couper, Mick. P. 1997. “Survey Introductions and Data Quality.” *Public Opinion Quarterly* 61:317-338.

Couper, Mick P., Eleanor Singer, and Dick A. Kulka. 1998. “Nonresponse to the 1990 Census: Politics, Privacy, and/or Pressures.” *American Politics Quarterly* 26:59–80.

Curtin, Richard, Stanley Presser, and Eleanor Singer. 2005. “Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69:87-98.

Curtin, Richard, Eleanor Singer, and Stanley Presser. 2007. “Incentives in Random Digit Dial Telephone Surveys: A Replication and Extension.” *Journal of Official Statistics* 23:91-105.

Daigneault, Sylvie, Claude M. J. Braun, and Harry A. Whitaker. 1992. “Early Effects of Normal Aging on Perseverative and Non-perseverative Prefrontal Measures.” *Developmental Neuropsychology* 8: 99-114.

Dayton, James, Emily Burns, Arnold H. Levinson, Naomi Freedner, Kristie Hannah, and Beth Tarallo. 2009. “Improving Response Rates on Both Landline and Cell Surveys Through Strategic Use of Caller ID.” Paper presented at the Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

de Leeuw, Edith, and Wim De Heer. 2002. “Trends in Household Survey Nonresponse: A Longitudinal and International Comparison.” In R. Groves, D. Dillman, J. Eltinge, and R. J. A. Little (Eds.) *Survey Nonresponse* New York: Wiley.

Dempster, Frank N. 1991. “Inhibitory Processes: A Neglected Dimension of Intelligence.” *Intelligence* 15:157-173.

____. 1992. “The Rise and Fall of the Inhibitory Mechanism: Toward a Unified Theory of Cognitive Development and Aging.” *Developmental Review* 12: 45-75.

DiCamillo, Mark. 2008. “Surveying Voters with Cell Phone Listings from Registration-Based Voter Samples in California.” Paper presented at the Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.

Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley and Sons.

Dillman, Don. A., Michael D. Sinclair, and Jon R. Clark. 1993. "Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys." *Public Opinion Quarterly* 57:289-304.

Dimock, Michael, Leah Christian, and Scott Keeter. 2009. "Do Landline RDD Samples Adequately Cover the 'Wireless Mostly'?" Paper presented at the Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

Dove, Anja, Stefan Pollman, Torsten Schubert, Christopher J. Wiggins, D. Yves von Cramon. 2000. "Prefrontal Cortex Activation in Task Switching: An Event-Related fMRI Study." *Cognitive Brain Research* 9:103-109.

Dumontheil, Iroise, Paul W. Burgess, and Sarah-Jayne Blakemore. 2008. "Development of Rostral Prefrontal Cortex and Cognitive and Behavioural Disorders." *Developmental Medicine and Child Neurology* 50:168-181.

Dunkelberg, William C. and George S. Day. 1973. "Nonresponse Bias and Callbacks in Sample Surveys." *Journal of Marketing Research* 10:160-168.

Durbin, J. and Stuart, A. 1951. "Differences in Response Rates of Experienced and Inexperienced Interviewers." *Journal of the Royal Statistical Society, Series A* 114:163-206.

Ericsson, K. Anders and Herbert A. Simon. 1993. *Protocol Analysis: Verbal Reports as Data, Revised Edition.* Cambridge, MA: MIT Press.

Everett, Stephen E. and Shu-Ling C. Everett. 1989. "Effects of Interviewer Affiliation and Sex Upon Telephone Survey Refusal Rates." Paper presented at the Annual Conference of the Midwest Association for Public Opinion Research, Chicago.

Fahimi, Mansour, Dale Kulp and J. Michael Brick. 2008. "Bias in List-Assisted 100-Series RDD Sampling." *Survey Practice,* September 28, 2008.

Freedman, Michelle J, Kristin M. Lester, Cecelia McNamara, Jesse B. Milby, and Joseph E. Schumacher. 2006. "Cell Phone for Ecological Momentary Assessment with Cocaine-Addicted Homeless Patients in Treatment." *Journal of Substance Abuse Treatment* 30: 105-111.

Fleeman, Anna. 2006. "Merging Cellular and Landline RDD Sample Frames: A Series of Three Cell Phone Studies." Paper presented at the Second International Conference on Telephone Survey Methodology, Miami, FL.

____. 2007. “Survey Research Using Cell Phone Samples: Important Operational and Methodological Considerations.” Paper presented at the Annual Conference of the American Association for Public Opinion Research, Anaheim, CA.

____. 2009. “Using Address Frames to Identify Cell-Phone-Only and Cell-Phone-Primary Households.” Paper Presented at the Joint Statistical Meetings, Washington, DC.

Fleeman, Anna and Nicole Wasikowski. 2008. “Identifying Cell-Phone-Only Households Using an Address-Based Sample Frame.” Paper presented at the Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.

Fricker, Scott. 2007. “The Relationship between Response Propensity and Data Quality in the Current Population Survey and the American Time Use Survey.” Paper presented at the Annual Conference of the American Association of Public Opinion Research, Miami Beach, FL.

Fricker, Scott, Mirta Galesic, Roger Tourangeau, and Ting Yan. 2005. “An Experimental Comparison of Web and Telephone Surveys.” *Public Opinion Quarterly* 69:370-392.

Friedman, Alinda, Martha C. Polson, Cameron G. Dafoe, and Sarah J. Gaskill. 1982. Dividing Attention Within and Between Hemispheres: Testing a Multiple Resources Approach to Limited-Capacity Information Processing. *Journal of Experimental Psychology: Human Perception and Performance* 8:625-650.

Frisina, Laurin T. and Randall K. Thomas. 2008. “Response Nondifferentiation and Response Styles in Web-based Surveys: Causes and Consequences.” Paper presented at the Annual RC-33 Logic and Methodology in Sociology Conference, Naples, Italy.

Fuchs, Marek. 2008. “Communicative Channels, Cognitive Processes, and Question Understanding: Results from a Randomized Field Experiment with Video-enhanced Web Surveys” Paper presented at the Annual RC-33 Logic and Methodology in Sociology Conference, Naples, Italy.

Furnham, Adrian and Anna Bradley. 1997. “Music While You Work: The Differential Distraction of Background Music on Cognitive Test Performance of Introverts and Extraverts.” *Applied Cognitive Psychology* 11:445-455.

Fuster, Joaquin M. 1980. *The Prefrontal Cortex (1st ed).* New York: Raven Press.

Galesic, Mirta and Michael Bosnjak. 2009. “Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey.” *Public Opinion Quarterly* 73:349-360.

Galesic, Mirta and Roger Tourangeau. 2007. "What Is Sexual Harassment? It Depends on Who Asks! Framing Effects on Survey Responses." *Applied Cognitive Psychology* 21:189-202.

Galloway, Gantt P., Ryne Didier, Kathleen Garrison, and John Mendelson. 2008. "Feasibility of Ecological Momentary Assessment Using Cellular Telephones in Methamphetamine Dependent Subjects." *Substance Abuse: Research and Treatment* 1:9-14.

Gilbert, Sam J., Stephanie Spengler, Jon S. Simons, J. Douglas Steele, Stephen M. Lawrie, Christopher D. Frith, and Paul W. Burgess. 2006. "Functional Specialization Within Rostral Prefrontal Cortex (Area 10): A Meta-analysis." *Journal of Cognitive Neuroscience* 18:932–948.

Goldman-Rakic, Patricia S. 1987. "Circuitry of Primate Prefrontal Cortex and Regulation of Behavior by Representational Memory." *Handbook of Physiology—The Nervous System* 5:373-417.

Goyder, John. 1987. *The Silent Minority: Nonrespondents on Sample Surveys*. Boulder, CO: Westview Press.

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70:646-675.

Groves, Robert M. and Mick P. Couper. 1998. *Nonresponse in Household Surveys*. New York: Wiley.

Groves, Robert M., Mick P. Couper, Stanley Presser, Eleanor Singer, Roger Tourangeau, Giorgina Piani Acosta and Lindsay Nelson. 2006. "Experiments in Producing Nonresponse Bias." *Public Opinion Quarterly* 70:720-726.

Groves, Robert M. and Nancy H. Fultz 1985. "Gender Effects among Telephone Interviewers in a Survey of Economic Attitudes." *Sociological Methods and Research* 14:31-52.

Groves, Robert M. and James M. Lepkowski. 1985. "Dual Frame, Mixed Mode Survey Designs." *Journal of Official Statistics* 1:263-286.

Groves, Robert M., Stanley Presser and Sarah Dipko. 2004. "The Role of Topic Interest in Survey Participation Decisions." *Public Opinion Quarterly* 68:2-31.

Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. "Leverage-Salience Theory of Survey Participation: Description and an Illustration." *Public Opinion Quarterly* 64:299-308.

Groves, Robert M., Roger Tourangeau, Stanley Presser, Mick P. Couper, Eleanor Singer, Christopher Toppe, Brady West, and John LaFrance. 2009. "A Randomized

Experiment Investigating a Common Cause Model of Survey Nonresponse Error and Survey Measurement Error." Paper presented at the Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

Han, Daifeng and David Cantor. 2008. "Cell Phone-only Households in a National Mail Survey – Who Are They?" Paper presented at the Joint Statistical Meetings, Denver, CO.

Hansen, Kasper M. 2006. "The Effects of Incentives, Interview Length, and Interviewer Characteristics on Response Rates in a CATI-Study." *International Journal of Public Opinion Research* 19:112-121.

Hansen, Morris H., William N. Hurwitz, and Max A. Bershad. 1961. "Measurement Errors in Censuses and in Surveys." *Bulletin of the International Statistical Institute* 38:359-374.

Hansen, Morris H., William N. Hurwitz, and Leon Pritzker. 1964. "The Estimation and Interpretation of Gross Differences and the Simple Response Variance." Contributions to the Statistics Presented to Professor P.C. Mahalanobis on the Occasion of His 70th Birthday. Pergamon Press, Calcutta, India.

Harmon-Jones, Eddie and John J.B. Allen. 2001. "The Role of Affect in the Mere Exposure Effect: Evidence from Psychophysiological and Individual Differences Approaches." *Personality and Social Psychology Bulletin* 27:889-898.

Harris-Kojetin, Brian and Clyde Tucker. 1999. "Exploring the Relation of Economic and Political Conditions with Refusal Rates to a Government Survey." *Journal of Official Statistics* 15:167-184.

Heberlain, Thomas A. and Robert Baumgartner. 1978. "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Publish Literature." *American Sociological Review* 43:447-462.

Holbrook, Allyson L., Jon A. Krosnick, and Alison Pfent. 2008. "The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms" In Lepkowski, J., C. Tucker, M. Brick, E. De Leeuw, L. Japec, P. J. Lavrakas, M. Link, and R. Sangster (Eds.) *Advances in Telephone Survey Methodology*. New York: John Wiley & Sons.

Holtgraves, Thomas. 2004. "Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding." *Personality and Social Psychology Bulletin* 30:161–172.

House, James S. and Sharon Wolf. 1978. "Effects of Urban Residence on Interpersonal Trust and Helping Behavior." *Journal of Personality and Social Psychology* 36:1029-1043.

Howes, Cynthia, Angela DeBello, Kirk Wolter, and Karen Wooten. 2008. “Practical Steps to Conducting Cellular Telephone Surveys.” Paper presented at the Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.

Hyman, Herbert H. and Paul B. Sheatsley. 1950. “The Current Status of American Public Opinion.” In J.C. Payne (Ed.) *The Teaching of Contemporary Affairs: Twenty-first Yearbook of the National Council for Social Studies* (pp. 11-34) New York: National Education Association.

Jersild, Arthur T. 1927. “Mental Set and Shift.” *Archives of Psychology* Whole No. 89.

Jobber, David and Daragh O’Reilly. 1998. “Industrial Main Surveys: A Methodological Update.” *Industrial Marketing Management* 27:95–107.

Jones, Jeff. 2008. “Cell Phones in Primary Pre-Election Surveys” Paper presented at the Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.

Jones, Wesley H. 1979. “Generalizing Main Survey Inducement Methods: Population Interactions with Anonymity and Sponsorship.” *Public Opinion Quarterly* 43:102-111.

Kahneman, Daniel. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Kane, Emily W. and Laura J. Macaulay. 1993. “Interviewer Gender and Gender Attitudes.” *Public Opinion Quarterly* 57:1-28.

Keeter, Scott. 2006. “The Impact of Noncoverage Bias on Polling in the 2004 Presidential Election.” *Public Opinion Quarterly* 70:88-98.

Keeter, Scott, Michael Dimock, Courtney Kennedy, Jonathan Best, and John Horrigan. 2008. “Costs and Benefits of Dual-frame Telephone Survey Designs.” Paper presented at the Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.

Keeter, Scott, Courtney Kennedy, April Clark, Trevor Tompson and Mike Mokrzycki. 2007. “What’s Missing from National Landline RDD Surveys? The Impact of the Growing Cell-only Population.” *Public Opinion Quarterly* 71:772-792.

Kennedy, Courtney. 2007a. “Assessing Measurement Error in Landline and Cell Phone RDD Surveys.” Paper presented at the Annual Conference of the Midwest Association for Public Opinion Research, Chicago.

____. 2007b. “Evaluating the Effects of Screening for Telephone Service in Dual Frame RDD Surveys.” *Public Opinion Quarterly* 71:750-771.

Kennickell, Arthur and Douglas McManus. 1993. “Sampling for Household Financial Characteristics Using Frame Information on Past Income.” *Proceedings of Survey Research Methods Section of the American Statistical Association*, pp. 88–97. Alexandria, VA: American Statistical Association.

Kim, Sun-Woong and James M. Lepkowski. 2002. “Telephone Household Non-coverage and Mobile Telephones.” Paper presented at the Annual Conference of the American Association for Public Opinion Research, St. Petersburg, FL.

King, James. 2009. “Telemarketing Scam Bilking Phoenix Seniors Revealed in Federal Court” Posted on Phoenix New Times.com August 5, 2009. Available at http://blogs.phoenixnewtimes.com/svalleyfever/2009/08/telemarketing_scam_robbing_pho.php. Accessed August 12, 2009.

Klingberg, Torkel. 1998. “Concurrent Performance of Two Working Memory Tasks: Potential Mechanisms of Interference.” *Cerebral Cortex* 8: 593-601.

Knäuper, Bärbel. 1999. “The Impact of Age and Education on Response Order Effects in Attitude Measurement.” *Public Opinion Quarterly* 63:347-370.

Koechlin, Etienne, Gianpaolo Basso, Pietro Pietrini, Seth Panzer, and Jordan Grafman. 1999. “The Role of Anterior Prefrontal Cortex in Human Cognition.” *Nature* 399:148–151.

Kray, Jutta and Ulman Lindenberger. 2000. “Adult Age Differences in Task Switching.” *Psychology and Aging* 15:126–147.

Krosnick, Jon A. 1991. “Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys.” *Applied Cognitive Psychology* 5:213-236.

Krosnick, Jon A. and Duane F. Alwin. 1987. “An Evaluation of a Cognitive Theory of Response Order Effects in Survey Measurement.” *Public Opinion Quarterly* 51:201-219

Kulka, Richard A. 1994. “The Use of Incentives to Survey ‘Hard-to-Reach’ Respondents: A Brief Review of Empirical Research and Current Practice.” Alexandria, VA: Council of Professional Associations for Federal Statistics (COPAFS).

Kulka, Richard. A., Jennifer J. McNeill, and Arthur J. Bonito. 1982. “On the Manifest Designation of Key Items: A Cost Effective Procedure for Improving the Collection and Processing of Survey Data.” Paper presented at the Annual Conference of the American Association for Public Opinion Research, Hunt Valley, MD.

Kulp, Dale W. 2004. “Meeting the Challenge of the Telephone Consumer Protection Act and the FCC Telecommunications Act: Development of a Process to Identify

Landline Telephone Numbers Ported to Cellular Phones" Paper presented at the Joint Statistical Meetings, Toronto.

Kuusela, Vesa, Mario Callegaro, and Vasja Vehovar. 2008. "The Influence of Mobile Telephones on Telephone Surveys." In Lepkowski, J., C. Tucker, M. Brick, E. De Leeuw, L. Japec, P. J. Lavrakas, M. Link, and R. Sangster (Eds.) *Advances in Telephone Survey Methodology*. New York: John Wiley & Sons.

Lavrakas, Paul J., Charles D. Shuttles, Charlotte Steeh and Howard Fienberg. 2007. "The State of Surveying Cell Phone Numbers in the United States: 2007 and Beyond." *Public Opinion Quarterly* 71:840-854.

Lavrakas, Paul J., Trevor N. Tompson, and Robert Benford. 2009. "Investigating Data Quality in Cell Phone Surveying." Paper presented at the Annual Conference of the Midwest Association for Public Opinion Research, Chicago.

Lee, Angela Y. and Aparna Labroo. 2004. "Effects of Conceptual and Perceptual Fluency on Affective Judgment." *Journal of Marketing Research* 41:151-165.

Lepkowski, James M. and Mick Couper. 2002. "Nonresponse in the Second Wave of Longitudinal Household Surveys." In Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. Little (Eds.) *Survey Nonresponse*, New York: John Wiley & Sons, Inc.

Lessler, Judith T. and William D. Kalsbeek. 1992. *Nonsampling Error in Surveys*. New York: John Wiley & Sons, Inc.

Lin, I-fen and Nora Cate Schaeffer. 1995. "Using Survey Participants to Estimate the Impact of Nonparticipation." *Public Opinion Quarterly* 59:236-258.

Link, Michael W., Michael P. Battaglia, Martin R. Frankel, Larry Osborn and Ali H. Mokdad. 2007. "Reaching the U.S. Cell Phone Generation: Comparison of Cell Phone Survey Results with an Ongoing Landline Telephone Survey." *Public Opinion Quarterly* 71:814-839.

____. 2008. "A Comparison of Address-Based Sampling (ABS) Versus Random-Digit Dialing (RDD) for General Population Surveys." *Public Opinion Quarterly* 72:6-27.

Link, Michael W., Gail Daily, Charles D. Shuttles, Christine Bourquin, and Tracie Yancey. 2008. "Addressing the Cell Phone-Only Population: Cell Phone Sampling Versus Address Based Sampling." Paper presented at the Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.

Link, Michael W., Gail Daily, Charles D. Shuttles, L. Tracie Yancey, Anh Thu Bursk, and Christine Bourquin. 2009. “Building a New Foundation: Transitioning to Address Based Sampling After Nearly 30 Years.” Paper presented at the Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

Link, Michael W. Ali H. Mokdad, Dale Kulp, and Ashley Hyon. 2006. “Has the National Do Not Call Registry Helped or Hurt State-level Response Rates?: A Time Series Analysis.” *Public Opinion Quarterly* 70:794-809.

Link, Michael W. and Robert W. Oldendick. 1999. “Call Screening: Is It Really a Problem for Survey Research?” *Public Opinion Quarterly* 63:577-589.

Lohr, Sharon, Burke Grandjean, and Patricia A. Taylor. 2009. “Exploring the Overlap Domain in a National Dual-frame (Cell/Landline) Survey: Results from the National Park Service Comprehensive Survey of the American Public.” Paper Presented at the Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

Lynn, Peter. 2003. “PEDAKSI: Methodology for Collecting Data about Survey Non-respondents.” *Quality and Quantity* 37:239 – 261.

Mazzae, Elizabeth N., Thomas A. Ranney, Ginger S. Watson, and Judith A. Wightman. 2004. “Hand-held or Hands-free? The Effects of Wireless Phone Interface Type on Phone Task Performance and Driver Performance.” *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 2218–2221). Santa Monica, CA: Human Factors and Ergonomics Society.

McCarthy, Jaki Stanley, Daniel G. Beckler, and Suzette M. Qualey. 2006. “An Analysis of the Relationship Between Survey Burden and Nonresponse: If We Bother Them More, Are They Less Cooperative?” *Journal of Official Statistics* 22:97-112.

Menon, Geeta. 1993. “The Effects of Accessibility of Information in Memory on Judgments of Behavioral Frequency.” *The Journal of Consumer Research* 20:431-440.

Meyer, David E. and David E. Kieras. 1997. “A Computational Theory of Executive Cognitive Processes and Multiple-task Performance: Part 1. Basic Mechanisms.” *Psychological Review* 104:3-65.

Mokrzycki, Michael, Scott Keeter, and Courtney Kennedy. 2010. “Cell-Phone-Only Voters in the 2008 Exit Poll and Implications for Future Non-coverage Bias.” *Public Opinion Quarterly* (Accepted).

Monahan, Jennifer L., Sheila T. Murphy and Robert B. Zajonc. 2000. “Subliminal Mere

Exposure: Specific, General, and Diffuse Effects." *Psychological Science* 11:462-466.

Moreland, Richard L. and Scott R. Beach. 1992. "Exposure Effects in The Classroom: The Development of Affinity Among Students." *Journal of Experimental Social Psychology* 28:527-555.

National Academy of Sciences. 1979. "Panel of Privacy and Confidentiality as Factors in Survey Response." Washington DC: National Academy Press.

Nedungadi, Prakash. 1990. "Recall and Consumer Consideration Sets: Influencing Choice without Altering Brand Evaluations," *Journal of Consumer Research* 17:263–276.

Neely, James H. 1977. "Semantic Priming and Retrieval from Lexical Memory: Roles of Inhibitionless Spreading Activation and Limited Capacity Attention." *Journal of Experimental Psychology: General* 106:226-254.

Norman, Donald A. and Daniel G. Bobrow. 1975. "On Data-limited and Resource-limited Processes." *Cognitive Psychology* 7:44-64.

Norman, Greg and Richard Sigman. 2009. "Using Addresses as Sampling Units in the 2007 Health Information National Trends Survey." Paper presented at the Joint Statistical Meetings, Washington, DC.

Norris, Douglas A. and Jeff Hatcher. 1994. "The Impact of Interviewer Characteristics in a National Survey of Violence Against Women." *Proceedings of the Survey Research Methods Section of the American Statistical Association*. Alexandria, VA: American Statistical Association.

Oksenberg, Lois and Charles Cannell. 1988. "Effects of Interviewer Vocal Characteristics on Nonresponse." In R. M. Groves, P. P. Biemer, and L. E. Lyberg (Eds.) *Telephone Survey Methodology* (pp. 257-269) New York: Wiley.

Olson, Kristen. 2006. "Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Survey Error." *Public Opinion Quarterly* 70:737-758.

Olson, Kristen. 2007. An Investigation of the Nonresponse/Measurement Error Nexus. Doctoral dissertation. University of Michigan, Ann Arbor.

O'Muircheartaigh, Colm. 1991. "Simple Response Variance: Estimation and Determinants." In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (Eds.) *Measurement Error in Surveys*. New York: Wiley.

Ottati, Victor C., Ellen J. Riggle, Robert S. Wyer, Norbert Schwarz, and James Kuklunski. 1989. "Cognitive and Affect Bases of Opinion Survey Responses." *Journal of Personality and Psychology* 57:404-415.

Pickery, Jan and Geert Loosveldt. 2002. "A Multilevel Multinomial Analysis of Interviewer Effects on Various Components of Unit Nonresponse." *Quality & Quantity* 36:427–437.

Putnam Robert D. 1995. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Rauscher, Frances H, Gordon L. Shaw, and Catherine N. Ky 1993. "Music and Spatial Task Performance." *Nature* 365:611.

Reimers, Stian and Elizabeth A. Maylor. 2005. "Task Switching Across the Life Span: Effects of Age on General and Specific Switch Costs." *Developmental Psychology* 41:661–671.

Remington, Todd D. 1992. "Telemarketing and Declining Survey Response Rates." *Journal of Advertising Research, Research Currents*, May–June, pp. RC6–RC7.

Rips, Lance J., Frederick G. Conrad, and Scott S. Fricker. 2003. "Straightening the Seam Effect in Panel Surveys." *Public Opinion Quarterly* 67:522-554.

Rizzo, Lou, Graham Kalton, and J. Michael Brick. 1996. "A Comparison of Some Weighting Adjustment Methods for Panel Nonresponse." *Survey Methodology* 22:43–53.

Rogelberg, Steven G., Gwenith G. Fisher, Douglas C. Maynard, Milton D. Hakel, and Michael Horvath. 2001. "Attitudes Toward Surveys: Development of a Measure and its Relationship to Respondent Behavior." *Organizational Research Methods* 4:3-25.

Salvanto, Anthony M. 2008. "The Accidental Cell: A New Mitofsky-Waksberg Experiment and Analysis of CBS News' Cell Phone Respondents in National Surveys." Paper presented at the Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.

Schaeffer, Nora Cate and Douglas W. Maynard. 2002. "Occasions for Intervention: Interactional Resources for Comprehension in Standardized Survey Interviews." In *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, Douglas W. Maynard, Hanneke Houtkoop-Steenstra, Nora Cate Schaeffer, and Johannes van der Zouwen (Eds.) New York: Wiley.

Schleifer, Stephen. 1986. "Trends in Attitudes Toward and Participation in Survey Research." *Public Opinion Quarterly* 50:17-26.

Schroeder, Paul and Brian Meekins. 2009. “Have I Caught You at a Bad Time? Call Protocol and Bias in the 2008 BRFSS Cell Phone Sample.” Paper presented at the Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

Schuman, Howard and Jean M. Converse. 1971. “The Effects of Black and White Interviewers on Black Responses in 1968.” *Public Opinion Quarterly* 35:44-68.

Schuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitudes Surveys*. New York: Academic Press.

Schwarz, Norbert and Julia Bienias. 1990. “What Mediates the Impact of Response Alternatives on Frequency Reports of Mundane Behaviors?” *Applied Cognitive Psychology* 4:61-72.

Schwarz, Norbert, Hans-J. Hippler, Brigitte Deutsch, and Fritz Strack.1985. “Response Scales: Effects of Category Range on Reported Behavior and Subsequent Judgments.” *Public Opinion Quarterly* 49:388-395.

Schwarz, Norbert and Seymour Sudman. 1992. *Context Effects in Social and Psychological Research*. New York: Springer-Verlag.

Schwarz, Norbert and Gerald L. Clore. 1983. “Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States.” *Journal of Personality and Social Psychology* 45:513-523.

Schyberger, Bo W. 1967. “A Study of Interviewer Behavior.” *Journal of Marketing Research* 4:32-35.

Shtatland, Ernest S., Ken Kleinman, and Emily M. Cain. 2004. “A New Strategy of Model Building in PROC LOGISTIC With Automatic Variable Selection, Validation, Shrinkage and Model Averaging.” Presented at the SAS Users Group International Conference, Montreal, Canada.

Simon, Herbert. A. 1957. *Models of Man: Social and Rational*. New York: Wiley.

Singer, Eleanor, Martin R. Frankel and Marc B. Glassman. 1983. “The Effect of Interviewer Characteristics and Expectations on Response.” *Public Opinion Quarterly* 47:171-188.

Singer, Eleanor, John Van Hoewyk, and Mary P. Maher. 2000. “Experiments with Incentives in Telephone Surveys.” *Public Opinion Quarterly* 64:171–188.

Singer, Eleanor, John Van Hoewyk, and Randall J. Neugebauer. 2003. “Attitudes and Behavior: The Impact of Privacy and Confidentiality Concerns on Participation in the 2000 Census.” *Public Opinion Quarterly* 67:368-384.

Singer, Eleanor, Dawn R. Von Thurn, and Esther R. Miller. 1995. “Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature.” *Public Opinion Quarterly* 59:66-77.

Skalland, Benjamin, Martin Barron, and Karen Wooten. 2009. “Operational Results from an Address-Based Sampling Pilot for the National Immunization Survey.” Paper presented at the Joint Statistical Meetings, Washington, DC.

Slymen, Donald J., Joseph A. Drew, Bridget L. Wright, John P. Elder, and Stephen J. William. 1994. “Item Non-response to Lifestyle Assessment in an Elderly Cohort.” *International Journal of Epidemiology* 23:583-591.

Spector, Amos and Irving Biederman. 1976. “Mental Set and Shift Revisited.” *American Journal of Psychology* 89:669-679.

Steeh, Charlotte. 2004. “A New Era for Telephone Surveys.” Paper presented at the Annual Conference of the American Association for Public Opinion Research, Phoenix, AZ.

Stone, Arthur A. and Saul Shiffman. 1994. “Ecological Momentary Assessment in Behavioral Medicine.” *Annals of Behavioral Medicine* 16:199–202

Strack, Fritz, Norbert Schwarz, and Michaela Wankë. 1991. “Semantic and Pragmatic Aspects of Context Effects in Social and Psychological Research.” *Social Cognition* 9:111-125.

Strayer, David L., Frank A. Drews, and Dennis J. Crouch. 2006. “A Comparison of the Cell Phone Driver and the Drunk Driver.” *Human Factors* 48:381-391.

Thorne, Jeremy D. and Stefan Debener. 2008. “Irrelevant Visual Stimuli Improve Auditory Task Performance.” *NeuroReport* 19:553-557.

Tortora, Robert D. 2004. “Response Trends in a National Random Digit Dial Survey.” *Metodološkizvezki* 1:21–32.

Tourangeau, Roger. 1984. “Cognitive Science and Survey Methods.” In T. Jabine, M. Straf, J. Tanur, and R. Tourangeau (Eds.), *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines* (pp. 73-100). Washington, D.C.: National Academy Press.

____. 2004. “Survey Research and Societal Change.” *Annual Review of Psychology* 55:775-801.

Tourangeau, Roger, Robert M. Groves, Courtney Kennedy, and Ting Yan. 2009. "The Presentation of a Web Survey, Nonresponse and Measurement Error among Members of Web Panel "*Journal of Official Statistics* 25:299-321.

Tourangeau, Roger, Lance J. Rips, and Kenneth A. Rasinski. 2000. *The Psychology of Survey Response.* Cambridge: Cambridge University Press.

Tourangeau, Roger and Kenneth A. Rasinski. 1988. "Cognitive Processes Underlying Context Effects in Attitude Measurement." *Psychological Bulletin* 103:299-314.

Tourangeau, Roger and Kenneth A. Rasinski, Norman Bradburn, and Roy D'Andrade. 1989. "Belief Accessibility and Context Effects in Attitude Measurement." *Journal of Experimental Social Psychology* 25:401-421.

Traugott, Michael W. and Sung-Hee Joo. 2003. "Differences in the Political Attitudes and Behavior of Cell and Landline Telephone Users." Paper presented at the Annual Conference of the American Association of Public Opinion Research, Nashville, TN.

Traugott, Michael W., Sung-Hee Joo, Richard Ling, and Ying Qian. 2006. "On the Move: the Role of Cellular Communications in American Life." Pohs Report on Mobile Communication. Ann Arbor, MI: University of Michigan.

Tuckel, Peter and Harry O'Neill. 1996. "New Technology and Nonresponse Bias in RDD Surveys." *Proceedings of the Survey Research Methods Section of the American Statistical Association*. Alexandria, VA: American Statistical Association.

Tucker, Clyde, J. Michael Brick, and Brian Meekins. 2007. "Household Telephone Service and Usage Patterns in the United States in 2004: Implications for Telephone Samples." *Public Opinion Quarterly* 71: 3-22.

Turakhia, Chintan, Mark A. Schulman, and Seth Bohinsky. 2008. "Experiments on Use of Internet Cell Phone Only Panelists to Supplement RDD Samples." Paper presented at the Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.

Vehovar, Vasja and Mario Callegaro. 2007. "Social Aspects of Mobile Phone Usage and Their Impact on Survey Participation." Paper presented at the Annual Conference of the American Association of Public Opinion Research, Anaheim, CA.

Voogt, Robert J.J. and Hetty van Kempen. 2002. "Nonresponse Bias and Stimulus Effects in the Dutch National Election Study." *Quality and Quantity* 36:325-345.

Wallis, Claudia. 2006. "The Multitasking Generation." TIME. Available at http://www.time.com/time/printout/0,8816,1174696,00.html.

Whittlesea, Bruce W.A. 1993. “Illusions of Familiarity.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:1235-1253.

Wickens, Christopher D. 1980. “The Structure of Attentional Resources” In R. Nickerson (Eds), *Attention and Performance, Vol. VIII.* Hillsdale, NJ: Erlbaum.

Wilson, Timothy D., Dolores Kraft, and Dana S. Dunn. 1989. “The Disruptive Effects of Explaining Attitudes: The Moderating Effect of Knowledge about the Attitude Object.” *Journal of Experimental Social Psychology* 25:379-400.

Witt, Lindsey, Randall ZuWallack, and Frederica Conrey. 2009. “Out and About: An Evaluation of Data Quality in Cell Phone Surveys.” Paper presented at the Annual Conference of the American Association of Public Opinion Research, Hollywood, FL.

Zabel, Jeffrey E. 1998. “An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior.” *Journal of Human Resources* 33(2):479–506.

Zajonc, Robert B. 1968. “Attitudinal Effects of Mere Exposure.” *Journal of Personality and Social Psychology: Monograph Supplement* 9:1-27.

Zaller, John and Stanley Feldman. 1992. “A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences.” *American Journal of Political Sciences* 36:579-616.