A Framework for Survey Breakoffs

Andy Peytchev, Survey Methodology Program, University of Michigan
Institute for Social Research, 426 Thompson St., Rm. 4062, Ann Arbor, MI 48104, andrey@umich.edu

Keywords: Breakoffs, Nonresponse, Measurement Error Model, Theoretical Framework, Web surveys.

1 Abstract

There is very little that is known about breakoffs as a form of nonresponse. Without understanding the mechanisms generating them, breakoffs tend to be treated together with unit-nonresponse. Understanding the unique and the common causes of breakoffs would allow for reduction of breakoffs through survey design, as well as for better informed imputation and weighting adjustment models. Studying breakoffs could also provide insight into factors affecting cooperation decisions in a manner that could not be done before – in mail surveys, for example, only *a priori* hypotheses have been possible due to a single binary outcome being observed.

A framework for breakoff and associated response behaviors in web surveys is proposed. The framework incorporates respondent and survey related factors, the multiple response decisions that are possible in web surveys, and presents survey breakoff as a participation decision that is continuously reevaluated throughout the survey.

An empirical test was conducted on two similar large scale web survey experiments with 6,000 completed surveys and a 12% overall breakoff rate. A discrete hazard survival model with page-varying and page-invariant covariates was used to estimate the risk of breakoff. The size of the task indicated by the number of questions in a page, comprehension (long questions), mapping (open-end response format), and topic & commitment (introductory screens) led to higher risk of breakoff. There was some indication that questions requiring more judgment processes to cause breakoffs, and no effect was found of retrieval-intensive and editing-intensive (sensitive or threatening) questions. The effect of respondent education was consistent with a hypothesis of cognitive ability. Contrary to previous beliefs, those who break off tend to be more careful respondents who may be finding various tasks too difficult, rather than being careless respondents who haphazardly break off.

2 Background and Significance

Survey breakoff is when a respondent starts the survey but stops providing answers at some point during the survey. Like unit-nonresponse and other behaviors causing errors of nonobservation, breakoffs present a threat to survey inference. This threat is large in web surveys, with meta-analyses finding median reported breakoff rates of 16% and 34% (Lozar-Manfreda and Vehovar, 2002; Musch and Reips, 2000). Despite the magnitude of the problem, breakoffs have received very little attention in the literature.

Breakoffs are not unique to web surveys (Blumberg et al., 2005; Catania et al., 1996; Groves and Kahn, 1979; Olson et al., 2004; Stussman, Taylor and Riddick, 2003). They could also be manipulated - allowing respondents in a telephone survey on sex topics to choose the gender of the interviewer significantly reduced breakoffs (Catania et al., 1996). However, using longer questions increased breakoffs in a manner we would expect as in self-administered surveys.

The findings from telephone and face-to-face methods of data collection are not directly relevant to web surveys. The nature of the social interaction in interviewer-administered surveys may attenuate the effect of causes of breakoffs (that could otherwise play a large role in web surveys), such as lack of interest in the topic. Therefore most of the breakoffs in interviewer-administered surveys occur during the screening and respondent selection stages, such as in the Behavior Risk Factor Surveillance System survey (Osborn, Blumberg and 0lson, 2000), the National Health Interview survey (Stussman et al., 2003), the National Survey of Children's Health Interview survey, and the National Immunization Survey (Olson et al., 2004), which is not very informative about the mechanisms producing breakoffs throughout surveys.

Mail surveys lack the social interaction and therefore questionnaire characteristics can more easily affect respondent cooperation. Unfortunately, in mail surveys only a binary outcome is observed – the respondent either returned the survey or failed to do so. The questionnaire characteristics that affect the participation decision can only be hypothesized *a priori* and tested through split-ballot experiments. For example, a researcher can hypothesize that white space can make a survey look easier and vary this feature to random parts of the sample and compare their response rates (e.g., Champion and Sear, 1969). This is not only inefficient as each sampled respondent has only one of two outcomes in a between-subjects analysis, but more importantly, relies on survey methodologists to be able to come up with all design characteristics affecting survey cooperation in order to later test them in targeted experiments.

The rare use of screening and respondent selection, the absence of a social interaction with an interviewer, or the likely combination of these two and other factors leads to higher rates of breakoff throughout the entire questionnaire in web surveys. That is, in web surveys respondents seem to be reevaluating their decision to participate on every page and possibly on every question, and the outcome of each decision can be recorded. This view of multiple decisions to participate in a survey is valuable in beginning to understand participation decisions. The rate of breakoff by respondents with particular characteristics on pages with certain cognitive tasks is analogous to studying reasons for nonresponse through costly split-ballot experiments, but does not require explicit experimentation. Furthermore, it permits the discovery of factors that have not been considered and tested before.

Unfortunately, breakoffs are treated as unit nonrespondents, commonly combining the two when studying reasons for nonresponse and postsurvey adjustments fail to differentiate them. This is particularly erroneous if we find that the same reasons that induce unit nonresponse do not induce breakoffs. Researchers may be wrong in adopting a framework for unit nonresponse such as Groves and Couper's (1998) for

The surveys used here were supported by grants from the National Science Foundation (SES-0106222) and NIH (R01 HD041386-01A1) to Roger Tourangeau, Mick Couper, Fred Conrad, and Reg Baker. Any opinions, findings, or conclusions expressed here are those of the author and do not reflect the views of the National Science Foundation or the National Institutes of Health. The author wishes to thank the Primary Investigators for the use of these data and the opportunity to have been involved on these projects.

breakoffs that may imply respondent-level characteristics. Instead, there may be causes for item nonresponse as identified in Beatty and Herrmann's framework (2002) such as respondent's inability to retrieve information, that are also associated with breakoffs. For example, one way to minimize item-nonresponse in web surveys is to use prompts. Mooney, Rogers, and Trunzo (2003) found that while prompts decreased item-nonresponse, their use may increase breakoff rates. Such relationships between survey behaviors are generalizable when the causes are identified.

Groves and Couper (1998) provide a framework for a single cooperation decision in household surveys, classifying factors into environment, respondent, survey, and interviewer. Yet decisions to breakoff are conditional on the respondent having made the decision to start, and are likely to be caused by individual features within the questionnaire. Beatty and Herrmann's framework (2002) for item nonresponse may be more relevant here as one reason they identify as a cause of item nonresponse is the respondent's inability to retrieve the information needed to answer the question. As respondents continuously reevaluate their participation in the survey, an

incentives, there are also unique factors such as questionnaire layout, respondent equipment, etc. The household interview cooperation model has a single decision, whether to cooperate or not, but a model for web survey participation has to accommodate a decision to start the survey that can result in unit-nonresponse, multiple decisions to continue observable as breakoffs on different pages, decision to answer a particular question within a page that can result in item-nonresponse, and a decision on how to answer the question affecting the measurement properties of the response.

There are three sets of factors: respondent, survey design[1], and questionnaire characteristics. Respondent factors include factors such as environment, socio-demographic characteristics, predispositions, previous survey experience, and response behavior. The common examples in the literature are age, gender, race and ethnicity, and various study specific categorizations that are related to participation decisions, such as teachers vs. students in academic populations. Most of these do not have a causal relationship with response outcomes, but are proxies for many of the causes that cannot be measured. Education, for example, is a common proxy for the respondent's cognitive ability. Although not always causally, they provide the ability to predict respondent behaviors. Such variables are useful prior to data collection for sample stratification and using different survey designs between strata, and after the study, for postsurvey adjustments such as weighting. More respondent-specific factors are much more difficult to study, to a large degree because frames seldom provide data at the respondent level that are strongly related to response decisions, such as previous survey experience and previous response outcomes. Such factors include predispositions toward surveys, web surveys, and can be specific to a survey, such as topic involvement (e.g. Goyder, 1987; Groves, Presser and Dipko, 2004). To the extent that respondents are consistent in their response decisions across surveys and this consistency cannot be explained sufficiently with socio-demographic covariates, better attempts at measuring these predispositions are needed.

Another set of factors are survey design features. Those have received the most attention as many are under the researcher's control and can be manipulated at relatively low cost in web surveys. These features include the announced survey length (e.g. Crawford, Couper and Lamias, 2001; Trouteaud, 2004), sending prenotification (e.g. Kaplowitz, Hadlock and Levine, 2004), sending reminders (e.g. Deutskens, de Ruyter, Wetzels and Oosterveld, 2004; Kaplowitz et al., 2004), content of invitation (e.g. Trouteaud, 2004), timing of reminders (e.g. Crawford et al., 2001), type and amount of incentives (Bosnjak and Tuten, 2003; Deutskens et al., 2004). Other survey design features are also known to have an impact on cooperation decisions and in a manner that introduces systematic bias in the survey estimates (such as the survey topic and the sponsor of the survey).

Questionnaire characteristics do not affect initial nonresponse decisions as shown in Figure 1, but can affect decisions on whether to continue the survey. Such characteristics are seen by the respondent only after starting the survey. These include the type of questions asked, such as open-ended (Crawford et al., 2001), the actual length of the survey (e.g. Deutskens et al., 2004; Hogg and Miller, 2003; MacElroy, 2000), use of grid layout (O'Neil, Penrod and Bornstein, 2003), and technical difficulties (e.g. Schwarz and Reips, 2001).

This framework is needed not just to illustrate the possibility of common factors affecting different response outcomes (and creating correlations between outcomes), but also to delineate that they are different outcomes resulting from multiple decision processes and attention should be devoted to each one. The framework also identifies the place for breakoffs within the participation process in web surveys. This paper attempts to address one of these decision processes within this framework, studying respondent and questionnaire factors affecting the decisions to breakoff throughout the survey, bolded in Figure 1.

There are different patterns for breakoff that could be postulated. Respondents may still be making a decision in the alternative to item nonresponse is breakoff. But this framework is limited to reasons for answering or not answering a question. Factors affecting how a question is answered may also affect breakoff. In the most general view, the survey response process model (Cannell, Miller and Oksenberg, 1981; Cannell, Marquis and Laurent, 1977; Strack and Martin, 1987; Tourangeau, 1984; 1987) provides a theoretical structure. How the respondent comprehends the question, retrieves the information, uses it to make a judgment, maps it onto the response options, and edits it based on sensitivity or threat, affects measurement properties and is likely to be also affecting item nonresponse and breakoffs.

For these reasons there is a need for a new framework for studying participation decisions in web surveys: there are *multiple participation decisions* leading to *multiple outcomes* in web surveys that are made *throughout* the survey administration, *multiple factors* that can be *common* to more than one decision, and some of these factors are *unique to web surveys*. While some factors can be shared with those for household interview surveys, such as topic of the survey and

[1] Survey design is often confused with questionnaire design. Survey design refers to a set of study characteristics, such as topic, sponsor, and data collection protocol.

initial questions whether to do the survey, thereby breakoffs occurring at the beginning. It could also be fatigue, leading to increasing breakoffs towards the end of the survey, given that the survey is of sufficient length to observe this phenomenon. Thirdly, breakoffs may be a somewhat random phenomenon with a somewhat uniform rate of breakoffs throughout the survey.

I propose that breakoffs are instead a function of more complicated mechanisms that involve the different cognitive tasks in each page in the questionnaire, burden accumulated over the course of the survey, overall survey design features, respondent characteristics, respondent environment, and respondent behavior. Understanding the mechanisms for survey breakoffs would allow better design of surveys. For example, a set of questions that rely on the respondents' ability to retrieve information may induce higher rates of breakoff. Placing them at the beginning of the survey will result in collecting fewer data on these respondents. To the extent that those who broke off are different from those who completed the survey on key variables, biases in survey estimates will be higher than if the retrieval-intensive questions had been placed at the end of the survey. Furthermore, to the extent that different types of questions lead to different respondents breaking off, order of the questions or sets of questions within the questionnaire will influence the mix of respondents and affect survey estimates.

Reduction and adjustment for survey breakoff requires knowledge of the causes or at least the covariates of breakoff. Studying breakoffs systematically within a framework is essential in identifying a more comprehensive set of factors needed to reduce and/or adjust for this nonrandom form of nonresponse within a questionnaire.

One expectation is that long questions place more burden on respondents in terms of comprehension and will be associated with higher breakoffs, controlling for other features of that question and page. Similarly, questions that ask respondents to recall behaviors will be retrieval intensive, questions that ask for attitudes to be constructed will be judgment intensive, numeric questions that require the respondent to type and even more so for open-ended questions will require more effort in providing the answer, and sensitive or threatening questions will affect the respondent's editing of the response. The number of questions per page may affect how burdensome the task is perceived. As introductory screens often switch topics, and also announce the beginning of a new part of the survey which can be similar to the initial survey participation decision, they are expected to increase the risk of breakoff just as in earlier findings in telephone surveys (e.g., Groves and Kahn, 1979).

If respondents keep reevaluating their decision to participate at the beginning of the survey, the cumulative number of questions asked should be negatively associated with the likelihood of breakoff. However, if respondents break off due to burden or fatigue that cumulates over the course of the survey, there would be a positive association between the number of questions asked and the likelihood of breakoff. If the individual page and the respondent characteristics are the dominant factors for breakoff, the cumulative number of questions that have been asked should not be associated with breakoff.

Since we do not know respondent characteristics of those who breakoff, based on findings on unit nonresponse, nonwhite, less educated, and younger respondents would be expected to break off at higher rates. That is, question and page characteristics are likely the driving force for survey breakoffs, but respondent characteristics associated with unit nonresponse may show a limited link between unit nonresponse and breakoffs.

In the interest of adjustment models, paradata can be collected on those who start the survey, such as the computer's operating system, screen size, and internet browser type. While these variables are difficult to interpret as they form both the survey environment (how the survey is rendered to the respondent) and the respondent's choice in technology, this information is collected on those who complete and those who breakoff. To the extent that these paradata are also associated with variables of interest, valuable bias-reducing postsurvey adjustments can be made.

Finally, if we can obtain behavioral measures on those who complete and those who breakoff, not only could we improve adjustment models, but also better understand why some respondents are more likely to breakoff. It is a widespread assumption that those who break off are respondents who answer the survey haphazardly and are not putting in the effort anyway. If they are less committed to the survey task, then they will tend to answer questions faster. However, if they are actually the respondents who put in effort but have difficulty with the task at hand causing them to break off, they will be answering the questions slower than the rest of the respondents.

3 Data and Methods

Two web survey experiment studies were conducted in December 2003 – January 2004 (Study 1) and December 2004 – January 2005 (Study 2). These studies had the objective of testing various visual design features in web surveys, which also provides a needed variety of questions for the current analysis of breakoffs. The surveys have identical design as part of the same program of research, conducted by Market Strategies Inc. and designed by the same primary investigators with considerable overlap (replication in the second survey) of experiments. The samples were drawn in the same way from the same pool of respondents, just over half from the Survey Sciences Inc. (SSI) national web panel and the rest recruited through pop-up messages in America On-Line (AOL) web pages, shown in the last column in Table 1. Among the 3,195 respondents who started the first study, 478 broke off (15%), and among the 2,831 who started the second study, 244 broke off (9%). Study 1 was longer, as the median time for completing it was 21.3 minutes versus 18.3 minutes for Study 2. Apart from a different mix of questions, there was an experiment unique to Study 1 that produced high breakoff rates on a few questions, which analyses can account for.

The questionnaire involved no skip patterns, so that exposure to particular questions is not dependent on substantive responses and respondent characteristics, needed for unbiased estimation of causes for breakoff. There was random assignment of the order of questions, as some conditions randomly assigned different versions of questions and also placement of questions within the questionnaire. Surveys commonly confound the page and question characteristics with

location in the questionnaire as questions follow a predefined order in the questionnaire. The random assignment of order allows the separation of effects of characteristics from location and context so that the fifth page for one respondent may have a single sensitive closed-ended question that requires retrieval, while for another respondent it is randomly assigned to have eight nonsensitive closed-ended questions that require more judgment processes, and yet other combination of features assigned to other respondents.

Each sequential page was coded for each respondent (i.e., accounting for all possible conditions that manipulated which questions were displayed and in what order in the questionnaire for different respondents) [2]. The variables that were coded for each page included: number of questions in the page, number of questions in the previous page, cumulative number of questions, whether the question was long, required a closed-ended, a numeric, or a verbal response, whether the page was an introduction to a new section of the questionnaire, whether any of the questions were on sensitive topics, required retrieval (typically behavioral questions), required judgment (typically attitudinal questions), whether definitions were provided for key terms in the questions, and whether a Java-applet slider bar was used as a response scale[3]. Another manipulation was coded at the questionnaire level, which altered respondents' perception of the length of the questionnaire by displaying fast progress at the beginning, slow progress at the beginning, actual progress, or no progress at all.

Auxiliary respondent data was available only for the SSI respondents as they had provided key demographic information when they became members of the panel. These variables included age, gender, race, marital status, and occupation.

Paradata collected about the respondents' computers through a javascript and collected after they passed the survey introduction screen included the type of device, operating system, browser type, browser dimensions, and screen dimensions.

In order to use time spent on the first question as a measure of respondent behavior, time was standardized within each survey and experimental condition, as although the first questions in the two studies were almost identical, respondents in Study 2 were randomly assigned to answer a different part of the survey first, changing the first question seen by the respondent.

While the variation of question types and varying question order for respondents allows the page characteristics to be separated from location in the survey, it requires a more complex model. A discrete hazard survival model with page-varying covariates is employed, where the page-characteristics are allowed to vary for across pages and respondents, and questionnaire-level experiments, respondent characteristics, respondent paradata, and respondent behavior are added sequentially to the model as page-invariant covariates. In the model below,

$$\ln\left(\frac{P_{iq}}{1-P_{iq}}\right) = \alpha_q + \beta_1 X_{i1} + \beta_2 X_{i2}(q)$$

P_{iq} is the probability for respondent "i" to breakoff on page "q",
α_q is the baseline hazard of breaking off on page "q",
X_{i1} is a vector of page-invariant covariates for respondent "i", and
$X_{i2}(q)$ is a vector of page-varying covariates for respondent "i" for page "q".

The covariates are sequentially entered in sets of represented factors. The first four models use only the SSI data as it has respondent socio-demographic characteristics on all respondents. The following three models will not include these characteristics but will use both SSI and AOL samples.

4 Results

Figure 2 shows the pattern of breakoffs throughout the first study, labeling the characteristics of the pages with higher rates of breakoff. For example, some of the higher breakoff rates are on pages with grids of multiple questions, section introduction pages, a question on amount spent on alcohol, a page with multiple open-ended questions, and pages with Java slider bar scales. At the questionnaire-level, the respondents who were assigned to no progress indicator or presented slow progress at the beginning seem to breakoff at higher rates throughout the entire survey. However, this figure provides an incomplete picture and confounds many factors, displaying only one selected page characteristic at a time.

Combining the two studies, additional variation in question/page characteristics is achieved, with a total sample size of 6,026 respondents who started a survey, and a maximum of 87 pages.

A model predicting breakoff was first estimated with page and questionnaire characteristics, presented in Table 1. The relative risk of breakoff increases by 22% for each additional question in a page, but neither the number of questions in the previous page, nor the cumulative number of questions asked to that point have an effect on breakoff. Open questions and introduction pages increase the risk by almost two and a half times, while long questions (typically consisting of more than one sentence in the question stem) triple the risk of breakoff. There is some indication that pages with questions that require retrieval tasks induce breakoffs, but the coefficient is not significant at the .05 level. Respondents who were shown slow progress early in the questionnaire had more than twice the risk of breakoff relative to those who were not shown a progress indicator at all.

Among the respondent characteristics available for the SSI panel samples, white, older, and respondents with at least some graduate education were less likely to break off, Table2, Model 2.

In terms of paradata added in Model 3, predictors were rather weak with only an indication that those not using a Microsoft Windows operating computer are at a higher risk of breakoff.

Respondent behavior in terms of time spent on the first question was significantly associated with risk of breakoff, with

[2] While it is preferable to discuss question characteristics as a lower-level unit of analysis rather than page characteristics, when a respondent breaks off only the page is captured by the survey system. However, the two surveys typically placed one question per page, and when multiple questions were displayed, they tended to be very similar and the characteristics of all the questions were coded.

[3] High breakoff rates were observed for respondents who were presented these questions. It is unknown whether it displayed properly for all respondents causing the breakoffs, but analyses need to account for it to avoid bias in other parameter estimates.

each standard deviation slower responding resulting in 34% higher risk of breakoff, shown in the last row for Model 4.

In order to use the respondent demographic information, the AOL sample was ignored in Models 1-4. Table 3 shows models based on the full set of data from both SSI and AOL samples. A key difference is that the breakoff risk on a page with judgment questions is almost twice as high relative to non-judgment questions, and the effect remains significant after controlling for respondent environment in Model 6 and respondent behavior in Model 7.

5 Discussion and Conclusions

The framework allows the study of different response behaviors as possibly interrelated through causes. It helps in the classification of factors, their interactions, and how they can affect multiple outcomes. The present study focused on breakoff behavior, and found it to be affected by questions requiring more comprehension, judgment, and response formatting. In addition, the amount of the demands in a page, and the switching of topic also increased the likelihood of breakoff.

The number of questions increased the risk of breakoff and the effect remained stable in all models. However, the relative risk of breakoff increased by only 22% for each additional question. This finding needs to be taken with caution, given the limited number of pages that had multiple questions in these studies. The effect is immediate, as the number of questions in the previous page did not have an effect on breakoffs. The overall accumulated burden measured by the cumulative number of questions asked, also had no effect on breakoff, when controlling for page characteristics.

Long questions require more effort on behalf of the respondent to comprehend the question and were associated with an increased risk of breakoff. While the availability of definitions for some questions did not affect breakoff, it is likely that respondents make the distinction that unlike the question itself, the definitions are not essential to the task of answering a question. Open-ended questions require both formulation of an answer and typing it in and also led to higher risk of breakoff. This effect was smaller when the open-ended response was numeric.

The hypothesis that breakoffs are not just a phenomenon resulting from respondents reaching some pre-existing threshold for burden, but rather the continuous reevaluation of the decision to participate was further supported by finding that the risk of breakoff was 2-4 times greater on section introductory screens. These pages did not have any questions and the text was typically limited to a single sentence. It seems that these are points at which respondents reevaluate whether they want to begin the next section of questions.

While none of the models found the sensitive topic of questions to be associated with breakoff, it is also possible that more sensitive questions could produce such effects, or that this effect would be stronger for non-volunteering respondents.

The first set of models that were fit to the data from the SSI panel respondents showed some indication for retrieval questions to increase breakoff risk, compared to other mostly factual questions. This was not even marginally significant after controlling for respondent behavior and also in the analyses of the combined SSI and AOL data. However, judgment questions were associated with almost twice the risk of breakoff in the combined data. A key difference between the two samples is that the AOL respondents were not members of a volunteer panel and had less prior survey experience. Furthermore, after controlling for respondent behavior, i.e., respondents spending more time on the first question were also more likely to break off, the effect of the judgment processes was reduced (Model 7). This interaction is key for future research, as it means that the respondents who may be putting in effort in the survey and/or finding it more difficult than other respondents, are also more likely to break off on questions relying on judgment.

Making respondents believe that the survey is longer by presenting a progress indicator that changes very little at the beginning more than doubles the risk of breakoff in all models, relative to not showing any progress indicator. This is also an example of how overall questionnaire design features could affect breakoffs in a survey, and the need for understanding the consequences of such design features.

Some respondent characteristics were associated with breakoffs – white, older, and more educated respondents were less likely to abandon the survey, somewhat consistent with findings on unit-nonresponse in other self-administered modes. However, these respondent socio-demographic variables did not affect the relationship between any of the question and questionnaire characteristics, and breakoff.

Some of the paradata collected on the respondents' environment was predictive of the likelihood of breakoff in the combined sample (Models 6 and 7). Respondents using less popular hardware and software were more likely to break off, specifically those using Macintosh and some indication for a smaller difference for those using PCs with older versions of Microsoft Windows, relative to PCs with the currently popular Windows XP. This finding seems atheoretical, but it could be proxy information for aspects that we have greater difficulty measuring, such as income, internet familiarity, connection speed, etc., as these paradata were not significantly predictive of breakoffs when age and education were in the models with the SSI data (Model 3) and even less predictive when time on the first question was added (Model 4).

While some may believe that respondents who breakoff are simply not committed to the survey task, it seems to be quite the contrary – it is those who spend more time on answering questions that may be finding the task difficult and therefore breaking off. In both the SSI sample and the combined sample those who spent a standard deviation longer on the first page with a question were 37% more likely to break off. This finding needs further research, but one likely hypothesis is that those respondents try to answer questions carefully, hence we should try to keep them in the survey whether by assisting them, by avoiding questions that they find difficult, or varying the location of these questions in the survey.

There are three important implications from these findings. Some of these factors identified here as causes and covariates of breakoffs are also associated with other response outcomes, including unit nonresponse and item nonresponse. This means that tradeoffs need to be considered when reducing breakoffs through survey design.

Secondly, there are distinct respondent, question and questionnaire features inducing breakoffs. We should not be talking generally about difficult questions or poor design, but about cognitive demands, comprehension problems, retrieval

difficulties, judgment processes, layouts with multiple questions, as distinct causes that can make different respondents break off at different points in the survey.

Lastly, we need to think of solutions that are informed from these models. If some questions induce higher rates of breakoff by particular respondents, while other question characteristics lead to breakoff by different respondents, then placement of modules of questions within the questionnaire can be varied. This way particular respondents will not be breaking off at the beginning of the survey, that could have had different responses to later questions. This would also allow for better imputation models that will not be consistently missing information on a distinct part of the sample. Another benefit from using this framework is using it to identify other informative variables that are associated with both the dependent variables and breakoff processes. One such example are questions that rely on judgment processes – respondents who break off on such questions can be different in terms of responses regardless of the topic involved, and using a wider framework will lead to the use of other covariates like time spent on responding that will improve adjustment models through imputation and/or weighting.

6 References

Beatty, P. and D. Herrmann (2002). To Answer or Not to Answer: Decision Process Related to Survey Item Nonresponse. Survey Nonresponse. R. M. Groves, D. A. Dillman, J. L. Eltinge and R. J. A. Little. New York, Wiley**:** 71-85.

Blumberg, S. J., L. Olson, et al. (2005). "Design and operation of the National Survey of Children's Health, 2003." National Center for Health Statistics, Vital Health Statistics **43**(1).

Bosnjak, M. and T. L. Tuten (2003). "Prepaid and promised incentives in web surveys - An experiment." Social Science Computer Review **21**(2): 208-217.

Cannell, C., P. Miller, et al. (1981). "Research on Interviewing Techniques." Sociological Methodology: 389-437.

Cannell, C. F., K. H. Marquis, et al. (1977). A Summary of Studies of Interviewing Methodology, Vital Health Stat. No. 69.**:** 77-1343.

Catania, J. A., D. Binson, et al. (1996). "Effects of Interviewer Gender, Interviewer Choice, and Item Wording on Responses to Questions Concerning Sexual Behavior." Public Opinion Quarterly **60**(3): 345-375.

Champion, D. J. and A. M. Sear (1969). "Questionnaire response rate: A methodological analysis." Social Forces **47**: 335-339.

Crawford, S. D., M. P. Couper, et al. (2001). "Web surveys - Perceptions of burden." Social Science Computer Review **19**(2): 146-162.

Deutskens, E., K. de Ruyter, et al. (2004). "Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study." Marketing Letters **15**(1): 21.

Goyder, J. (1987). The Silent Minority: Nonrespondents on Sample Surveys, Polity Press/Westview Press.

Groves, R. and R. L. Kahn (1979). Surveys by Telephone. New York, Academic Press.

Groves, R. M. and M. P. Couper (1998). Nonresponse in Household Inteview Surveys. New York, John Wiley and Sons.

Groves, R. M., S. Presser, et al. (2004). "The role of topic interest in survey participation decisions." Public Opinion Quarterly **68**(1): 2-31.

Hogg, A. and J. Miller (2003). Watch out for Droupouts.

Kaplowitz, M. D., T. D. Hadlock, et al. (2004). "A comparison of Web and mail survey response rates." Public Opinion Quarterly **68**(1): 94-101.

Lozar-Manfreda, K. and V. Vehovar (2002). Survey Design Features Influencing Response Rates in Web Surveys. ICIS.

MacElroy, B. (2000). Variables Influencing Dropout Rates in Web-based Surveys.

Mooney, G. M., B. Rogers, et al. (2003). Examining the Effect of Error Prompting on Item Nonresponse and Survey Nonresponse in Web Surveys. American Association for Public Opinion Research, 58th Annual Conference, Nashville, Tennessee.

Musch, J. and U.-D. Reips (2000). A brief history of web experimenting. Psychological experiments on the Internet. M. H. Birnbaum. San Diego, CA, Academic Press**:** 61-85.

O'Neil, K. M., S. D. Penrod, et al. (2003). "Web-based Research: Methodological Variables' Effects on Dropout and Sample Characteristics., 35. 217-236." Behavior Research Methods, Instruments, and Computers **35**(2): 217-236.

Olson, K., M. C. Murray, et al. (2004). Breaking Off in the Middle: Partial Interviews in a Large Telephone Survey. Paper presented at the Midwest Association for Public Opinion Research Annual Conference, Chicago, IL.

Osborn, L., S. J. Blumberg, et al. (2000). Assessing a New Data Collection Methodology for Obtaining Sensitive Data Using an RDD Telephone Survey. Paper Presented at the 55th Annual Conference of the American Association for Public Opinion Research & World Association for Public Opinion Research, Portland, Oregon.

Schwarz, S. and U.-D. Reips (2001). CGI versus JavaScript: A web experiment on the reversed hindsight bias. Dimensions of Internet science. U.-D. Reips and M. Bosnjak. Lengerich, Germany, Pabst Science**:** 75-90.

Strack, F. and L. L. Martin (1987). Thinking, Judging and Communicating: A Process Account of Context Effects in Attitude Surveys. Social Information Processing and Survey Methodology. H.-J. Hippler, N. Schwarz and S. Sudman. New York, Springer-Verlag.

Stussman, B. J., B. Taylor, et al. (2003). Partials and Break-offs in the National Health Interview Survey, 2002. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Arlington, VA.

Tourangeau, R. (1984). Cognitive Science and Cognitive Methods. Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. T. Jabine, M. L. Straff, J. M. Tanur and R. Tourangeau. Washington, DC, National Academy Press**:** 73-100.

Tourangeau, R. (1987). Attitude measurement: A cognitive perspective. Social Information Processing and Survey

Methodology. H.-J. Hippler, N. Schwarz and S. Sudman. New York, Springer-Verlag: 149–162.

Trouteaud, A. R. (2004). "How you ask counts: A test of Internet-related components of response rates to a web-based survey." Social Science Computer Review **22**(3): 385-392.

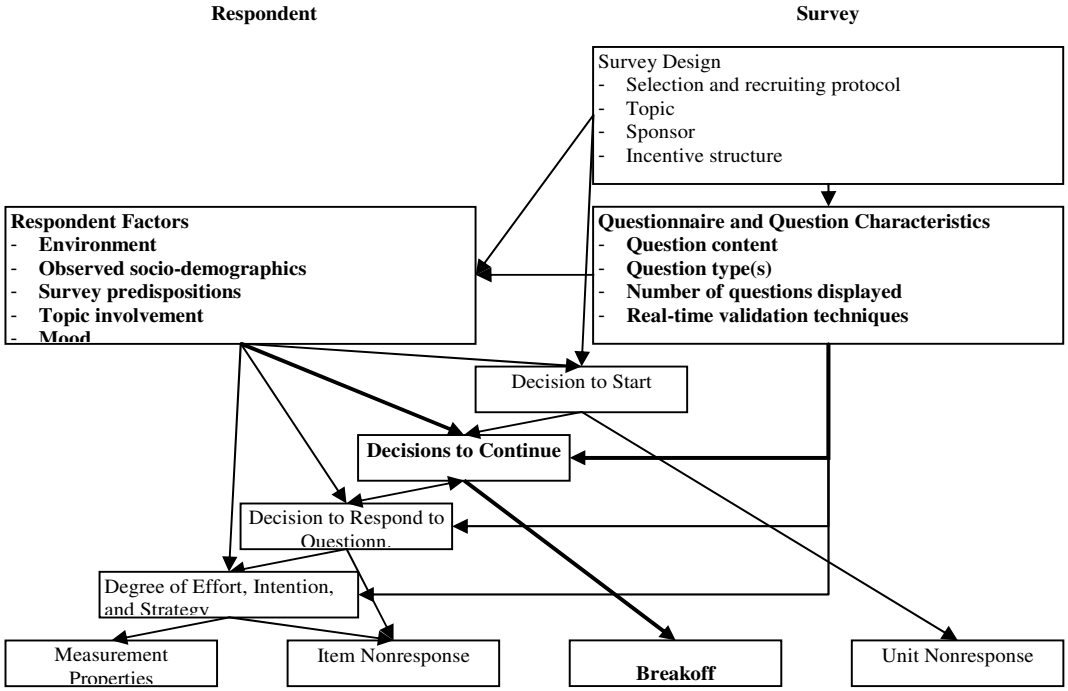

Figure 1: Framework for Participation Decisions in Web Surveys.

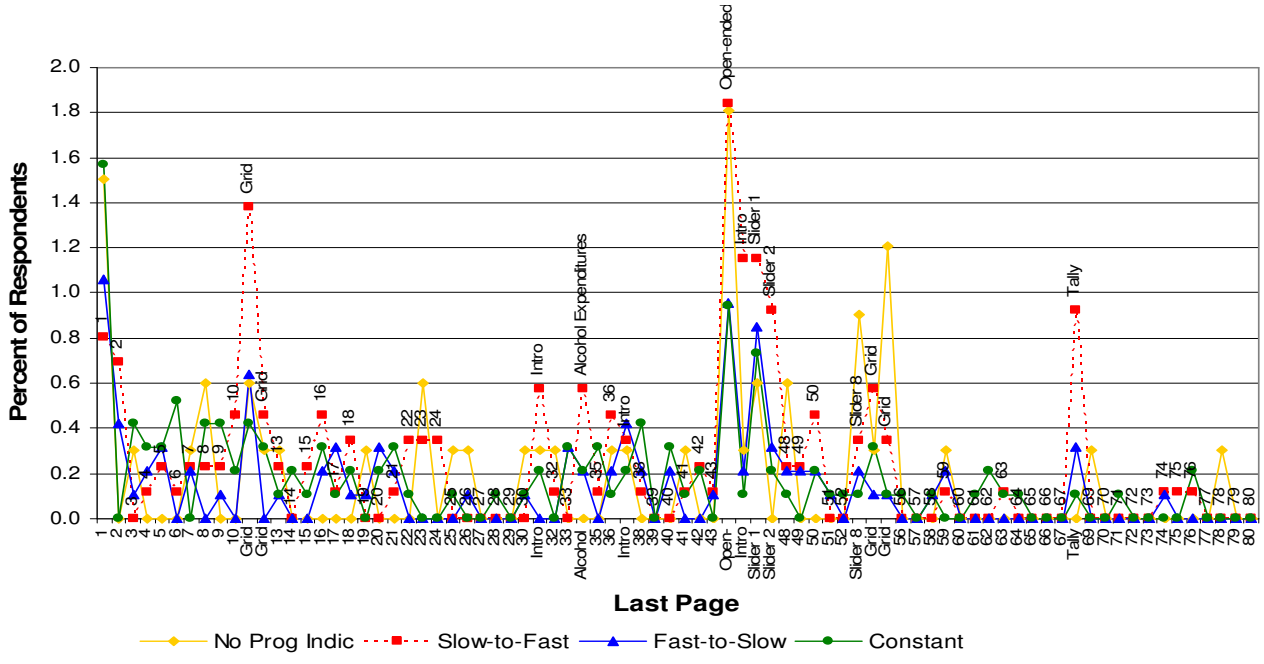

Figure 2: Breakoff by Page Characteristics and Experimental Condition in Study 1.

Table 1: Response Dispositions for Both Studies by Sample Source.

		Breakoffs		Completes		Total	
Study 1	SSI	268	15.8%	1427	84.2%	1695	100.0%
	AOL	210	14.0%	1290	86.0%	1500	100.0%
Study 2	SSI	137	9.1%	1361	90.9%	1498	100.0%
	AOL	107	8.0%	1226	92.0%	1333	100.0%
Total		722	12.0%	5304	88.0%	6026	100.0%

Table 2: Discrete Hazard Survival Models with Page-Varying Covariates Predicting Breakoff in the Combined Studies, SSI samples.

		Model 1			Model 2			Model 3			Model 4		
Factor	Predictor	Relative Risk	p-value	p-value	Relative Risk	p-value	p-value	Relative Risk	p-value	p-value	Relative Risk	p-value	p-value
Page and Question Characteristics	Number of questions	1.223	0.001		1.222	0.001		1.223	0.001		1.230	0.001	
	Lagged # of qns	0.995	0.931		0.999	0.982		0.981	0.759		1.004	0.956	
	Cumulative # of qns	1.008	0.441		1.008	0.413		1.008	0.408		1.009	0.376	
	Numeric question	1.175	0.439		1.208	0.368		1.250	0.288		1.134	0.566	
	Open question	2.398	0.024		2.420	0.023		2.280	0.035		2.270	0.037	
	Long question	3.035	0.024		2.995	0.026		3.080	0.022		3.228	0.019	
	Introduction	2.498	0.009		2.530	0.009		2.483	0.011		2.043	0.054	
	Sensitive question	0.761	0.440		0.743	0.404		0.729	0.378		0.687	0.303	
	Retrieval question	1.528	0.075		1.533	0.075		1.548	0.070		1.496	0.100	
	Judgment question	1.479	0.165		1.503	0.151		1.501	0.155		1.362	0.287	
	Definitions	1.070	0.832		1.066	0.842		1.156	0.656		1.115	0.740	
	Slider-bar question	5.170	<.001	<.001	5.189	<.001	<.001	4.997	<.001	<.001	5.003	<.001	<.001
Questionnaire Characteristics	Slow to fast progress	2.265	<.001		2.202	<.001		2.134	<.001		2.219	<.001	
	Fast to slow progress	0.816	0.277		0.786	0.202		0.770	0.168		0.789	0.218	
	Constant progress	1.359	0.058	<.001	1.313	0.095	<.001	1.280	0.133	<.001	1.302	0.114	<.001
Respondent Characteristics	White				0.752	0.041	0.041	0.743	0.034	0.034	0.790	0.096	0.096
	Some college				1.131	0.355		1.155	0.280		1.235	0.121	
	college				0.832	0.253		0.837	0.272		0.916	0.599	
	Grad school or graduate				0.532	0.002	<.001	0.515	0.001	<.001	0.597	0.014	<.001
	Age				0.991	0.008	0.008	0.989	0.002	0.002	0.986	<.001	<.001
Respondent Environment	Windows 95, 98, NT							1.225	0.083		1.128	0.311	
	Windows 2000							1.007	0.971		0.999	0.997	
	Macintosh							2.051	0.011		1.853	0.027	
	WebTV							1.373	0.389	0.079	0.933	0.859	0.252
	Netscape							0.763	0.317	0.317	0.808	0.426	0.426
Resp. Behavior	Time on first question										1.339	<.001	<.001

Number at risk/number of breakoffs in each model: 3193/392; 3192/390; 3190/388; 3164/377.

Table 3: Discrete Hazard Survival Models with Page-Varying Covariates Predicting Breakoff in the Combined Studies, SSI and AOL samples.

		Model 5			Model 6			Model 7		
Factor	Predictor	Relative Risk	p-value	p-value	Relative Risk	p-value	p-value	Relative Risk	p-value	p-value
Questionnaire Page and Question Characteristics	Number of questions	1.222	<.0001		1.223	<.0001		1.219	<.0001	
	Lagged # of qns	0.982	0.713		0.971	0.563		1.004	0.934	
	Cumulative # of qns	1.010	0.180		1.010	0.167		1.011	0.141	
	Numeric question	1.449	0.016		1.478	0.011		1.318	0.086	
	Open question	2.492	0.003		2.417	0.004		2.344	0.006	
	Long question	2.144	0.059		2.181	0.053		2.381	0.031	
	Introduction	3.809	<.0001		3.788	<.0001		2.517	0.001	
	Sensitive question	1.142	0.591		1.134	0.611		0.977	0.926	
	Retrieval question	1.364	0.096		1.367	0.095		1.355	0.107	
	Judgment question	1.879	0.003		1.879	0.003		1.586	0.035	
	Definitions	1.220	0.421		1.280	0.323		1.205	0.460	
	Slider-bar question	4.658	<.001	<.001	4.568	<.001	<.001	4.824	<.001	<.001
Survey Design	Slow to fast progress	2.094	<.001		2.067	<.001		2.292	<.001	
	Fast to slow progress	1.069	0.617		1.064	0.644		1.189	0.204	
	Constant progress	1.400	0.007	<.001	1.389	0.009	<.001	1.516	0.001	<.001
Respondent Environment	Windows 95, 98, NT				1.169	0.067		1.065	0.472	
	Windows 2000				1.066	0.675		1.111	0.511	
	Macintosh				1.993	0.002		1.630	0.028	
	WebTV				1.464	0.260	0.016	0.937	0.857	0.273
	Netscape				0.685	0.119	0.119	0.791	0.335	0.335
Resp. Behavior	Time on first question							1.373	<.001	<.001

Number at risk/number of breakoffs in each model: 6026/673; 6020/671; 5939/639.