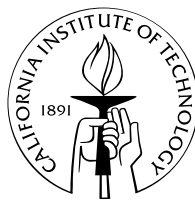# The Shirking Model —
A Theory of How People Answer Survey Questions

Thesis by

Carla VanBeselaere

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2004

(Submitted May 28, 2004)

ii

Acknowledgements

The members of my dissertation committee, Mike Alvarez, Robert Sherman, Ed McCaffery, Rod Kiewiet, and David Grether, have greatly shaped the course my research has taken, and I thank them for the time and energy that they have dedicated to my academic and professional development. Mike Alvarez taught me most of what I know about political science and survey research; the ideas in the following chapters were a product of our extensive discussions. Bob Sherman provided me with new insights into empirical methods that were instrumental to the analysis in the chapters that follow. Ed McCaffery and Rod Kiewiet were also helpful in providing me with conceptual ideas and practical perspectives on my work. The financial support from the Division of Humanities and Social Sciences at Caltech, the Center for the Study of Law and Politics, and the Fulbright Program are gratefully acknowledged. Parts of this work have been presented in a number of seminars throughout the last three years. I am grateful to seminar participants at the American Political Science Association (APSA) Annual Meetings, Harvard, Mount Allison, and Caltech for their input.

I would also like to thank everyone who helped me through my five years of graduate study. My sister (Paula), my friends (Steffanie, Rachel, Levon, Irit, Alice, Divi, Jeannie, and Eden), my classmates (Maggie, Dan, and Ben), other HSS students, and Caltech staff members (especially Karen and Chris) gave me perspective on my work and helped me out in what ever way they could. I would also like to give special thanks to my parents, Duane and Dagmar VanBeselaere, who provided both moral support and helpful advice. But most of all I want to thank my husband, Jonathan Parsons, for his continued patience and support during my struggles. I could not have achieved this work without the love and support of my family and friends.

Abstract

This thesis explores survey responses in an attempt to uncover behavior that might indicate that respondents are not exerting sufficient effort to ensure reliable responses. By developing a rational choice theory about survey response behavior, new tests for examining how respondents behave when answering opinion questions are established. The first and most important problem addressed in this work is the identification of respondents who shirk by providing responses that are not fully thought-out. Assuming that shirking behavior is indicated by "no opinion" responses, this work explores the frequency and sources of this behavior. The results from this analysis provide useful recommendations for survey design. However, to further explore the implications of shirking behavior a new measure of shirking is proposed. This measure, based on survey response times, enables an examination of the implications of shirking behavior and proposes possible ways by which shirking behavior might be accounted for in the analysis of survey data. The results from this work suggest that if analysts want to ensure that accurate results drawn from survey data, consideration should be given to the likelihood that respondents shirk in answering survey questions.

Contents

List of Figures

List of Tables

Chapter 1   Introduction to Shirking in Survey Answers

> Just as an experienced cook can tell how a large pot of soup will taste
> by trying only a spoonful, it is possible to learn the opinions of a very
> large group of people by talking to a small number of them (Groves 1989,
> p. 239).

Surveys are based on the concept of obtaining information by asking questions (Tourangeau, Rips & Rasinski 2000).[1] But the data collected from questionnaires is "only as meaningful as the answers the survey respondents provide" (Schwarz & Sudman 1996, pg. 1). In other words, respondent behavior will determine the reliability or accuracy of survey responses.

The potential that surveys might inaccurately measure beliefs or behaviors, introduces doubt into the validity of survey data and the analytical results from this data. To address this concern, this thesis considers how respondents deal with answering survey questions. A formal rational choice model of survey behavior that accounts for the cognitive costs and benefits of survey participation is proposed. Borrowing from the labor economics literature, the behavior posited by this theory is called shirking. By modeling respondents as rational agents who optimally select how much effort to exert in answering survey questions, new insights into survey response behaviors are obtained.

Four issues are of particular interest in the study of shirking behavior: (i) identifying of behaviors that indicate when respondents are not providing carefully thought-out responses, (ii) examining the factors that affect respondent behavior, (iii) determining whether shirking behavior has a significant effect on survey data, and (iv) determining what can be done about shirking behavior among survey respondents.[2]

---

[1] For simplicity, the male pronoun will be used to refer to the respondent and the female pronoun to refer to the researcher or interviewer.

[2] The first two of these issues are discussed extensively in this work although conclusive results have not yet been obtained on the identification of shirking due to data limitations. The latter two

For surveys to achieve the goal of understanding the world a little better, a necessary condition is that they collect accurate or reliable information (i.e., information that truthfully represents what respondents know or believe), it is important that appropriately selected respondents provide carefully thought-out responses to survey questions. For factual or behavioral questions, it may be possible to validate survey responses using auxiliary data but this is generally not an option in opinion studies. As discussed below, identifying whether opinion expressions are reliable is complicated by the possibility that opinions change. To address the identification problem, this work draws on a variety of data sources to identify observable behaviors that might be indicative of shirking.[3]

Assuming that shirking corresponds to observable behaviors, such as answering questions quickly or providing a "no opinion" response, attention can be turned to examining the factors which are likely to affect this behavior. Because the shirking model has clear empirical implications, it is straightforward to test which factors impact respondent behavior. Results from this analysis provide recommendations about how to maximize the reliability of survey responses.

Given that shirking exists, the final question is what can be done about it? To answer this question, measures of shirking are incorporated into several substantive models of interest to test whether respondents who are identified as shirkers behave differently than their non-shirking counterparts. Based on evidence from these tests, empirical techniques are proposed to ensure that survey data is analyzed in an appropriate fashion.

Survey researchers should be especially concerned about the possibility that respondents shirk when answering survey questions; meaningful conclusions can only be drawn from survey data if appropriately selected respondents provide carefully

issues are discussed briefly because the significance and correction of shirking cannot be resolved without reliable measures of shirking. A discussion about the significance of shirking will depend on the particular survey question under consideration and significance is defined.

[3]Although interview data is explored in an attempt to uncover behaviors that might be indicative of shirking, no formal test of the identification problem are feasible because there is insufficient information. Perhaps by asking respondents to indicate how much effort they exerted in answering survey questions.

thought-out answers. The issue of sample selection has been extensively discussed by statisticians and survey methodologists (e.g., Cochran 1953, Brehm 1993) but the question of obtaining reliable survey responses has primarily been viewed as "an art" (Sudman, Bradburn & Schwarz 1995, Tourangeau, Rips & Rasinski 2000). Thus, very few systematic and comprehensive studies have been done in the area of respondent behavior. While a substantial literature exists in survey methodology concerning respondent behaviors, most of this work involves the piecemeal examination of respondents behaviors in specific settings. A unifying formal theory with clear empirical implications is still missing.

By modeling survey response behavior as a function of effort, this thesis attempts to provide a concrete environment within which to consider survey response behavior. Assuming that respondents are able to assess costs and benefits from exerting effort to answer survey questions, the shirking model proposes that rational respondents will engage in answering survey questions only if the rewards outweigh the demands. Thus, if the benefits from answering survey questions are not sufficient, some respondents may choose to shirk by shortcutting their cognitive processes and providing responses that are not fully thought-out.

This work draws upon ideas theories about comprehension, memory, judgment, cognitive psychology, and rational choice behavior in order to model the cognitive process of survey response.[4] But

about what is happening in respondents minds when they answer survey questions from in-depth interviews with survey respondents and from theories of comprehension, memory, and judgment. Given evidence that respondents recognize the costs and benefits of answering survey questions, the effort exerted in answering these questions will be affected by factors that determine respondents' net benefits. These factors, classified as task difficulty, respondent ability, and respondent motivation,

---

[4]Surveys of this survey methodology literature can be found in Jabine, Straf, Tanur & Tourangeau (1984), Jobe & Loftus (1991), Hippler, Schwarz & Sudman (1987), Sudman, Bradburn & Schwarz (1995), Schwarz & Sudman (1996). Sudman, Bradburn & Schwarz (1995), Zaller (1992), Krosnick (1991), Tourangeau, Rips & Rasinski (2000), and Alvarez & Brehm (2002) summarize some of the fundamental ideas about the cognitive and communicative processes involved in survey participation.

will determine how diligently a respondent works at answering a question thereby impacting the reliability of survey responses. Thus, with appropriate measures of ability, motivation, and task difficulty, it is possible to evaluate the determinants of shirking behavior.

The first topic confronted by this analysis is the identification of shirking. To obtain insight into this issue, survey participants were interviewed to get an idea of the ways in which people deal with answering complex survey questions.[5]

Since shirking cannot be observed directly, this thesis considers several different means of identifying shirking. The first approach uses accuracy and response times for factual questions to identify respondents who might be shirking. Next, "no opinion" responses (indicated by the selection of the "don't know" or "no opinion" response options) are explored as indicators of shirking (Krosnick 1991).[6] Because respondents may choose a "no opinion" response despite exerting significant effort, response time data is also explored as an alternative means of identifying shirking.

If shirking is demonstrated to be significant, its existence will yield important implications for both survey design and data analysis. First, surveys should be designed to collect data that enables researchers to identify when respondents are shirking. In terms of data analysis, if shirking is perceived to be a significant problem, researchers will need to account for this in any substantive analysis or face the possibility that conclusions drawn from the data are invalid. Once responses are associated with shirking, they should be treated differently from those coming from respondents who are exerting enough effort to provide reliable responses.[7]

The first four chapters of this thesis explore the different components of the survey response process. From interviews with survey respondents, discussed in Chapter 2, and theories of comprehension, memory, and judgment, reviewed in Chapter 3, the

---

[5]The analysis of these interviews is contained in Chapter 2.

[6]Alternatively, "no opinion" responding might be viewed as a noisy signal of shirking since some respondents might not be able to establish their opinions no matter how much effort they exert. To proceed with this type of analysis some means of distinguishing which respondents really do not have an opinion has to be developed. Current data sources do not contain sufficient information to establish this distinction thus this type of analysis is beyond the scope of this thesis.

[7]In Chapter 7, different ways of accounting for shirking behavior are explored.

foundations for the rational choice model of shirking are established. Observations of response unreliability and instability are reviewed in Chapter 4, together with some existing hypotheses explaining this behavior. Finally, the shirking model of survey response behavior is developed in Chapter 5.

With the rational shirking framework in mind, the subsequent chapters in this work explore the use of different indicators of shirking behavior, test the determinants of shirking behavior, and determine the effect of this behavior on substantive survey results. Beginning with the idea that shirking behavior is indicated by respondents who quickly provide incorrect response to factual questions, Chapter 6 examines the frequency and implications of this behavior. This chapter also implements an analysis of shirking as indicated by the selection of "no opinion" responses in the 1980 and 1984 National Elections Studies (NES) Pre-Election Surveys. Given that objective measures of question difficulty and measures of respondent ability and motivation cannot fully explain "no opinion" response behavior, the analysis of "no opinion" responding is extended to incorporate measures of question difficulty. Collecting auxiliary difficulty data and using data available from the 1998 Multi-Investigator Study, the effect of question difficulty on shirking behavior is explored. Chapter 7 extends the analysis of shirking, by implementing new measures of shirking which rely on response times. With the measures of shirking identified from response times, it is possible to examine the implications of shirking for the analysis of substantive issues. Together with a summary of these new insights into respondent behavior, the concluding chapter suggests what type of data would be useful for the further exploration of shirking behavior in survey answers.

Chapter 2 Talking about Answering Surveys

> To speak with precision of public opinion is a task not unlike coming to grips with the Holy Ghost (Key, V.O. Jr. 1966, p. 8).

2.1 Setting up the Conversation

Do people have well-established predispositions (Alvarez & Brehm 2000) on which they base their responses to opinion questions or are survey responses derived from considerations that are at "the top of their minds" (Zaller 1992)? Are survey responses reliable, in the sense that they accurately represent what the respondent knows, thinks, or believes? Do respondents behave in such a way as to ensure that their answers do not change as a result of minor alterations to how questions are asked?

To try and address these questions and to better understand the factors that determine survey respondent behavior, this chapter examines data collected through in-depth interviews with subjects who completed the Caltech Internet Survey Project (CISP) 2003 Attitudes Survey.[1] These interviews provide a unique opportunity to examine how much effort respondents exert in answering survey questions. From this data, it is possible to identify different response behaviors that might be indicative of survey answers that are not providing carefully thought-out or reliable.

2.2 The Pros and Cons of Interview Data

Although cognitive psychology has made inroads into understanding the processes that respondents might use to understand and answer questions, data about what re-

---

[1]For simplicity respondents who participated in interviews will be referred to using the male pronoun despite the fact that both men and women participated in the in-depth surveys. More details about the Caltech Internet Survey Project (CISP) can be found in Appendix A.1 and Alvarez, Sherman & VanBeselaere (2003).

spondents are really thinking about when they answer opinion questions is limited.[2] To overcome this deficiency, eleven Caltech undergraduate students who participated in the 2003 Attitudes Survey were selected to take part in one-on-one face-to-face interviews. The data from these interviews provides new insights into the process of answering questions about complex social policy issues. These in-depth interviews offer a means of collecting data that cannot be easily uncovered through traditional survey formats. The interviews, undertaken as part of this study, provided respondents the opportunity to explain what they were thinking about when they answered the survey and to describe how they established their opinions on the issues addressed by the survey.

The interview setting allowed respondents to expand on their survey responses and to describe their approach to answering survey questions. Since these one-on-one interviews were undertaken face-to-face, both verbal and non-verbal communications were collected.[3] By facilitating an open discussion about the survey taking experience, these interviews were able to uncover a new wealth of information concerning how respondents form opinions and how they deal with answering survey questions.

Unfortunately in-depth interviews are not generally feasible. Interviewing all survey participants is likely to be prohibitively expensive when the subject pool is large. In addition, because the analysis of the interview results depends on the interpretation of both verbal and non-verbal cues, it was important to provide a comparable interview environment for all participants; maintaining this type of consistency is particularly difficult if subjects are geographically dispersed and multiple interviewers are involved. As a result of these difficulties, in-depth interview data tends to be

---

[2]Survey questions designed to collect information about predispositions or core values have been included in surveys, such as the General Social Survey (GSS), but this data does not reveal when these predispositions influence survey responses. Another approach to collecting information on how respondents form options has been through the use of open-ended questions; the NES has used this approach by asking respondents to identify positive and negative aspects of candidates that might influence candidate preference. Unfortunately, the structure of these questions encourages respondents to justify their responses rather that explain how they would actually make a preference judgment.

[3]Non-verbal cues were occasionally useful for determining the level of subject cooperation and participation.

limited in sample size and scope. In the use of interviews, the trade-off is between the breadth of information obtainable and the sample size.

In the subsequent analysis, interview data was obtained from eleven subjects who participated in the 2003 Attitudes Survey. Since interview participants were Caltech students, it was possible to maintain a similar interview experience for each participant—all participants were interviewed by the same interviewer in the same setting.[4] The small size also allowed the interviewer to collect as much information as possible from each respondent. The trade-off is that the sample is not necessarily representative of survey respondents in general. Of course the data does provide some insights into how different respondents might deal with answering survey questions.

2.3 The 2003 Attitudes Survey

The 2003 Attitudes Survey, conducted in January 2003, was designed to capture general attitudes towards social policy issues and the American government. The primary focus of this survey was determining attitudes towards government policies concerning medical marijuana and euthanasia. Additional information about respondents' beliefs was collected in an attempt to establish the sources of the opinion expressed. The questions asked in this survey were chosen to present topics that were both controversial, complex, and currently debated by political elites. To capture respondents' opinions about these issues the following questions were asked:

- Do you think the possession of small amounts of marijuana should or should not be treated as a criminal offense?

- Regardless of what you think about the personal non-medical use of marijuana, do you think doctors should or should not be allowed to prescribe marijuana for:

  - patients with terminal cancer?
  - patients in the later stages of AIDS/HIV?

---

[4]The interviewer's familiarity with Caltech also facilitated these interviews and ensured the interview environment accommodated subjects' needs.

- – patients with long-term chronic pain?
- – patients with chronic migraine headaches?

- When a person has a disease that cannot be cured, do you think doctors should be allowed by law to end a patient's life by some painless means if the patient and his family request it?

- Regardless of what you think about doctor assisted suicide, do you think a person has a right to end his or her own life if this person:

  - – has an incurable disease?
  - – has gone bankrupt?
  - – is in chronic pain?
  - – has dishonored his or her family?
  - – is tired of living and ready to die?

In order to uncover the source of respondents' opinions, a series of questions concerning core values or predispositions relevant to medical marijuana and euthanasia were also included in this survey. These questions were designed to help establish respondents' perceptions about society, laws, and God—values deemed to be relevant to the discussion of marijuana policy and euthanasia/suicide.[5] Responses from these questions were used in the construction of social-trust, libertarianism, and fear-of-God value scales. To encourage respondents to reflect on these issues when answering questions about marijuana and euthanasia policies, the following questions about society, laws, and God were asked at the beginning of the survey:

- Generally speaking, do you think that most people

  - – Can be trusted or you can't be too careful?
  - – Try to take advantage or try to be fair?
  - – Try to be helpful or are just looking out for themselves?
  - – Work hard to get ahead or just get lucky?

- There are different ways of perceiving the role of laws in society. We would like to know what principles you think should guide the application of laws.

---

[5]For a discussion about the values debated in reference to these issues, see, for example, Alvarez & Brehm (2002) and Stein (Nov. 4, 2002).

- – Security or liberty?

- – Prevention or punishment?

- – Adherence to the law or flexibility in the law?

- – Strict enforcement or liberal interpretation?

- – Individual freedom or respect for authority?

- There are many different ways of picturing God (or The Divine Power). We would like to know the kinds of images you are most likely to associate with God (or The Divine Power). Below are some contrasting images, we would like to know which most closely describes your image of God (or The Divine Power).

  - – Mother or father?

  - – Lover or judge?

  - – Creator or healer?

  - – King or friend?

Furthermore, to capture measures of how politically informed respondents were, several factual knowledge questions were included. Data from these questions was used to measure levels of respondent informedness about both the specific issues under consideration and politics in general. The following questions were geared towards respondents' levels of political informedness:

- Approximately what proportion of marijuana users have also used "harder" drugs such as cocaine and heroin?

- When did the U.S. outlaw possession of marijuana?

- Please identify the job or political office held by the following:

  - – Tom Ridge

  - – Kofi Annan

  - – John Walters

  - – Wolf Blitzer

This survey was available on the Caltech Internet Survey Project (CISP) website and subjects who registered to become members for the Caltech Internet survey panel

were directed to complete this survey.[6] In addition, subjects were recruited from among Caltech undergraduate students enrolled in political science classes. These students were also encouraged to participate in in-depth interviews. In total 1,176 responses were provided to the 2003 Attitudes survey, including the eleven Caltech students who were involved in the in-depth interviews.[7]

2.4   Interviews about Answering Surveys

Immediately following the completion of the CISP 2002 Attitudes Survey, eleven Caltech students participated in in-depth interviews designed to supplement data available from survey responses.[8] During these interviews, the participant and interviewer reviewed the survey and discussed what the respondent was thinking about when answering each of the above mentioned survey questions. Discussions during these interviews were intended to uncover how respondents perceive survey questions, what they think about when providing answers, and how they feel about the overall survey experience. The goals of the in-depth interviews were to obtain more insight into how people behave when they are asked to provide their opinions on complex issues, ways of identifying respondent behaviors, and factors that influence respondent behavior. Since standard survey questionnaires do not provide respondents sufficient opportunity to explain in detail how they feel about different issues nor do they provide researchers the breadth of information necessary to fully understand how respondents perceive different questions, interviews were determined to be the most appropriate means of uncovering information useful for studying the amount of effort respondents exert when answering survey questions. Results from these surveys provide insights the identification of shirking behavior.

---

[6]See Appendix A.1 and Alvarez, Sherman & VanBeselaere (2003) for a further discussion of the Caltech Internet Survey Project and the recruitment of panel members.

[7]Attention in this work will focus only on the responses from the eleven Caltech students who participated in the in-depth interviews.

[8]These subjects were not randomly selected and thus their data is only intended to be illustrative and not representative. To ensure anonymity of respondents, each was assigned a randomly generated subject number between one and ninety-nine. These subject numbers allowed interview data to be linked with survey responses and will be used in the discussion to identify the respondents.

During the interviews, respondents were asked to describe their survey response process. The interviewers reviewed the questions (as presented in the survey) with the respondents and asked them to explain what they were thinking about when they answered each question. Subjects were encouraged to pursue their stream-of-thought about the survey taking experience during the interview. The following questions were offered as guidelines for structuring the interviews:

- How do you feel about the survey?
- How do you feel about (specific questions)?
- What were you thinking about when you answered (specific questions)?
- Why do you think you were thinking this way?
- Did you think about pros and cons?
- Can you list some pros and cons you were thinking about concerning (marijuana/euthenasia)?

Interview subjects completed the survey in approximately 15 minutes and the interviews lasted between ten and fifteen minutes. To ensure accurate reporting of interview results, the interviews were recorded.[9] Independent reviewers listened to these tapes to assess the nature of subjects' behaviors and to categorize respondent behaviors.

Although the interviews allowed the subjects to respond to questions as they wished, a basic format was followed by the interviewer. The first questions asked for subjects' general reaction to the survey and then the specific survey questions were discussed individually. The interviewer probed subjects a number of times to focus respondents on what they were thinking as they tried to express their opinion—occasionally it was impossible to elicit this information because the subject spent the interview trying to justify his answers or he had forgotten how he arrived at his opinion.

---

[9]Tapes of these interviews will be maintained by the Caltech Internet Survey Project staff for no more than ten years upon which time they will be disposed of appropriately.

The information available from the interviews helps to illuminate different respondent behaviors and identify factors that might influence these behaviors. By reviewing the information from these interviews, it was possible to categorize subjects according to how much effort they appear to have exerted. Based on the evidence from interviews, respondents were assigned to one of three categories: those who exerted sufficient effort to express their opinions reliably, those who did not exert enough effort, and those who fell somewhere in between.[10] The classification of respondents into these categories was determined primarily by the nature of respondents' discussions during the interviews. Respondents who were unable to explain their responses, who constantly indicated that they did not know the basis of their survey responses, and who did not participate fully in the interviews, were determined to not have exerted enough effort. At the other extreme, respondents who expressed a clear understanding of the issues and demonstrated that their responses were based on careful consideration were found to have exerted exerting significant amounts of effort. The remaining subjects, who had some ideas of how they determined their opinions but were unable to clearly explain the source of their survey responses were classified in the middle category. Subject behavior was considered separately for the two sets of questions about marijuana policy and euthanasia/suicide. Table 2.1 indicates how respondent effort levels were distributed across both the marijuana and euthanasia/suicide questions.

Table 2.1: Effort Exerted by Subjects on the Marijuana and Euthanasia Questions

Categories of Effort Level	Marijuana Questions		Euthanasia/Suicide Questions	
	Frequency	Percent	Frequency	Percent
Sufficient Effort	6	54.54	5	45.45
In Between	3	27.27	4	36.36
Not Enough Effort	2	18.18	2	18.18

---

[10]It should be noted, however, that interview subjects may have exerted more effort because they knew they were being observed, i.e., a Hawthorne effect (Mayo 1933).

The data in Table 2.1 suggests that most of the interview participants appear to have exerted a fair amount of effort in answering the survey questions. Only in about eighteen percent of the responses was it obvious that respondents were not exerting much effort.

Different indicators were used to categorize respondents according to the effort they exerted. For example, subject 16 was classified as exerting sufficient effort on both these questions because, despite indicating that he provided responses "off the top of the head," he was able to clearly explain the source of his beliefs on medical marijuana and euthanasia/suicide.[11] The interview discussions with subject 26 also indicated that he had thought about medical marijuana extensively, but, when asked about euthanasia/suicide, subject 26 vacillated on the responses he provided. This subject was therefore classified as having exerted sufficient effort on the medical marijuana questions but placed in the middle category on the euthanasia/suicide questions. At the other extreme was subject 42, who spent most of the interview saying "don't know," avoiding the discussion of how his survey responses were attained, and refusing to look directly at the interviewer. Subject 96, despite mentioning that the survey was not difficult to answer, also said "don't know" frequently during the interview and was easily swayed from his opinion expressions. Thus, these latter two subjects were classified as not having exerted much effort on both sets of questions.

In order to identify different response behaviors that might be indicative of unreliable responses, it is useful to look at the survey responses of interview subjects determined to have not exerted much effort. Subjects 96 and 42, for example, provided mainly "no opinion" responses to the medical marijuana and euthanasia/suicide questions. However, other subjects, such as subject 37, who vacillated during the interview and thus was classified as having exerted a medium amount of effort, selected substantive responses rather than "no opinion" responses. Taken together, the behavior of these three subjects suggests that, in an attempt to minimize the effort

---

[11]To ensure anonymity of respondents, each was assigned a randomly generated subject number between one and ninety-nine. These subject numbers allowed interview data to be linked with survey responses.

exerted, a subject may say that he has "no opinion" or he may select a response option that does not really reflect his opinions. Furthermore, the fact that these three respondents did not take significantly longer than other subjects to answer the survey suggests that these respondents did not take enough time to carefully answer the survey questions.[12] Given that some respondents appear to have answered quickly without careful consideration, short response times can provide an alternative means of identifying respondents who do not exert enough effort to provide reliable answers to survey questions.

In order to determine when respondents are likely to provide unreliable survey responses, it is useful to examine the factors that might influence how much effort respondents choose to exert. Evidence from the interviews suggests that both question and respondent characteristics are likely to affect how much effort a respondent chooses to exert. For example, interview participants who indicated that they were interested in politics or for whom these specific issues were of particular interest were more likely to clearly explain the source of their opinions and did not vacillate in their responses. Also, some respondents were unable to reconcile their opinions about one particular issue but had no difficulty with the others. This suggests that certain questions may be more difficult for some respondents.

These in-depth interviews also provided evidence that respondents recognize the rewards of survey participation. For example, subject 29 indicated that he "appreciated the opportunity to take part in this type of survey." Other respondents, including subject 21, thought that the survey questions were particularly important. In the interviews, these respondents demonstrated real enthusiasm for the survey and the issues under consideration; these respondents were also able to speak extensively about their opinions and the sources of their opinions. This supports the idea that respondents who find surveys rewarding will be more motivated to answer survey questions. On the other hand, respondents indicated that answering some of the questions was demanding; several subjects mentioned that they found many of the

---

[12]Unfortunately, response times were not collected during the 2003 Attitude Survey so it is impossible to establish the relationship between effort and response time using this interview data.

survey questions difficult to answer. For example, subject 72 indicated that the question about euthanasia was difficult because it required him to consider his religious beliefs, how he would feel if a family member's suffering was so acute that the only respite was suicide, as well as recent media reports about cases of euthanasia. Together this evidence suggests that respondents consider both the costs and benefits of survey participation.

Insight from these interviews provides context for the analysis of survey respondent behavior. First, the evidence here suggests that some people may provide unreliable survey responses because they do not exert much effort. Second, respondents may display this behavior by selecting the "no opinion" response option or by providing answers quickly and without careful consideration. Furthermore, the amount of effort respondents exert in answering questions appears to depend on both respondent and question characteristics. This evidence, together with theories of comprehension, memory, and judgment, reviewed in the next chapter, provide a justification for the shirking model developed in Chapter 5.

Chapter 3 Demands and Rewards from Answering Surveys

> I know [patriotism] exists, and I know it has done much in the present contest. But a great and lasting war can never be supported on this principle alone. It must be aided by a prospect of interest, or some reward (Washington April 21, 1778).

Surveys, with "a stranger on your doorstep or in your living room" carefully recording answers to a wide variety of questions, have been described as an "odd social experience" that many people "may find flattering" (Mueller 1973, p. 1).[1] An alternative view is that surveys are just a different form of social exchange (Dillman 2000, Tourangeau, Rips & Rasinski 2000). Both of these views, however, emphasize the importance of the social interaction between a surveyor and a respondent. As with most social exchanges, individuals will be motivated by the rewards of their actions (Blau 1964, Gallegos 1974, Dillman 1978, Goyder 1987). This chapter reviews theories of comprehension, memory, and judgment to establish the potential demands and rewards from answering survey questions.[2]

Answering questions places cognitive demands on respondents. The descriptive model of the survey response process identifies four cognitive steps of answering a survey question: comprehension, retrieval, judgment, and response selection (Tourangeau, Rips & Rasinski 2000). Comprehension encompasses such processes as attending to the question, accompanying instructions, and identifying the information sought (Clark 1985, Graesser, Singer & Trabsso 1994, Lehnert 1978). Retrieval and judgment involve recalling relevant information or ideas from memory and filling

---

[1]The survey taking experience may be more familiar today since the use of survey research is extremely widespread. Modern surveys are also more likely to be administered over the telephone.

[2]A related issue is how the demands and rewards of surveys affect the decision to participate in the survey. Individuals who expect a survey not to be sufficiently rewarding are likely to refuse to participate in the survey. This behavior is described as unit non-response and is distinct from the "no opinion" response behavior, often referred to as item non-response, examined in this work. For a discussion of how the costs and benefits of survey participation affect unit non-response, see Brehm (1993).

in incomplete memories or opinions (Tourangeau, Rips & Rasinski 2000, Tourangeau & Rasinski 1988a, Tourangeau, Rasinski, Bradburn & D'Andrade 1989). Finally, answering requires that a respondent edit his retrieved judgments and, if response options are provided, map these judgments into the provided response categories (Tourangeau & Rasinski 1988b). Effort is required to carefully perform these various stages of answering questions. The effort required to reliably answer questions will depend on how the question is worded, the type of information being requested, how the respondent has stored information in his memory, and his ability to draw information from his memory. Different types of questions will place different demands on respondents. For example, if asked how many times he visited the doctor last year, the respondent must identify the one-year time period under consideration, recall his behavior, and then map these behaviors into responses. However, if asked how he feels about abortion, a respondent has to evaluate his thoughts, beliefs, and feelings about abortion (possibly including the specifics of what abortion entails, his religious or moral values, and his feelings about the rights of the women and the unborn fetus involved) in order to determine how to express an opinion. Given the various components to answering surveys, respondents "may, quite reasonably, view the interview as a task" (Brehm 1993, p. 59).

Despite the demands or costs of survey participation, some people do agree to participate. For these individuals, perceived rewards must outweigh the costs (Brehm 1993). Although the rewards or benefits from participating in surveys are rarely pecuniary in nature, respondents might derive benefits from the opportunity for self-expression, for interpersonal exchange, for intellectual challenge, for self-understanding, or for emotional catharsis (Dillman 2000). People might also participate in surveys because they feel it is their civic duty or because they think their responses might help governments and businesses make better-informed decisions (Riker & Ordeshook 1968, Krosnick 1991). When asked why they participated in the 1989 Detroit Area Studies (DAS), almost half of the respondents talked about the usefulness of surveys or their personal enjoyment from survey participation. The decision to

participate in a survey, however, is only the first step in the survey response process. Conditional on participation, respondents still have the opportunity to determine how much effort they will exert when answering individual questions. However, if the perceived demands and rewards of survey participation are a function of how much effort respondents anticipate exerting in answering the survey questions, these demands and rewards will affect both the survey participation decision and response behavior on each question. Hence both the decision to participate and the choice of how much effort to exert in answering each question will depend on the same factors.

By conceptualizing survey response behavior as a function of the demands and rewards from answering survey questions, it is possible to obtain specific predictions about survey respondent behavior. For example, if a question places significant demands on respondents, a cost-benefit model of survey response behavior would predict that responses to this question will be less reliable. Assuming that, as discussed in Section 5.5, respondents who choose the "no opinion" response option or who answer questions quickly are not exerting sufficient effort to provide reliable responses, it is possible to test which respondent and survey characteristics affect respondent behavior and response reliability. These results provide new insights into why survey responses might not be reliable and offer suggestions on how to ensure that survey results are meaningful.

The various cognitive components of the question-answer process are examined below to help determine the demands placed on respondents during surveys. Subsequently, the rewards from survey participation are considered in an effort to establish why respondents would be motivated to answer survey questions. The factors that affect the demands and rewards of survey participation are examined closely in an attempt to uncover possible causes of unreliable survey responses.

3.1   Costs or Demands

The costs of survey participation are primarily a function of the survey questionnaire and respondents' interest or ability. As discussed above, the demands of any particular

survey question will depend on the type of information requested, whether or not the respondent has acquired relevant information, how this information has been stored in memory, and the respondent's ability to use recalled information in answering the survey question. Researchers, recognizing these limitations, have tried to piece together cognitive theories in order to explain how people answer different types of questions.

As mentioned in the introduction to this chapter, the accepted view of the survey response process involves four steps: understanding, retrieval, assessment and integration, and response selection. Cognitive psychology provides insights into the demands these steps place on respondents. Unifying ideas about comprehension, memory, judgment, and communication, survey behaviorists have developed various explanations for different types of respondent behavior. Although these behavioral interpretations are primarily descriptive, they provide insights that are incorporated into the shirking model of survey response. In particular, this work helps establish a better understanding of the demands and costs associated with different types of survey questions.

3.1.1 Understanding

The first requirement for social exchange, such as survey participation, is understanding. To answer questions accurately, a respondent must pay attention to the instructions and questions in order to identify the information sought (Clark 1985, Graesser, Singer & Trabsso 1994, Lehnert 1978). Many reporting problems in survey responses arise because respondents misunderstand the questions (Tourangeau, Rips & Rasinski 2000). Research on the pragmatics of language and conversation emphasizes the reasoning involved in the comprehension of questions (Grice 1989, Lewis 1975, Sperber & Wilson 1986, Clark & Schober 1992).

After interpreting the instructions, a respondent must determine what the question is asking. The grammatical and phonological structure of questions or interrogative sentences is different from that of typical declarative statements—the gram-

matical structure of a question is usually such that emphasis is placed on the construct for which information is being sought. This often occurs by inverted word order: "What is your opinion on abortion?" rather than "Your opinion on abortion is what?" Theories about the semantics of questions suggest that questions can be viewed as a set of possible answers (Tourangeau, Rips & Rasinski 2000, Groenendijk & Stokhof 1997, Higginbotham 1996). In this context, the objective of survey methodologists is to ensure that questions are worded so that respondents can link the key terms and relevant concepts to provide the correct response from the set of possible answers.

According to current generative theories of grammar, interrogative statements contain an unpronounced grammatical marker or trace that identifies what the question is seeking (Radford 1997). The understood trace in yes/no questions like "Do you approve of George Bush?" is the yes/no response. In questions beginning with wh-words (i.e., who, where, when, why, what, which, and how), identifying the trace can be more complicated, especially when there are several modifiers in the question (Tourangeau, Rips & Rasinski 2000). For example, the question "When did you last visit the doctor?" requires the respondent to indicate the date of his most recent doctors visit. However, the question "When did you last speak with your doctor?" might include both doctor's visits and telephone conversations. Ambiguity and complexity introduce syntactical difficulties that make answering questions harder. Furthermore the resolution of ambiguity or complexity may vary across respondents.

Semantic effects such as presupposition, unfamiliarity, and vagueness are also likely to complicate the determination of the appropriate set of possible answers (Tourangeau, Rips & Rasinski 2000). For example, the question about visiting the doctor assumes that the respondent actually does visit a doctor. If, instead, the primary health provider is not a doctor, the correct response set may be difficult to determine. Unfamiliar or vague terms may also complicate the response process; even terms that are commonly used may introduce uncertainty. For example, in questions that ask respondents to rate their feelings (i.e., "Would you say that you and your

family are much better off, better off, just about the same, worse off, or much worse off financially as you were a year ago?"), the interpretation of the response scales may vary across respondents (Aldrich & McKelvey 1977).

The demands that comprehension places on respondents are a component of the cost incurred if respondents choose to answer questions carefully. Understanding how different respondents perceive this source of difficulty enable the identification of different respondent behaviors.

3.1.2 Memory and Recall

The demands placed on respondents are not limited to simply understanding the question. Once a question has been interpreted (correctly or not), respondents must retrieve relevant information. Memory consists of various kinds of information—details specific to events or issues, event tallies, generic knowledge or stereotypes, and feelings or impressions (Rosch & Lloyd 1978, Schank & Abelson 1977, Schank 1982, Smith 1991, Smith, Jobe & Mingay 1991, Pillemer, Goldsmith, Krensky, Kleinman & White 1991, Tulving 1983, Kolodner 1985, Conway 1996, Lodge, McGraw & Stroh 1989). But memories are not always completely accurate, especially as they fade over time, nor are they always organized in a fashion that is amenable to answering survey questions.

Contemporary theories of forgetting suggest that the experience of similar events combined with the passage of time will interfere with memory about a specific event. Repetition and lack of distinction are likely to result in a person encoding multiple memories in the same index (Kolodner 1985). Studies of very-long-term memory demonstrate that the rate of forgetting varies with the type of information (Bradburn, Rips & Shevell 1987). Further studies have demonstrated that things that occur close to significant life events or temporal landmarks (i.e., the first day of college) are easier to recall (Kurbat, Shevell & Rips 1994). These differences are attributed to the way in which respondents store and retrieve different memories and thoughts.

The process of recalling information is likely to be somewhat circular (Tourangeau, Rips & Rasinski 2000). For example, if asked when he last visited the doctor, a respondent might take various approaches to recalling relevant information or ideas. Figure 3.1 demonstrates some possible steps. The respondent might remember that on his 40th birthday in April he did not feel very well. He might then remember that a few days after his birthday he went on a golf vacation and had to cancel a few rounds of golf. From this he might recall that he waited several months to see a doctor. Alternatively, he might remember that his new job, which started in July, provided him with a new insurance policy through which he found a new doctor. These results combine to provide an estimate concerning the date when the doctor was visited. The figure emphasizes that there may be multiple routes to recalling memories and that linkages between references need not be linear in time (Friedman & Wilkins 1985). When memories or thoughts are indistinct or non-unique, respondents will have to make judgments or inferences and this will require additional effort.
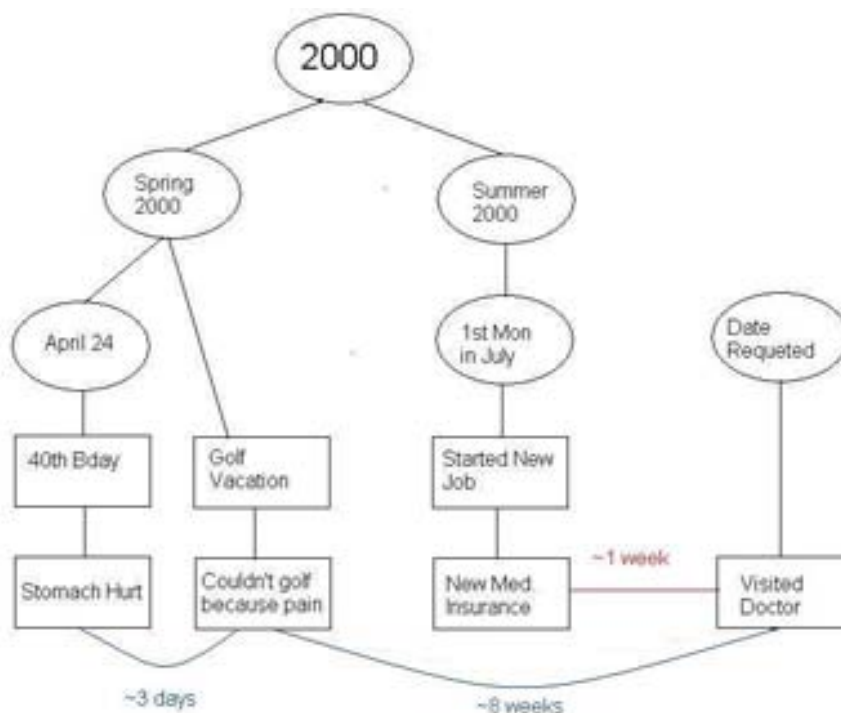
Figure 3.1: Memory and Recall for Doctor Visit Event

When asked to provide opinions, respondents are unlikely to have specific memories concerning the issue under consideration. Instead, respondents are likely to base their attitude expressions on a variety of components including general feelings, likes and dislikes, overall political leanings, and tallies of events.[3] For example, in determining voting intentions a voter might "canvas his likes and dislikes of the leading candidates and major parties involved in an election" (Kelley & Mirer 1974, p. 574). Alternatively, voters may draw on a summary tally to inform their evaluations when asked to indicate their candidate preferences (Lodge, McGraw & Stroh 1989).

Information relevant to opinions will be stored and accessed the same way as other memories and effort will be required to retrieve this information. Unlike factual or event recall questions, however, opinions are likely to be based on many different types of memories. The way in which these are retrieved and related to the question will be affected by the interview process and characteristics of the respondent. For example, one respondent may be motivated by the interviewer to exert significant amounts of effort to remember all his ideas about medical marijuana, a second respondent may already have a crystalized opinion on this issue, while a third might not wish to consider the issue in enough depth to express his opinion accurately. Differences across respondents and questions must be examined in order to establish the likelihood the a respondent will provide a reliable response.

3.1.3 Retrieval and Judgment for "Factual" Questions

Because searching memory for information is an imperfect process, respondents may retrieve different types of information and make judgments based on multiple pieces of information. For example, when asked to answer a factual question about the date of the most recent doctor's visit, the respondent must try to retrieve any mental record related to this event. In the best case, the respondent will be able to recall their last doctor's visit directly but, as suggested by Figure 3.1, in many situations he may have to rely on circumstantial or indirect evidence.

---

[3]These components on which opinions are based will subsequently be referred to as belief constructs.

Factors that affect individuals' ability to retrieve information about events include how important the event was, temporal landmarks related to the event, and the distinctiveness of the event (Kolodner 1985, Bradburn, Rips & Shevell 1987, Kurbat, Shevell & Rips 1994, Tourangeau, Rips & Rasinski 2000). Question design issues, such as the length of period under consideration and the pace of the interview, are also likely to affect retrieval. Finally, characteristics of the respondent will affect how much effort would be required to retrieve information about events accurately. Respondents who are adept at performing complex mental operations will be better able to retrieve information from their memories (Krosnick 1991, Krosnick & Alwin 1987, Schuman & Presser 1981).[4] The amount of cognitive practice a respondent has with the topic of a particular survey may also facilitate answering questions, i.e., the more an individual has thought about a topic, the easier it should be to retrieve the relevant information (Zaller 1992).

After remembering relevant information, a respondent may have to "tote up, average, combine, or summarize information" before being able to answer the question (Tourangeau, Rips & Rasinski 2000, pg 136). This suggests that a resulting survey response might best be represented by a weighted average of judgments,

$$\sum_J \text{Weight}_j * \frac{\text{Judgment}_j}{J} \quad \text{with} \quad \sum_J \text{Weight}_j = 1, \tag{3.1}$$

where the weights are determined based on factors such as the strength or reliability of the judgment.[5] This weighting procedure offers a mathematical summary of how respondents might determine their responses even if judgments are conflicting.[6] Respondents will have to exert different amounts of effort depending on how many judgments they retrieve and how complicated these judgements are to reconcile.

---

[4]This aptitude has been referred to as cognitive sophistication in the survey methodology literature.

[5]If only one judgment is determined to be relevant then the weight on this judgment would be 1.

[6]The multiplicity of judgments is also relevant for attitudinal questions. Section 3.1.4 discusses the possibility of coincidence of ideas in relation to opinion formation and expression.

3.1.4 Retrieval and Judgment for Attitude Questions

Attitude questions, like many factual or behavioral questions, will require respondents to judge the information they recall in order to respond to a question. As part of this process respondents must "form" an opinion.[7] Whether respondents form their opinions "on the door-step" (Converse 1964, Zaller 1992, Zaller & Feldman 1992) or from underlying predispositions (Alvarez & Brehm 2002), the retrieval and judgment process for attitude questions can be quite involved. In fact, on average, self-relevant factual questions generally take less than one second to answer while typical response times for attitude questions are between two and three seconds (Bassili & Fletcher 1991, Fazio & Williams 1986).

Retrieval and judgment for attitude questions requires respondents to undertake a multistage process. First they must evaluate and incorporate information. This information can then used to form considerations, beliefs, predispositions, and values. Then these considerations, beliefs, predispositions, and values must be reconciled in the formation of opinions. And finally, all the thoughts identified during the retrieval process have to be evaluated in the expression of opinions.

Information

Information plays a very important role because it enables respondents to shape their beliefs by persuading or cueing (Zaller 1992). A persuasive message provides information that can induce a person to change or solidify his opinions. "Medical marijuana has alleviated the suffering of patients with AIDS, cancer, and glaucoma" is an example of a persuasive message. A cueing message consists of contextual information that activates existing beliefs. For example, in the 1987 NES pilot study, respondents were asked about their opinions concerning U.S. support for the Contra rebels in Nicaragua; when questions contained cues about the Contras association with the fight against communism in Central America, different beliefs were triggered than

---

[7]Respondents might have pre-existing opinions on certain issues, in which case they "form" their opinions by simply calling to mind their pre-established opinions.

26

when respondents were simply asked whether they supported increasing aid to the Contras (Zaller 1992). Persuasion is a direct effect of information on opinions, while belief activation, as a result of cueing, represents an indirect effect of information. Another way that information might influence a person is by reducing their uncertainty (Alvarez & Brehm 2002) thereby making the response selection easier.

According to Zaller (1992, p. 6), opinions arise from the marriage of information and predispositions.[8] In other words, new information will be integrated with an individual's existing belief structures.[9] Although much of the existing literature on opinion formation is ambiguous with regards to the way in which this integration occurs.[10]

Opinion Formation and Expression

That people do not walk around with fully articulated opinions on every issue is acknowledged (Achen 1975). Instead, researchers search for underlying structures that

---

[8]In Zaller (1992) predispositions appear to describe individuals' inclinations or leanings while Alvarez & Brehm (2002) define predispositions as moderately durable belief constructs.

[9]The term belief structures is used to denote fundamental components of respondents beliefs. These might include predispositions, considerations, or attitudes. This terminology is used to distinguish these constructs from the reported beliefs that respondents provide in response to opinion survey questions.

[10]Whether belief updating takes place as a result of simple averaging (Tourangeau, Rips & Rasinski 2000), as a result of filtered elite messages (Zaller 1992), or by incorporation of elite information based on the degree of reliability or certainty (Alvarez 1998, Alvarez & Brehm 2002), the formation of opinions is likely to be quite involved. According to Zaller (1992), people resist arguments that are inconsistent with their predispositions, thus only certain types of messages will be received. If respondents perfectly filtered information based on their predispositions, their opinion would not change since they would only receive messages that support their current opinion. However, if filtering is less than perfect, information that contradicts predispositions will be received but at a lower rate than confirmatory information. These rates of information inflow together with the volume of both contradictory and confirmatory information will affect the probability that an individual will change his opinion. Although both Zaller (1992) and Alvarez & Brehm (2002) agree that predispositions are likely to derive from elite discourse, there appears to be relatively little discussion about how these predispositions develop over time. The in-depth interviews, for example, suggested that predispositions or values were determined from upbringing and past experiences. This suggests that these predispositions are learned early in life and are not easily updated since predispositions affect the information received. A fully developed model of information filtering and opinion change would have to take into account the history of an individual. In such a model, opinions will be path dependent in such a model. While this idea has merit, the objective here is simply to recognize the potential complexity of opinion formation.

enable respondents to render judgment on opinion questions.[11] Proposed sources of opinions include: considerations (Zaller 1992) and predispositions (Alvarez & Brehm 2002). Zaller (1992, p. 40), defines considerations as "any reason that might induce an individual to decide a political issue one way or the other." This necessarily admits the possibility that preferences may be based on temporary assessments. Tourangeau, Rips & Rasinski (2000), building on Zaller's idea, identify three sources of attitudes: specific beliefs or feelings, general attitudes or values, and impressions or stereotypes. While this insight does not eliminate the possibility that opinion expressions are malleable, some of these sources of opinions are of a more enduring nature. On the other hand, in Alvarez & Brehm (2002), predispositions are defined as durable moderately specific belief structures with object-relevant contents, like identities, classifications, and standards; they include values, group attachments, affective judgments, and expectations.

While disagreement remains concerning the cognitive structures underlying the expression of opinions, the consensus is that opinions are pieced together from various elements. For example, when asked whether he approves or disapproves of President Clinton, a respondent is likely to consider many things, such as knowledge about the economic conditions during the Clinton administration, core values concerning morality, feelings about the Middle East Peace Process, or expectations concerning Clinton's trustworthiness. As with judgments on factual questions, the process by which considerations or predispositions are combined is poorly understood. However, it is clear that the demands placed on a respondent will depend on the number of considerations or predispositions activated by a survey question and the relationship among these belief constructs.[12]

The coincidence of predispositions can induce three possible states of mind: uncertainty, equivocation and ambivalence (Alvarez & Brehm 2002, p. 58). Uncertainty

---

[11]See, for example, Zaller (1992), Zaller & Feldman (1992), Tourangeau, Rips & Rasinski (2000), Lippmann (1922), and Alvarez & Brehm (2002). Lippmann (1922, p. 59) summarized this idea when he said that opinions are "to be pieced together out of what others have reported and what we can imagine."

[12]A belief construct is activated if a respondent identifies it as relevant to the issue under consideration.

is described as a situation where multiple predispositions do not affect response variability and additional information reduces response variability. Equivocation occurs when multiple predispositions are reinforcing—response variability is reduced by both coincidence of predispositions and the level of information. Finally, respondents are deemed ambivalent if both information and the coincidence of predispositions increase response variability.

Returning to the Clinton approval question, respondents may have difficulty determining their feelings because they are ambivalent—they may, for example, be torn between their concerns over Clinton's Middle East policy and their support for his economic policies. On the other hand, in cases of equivocation (when respondents have reinforcing beliefs or feelings concerning Clinton), respondents will be more easily able to express their opinions ceteris paribus. Additional information is also likely to affect respondents' feelings concerning Clinton. More information may make it easier for individuals who were uncertain to provide reliable opinion expressions but more information (particularly contradictory information) can also increase ambivalence thereby making the response process more difficult. In other words, the difficulty of a question will depend on the situation and the specific respondent involved. Because a multiplicity of belief constructs (i.e., considerations, predispositions, core values, etc.) may be activated by a survey question, respondents will have to carefully combine and access these belief constructs in order to provide reliable answer.s For example, in determining for whom to vote, a voter must retrieve relevant memories or ideas about the candidates and parties. Whether a respondent's belief constructs are pre-formulated tallies or itemized likes and dislikes will affect how much effort the respondent will be required to exert in order to provide a reliable opinion expression.

Wilson & Hodges (1992) describe the process of answering an opinion question as a matter of consulting appropriate mental files. But memory has been shown to be systematically biased (Lodge, McGraw & Stroh 1989); anecdotal evidence also indicates that information is not conveniently organized in files that can be easily retrieved when asked to express opinions (Tourangeau, Rips & Rasinski 2000).

Instead, opinion expressions will be determined by the accessibility of belief constructs (Zaller 1992, Tourangeau, Rips & Rasinski 2000). The activation of belief constructs is likely to be affected by both elite discourse and survey design (Gamson & Modigliani 1987, Kinder & Sanders 1990). Furthermore, people might arrive at their responses using a "high-road" or a "low road" (Cannell, Miller & Oksenberg 1981, Krosnick 1991, Krosnick & Alwin 1991, Stack & Martin 1987). For example, Krosnick (1991) proposes that respondents "satisfice" by shortcutting their cognitive processes and providing responses that are not carefully thought-out.[13] None of these models, however, contains an explicit explanation of why certain belief constructs are more likely to be activated.

Examining considerations and predispositions activated by questions is important to interpreting respondent behaviors. As with factual questions, both individual characteristics and question design will affect how opinions are formed and expressed. A better understanding of the demands respondents identify relative to specific questions can provide insight into the response process and help explain why respondents report unreliable opinions.

3.1.5 Reporting Responses

"Even when respondents have a clear answer to report, it may not be clear to them how to report it" (Tourangeau, Rips & Rasinski 2000, p. 13). Once a respondent has determined his answer to a particular question, be it a factual, behavioral, or attitude question, this answer has to be reported. When faced with questions containing fixed response options, respondents must map their judgments into the scale provided.

As discussed in Section 3.1.1, response options may be vague or unclear. For example, when asked to identify their position on a scale, different respondents use different anchors to identify the endpoints of the scale (Aldrich & McKelvey 1977). Also, more than one response option may be appropriate and forcing respondents to choose can introduce unreliability of responses. Even if the question offers an

---

[13]See Section 4.2.2 for a more thorough discussion of the satisficing model.

open-ended response, respondents are likely to edit their answers for consistency or acceptability. Again, the demands of the survey, which depend on the structure of the survey and respondent characteristics, will determine how respondents behave when reporting their responses.

3.2   Benefits or Rewards

Having identified the possible demands of survey participation, the next stage to understanding respondent behavior is to recognize the benefits derived from survey participation.

As discussed at the beginning of this chapter, respondents will be encouraged to incur the cognitive costs of survey participation if they view the survey process as rewarding. Given that survey researchers ask participants to exert effort and incur the costs of answering questions, it is important to compensate respondents for their efforts. If survey researchers are concerned about the reliability of survey responses, compensation for survey participation should ideally be contingent upon the cognitive effort exerted by the respondent.

3.2.1   Financial Rewards

Financial rewards have been discussed as a means of compensating respondents for participating in surveys.[14]. The debate over how much and when to pay respondents has focused primarily on ensuring higher response rates (Dillman 2000, James & Bolstein 1992, Johnson & McLaughlin 1990) but appropriately designed financial rewards may also have an affect on the amount of effort respondents exert in answering individual questions—recall that the factors influencing the decision to participate are directly linked with those that affect response behavior on individual questions. Thus, a consideration of how to compensate respondents for exerting effort in answering questions will have direct implications for encouraging higher survey response rates.

---

[14]see, for example,, Church (1992)

Typically, financial compensation for survey participation entails some form of lump sum payment but this will not guarantee that respondents exert enough effort to provide reliable responses. Payments contingent upon effort would encourage respondents to more exert effort but this may not be an option as it is difficult, if not impossible, to directly measure effort and link it to payments.

3.2.2 Non-Pecuniary Rewards

Non-pecuniary rewards, such as contributing to policy discussions, however, may be more closely tied to effort. For example, if a respondent considers the issue of medical marijuana very important and thinks that his survey responses are likely to affect policy, he should be willing to exert enough effort to provide reliable responses to questions about medical marijuana. However, if a respondent views an issue as irrelevant or thinks that his responses are unlikely to have any effect on relevant policy outcomes, he may prefer not to exert much effort. Thus, these forms of non-pecuniary rewards will play an important role in determining the effort respondents exert in answering survey questions and, ultimately, the reliability of their survey responses.

Unfortunately, the understanding of the non-pecuniary rewards derived from survey participation is limited. As mentioned above, it has been proposed that respondents might derive benefits from the opportunity for self-expression, for interpersonal exchange, for intellectual challenge, for self-understanding, or for emotional catharsis (Brehm 1993, Dillman 2000).[15] Rewards may also be incurred because people feel that they are "performing their civic duty" or "informing government and business decisions" (Riker & Ordeshook 1968, Krosnick 1991).

---

[15]Dillman (2000) identifies "best-practices" in survey design and as a component of this he outlines several ways of minimizing the demands on respondents and offering appropriate non-pecuniary compensation and encouragement. In particular, Dillman (2000, pp. 15-21) provides a list of possible ways to "reward" respondents.

3.2.3 Some Evidence of Rewards

In an attempt to understand what motivates survey participation, the 1989 Detroit Area Study (DAS) asked respondents why they agreed to be interviewed. The factors that influence survey participation have been categorized as functions of the respondent's relationship to the interview, relationship to interviewer, and relationship to self (Brehm 1993). The top four reasons for participating in the survey were as follows: (1) curious about/interested in surveys, (2) believe surveys help solve problems and let government know what people want, (3) the respondent sees himself as a helpful or nice person, and (4) the respondent enjoys expressing opinions. Since almost half of the respondents talked about the usefulness of surveys and their personal enjoyment of the survey participation process, this indicates that the rewards from survey participation are related to emotional catharsis and easing of social conscience.

Chapter 2 also identified some preliminary evidence suggesting that respondents are aware of the rewards from survey participation. Several of the interview participants who completed the 2003 Attitudes Surveys mentioned that they valued the opportunity to express their opinions on the issues of abortion and euthanasia. Subjects who viewed these issues as particularly important were also more likely demonstrate that they had exerted a significant amount of effort in determining their opinions on these issues and providing responses to the survey questions.

3.3 The Net Benefits

The theories about comprehension, memory, and judgment discussed above suggest that respondents will be required to exert effort in order to provide reliable responses. In fact the evidence discussed above suggests that carefully interpreting questions, retrieving relevant memories, accessing judgments, and selecting responses may be quite demanding. Also, as discussed above, survey participants may receive non-pecuniary rewards from survey participation—respondents might enjoy expressing their opinions and contributing to government and corporate policy. Together with

the evidence from in-depth interviews discussed in Chapter 2, this evidence suggests that modeling survey response behavior in a rational choice framework that accounts respondents' choice of effort may be appropriate.

By examining survey responses in the rational choice framework, appropriate empirical tests can be developed to determine the factors that influence the reliability of survey responses. With this in mind, the next chapter reviews evidence concerning the lack of reliability and stability in survey responses and explores some existing hypothesis explaining this behavior. Chapter 5 expands existing theories of respondent behavior by developing a rational choice model that explicitly incorporates effort. Using this model, Chapters 6 and 7 implement tests of how the costs and benefits of survey participation affect respondent behavior.

Chapter 4 Instability and Reliability of Opinion Survey Answers

> It is widely documented that question alterations can produce changes in responses (Krosnick 1991, p. 213).

Determining when survey responses are reliable should be of particular interest to survey researchers who wish to ensure that results from their surveys are meaningful and predictive of behavior. Fortunately, the accuracy of factual or behavioral responses can been established by supplementing survey data with secondary sources of information such as diaries, doctors' records, or voting records.[1] Discussing the reliability of opinion questions, however, is much more complicated since there is no way to verify the accuracy of respondents' opinions. Researchers have pointed to data indicating that "even slight variations in the way an attitude question is asked can significantly change answers" (Krosnick 1991, p. 213).

The first step to answering this question is to identify when survey responses appear to vary. Survey methodologists have undertaken extensive studies of response instabilities (Achen 1975, Bradburn, Rips & Shevell 1987, Clark & Schober 1992, Converse 1964, Dillman, Brown, Carlson, Carpenter, Lorenz, Mason, Saltiel & Sangster 1995, Fletcher & Chalmers 1991, Kinder & Sanders 1990, Krosnick & Alwin 1991, Krosnick & Alwin 1987, Krosnick 1991, Narayan & Krosnick 1996, Schuman 1992, Schuman & Presser 1981, Tourangeau, Rips & Rasinski 2000, Wilson, Kraft & Dunn 1989, Zaller 1992). Having identified situations where survey responses vary with the wording of questions, researchers have sought to explain and account for these instabilities. Unfortunately, the explanations for these behaviors tend to

---

[1]Researchers have documented many different sources of inaccuracy in behavioral recall questions. For example, when asked to recall events during a certain period of time, respondents often omit some events and include others from outside the reference period; such behaviors are called telescoping errors. For an extensive discussion of the different sources of inaccuracies induced by different types of factual or behavioral question questions see, for example, Tourangeau, Rips & Rasinski (2000).

be ad hoc and limited in scope. Belief constructs such as considerations and predispositions provide potential avenues to understanding such behaviors but work still needs to be done to fully integrate these ideas into a comprehensive model of survey response behavior. The following review of some of the studies in which response instability was observed along with a critical examination of proposed explanations, demonstrates how little is understood concerning respondent behavior.

4.1 Observed Attitude Instability

In his seminal work on attitudes among mass publics and elites, Converse discovered high levels of respondent opinion variability across both time and across related topics (Converse 1964).[2] He found rank-order correlations of approximately 0.30 between individuals' positions on foreign aid in 1958 and in 1960; similarly weak results were obtained for isolationism, federal support for employment and education, and school desegregation (Converse 1964, p. 240). Comparing pairs of attitudinal questions asked on related issues in the same survey, extremely low rank-order correlations were also found (Converse 1964, p. 228). These observations were inconsistent with Converse's idea of cohesive attitudes, thus, he concluded that large proportions of the electorate did not have meaningful beliefs, even on significant issues.

Survey researchers have also demonstrated that the same questions often produce quite different answers depending on the "context" in which they are asked.[3]. Context effects have been extensively examined through survey experiments. By changing the order in which questions are presented, several studies have found high levels of response variability. For example, the level of support expressed for legal abortion when the "woman is married and does not want any more children" is lower when this

---

[2]The data used in by (Converse 1964) was from the 1956, 1958, and 1960 University of Michigan Survey Research Center (SRC) election studies. In 1958 and 1960 SRC re-interviewed respondents from the original 1956 sample, providing repeated observations for the same individuals.

[3]See, for example,, Bradburn, Rips & Shevell (1987), Schuman & Presser (1981), Tourangeau, Rips & Rasinski (2000)

question follows a question concerning support for abortion "when there is a strong chance of serious defect in the baby" (Schuman 1992, Schuman & Presser 1981).

Survey experiments in which question frames (organizing ideas or stories that provide meaning by identifying the essence of issues) are varied have also been used to study how responses change in different contexts.[4] Kinder & Sanders (1990) report that the question frames determines whether racial resentment attitudes or equal opportunity preferences are active in the opinion expression. Different question frames, however, do not always induce different responses; Alvarez & Brehm (2000) found only marginal impacts on survey responses when they varied question frames on social policy questions in the 1998–1999 Multi-Investigator Study.

4.2 Non-Attitudes vs. Durable Belief Constructs

Many explanations have been proposed to account for the observed instability of opinion survey responses. The first, and ever present, explanation is that respondents do not have meaningful attitudes. This view, attributed to Converse (1964), has instigated much of the current research in survey response behavior.

As discussed in Chapter 3, most researchers agree that respondents do not possess fully formed opinions on every topic. Instead, the opinions expressed on surveys are predicted to derive from fundamental belief constructs such as considerations, predispositions, or core values (Alvarez & Brehm 2002, Zaller 1992). Whether these belief constructs are temporary or durable has not been conclusively established but it is clear that how these constructs are activated and accessed will influence the response provided. As information, considerations, and predispositions vary so will survey responses. Thus survey response instability may result from changes in the respondent's environment.[5]

---

[4](Gosnell 1927) is attributed as having introduced experimental methods into political process analysis by assigning Chicago residents to receive postcards reminding them of their sacred right to vote; come election day, about eight percent more of the experimental group than the control group made it to the polls.

[5]Given this variable nature of opinion expressions, it may be difficult to distinguish real opinion change from response variability that results because survey answers do not accurately capture

4.2.1 Belief Construct Pluralism

While there remains some dispute concerning the nature of belief structures underlying responses to opinion questions (Alvarez & Brehm 2002, Zaller 1992, Zaller & Feldman 1992, Tourangeau, Rips & Rasinski 2000), most researchers agree that respondents may consider many things when answering a question. If respondents identify several relevant belief constructs, conflicts are likely to arise and thus judging and reconciling these conflicts will be a necessary component of the survey response process.[6] For example, in examining support for affirmative action quotas, conflict may arise between values of equality and individual merit. Confronting respondents with arguments running counter to the view they expressed in their initial response has been shown to induce respondents to change their opinion expressions (Fletcher & Chalmers 1991).[7] However, attempts by Alvarez & Brehm (2000) to trigger conflicting belief constructs through the use of question frames did not confirm the sensitivity of responses.[8] This evidence suggests that belief construct pluralism may offer some insight into the variability of survey responses.[9]

Conflicting considerations, predispositions, or core values are likely to increase the demands placed on respondents who try to resolve these conflicts. How respondents deal with these demands will determine the quality of the response that they provide.

---

opinions using inter-temporal studies. By using "no opinion" responding and response times as indicators of shirking, this work avoids this issue.

[6]Evidence indicates that respondents are capable of simultaneously maintaining both positive and negative evaluations of an object (Cacioppo & Gardner 1999, Holbrook, Krosnick, Visser, Gardner & Cacioppo 2001).

[7]Respondents who supported quotas were asked if they would feel the same way even if this meant not hiring the best person for the job, whereas respondents who opposed quotas were asked if they would feel the same even if it meant that women remained economically unequal. Fletcher & Chalmers (1991) found that over fifty percent of respondents (movers) changed their opinions when the countervailing argument was introduced.

[8]Alvarez & Brehm (2002) used seven different question stems in attempt to induce egalitarian, individualism, and anti-black values. Comparing results across the different treatments, no conclusive evidence was found about how responses depend on how questions were framed. A difficulty with this approach is that inducing values, considerations, or predispositions may not be as straightforward as changing the way in which questions are asked.

[9]Unfortunately, it is difficult to identify when conflicting predispositions are triggered. The in-depth interviews discussed in Chapter 2 offered some evidence that conflict induced respondents to provide neutral or unreliable responses. But this sample is relatively limited and the subjects were not re-interviewed so it is impossible to make definitive statements about how predisposition conflict affects response stability or how conflict can be distinguished from shirking.

If respondents shortcut their cognitive processes by incompletely evaluating their belief constructs, respondents are less likely to provide reliable, cohesive, and stable answers.

4.2.2 Shortcutting Cognitive Processes

Respondents may choose to shortcut their cognitive processes when they are tired, uninterested, or distracted Krosnick (1991).[10] Unfortunately, it is difficult to directly determine when respondents are shortcutting their cognitive process. Krosnick (1991) lists behaviors that he claims are "typical" of different degrees of satisficing but very little evidence is given to support the use of these measures indicators of cognitive shortcutting.[11] Using these indicators as measures of satisficing does not distinguish respondents who are exerting significant effort from those who are shortcutting their cognitive processes.

Krosnick (1991, p. 221) describes three conditions that foster this behavior: the difficulty of the task, the respondent's ability to perform the task, and the respondent's motivation. These effects are combined and the probability of satisficing is defined as (Krosnick 1991, p. 225):

$$Pr[\text{Satisficing}] = \frac{a_1(\text{Task Difficulty})}{a_2(\text{Ability}) * a_3(\text{Motivation})}. \tag{4.1}$$

This equation suggests, for example, that if a task is especially difficult and the respondent is not very motivated or able, the likelihood of observing satisficing behavior is high. While this relationship amongst these factors is reasonable, there is no theoretical foundation for the specification proposed. In Chapter 5 a model that produces

---

[10]Borrowing from the economic literature, Krosnick (1991) termed this behavior shirking. The term satisficing was coined by Simon (1957) who suggested, that, when faced with demanding information-processing or decision-making tasks, individuals expend only the effort necessary to make a satisfactory or acceptable decision. The shirking model has concepts similar to those in the satisficing model but in the shirking model cognitive shortcutting is fully rational and this makes it inappropriate to refer to this as satisficing.

[11]Krosnick (1991) suggests that the following behaviors are indicative of satisficing: selecting the first reasonable response, acquiescence, endorsing the status quo, non-differentiation among questions with the same rating scales, and saying don't know.

behavior similar to that suggested by Krosnick (1991) is developed. By explicitly modeling respondents as rational agents who optimally determine how much effort to exert in answering survey questions, the shirking model has testable implications about how ability, motivation, and task difficulty will affect respondent behavior. But before developing this model, it is important to understand that all researchers interested in analyzing the reliability opinion expressions struggle to identify when responses are likely to be unreliable.

4.3   Identifying Reliable Responses

Because it is difficult to identify when a respondent has reliably answered an opinion question, much of the existing work on opinion reliability has focused the correlation of opinion expressions across time, across related issues, and across similarly worded questions. But as discussed above, changes in opinion expressions cannot necessarily be attributed to a lack of reliability of survey responses. If, as a result of uncertainty, opinions might best be expressed as distribution over a policy space, then repeated questioning will generate a certain amount of variation in responses (Achen 1975, Alvarez & Brehm 2002). The coincidence of relevant predispositions or considerations may also induce response variability; responses might vary because different predispositions or considerations are activated in different settings. The fundamental problem is to distinguish unreliable opinions from those which have been established after thorough contemplation.

Unfortunately, formal tests of whether respondents are reliably expressing their opinions on survey questions is not possible given the available data. Instead, using an approach similar to that of Krosnick (1991), certain observable behaviors are assumed to correspond with responses that are provided without sufficient consideration— evidence supporting the use of these indicators is drawn from the in-depth interviews discussed in Chapter 2.

Assuming that respondents who do not exert enough effort to provide reliable responses can be identified by observable behaviors, Chapters 6 and 7 test the factors

that influence this behavior. Because this work does not use inter-temporal measures to identify responses that are unreliable, confounding effects from opinion changes over time or across different question wordings are avoided. Instead, this work attempts to directly address the issue of reliability.[12]

---

[12]This work does not resolve issue of identification although the evidence from the in-depth interviews provides some indication that "no opinion" responding and response times might be used to identify respondents who are not providing reliable responses.

Chapter 5   The Shirking Model of Answering Surveys

> It is hard enough to remember my opinions, without also remembering my reasons for them! (Nietzsche 1883)

Taken together, the above discussion about the demands and rewards of survey participation, the observation that survey responses are unstable and unreliable, and the lack of a unified model which can explain respondent behavior, suggest a need for a formal theory of the survey response behavior which explicitly incorporates effort. Drawing from foundations in rational choice theory, this chapter proposes such a formal model—the shirking model. In Chapters 6 and 7, the implications of this model are tested. The fundamental problems confronted in this analysis are as follows:

- Identification of respondents who are shirking,

- Establishing whether shirking is a significant problem or not, and

- Determining the implications of shirking behavior for survey design and data analysis.

Because the shirking model is based on some basic assumptions, a concrete framework within which to examine respondent behavior is established. The shirking model also has direct implications for the specification of empirical tests.

5.1   Deriving the Shirking Model

As discussed in Chapter 3, answering surveys requires respondents to exert effort but they receive non-pecuniary rewards for this effort.

Assumption 1 *Effort is costly but surveys reward respondents for this effort.*

Given these costs and rewards to answering survey questions, respondents must be able to evaluate their optimal survey response behavior.

Assumption 2 *Respondents rationally select an effort level to maximize the difference between their costs and benefits (i.e., their net benefits).*

Assuming that respondents are rational allows us to predict how respondents will behave when faced with different costs and benefits to survey participation.

As outlined in Section 3.1, respondents incur costs from survey participation because accurately answering questions requires that they understand the question, retrieve relevant concepts from memory, make judgments, and finally select responses. Hence, the demands of different survey questions will depend on the question difficulty and respondent ability. Rewards for answering questions, on the other hand, will be affected by respondent motivation. Since rational respondents consider the demands and rewards of survey participation, the effort respondents choose to exert in answering survey questions will depend on ability, motivation, and difficulty.

The concept of shirking is based on the idea that the effort level chosen by rational respondents does not necessarily correspond with the amount of effort that would be required to provide reliable responses. The discrepancy arises because respondents will not necessarily incorporate the social benefits from survey participation into their net benefit evaluation.[1] Since this work is based on the assumption that respondents behave rationally, it is useful to fully describe rational behavior in the context of survey participation.

---

[1]Surveys may be viewed as producing externalities. This raises the idea that respondents may be encouraged to provide sufficient effort if they incorporated these externalities into their net benefit function. An examination of this possibility is deferred to future research.

5.2 Preferences and Rationality in the Survey Setting

Individual decision-making behavior can be represented by a binary preference relation, usually denoted by $\succeq$, on the set of choice alternatives. In the survey behavior setting, the decision set is the amount of effort the respondent could choose to exert. Standard behavioral theory assumes that individual preferences are rational.

Definition 1 *Rational Preferences:* complete and transitive preferences.

Definition 2 *Completeness:* for all $x, y \in X$, we have that $x \succeq y$, $y \succeq x$ or both.

Definition 3 *Transitivity:* for all $x, y, z \in X$, if $x \succeq y$ and $y \succeq z$, then $x \succeq z$.

The completeness assumption says that individuals have well-defined preferences between any two different levels of effort.[2] Transitivity, on the other hand, implies that no sequence of pairwise choices will cycle. In the context of choosing how much effort to exert, if exerting little effort is preferred to exerting a moderate amount of effort and exerting a moderate amount of effort is preferred to exerting a lot of effort, then transitivity implies that exerting a lot of effort cannot be preferred to exerting little effort (Mas-Colell, Whinston & Green 1995). The rational choice framework used in the survey setting implies that respondents can calculate the benefits and costs of exerting effort and choose effort to maximize their personal net benefits.

To enable an examination of behavior that is not primarily predicated on financial interactions, the inclusion of non-pecuniary rewards is essential. By admitting all factors that might influence actions, the rationality definition becomes almost tautological in that all behavior can be described as rational (Riker & Ordeshook 1968). The shirking model attempts to limit the scope of the rationality description by including only factors relating to ability, motivation, and difficulty.[3]

---

[2]This assumption implies that respondents can compare all possible levels of effort.

[3]By explicitly expressing factors that influence behavior, I hope to restrict the debate to substantive issues.

5.3 Choosing Effort

In taking a survey, respondents determine their preferences based on perceptions about the demands/costs and rewards/benefits from exerting effort in answering questions. Assuming that both costs and benefits are increasing in the effort exerted by respondents, benefits and costs to respondent $i \in N$ from answering a survey question $q \in Q$ can be expressed as non-decreasing functions of a continuous measure of effort, e.

Assumption 3 *Benefits:* $B(e_{iq})$ *with* $B'(e_{iq}) \geq 0$.

Assumption 4 *Costs:* $C(e_{iq})$ *with* $C'(e_{iq}) \geq 0$.

Net benefits, or utility, from each question can then be expressed as a function of effort.[4]

Assumption 5 *Utility:* $U_{iq} = B(e_{iq}) - C(e_{iq})$.

The rational respondent i, will choose an effort level e_{iq}^{r*} to maximize U_{iq}.[5]

5.4 Defining Shirking

In order to draw valid conclusions about beliefs and opinions, researchers need respondents to provide reliable expressions of their opinions.[6] To provide a reliable opinion on question q, respondent i would be required to exert an effort level of at least e_{iq}^{s*}.[7] However, a respondent i, who maximizes utility by choosing e_{iq}^{r*}, may exert

---

[4]This utility may be oversimplified but it is only intended as a sketch to illustrate how effort levels are chosen. While a thorough development of this rational choice model of survey response would be useful, the purpose of this analysis is to explicitly introduce the concept of effort into a formal model of survey response behavior.

[5]The superscript r indicates that this is the optimal level of effort for respondent i to exert on question q.

[6]Most research to date has focused on how responses change as a result of variations in how questions are asked.

[7]The s superscript is used to indicate that this effort is optimal from the surveyor's perspective.

less effort than what is optimal from the surveyor's perspective, i.e., $e_{iq}^{r*} < e_{iq}^{s*}$. By construction, this results in responses that are not be reliable.[8]

Definition 4 *Shirking: if $e_{iq}^{r*} < e_{iq}^{s*}$ respondent i is* shirking *on question q.*

By construction, a respondent can provide either too little effort or more than enough effort, thus respondent i either shirks on a question q, denoted $S_{iq} = 1$, or not, $S_{iq} = 0$.[9] Given the above definition of shirking, the probability that a respondent i shirks on question q can be expressed as

$$Pr[S_{iq} = 1] = Pr[e_{iq}^{s*} - e_{iq}^{r*} > 0]. \tag{5.1}$$

Unfortunately neither the effort level exerted by a respondent, e_{iq}^{r*}, nor the surveyor's optimal effort level, e_{iq}^{s*}, can be observed directly.[10] Instead efforts levels must be expressed as functions of observable respondent characteristics, X_i, and question characteristics, Z_q, affect optimal effort levels, e_{iq}^{r*} and e_{iq}^{s*}. As suggested by (Krosnick 1991), factors affecting respondents' choice over effort levels might best be classified under three headings: ability, motivation, and difficulty. Similarly, the optimal effort level from the surveyors perspective will also depend on question difficulty and respondent ability.[11] Individual characteristics, X_i, and survey characteristics, Z_q, attempt to capture these components.

If it is assumed that the optimal effort levels can be expressed as linear functions of the observable characteristics plus stochastic (unobserved) components, ϵ_{iq}^r and ϵ_{iq}^s,

---

[8]The surveyor's optimum may be greater than the respondent's because respondents do not internalize the benefits of reliable survey results and because the surveyor does not have to incur the costs of effort.

[9]For the purpose of the subsequent analysis shirking is assumed to be a binary variable but it is possible to extend the definition of shirking to admit different levels of shirking.

[10]The time that respondents take in answering a question has been proposed as a proxy for the effort level exerted. See below for a discussion of response times as a means of measuring respondent effort (e_{iq}^{r*}). Approximating the surveyor's optimal effort level, however, is more complicated. Establishing the reliability of survey responses requires the collection of auxiliary data for validation purposes but collecting such data for opinion questions is complicated; for example, if different belief structures are activated, then different opinion expressions cannot be compared to determine their reliability.

[11]Of course the surveyor cannot control respondent ability but it is possible to design questions that are less difficulty.

Assumption 6 *Respondent's optimal effort: $e_{iq}^{s*} = X_i'\beta^s + Z_q'\gamma^s + \epsilon_{iq}^s$*

Assumption 7 *Surveyor's optimal effort: $e_{iq}^{r*} = X_i'\beta^r + Z_q'\gamma^r + \epsilon_{iq}^r$*

then the probability a respondent i will shirk on question q, given X and Z, can be expressed as

$$Pr[S_{iq} = 1 \mid X, Z] = Pr[X_i'\beta + Z_q'\gamma + \epsilon_{iq}^{rs} > 0 \mid X, Z]. \tag{5.2}$$

Where $\beta = \beta^s - \beta^r$, $\gamma = \gamma^s - \gamma^r$, and $\epsilon_{iq}^{rs} = \epsilon_{iq}^s - \epsilon_{iq}^r$. If the distribution of the error term, $F(\epsilon_{iq}^{rs})$, is normal, then a binary probit model is appropriate (Greene 2000). However, if ϵ_{iq}^s and ϵ_{iq}^r are identically and independently distributed in accordance with the extreme value distribution then a binary logit model is implied (McFadden 1973).

The definition of shirking can be extended to admit different degrees of shirking.[12] In this case S_{iq} could be expressed as taking on ordinal values $0, 1, 2, ..., g$ indicating the levels of shirking. A simple extension of the probability model expressed above would suggest either an ordered logit or ordered probit model.

Unfortunately, the implications of this model can only be tested if shirking can be identified. As discussed above, to undertake the analysis of respondent behavior, this work identifies certain behaviors that correspond with shirking.

5.5 Indicators of Shirking

5.5.1 Survey Level Indicators of Shirking

The shirking model described above is designed to examine respondent behavior on each question separately but it may not always be feasible to identify whether a respondent is shirking on a particular question. Instead, survey level indicators of shirking might be implemented. These measures would identify respondents who do

---

[12]Degrees of shirking might be indicated by the frequency that respondents employ strategies indicative of shirking. For example, a respondent that provides "no opinion" responses on all survey questions might be coded as a strong shirker while someone who provides "no opinion" responses only occasionally might be perceived as a weak shirker.

not exert enough effort to provide reliable responses to the survey questions in general. One example of a respondent who might be identified as a shirker at the survey level is the person who provided "no opinion" responses on ninety-eight percent of the opinion questions in the 1984 NES Pre-Election Survey.

Since it is possible to establish the accuracy of factual responses, one definition of shirking would be to identify shirkers as those who quickly provided incorrect responses to these questions. Assuming that response time is positively correlated with effort and that with enough effort a respondent would be able to provide accurate responses to factual questions, respondents who provide wrong answer quickly are unlikely to be exerting sufficient effort. If behavior on factual questions is consistent with overall survey response behavior, then "quick-wrong" factual respondents can be taken as an indicator of shirking at the survey level. Unfortunately, to uncover shirking behavior at the question level, a different approach is necessary.

5.5.2 No Opinion Responding

As suggested by Krosnick (1991), if respondents do not wish to engage in the complex cognitive process of answering an opinion question they can simply indicate that they do not have an opinion.[13] Evidence from the in-depth interviews discussed in Chapter 2 suggests that respondents who are not exerting much effort may display "no opinion" response behavior. Furthermore, if it is assumed that, with enough effort, everyone is able to accurately express opinions, then a respondent who provides a "no opinion" response must be shirking.

The "no opinion" response measure can be implemented both at the survey and question level. Given the above assumptions, the proportion of "no opinion" responses provided through-out a survey provides another indicator of shirking at the survey level. Alternatively, the binary variable indicating whether respondents provided a "no opinion" response on a particular question offers a way of identifying respondents

---

[13]In this analysis, "no opinion" responses will include the selection of either the "no opinion" or "don't know" response option on an attitude question.

who might not be exerting enough effort to provide a reliable response to the question under consideration.[14]

5.5.3 Response Times

As mentioned above, the time that it takes respondents to answer a question provides an alternative way of trying to gage whether respondents are exerting enough effort to provide reliable responses. If longer response times are associated with more reliable responses, then responses provided relatively quickly would be indicative of shirking at the question level.

According to (Bassili & Fletcher 1991), response times (response latencies) provide new insights into survey respondent behavior. Studies have found that average response time is directly related to how much effort would be required to accurately answer questions. For example, respondents take less than one second on average to answer self-relevant factual questions while attitude questions have typical response times between two and three seconds (Bassili & Fletcher 1991, Fazio & Williams 1986). Also, questions that activate multiple conflicting considerations or predispositions take longer to answer (Fletcher & Chalmers 1991, Bassili & Fletcher 1991). In terms of ability and motivation, studies have found that education, days into a campaign, and partisan extremity are all negatively correlated with the time it takes respondents to answer a question about their party identification (Mulligan, Grant, Mockabee & Monson 2003). This evidence suggest that it might be reasonable to assume that respondents who answer questions relatively quickly are shirking.

5.6 Ability, Motivation, and Difficulty

To implement the model of shirking described above, respondent characteristics, X_i, and survey attributes, Z_q, need to be quantified. As mentioned previously, these

---

[14]Again, this is assuming that the "no opinion" response option is provided when respondents do not want to exert much effort. It is possible, however, that respondents select the "no opinion" response option only after careful contemplation. In this case the "no opinion" responses will likely be reliable and as such should not be used as an indicator of shirking.

characteristics can be described as respondent ability, respondent motivation, and task difficulty. These factors are related to respondents optimal effort, e_{iq}^{r*}, through the cost and benefit functions, $C(\cdot)$ and $B(\cdot)$. Respondent ability will affect the cost function; more able respondents will have lower survey participation costs and hence will be more likely to exert enough effort to provide reliable responses. Motivation, on the other hand, will affect the benefit function; respondents who are more motivated will obtain larger benefits from survey participation. Finally, both respondent ability and task difficulty will affect the amount of effort that would be required to provide a reliable response, e_{iq}^{s*}; able respondents faced with relatively easy questions will not have to exert much effort in order to provide reliable responses.

To understand how ability, motivation, and difficulty might affect survey response behavior, it is useful to define and describe these factors. With this understanding, determining appropriate measures for ability, motivation, and difficulty will be possible. Given these measures, empirical analysis of shirking behavior can be undertaken.

5.6.1 Defining Ability, Motivation, and Difficulty

Respondent ability concerns more than intelligence. It refers more broadly to how adept a respondent is at performing the four cognitive steps involved in answering survey questions (Krosnick 1991). The first component of ability is simple language skills; respondents able to understand the question and response options are more likely to provide accurate responses. Another component of ability involves the capacity to retrieve information and make judgments. Respondents adept at performing complex mental operations, i.e., respondents who are cognitively sophisticated, will be better able to retrieve and judge relevant information and ideas.[15] The amount of practice a respondent has at thinking about the topic under consideration will also affect his ability to answer the survey questions—the more an individual has thought about a topic, the easier it should be to retrieve the relevant considerations and express an opinion (Zaller 1992). Furthermore, the more crystalized a respon-

---

[15]The term cognitive sophistication is used prominently in the survey methodology literature. See, for example, Krosnick & Alwin (1987) and Schuman & Presser (1981).

dent's attitude towards the question the easier the cognitive process should be (Fazio & Williams 1986).

The next component determining how respondents behave when answering survey questions is motivation. Motivation is likely to result from an interaction of survey and individual characteristics. Respondents who enjoy confronting difficult cognitive tasks, for whom the topic is of interest, and who have a positive relationship with the interviewer or researcher are more likely to be motivated (Krosnick 1991). One way of viewing the components that affect respondent motivation is in terms of the following relationships: relationship to interview, relationship to interviewer, and relationship to self (Brehm 1993).[16]

Many features of the survey are likely to affect motivation. In determining how much effort to exert, respondents are likely to consider whether questions are useful and relevant or a waste of time—questions that are important to respondents will probably encourage more effort. The way in which interviewers ask survey questions may also affect motivation. For example, allowing extra time to formulate answers can influence the respondent's level of engagement in the survey and hence affect their motivation (Green, Krosnick & Holbrook 2002). Issues about the credibility of the survey, confidentiality of information, and anonymity can also influence respondents' motivation levels.

In addition, respondents will be motivated by factors internal to themselves, such as a need for self-expression, a sense of civic duty, or a desire for intellectual challenge. The Roper Organization Poll on Polls found that people who reported having been interviewed previously were more active in politics (Roper 1986). Respondents who reluctantly answered surveys also tended to be less politically interested and informed (Brehm 1993). Individuals who perceive themselves as helpful may also be more motivated to exert significant effort when answering survey questions—about fourteen percent of the 1989 Detroit Area Study (DAS) respondents indicated that they agreed

---

[16]Brehm (1993) discusses these factors in relation to survey response rates but, as mentioned previously, factors that affect the participation decision will influence behavior during the response stage.

to be interviewed because they considered themselves helpful. Brehm (1993) also found that respondents who did not have enough time, who were not interested in the survey, or who had negative previous experiences were reluctant to participate in surveys.

In his re-examination of the NES data used by Converse (1964), Achen (1975, p. 1229) concluded that response instability was "primarily the fault of the instruments."[17] Within the shirking framework, this problem is captured by the difficulty of the survey questions—the final factor that affects response behavior. Since the evidence from the in-depth interviews reviewed in Chapter 2 suggests that different respondents will find different questions difficult, a subjective measure of question difficulty may be most appropriate.

5.6.2 Measuring Ability, Motivation, and Difficulty

Having defined ability, motivation, and difficulty, the next task is to determine appropriate means of measuring these factors. As suggested by Green, Krosnick & Holbrook (2002) and Krosnick (1991), individual traits, such as educational attainment, political informedness, party membership, and level of party identification, provide useful proxies for respondent ability and motivation. Establishing appropriate measures of difficulty, however, is particularly complicated because difficulty tends to be extremely subjective. The search for appropriate measures of difficulty are further hampered by the fact that most surveys only supplement opinion questions with basic demographic questions. However, using newly collected data from third parties or survey participants, Chapter 7 explores alternative ways of measuring difficulty.

As discussed above, ability includes such things as cognitive sophistication, knowledge about political issues, and opinion crystallization. In examining cognitive sophistication, Krosnick & Alwin (1987) conclude that educational attainment is a good proxy. Since respondents' ability to answer factual political information questions,

---

[17]Achen (1975) proposed that opinion expressions are actually distributions around a central point—participants may answer the same question differently even if their underlying beliefs do not change because questions are vague.

called political informedness, has been shown to be highly correlated with awareness about political issues (Zaller 1992), measures of political informedness will also be used to capture ability.[18]

Motivation, on the other hand, will be affected by the interaction of respondent and survey characteristics. For example, political informedness, which captures respondent ability, is also an indicator of political involvement and, as such, offers some insight into respondent motivation; topic specific knowledge provides a similar measure of political involvement at the narrower topic specific level. Motivation might also be captured by other measures of a respondent's political involvement, including party membership and level of party identification. Politically disenfranchised people, such as racial minorities, women, and low income families, might also have different levels of motivation.[19] From the survey design perspective, issues such as the length of the survey and the relevance of the issues under consideration will affect respondent motivation. By combining these factors, it will be possible to test which characteristics of respondents affect shirking in surveys.

Until recently difficulty has not been conceptualized from an individual perspective. Most of the work on how task difficulty affects respondent behavior has focused on how different survey presentations affect the general behavior of respondents—the analysis of different survey modes (i.e., face-to-face, telephone, and self-completion) is a prominent example of this type of analysis (Green, Krosnick & Holbrook 2002). Unfortunately, question level difficulty measures are scarce and subjective measure of difficulty are even less available. To obtain subjective measures of difficulty, recent surveys have asked respondents to indicate how difficult they found certain questions to answer. This provides a new avenue through which to examine difficulty (Albertson, Brehm & Alvarez Forthcoming, 2004). Since this type of data is not

---

[18]Political opinion surveys often include factual knowledge questions designed to establish how familiar respondents are with politics.

[19]The direction of this effect is unclear since those who do not have much political influence through the standard political venues might view surveys as an alternative means of affecting political outcomes or they might simply accept their political marginalization and therefore have very little interest in politics.

available for most surveys, similar difficulty measures may be collected from auxiliary or third-party samples.

Assuming appropriate indicators of shirking can be identified, it is possible to test how ability, motivation, and difficult are related to the probability of shirking. The next two chapters use data collected in the 1996–1997 Indianapolis–St. Louis Election Study (Indianapolis 1996), the 1980 and 1984 National Elections Studies (NES) Pre-Election Surveys, and the 1998–1999 Multi-Investigator Study (MI1998) to explore different means of identifying shirking. Once shirking is identified, factors that influence this behavior can be examined. In Chapter 7, when response times are used to identify shirking, the impact that this type of shirking behavior has on analytic results is explored. In addition, different approaches to accounting for shirking behavior in empirical analysis are considered. Finally, the comparison in Chapter 8, of shirking behavior across survey modes, using data from the Caltech Internet Survey Project, provides another look at how different survey design factors might affect shirking behavior. This work culminates with a discussion the implications that observation of shirking behavior has for the design of surveys and for the analysis of survey data.

Chapter 6   Two Behaviors that Might be Indicative of Shirking

> If you choose not to decide, you still have made a choice (Rush 1990).

As suggested above, two alternative indicators of shirking are explored: respondents who provide incorrect answers quickly on factual questions and the proportion of "no opinion" responses. Using these measures, it is feasible to test the effect of respondent characteristics on shirking behavior but measuring difficulty at the survey level is complicated unless the survey was administered using multiple modes.[1]

Question level difficulty measures are explored as a means of explaining "no opinion" responding. Because difficulty measures were not collected in the 1980 and 1984 NES Pre-Election Surveys, auxiliary or third-party difficulty data was collected to supplement the existing data. These difficulty reports are explored below as is the correlation between these difficulty measures and "no opinion" response behavior. Unfortunately, these third-party difficulty reports do not appear to be very closely correlated with "no opinion" response rates. Hence, attention is turned to "no opinion" response behavior in the 1998–1999 Multi-Investigator Study (MI1998); the MI1998 collected respondent-reported difficulty measures for six policy questions. Aggregating respondent-reported difficulty across questions provides a measure of subjective survey difficulty that may be implemented with the survey level data from the MI1998.

6.1   The Quick and the Wrong

In many situations it may be impossible to identify shirking at the question level. If there is insufficient data or if it is impossible to determine what question level behavior

---

[1]Chapter 8 explores the differences in shirking behavior across telephone and Internet survey modes.

is associated as shirking, it may be easier to examine shirking at the survey level. Hence the first approach to identifying shirking explored here involves examining respondent behavior on factual question—respondents who answer factual questions quickly and provide incorrect responses will be identified with shirking. Assuming that response time is positively correlated with effort, that with enough effort a respondent would be able to provide accurate responses to factual questions, and that behavior on factual questions is indicative of overall survey response behavior, this "quick-wrong" measure provides a simple way to identify shirkers at the survey level. All that is needed is a survey that collected response times for factual questions.

Fortunately the 1996–1997 Indianapolis–St. Louis Election Study (Indianapolis 1996) collected extensive response time data.[2] In particular, latent timers, measured in hundredths of seconds, were collected for the three following political knowledge questions: Whose responsibility is it to determine if a law is constitutional or not? What are the first 10 amendments to the constitution called? How much of a majority is required for the U.S. Senate and House to override a presidential veto?[3]

The raw data indicated latent response times between zero seconds and over three minutes but, as mentioned above, many problems may occur with the coding of response times. Huckfeldt, Sprague & Levine (2000) recommend excluding from the analysis respondents with response times of less than one second (the minimum time that would be required to read the question) and more than three standard deviations above the sample mean.[4] Having considered several different transformations of response times, it appears that the distribution of the log of response times most closely resembles a normal distribution.[5]

---

[2]See Mulligan et al. (2003) for a discussion of latent and active response timers.

[3]See Appendix A.4 for a description of the data used from the Indianapolis–St. Louis Election Study, 1996–1997.

[4]This trimming process amounts to excluding 383 respondents, or fifteen percent of the sample, from the analysis. As long as timer inaccuracies are not correlated with responses, dropping respondents with inaccurate timer data should not affect consistency but may affect efficiency.

[5]Transformation considered include: the cubic, identify, square, square-root, natural log, inverse, inverse of the square, inverse of the cubic, inverse of the square-root.

Figure 6.1 contains distributional graphs for the log of the trimmed response times for three political knowledge questions.[6] Even after trimming of outliers, the response time data appears to be slightly skewed to the right but perhaps this can be explained by examining how response times correspond with accuracy.[7]
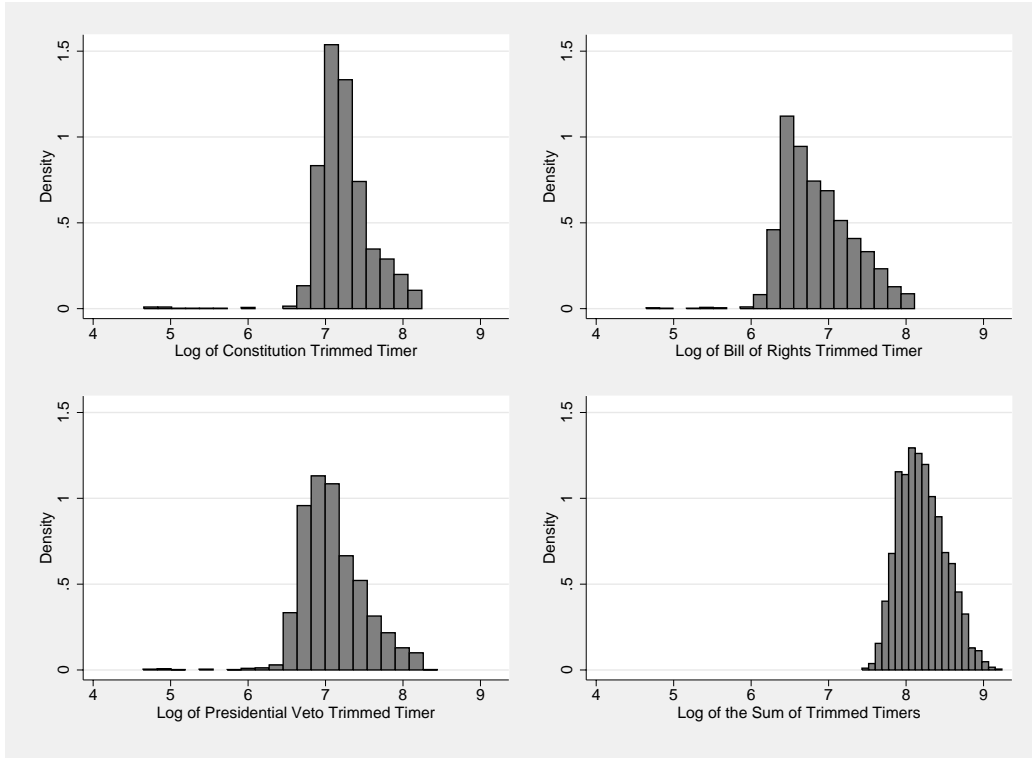
Figure 6.1: Histograms of Trimmed Latent Response Times for Factual Questions

6.1.1 The "quick-wrong" Indicator of Shirking

As indicated above, the Indianapolis 1996 survey contained three factual questions for which response times were obtained. There are a variety of ways that the "quick-wrong" indicator of shirking can be developed from this data. In an attempt to

---

[6]This analysis was also performed on the untrimmed data and the data with zeros recoded as 0.5 100^{th}s of a second. The results were very similar to those presented here.

[7]The lack of symmetry of the timers suggest that perhaps timers are not linearly related to effort. Several different transformations or response times have been considered and it appears that the log of response times has a distribution that more closely matches the normal distribution.

obtain a reasonable indication of respondents who answered these factual questions incorrectly, all respondents who were unable to answer at least one question correctly were classified as having incorrectly answered these questions.[8]

To identify the appropriate indicator of a quick responses, it is useful to examine the distribution of the log of the trimmed response times summed across all three factual questions. As indicated by the lower right histogram in Figure 6.1, these total responses times for all three factual questions range from 16.87 to 94.25 seconds (corresponding to logged values of 7.43 and 9.15). Because the variance of this distribution is relatively low, all responses time that were less than one standard deviation below the mean of the log of the trimmed sum of response times were identified as quick responses.[9]

Combining the above measures of quick and wrong, thirty-two respondents, or 1.47 percent of the sample, are identified as shirkers according to the "quick-wrong" measure of shirking. Table 6.1 indicates that respondents who are female and have a low level of education are more likely to be identified as shirking based on the "quick-wrong" measure. Republican party membership and long response times in answering demographic questions, on the other hand, are associated with statistically significantly lower probabilities of being identified as a "quick-wrong" shirker.

---

[8]Only 341 respondents got the wrong answer on all the factual questions. Combining this measure of wrong with alternative measures of quick identifies only between one and four respondents as "quick-wrong" shirkers.

[9]Other measures of quick, such as 1.96 standard deviations below the mean, were explored but they identified as few as fifteen respondents as quick.

Table 6.1: Logit Analysis of the "quick-wrong" Identifier of Shirking

	Coef.	Std. Err.	P-Value
Female	0.82[+]	0.44	0.06
Working	-0.02	0.38	0.96
College or Beyond Education Level	-1.68[*]	0.44	0.00
White	-0.23	0.45	0.61
Log of Time to Answer Demographic Questions	-0.51[+]	0.27	0.06
Interviewer Reported Resp. Knowledge	0.00[*]	0.00	0.01
Interested in Politics	0.12	0.12	0.31
Democrat	0.04	0.42	0.93
Republican	-1.11[+]	0.61	0.07
Constant	4.06	2.69	0.13
Number of Observations	2139		
Pseudo R-Squared	0.14		

[*] Statistically significantly different from zero at the $p \leq 0.05$ level.
[+] Statistically significantly different from zero at the $p \leq 0.10$ level.

6.1.2 Substantive Behavior of the "quick-wrong" Shirkers

Given the "quick-wrong" indicator of shirking, it is possible to explore how the behavior of these shirkers differs from their non-shirking counterparts. Table 6.2 contains a comparison of responses to a question asking respondents whether they favor or oppose the death penalty. According to this table, "quick-wrong" shirkers were more likely to provide neutral responses, strongly favor, and "no opinion" responses (don't know or refused). According to the Pearson Chi-Squared test comparing the response distribution between the "quick-wrong" and everyone, the null hypothesis that the distribution of responses is the same can be rejected. The distribution of responses provided by shirkers does appear to agree with the standard assumptions that shirkers are more likely to provide neutral responses or "no opinion" responses (Krosnick 1991).

This "quick-wrong" measure of shirking suggests that shirking is not a significant problem. But it is possible that this measure does not accurately identify respondents who are not exerting enough effort on the survey as a whole. It is possible that the

assumptions underlying this measure of shirking are not satisfied.[10] To take another approach in identifying shirking, the next section explores an alternative measure of shirking—"no opinion" responding.

Table 6.2: Comparing Support for the Death Penalty for "quick-wrong" Shirkers

	Percentages	
	"quick-wrong" Shirkers	Not "quick-wrong" Shirkers
Strongly Opposed	0	14.36
Opposed	0.00	8.30
Neutral	25.00	16.90
Favor	15.00	16.75
Strongly Favor	45.00	39.42
Don't Know	10.00	3.66
Refused	5.00	0.60

6.2 No Opinion Responding

Before embarking on an examination of shirking behavior as identified by "no opinion" responding, it is useful to consider the prevalence of this behavior. Using this measure of shirking suggests that shirking behavior is fairly prominent in certain types of questions. Focusing attention solely on opinion questions administered to all 1980 and 1984 NES Pre-Election Survey respondents, different levels of "no opinion" responses are observed both across questions and respondents. As mentioned in Section 5.5.2, the 1980 NES Pre-Election Survey, had "no opinion" responses rates ranging from less than one percent to over forty-two percent on different questions. There are even higher amounts of variation across individuals; in the 1980 NES Pre-Election Survey, some respondents provided opinions on all the questions while one respondent provided "no opinion" responses in ninety-two percent of the questions.

Figures 6.2 and 6.3 demonstrate how "no opinion" responses vary during the course of both the 1980 and 1984 NES Pre-Election Surveys. The height of each bar

---

[10]As mentioned previously, it is unfortunate that there no data available to validate the assumptions used to enable the identification of shirking.
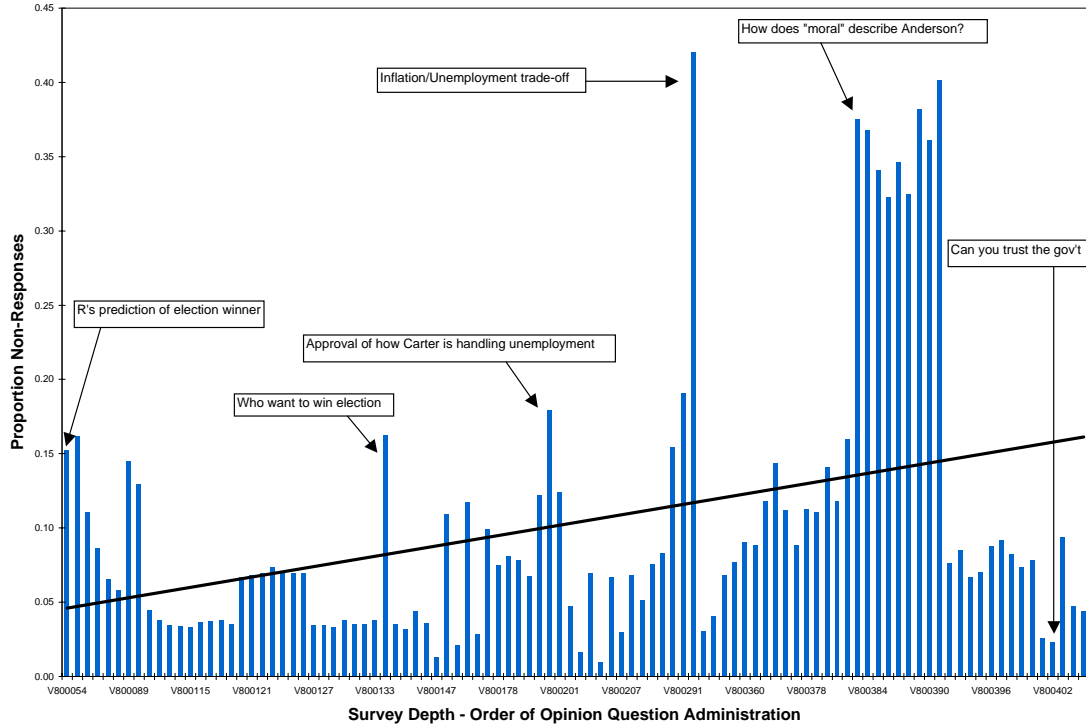
Figure 6.2: No Opinion Responses—NES 1980 Pre-Election Survey

indicates the fraction of respondents providing "no opinion" responses to each opinion question and questions are ordered as they were presented in the survey.[11] Thus the bars on the left end of the graph present the fraction of respondents who provided "no opinion" responses to opinion questions presented early in the survey while the right end contains "no opinion" response rates for opinion questions asked towards the end of the survey.[12]

---

[11]The 1980 and 1984 NES Pre-Election Surveys contained a mixture of opinion and factual knowledge question. Only opinion questions administered to all respondents are included in this graph. Thermometer score questions that ask respondents to rate their feelings towards various political figures and organization have been excluded from this analysis since the response options provided in these questions is very different from that of the other close-ended questions which usually have a maximum of 7 response options. A few questions that appear to be outliers in terms of "no opinion" response rate are noted on these tables. See the NES 1980 Codebook for a full description of these variables and their corresponding survey questions.

[12]On average, the 1980 and 1984 NES Pre-Election Surveys took over an hour to complete.
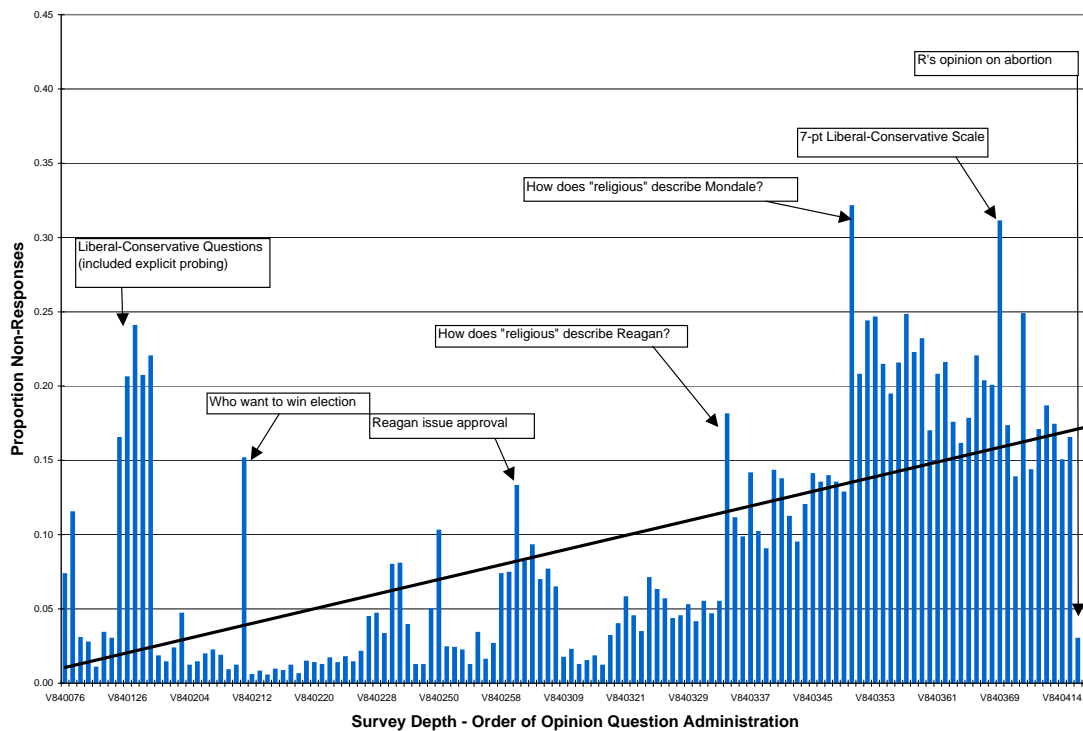
Figure 6.3: No Opinion Responses—NES 1984 Pre-Election Survey

The trend line in both figures suggests that "no opinion" response rates are increasing as the survey progresses.[13] These results are consistent with those found in other research that considered "no opinion" response rates or item non-response as a function the question position.[14] Unfortunately, the order of questions in these surveys was not randomized; certain types of questions appear at different stages of the survey. In fact, if the questions pertaining to character traits that describe John Anderson, an independent presidential candidate in the 1980 election, are excluded the apparent trend of increasing "no opinion" response disappears for the 1980 NES Pre-Election Survey.[15]

Figures 6.2 and 6.3 highlight questions for which the "no opinion" response rates are striking. For example, a relatively high number of respondents provided "no opinion"answers to questions asking how different traits, such as religiosity, describe the various presidential candidates; this is particularly true for candidates, such as Anderson and Mondale, who are not as well known. The 1984 NES Pre-Election Survey also has high "no opinion" response rates for questions asking respondents to place themselves and other political figures or organizations on a liberal-conservative scale. This occurs despite the fact that these questions are asked early in the survey and interviewers were required to probe for more information from respondents who refused to place themselves on this scale. In the 1980 NES Pre-Election Survey, the question asking respondents to indicate how they feel about the unemployment-inflation trade-off also received an exceptionally high number of "no opinion" responses. However, it should be noted that this question involved a seven point response scale and explicitly offered respondents the opportunity to indicate that they had "not thought much" about this issue. Questions concerning respondent support for presidential candidates

---

[13]The slope of the trend line for the 1980 NES Pre-Election Survey non-responses is 0.001 (Standard Deviation = 0.0003). For the 1984 data the slope of the trend line is 0.001 (Standard Deviation = 0.0001).

[14]see, for example,, Dickinson & Kirzner (1985).

[15]If the questions about Anderson are excluded from the analysis of the 1980 NES data, the slope is 0.0002 (Standard Deviation = 0.0002) which is not statistically significantly different from 0 at conventional levels.

and approval ratings of the current president also tend to display relatively high levels of "no opinion" responses.

Of equal note are the survey questions towards the end of the survey that received a very low proportion of "no opinion" responses. In the 1980 NES survey, for example, most respondents provided opinions to questions about their feelings concerning the government. The lowest "no opinion" response rate in this survey was for a question asking respondents how often they thought the government in Washington could be trusted to do what was right. Similarly, despite the fact a question concerning abortion policy was placed near the end of the 1984 NES survey, only three percent of respondents provided a "no opinion" response. This suggests that the position of a question in a survey is not the only factor that will affect the frequency of "no opinion" responses.

These patterns support the hypothesis that question characteristics affect respondent behavior. An attempt to examine how these factors affect "no opinion" responding is contained in Section 6.2.4. However, before trying to undertake this analysis, it is useful to consider how respondent characteristics affect survey level measures of "no opinion" responding. As mentioned in Section 5.5.2, one respondent provided non-opinions answers in ninety-two percent of the questions while, in the 1984 survey, one respondent indicated that he had "no opinion" on over ninety-eight percent of the opinion questions. This strongly supports the hypothesis that respondent characteristics affect shirking as identified by "no opinion" responding.

6.2.1 Proportion of No Opinion Responses as an Indicator of Shirking

Before beginning the analysis of respondent characteristics and their effects on shirking, it is useful to consider the fact that individual shirking behavior is observed across multiple questions. In this analysis, the proportion of questions for which each respondent provided a "no opinion" response will be used as an indicator of

shirking—the ith ($i \in N$) respondent's proportion of "no opinion" responses to the Q opinion questions is denoted by P_i.

Since the observed proportions of "no opinion" response, $P_1, ..., P_N$, provide an estimate of the population distribution, this problem can be treated as a simple one of sampling from a Bernoulli population.[16] Thus,

$$P_i = G(X_i'\zeta) + \varepsilon_i, \tag{6.1}$$

where $G(\cdot)$ is the cumulative distribution function of ϵ_{iq}^{rs} in Equation (5.2) and ε_i is the measurement error between the observed P_i and the true distribution at $X_i'\zeta$.[17]

Since the cumulative density function $G(X_i'\zeta)$ is strictly monotonic, it has an inverse. Let $g(t) = G'(t) \forall t \in \mathbb{R}$. Taking the inverse and a Taylor series expansion around $t = G(X_i'\zeta)$,

$$G^{-1}(P_i) = G^{-1}(G(X_i'\zeta) + \varepsilon_i) \approx X_i'\zeta + \frac{\varepsilon_i}{g(X_i'\zeta)}. \tag{6.2}$$

Recalling that Q is the number of questions respondent i is asked, this produces a heteroscedastic linear regression,

$$G^{-1}(P_i) = X_i'\zeta + u_i, \tag{6.3}$$

where

$$E[u_i|X_i] = 0 \quad \text{and} \quad Var[u_i|X_i] = \frac{G(X_i'\zeta)(1 - G(X_i'\zeta))}{Qg^2(X_i'\zeta)}. \tag{6.4}$$

If the distribution G is assumed to be logistic, then the log-odds are linear in the parameters,

$$\ln\left(\frac{P_i}{1 - P_i}\right) = X_i'\zeta. \tag{6.5}$$

---

[16]See Greene (2000, pp. 834–836) for a thorough development of the ordered logit and weighted least-squares regression on log-odds ratios.

[17]Since ε_i measures the error for individual i across several questions it will be independent of ε_j which is the error of individual j.

Letting $\Lambda_i = \frac{e^{X_i'\zeta}}{1+e^{X_i'\zeta}}$, Equation (6.3) can be estimated by a weighted least-squares regression on the log-odds ratios with the weights

$$w_i = Q\Lambda_i(1 - \Lambda_i). \tag{6.6}$$

6.2.2 Empirical Analysis using the No Opinion Measure of Shirking

As discussed in Section 5.6, the traits that affect shirking may be classified as ability, motivation, and difficulty. Respondent characteristics, including educational attainment, political informedness, party membership, and level of party identification provide proxies for ability and motivation. Incorporating these characteristics into the analysis of shirking behavior enables tests of the factors that are likely to influence this behavior.

The 1980 and 1984 NES surveys collected a significant amount of demographic and behavioral information about respondents. In addition to typical demographic variable such as region of residence, education level, employment status, religion, gender, age, and income, the NES surveys collected information about respondents' political leanings, their party membership, their levels of interest in politics, and their attentiveness to media coverage of government issues and the presidential campaign.[18] These individual characteristics are captured by X_i in Equation (5.2).

Tables 6.3 and 6.4 indicate the proportion of "no opinion" responses in the 1980 and 1984 NES Pre-Election Surveys. The bold rows in both tables contain the overall average proportion of "no opinion" responses. Subsequent rows consider subsets of the full population.

---

[18]See Appendix A.2 for a description of the NES variables included in this analysis.

66

Table 6.3: No Opinion Responding in the 1980 NES Pre-Election Survey

Demographic Group[†]	Average Non-Response Rate	Standard Deviation	Sample Size
Full Sample	**0.10**	**0.15**	**1614**
Low Education	0.13	0.15	1018
High Education	0.07	0.13	592
Low Poll Attentiveness	0.47	0.28	76
High Poll Attentiveness	0.14	0.14	484
Low Candidate Awareness	0.12	0.16	1229
High Candidate Awareness	0.07	0.10	385
Low TV Attention to Gov't	0.31	0.30	148
High TV Attention to Gov't	0.08	0.10	1466
Not Interested in Politics	0.10	0.12	1099
Interested in Politics	0.06	0.08	466

[†] See Appendix A.2 for a description of the variables used in this analysis.

Table 6.4: No Opinion Responding in the 1984 NES Pre-Election Survey

Demographic Group[†]	Average Non-Response Rate	Standard Deviation	Sample Size
Full Sample	**0.09**	**0.14**	**2257**
Low Education	0.12	0.15	1315
High Education	0.05	0.07	928
Low Candidate Awareness	0.11	0.15	1798
High Candidate Awareness	0.03	0.06	459
Don't Discuss Politics Frequently	0.09	0.12	511
Discuss Politics Frequently	0.05	0.08	999
Low TV Attention to Gov't	0.10	0.14	927
High TV Attention to Gov't	0.05	0.09	914
Not Interested in Politics	0.11	0.14	1612
Interested in Politics	0.05	0.10	638

[†] See Appendix A.2 for a description of the variables used in this analysis.

From these tables, it appears that demographic and behavioral characteristics affect the likelihood of providing "no opinion" responses. Less educated, less politically informed, and less politically attentive respondents select the "no opinion" response options more often than their highly educated, very politically informed, and politically attentive counterparts. These averages, however, provide only a partial picture.

To test which factors affect "no opinion" responding, it is useful to undertake a multivariate analysis.

The best approach to testing which variables affect respondent behavior is to regress the proportion of "no opinion" responses provided by the survey participants, P_i, on the characteristics that affect respondent behavior. The naive approach, and that employed in much of the survey response literature, involves running an ordinary least-squares (OLS) regression of the proportion of "no opinion" responses on the demographic characteristics (Green, Krosnick & Holbrook 2002). But, if the error terms have Type II extreme value distributions, as outlined above, the shirking model implies a weighted least-squares regression of the log-odds ratio on the exogenous variables.[19] Thus, if a weighted least squares regression of the log-odds is used, the estimates will be consistent. However, if the relationship between the proportion of "no opinion" responses and ability, motivation and difficulty is estimated using an OLS regression, the model is misspecified and the estimated coefficients will biased.[20]

Tables 6.5 and 6.7 contain the results for both a simple OLS regression and a weighted least-squares regression of the log-odds ratio.[21] In the OLS analysis, the dependent variable is the proportion of times that a respondent provided a "no opinion" response on the opinion questions that were administered to every survey participant. Since these models have different dependent variables, the estimated coefficients are not comparable. Instead, to compare the results from these two models, it is useful to examine the first differences, calculated as the change in predicted probability given a change in one of the explanatory variables. These first differences are provided in Table 6.6 and 6.8.

---

[19]The odds-ratio is calculated as $ln(\frac{P_i}{1-P_i})$ where P_i is the proportion of times respondent i provided "no opinion" responses on the Q questions under consideration.

[20]See Green, Krosnick & Holbrook (2002) for an example in which an OLS analysis of "no opinion" response behavior is undertaken.

[21]Probit, Poisson, and negative binomial models were also estimated for comparison. The results were very similar across these other models thus they are not included in this discussion.

Table 6.5: Multivariate Analysis of No Opinion Responding in the 1980 NES

"no opinion"Prop.	OLS Results		WLS Results[†]	
	Coefficient	Standard Error	Coefficient	Standard Error
South	0.03*	0.01	0.23*	0.07
West	0.00	0.01	-0.07	0.09
North Central	0.01+	0.01	0.16*	0.08
High Education	-0.04*	0.01	-0.50*	0.07
Female	0.02*	0.00	0.27*	0.05
White	-0.02*	0.01	-0.17*	0.07
Over 75 Years Old	0.08*	0.01	0.61*	0.081
Have Child(ren)	0.00	0.01	-0.13*	0.06
Republican	-0.01+	0.01	-0.19*	0.08
Democrat	-0.01	0.01	-0.06	0.06
Other Party Member	0.08+	0.05	0.23	0.32
Strength of Party Affiliation	0.00	0.01	-0.05	0.07
Candidate Awareness	-0.01*	0.01	-0.26*	0.07
TV Attention to Gov't	-0.03*	0.01	-0.27*	0.06
Social Class	0.00◇	0.00	-0.04*	0.01
Interested in Politics	-0.01*	0.01	-0.16*	0.07
Constant	0.12*	0.01	-1.52*	
Number of Observations	1413		1413	
Adjusted R-squared	0.19		0.20	

[†] Since the log of zero is undefined, $P_i = 0$ has been replaced by $P_i = 0.0001$.

* Statistically significantly different from zero at the $p \leq 0.05$ level.

◇ This coefficient is actually equal to -.004 and the standard error is .001 thus this coefficient is actually significantly different from zero at the $p \leq 0.05$ level.

+ Statistically significantly different from zero at the $p \leq 0.10$ level.

Table 6.6: First Differences from the No Opinion Responding Analysis of the 1980 NES

"no opinion"Prop.	OLS Results		WLS Results[†]	
	Mean[‡]	Standard Error	Mean[‡]	Standard Error
South	0.026*	0.007	0.035*	0.011
West	0.004	0.008	-0.010	0.016
North Central	0.013	0.007	0.023	0.012
High Education	-0.037*	0.006	-0.063*	0.009
Female	0.023*	0.005	0.043*	0.009
White	-0.022*	0.008	-0.024*	0.011
Over 75 Years Old	0.084*	0.010	0.109*	0.019
Have Child(ren)	0.000	0.005	-0.018	0.009
Republican	-0.014*	0.007	-0.026*	0.010
Democrat	-0.008	0.006	-0.008	0.010
Other Party Member	0.076	0.046	0.040	0.054
Strength of Party Affiliation	-0.005	0.006	-0.007	0.010
Candidate Awareness	-0.012*	0.006	-0.036*	0.009
TV Attention to Gov't	-0.026*	0.005	-0.036*	0.008
Social Class	-0.033*	0.011	-0.044*	0.016
Interested in Politics	-0.014*	0.006	-0.022*	0.009
Constant	.		.	
Number of Observations	1413		1413	
Adjusted R-squared	0.19		0.20	

[†] Since the log of zero is undefined, $P_i = 0$ has been replaced by $P_i = 0.0001$.

[‡] First differences are calculated as the difference between the predicted probability if all the explanatory variables set at their minimum and the predicted probability if the variable under consideration is increased to its maximum all else the same. Standard errors of the first differences are estimated using Monte Carlo simulation techniques similar to those suggested by King, Tomz & Wittenberg (1999) and King, Tomz & Wittenberg (2000).

[*] Statistically significantly different from zero at the $p \leq 0.05$ level.

[+] Statistically significantly different from zero at the $p \leq 0.10$ level.

Table 6.7: Multivariate Analysis of No Opinion Responding in the 1984 NES

No Opinion Prop.	OLS Results		WLS Results[†]	
	Coefficient	Standard Error	Coefficient	Standard Error
South	0.02*	0.01	0.22*	0.07
West	-0.01	0.01	0.06	0.08
North Central	-0.01	0.01-0.03	0.07	
High Education	-0.04*	0.01	-0.49*	0.06
Female	0.02*	0.00	0.20*	0.05
White	-0.03*	0.01	-0.35*	0.07
Over 75 Years Old	0.08*	0.01	0.54*	0.08
Social Class	0.00	0.00	0.00	0.01
Child(ren) in House	-0.01*	0.00	-0.21*	0.05
Republican	-0.02*	0.01	-0.13+	0.07
Democrat	-0.02*	0.01	-0.14*	0.06
Other Party Member	0.00	0.06	-0.46	0.56
Strength of Party Affiliation	-0.01	0.01	-0.12+	0.06
Candidate Awareness	-0.03*	0.01	-0.54*	0.07
TV Attention to Gov't	-0.02*	0.01	-0.27*	0.06
Interested in Politics	-0.02*	0.01	-0.23*	0.07
Constant	0.14*	0.01	-1.23	0.10
Number of Observations	2054		2054	
Adjusted R-squared	0.18		0.18	

[†] Since the log of zero is undefined, $P_i = 0$ has been replaced by $P_i = 0.0001$.
* Statistically significantly different from zero at the $p \leq 0.05$ level.
+ Statistically significantly different from zero at the $p \leq 0.10$ level.

Table 6.8: First Differences from the No Opinion Responding Analysis of the 1984 NES

No Opinion Prop.	OLS Results		WLS Results[†]	
	Mean [†]	Standard Error	Mean[‡]	Standard Error
South	0.016	0.007	0.040	0.013
West	-0.006	0.007	0.011	0.018
North Central	-0.010	0.007	-0.006	0.013
High Education	-0.044	0.005	-0.074	0.009
Female	0.022	0.005	0.037	0.009
White	-0.033	0.007	-0.056	0.012
Over 75 Years Old	0.080	0.009	0.108	0.018
Child(ren) in House	-0.011	0.005	0.001	0.003
Republican	-0.017	0.006	-0.033	0.009
Democrat	-0.018	0.006	-0.021	0.012
Other Party Member	-0.002	0.059	-0.024	0.010
Strength of Party Affiliation	-0.006	0.006	-0.056	0.075
Candidate Informedness	-0.030	0.006	-0.020	0.011
TV Attention to Gov't	-0.020	0.005	-0.080	0.011
Social Class	0.005	0.011	-0.191	0.024
Interested in Politics	-0.018	0.006	-0.037	0.011
Constant	.		.	
Number of Observations	2054		2054	
Adjusted R-squared	0.18		0.18	

[†] Since the log of zero is undefined, $P_i = 0$ has been replaced by $P_i = 0.0001$.

[‡] First differences are calculated as $G(X_j^1, X_{-j}^0, \hat{\beta} - G(X^0, \hat{\beta}$ where $G(X^0, \hat{\beta}$ is the predicted probability from the model with all the explanatory variables set at their minimum and $G(X_j^1, X_{-j}^0, \hat{\beta}$ is the predicted probability if X_j, the jth of the K explanatory variables, is increased to its maximum while the other $K - 1$ variables, X_{-j}, remain at their minimum.

[*] Statistically significantly different from zero at the $p \leq 0.05$ level.

[+] Statistically significantly different from zero at the $p \leq 0.10$ level.

From the models in Tables 6.5 and 6.7, it appears that respondents who are from the south, female and over 75 years old are more likely to provide "no opinion" responses. Both models also indicate that the more educated, politically active, and politically informed the less likely the respondent is to shirk. This supports the hypothesis that the more able (i.e., highly educated) and motivated (i.e., politically involved and interested) the respondent the less likely they are to shirk. Somewhat surprisingly, strength of party affiliation only appears marginally significant in the WLS regression on the 1984 NES data; this supports the claims by Zaller (1992) that

the best measures of political awareness are factual knowledge questions like those used to create the candidate awareness measure used in this analysis.

Although the two models indicate similar signs and levels of significance there are some notable differences. For example, in the 1980 NES sample, having children was not significant according to the OLS regression but was significant in the WLS model. Looking at the first differences, however, indicates that the two models have quite different predictions about shirking behavior. According to the WLS results, most of the variables have significantly larger impacts on the probability of shirking than those predicted by the OLS results. For example, according to the WLS results, increasing the level of political informedness from zero (unable to correctly identify any of the local congressional candidates) to one (able to correctly identify at least one candidate and the candidates' party) reduces the probability of shirking by 0.08 in the 1984 NES data—this is more than double the effect estimated by the OLS regression. In the 1980 NES data, the effect of candidate awareness as estimated by the WLS model is three times larger than that of the OLS model.

While these results support the idea that ability and motivation influence the probability of shirking in the form of "no opinion" responding in a predictable fashion, it is important to note that the maximum adjusted R-squared across all these models is 0.2. The R-squared results from the OLS regressions correspond with the following correlations between predicted and actual probabilities of shirking: 0.44 (1980) and 0.43 (1984). This compares with correlations of 0.45 (1980) and 0.45 (1984) for the weighted least-squares predictions. Thus the weighted least-squares model appears to be marginally better at predicting the probability of shirking.[22] But these relatively low levels of correlation suggest that respondent characteristics alone cannot fully explain "no opinion" response behavior.

---

[22]Also, the WLS estimates will ensure that $\hat{P}_i \in [0.1]$ which is not guaranteed in a standard OLS model.

6.2.3 Measuring Question Difficulty

As discussed in Section 5.6.2 shirking is likely to be affected by respondent as well as question characteristics such as question difficulty. Having concluded above that respondent characteristics alone cannot explain shirking behavior in the form of "no opinion" responding, it is useful to consider how one might obtain measures of question difficulty to be incorporated in the analysis of "no opinion" responding.

According to the discussion in Chapter 3, task difficulty will be influenced by how questions are worded. However, much of the existing analysis involving task difficulty has focused on survey level measures, such as survey mode, to capture difficulty since there is relatively little data about difficulty at the question level.[23] The recent use of questions asking respondents how difficult they found other questions provides an innovative way to measure difficulty at the question level. As mentioned previously, these direct measures attempt to capture respondents' perceptions of task difficulty which will be useful since the question difficulty is likely to be subjective (Albertson, Brehm & Alvarez Forthcoming, 2004). Unfortunately, these types of questions were not included in the 1980 and 1984 NES surveys. Instead, for the 1980 and 1984 NES Pre-Election Surveys, a small sample of subjects who did not answer the original survey were asked to indicate the difficulty of the opinion questions in the 1980 an d1984 NES surveys. Since this data was collected from individuals who did not participate in the original survey, hence this data will be referred to as third-party difficulty measures.

To obtain this data, participants were asked to rate questions in the 1980 and 1984 NES surveys on a scale from one (easy) to three (difficulty). In total, nine people provided difficulty ratings for each of the 1980 NES Pre-Election survey and four people provided difficulty ratings for each of the questions from the 1984 NES Pre-Election Survey. Participants were primarily Caltech undergraduate students.[24]

---

[23]See, for example,, Green, Krosnick & Holbrook (2002).

[24]A few volunteers were obtained from the outside the academic community. This sample was not designed to be random but only to provide an illustration of how difficulty might be collected for survey questions administered without respondent-reported difficulty measures.

This data was collected in January 2004, thus subjects were required to reflect on questions originally asked over 20 years ago.[25] Participants were told to "indicate how difficult" the questions were. To facilitate this process, participants were instructed to consider "how much effort would be required to answer" each question. Furthermore, respondents were specifically told that they did "NOT have to answer any of these questions."

With this difficulty rating data, it is possible to test which factors influence perceived response difficulty. Tables 6.9 and 6.10 contain the results from ordered logit regressions of the difficulty measures on different question characteristics and indicator variables for the different people who provided these difficulty ratings.[26] From these results, it is apparent that some people had significantly different difficulty ratings. For the 1980 NES data, Person 3 and Person 8 had statistically higher difficulty ratings than Person 1, while Person 5 and Person 9 had significantly lower difficulty ratings. In the 1984 NES data, Person 1 and Person 2 did not have statistically significantly different difficulty ratings but Person 3 and Person 4 had significantly higher difficulty ratings.

Characteristics of the questions also affected the difficulty ratings. Relative to questions with discrete choice options, questions with three point response scales and seven point response scales were rated as being significantly more difficult. In the 1980 NES data, long questions and questions with long explanations were rated as significantly more difficult, but in the 1984 data, long questions received lower difficulty ratings and questions with long explanations were not identified as statistically more difficult. However, the pseudo R-squared for the 1980 data, which is only 0.13, suggests that this model does not explain the variation in difficulty ratings very well. The model does a little better in 1984, where the pseudo R-squared is 0.32, but taken

---

[25]Some third-party subjects had difficulty with this task because they were unfamiliar with many of the issues and political figures discussed in these surveys from approximately twenty years ago. More recent data was not appropriate for this analysis, however, because the complex survey designs currently implemented by the NES do not allow a simple comparison of "no opinion" responding across subjects.

[26]The ordered logit model is used because the difficulty ratings are discrete ordering from one to three. The results are all relative to Person 1 whose dummy variable is excluded.

together the results from both these ordered logit regressions suggest that perceived difficulty cannot be entirely captured by identifiable question traits.[27]

Table 6.9: Ordered Logit Analysis of Difficulty Ratings in the 1980 NES

	Estimated Coefficient	Standard Error.	P-Value
Person 2	-0.11	0.29	0.72
Person 3	0.99*	0.29	0.00
Person 4	-0.02	0.29	0.94
Person 5	-0.62*	0.29	0.03
Person 6	-0.50$^+$	0.29	0.09
Person 7	-0.10	0.28	0.72
Person 8	0.60*	0.29	0.04
Person 9	-1.25*	0.30	0.00
Long Question	0.87*	0.22	0.00
Long Explanation	2.05*	0.21	0.00
Two Point Resp. Scale	0.45	0.28	0.11
Three Point Resp. Scale	1.59*	0.34	0.00
Four/Five Point Resp Scale	0.51$^+$	0.30	0.09
Seven Point Resp Scale†	2.99*	0.38	0.00
Cut 1	0.22	0.35	
Cut 2	3.45	0.37	
Number of Observations	900		
Pseudo R-squared	0.13		

† The excluded response scale type is discrete non-continuous response options.
* Statistically significantly different from zero at the $p \leq 0.05$ level.
$^+$ Statistically significantly different from zero at the $p \leq 0.10$ level.

---

[27]The higher R-squared statistic in the 1984 data in undoubtedly a reflection of the fact that only four subjects provided this data

Table 6.10: Ordered Logit Analysis of Difficulty Ratings in the 1984 NES

	Estimated Coefficient	Standard Error.	P-Value
Person 2	-0.53	0.40	0.18
Person 3	4.07*	0.35	0.00
Person 4	1.12*	0.31	0.00
Long Question	-1.24*	0.56	0.03
Long Explanation	-0.66	0.47	0.16
Two Point Resp. Scale	0.89	0.69	0.20
Three Point Resp Scale	2.89*	0.81	0.00
Four/Five Point Resp Scale	0.85	0.71	0.23
Seven Point Resp Scale†	1.69*	0.79	0.03
Cut 1	2.17	0.71	
Cut 2	3.92	0.74	
Number of Observations	520		
Pseudo R2	0.32		

† The excluded response scale type is discrete non-continuous response options.

* Statistically significantly different from zero at the $p \leq 0.05$ level.

6.2.4 Incorporating Difficulty Measures into the Analysis of the No Opinion form of Shirking

Since this analysis indicates that difficulty cannot be entirely captured by measurable question traits, the direct inclusion of third-party difficulty ratings into the analysis of "no opinion" response behavior might provide some additional explanatory power. Figures 6.4 and 6.5 combine third-party difficulty ratings with the "no opinion" response rates for survey participants. The bars represent the proportion of survey respondents who provided a "no opinion" response to each question, the dots indicate the average question difficulty rating, and the vertical lines associated with each dot provide the ninety-five percent confidence interval for the difficulty ratings. Dots that are between the two horizontal lines correspond to questions whose average difficulty ratings were between 2.5 and 3 (i.e., difficult questions). With the exception of the inflation/unemployment trade-off question asked in the 1980 NES survey, high "no opinion" response rates do not appear to correspond very closely with high average difficulty ratings. In fact, correlation between the average difficulty ratings and the "no opinion" response rate was 0.17 in 1980 and only 0.03 in 1984. This suggests

that these third-party difficulty measures are not very closely related to "no opinion" response rates. It is not clear, however, whether the low correlation between these measures can be attributed to the fact that the third-party samples were very small and the fact that these subjects were required to reflect on survey questions administered approximately twenty years ago. Furthermore, it is not clear that the average of third-party difficulty is the appropriate correlate or that "no opinion" responding is really indicative of shirking.
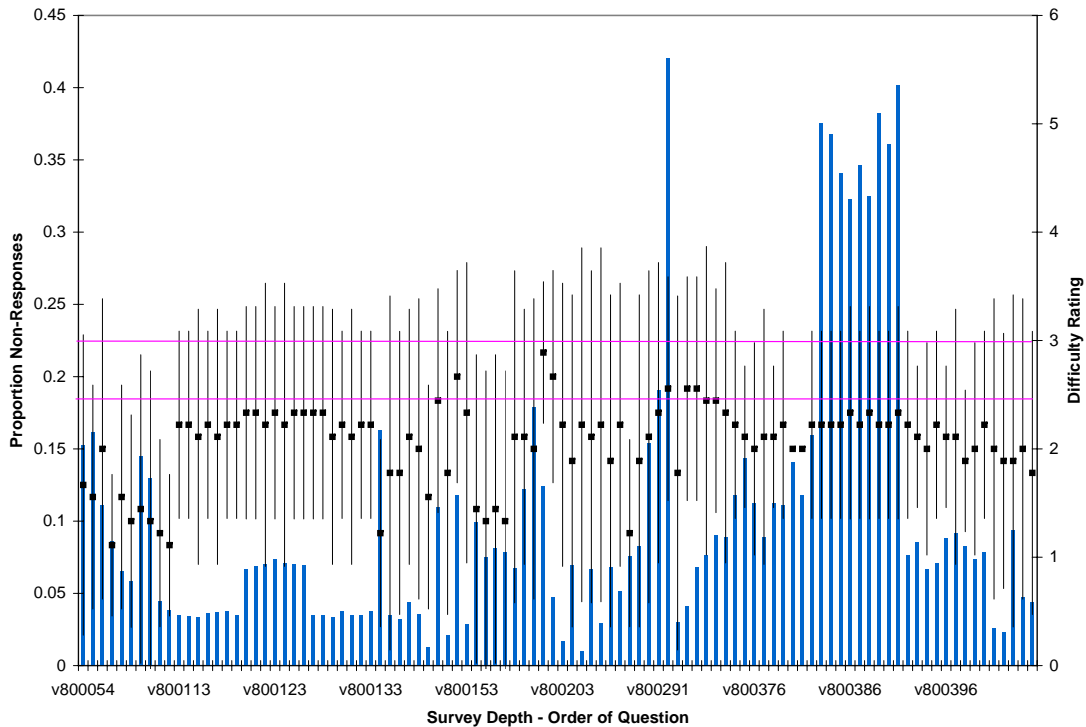
Figure 6.4: Combining Difficulty and Non-Response—NES 1980 Survey

According to Tables 6.9 and 6.10, respondents reported significantly different difficulty ratings. This suggest that difficulty ratings may be subjective, however, using third-party difficulty ratings provides an objective measure of difficulty however. Perhaps a better measure of difficulty can be obtained by asking each respondent to rate the difficulty of the opinion questions which they answer.[28] These respondent-

---

[28]An other alternative might be to randomly sample people to provide difficulty ratings at the same time as the survey is administered. Alternatively, difficulty ratings could be obtained from a
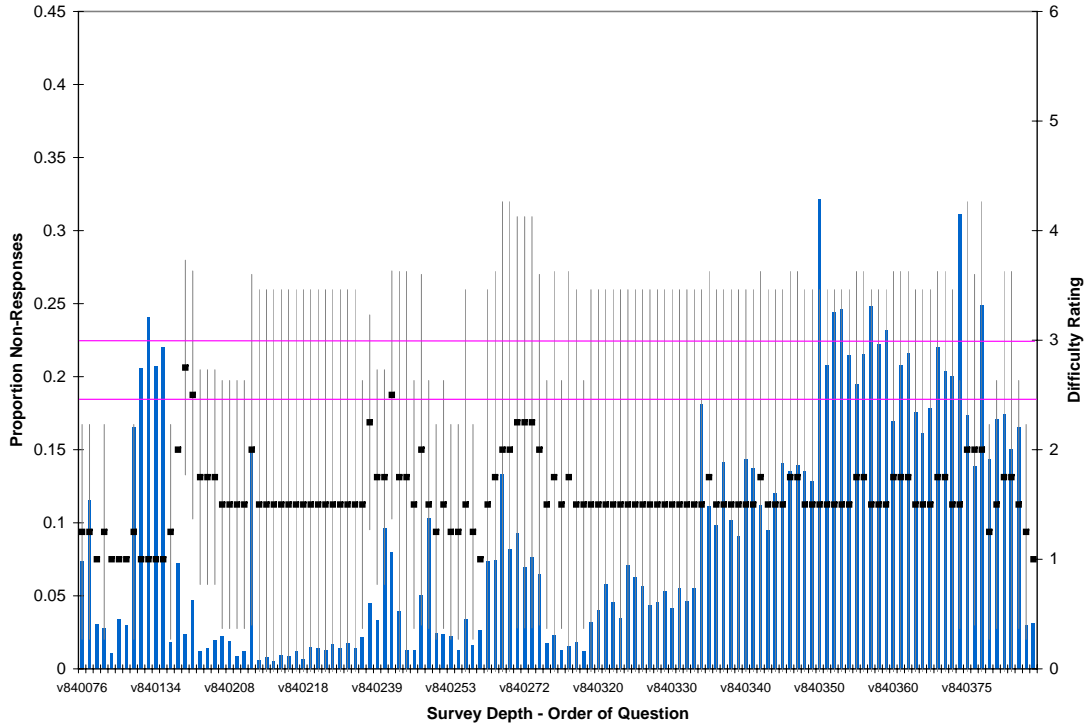
Figure 6.5: Combining Difficulty and Non-Response—NES 1984 Survey

reported difficulty measures would be able to capture variances of the perceived difficulty across survey participants.[29]

6.2.5   New Measures of Question Difficulty

The analysis in the above sub-sections suggests that respondent characteristics such as education and political informedness affect "no opinion" response behavior. These characteristics, however, are not fully able to explain the variance in "no opinion" response behavior suggesting that it might be important to include difficulty measures into the analysis. Because difficulty measures were not collected with the original NES

completely separate sample from those who answer the survey (third-party measure) or a subset of the survey participants might be asked to provide difficulty ratings in addition to answering opinion questions. There is a clear trade-off between the cost of asking all respondents to provide difficulty ratings on each opinion question and the power of the data provided by respondent-reported difficulty measures.

[29]Of course this is based on the assumption that respondents are able to assess the difficulty of questions. In Chapter 7, respondents' relatively high levels of variance in difficulty rating are used to support the idea that respondents are actually capable of assessing difficulty.

surveys, third-party difficulty ratings are analyzed. The results from this analysis suggest that there is not a very close relationship between third-party difficulty ratings and "no opinion" response behavior. However, the fact that difficulty ratings varied significantly across subject, even after controlling for objectively measurable question characteristics, suggests that objective measures of difficulty may not be appropriate. Instead, it may be useful to obtain objective measures of difficulty. With this in mind, attention is now turned to the 1998–1999 Multi-Investigator Study (MI1998) survey in which respondent-reported difficulty measures and response times were collected. These two measures provide alternative ways of capturing respondent perceptions of question difficulty. Assuming that "no opinion" responding is an indicator of shirking, by combining these new measures of question difficulty with respondent characteristics a better understanding of the causes of shirking can be established.

Difficulty Measures in the 1998–1999 Multi-Investigator Study

The 1998–1999 Multi-Investigator Study (MI1998) attempted to collect data about how people form their opinions.[30] As part of this effort, a series of six policy questions concerning affirmative action, welfare, and immigrant rights were paired with questions asking respondents to indicate how difficult these questions were to answer. For example, after answering the question "When it comes to setting aside a certain number of government construction contracts for businesses that are owned and operated by minorities, are you for or against this?" respondents were asked "How hard was it for you to make up your mind on that last question—not hard at all, not very hard, somewhat hard, or very hard?"

Table 6.11 indicates the distribution of difficulty ratings for these six policy questions. Respondents who answered "don't know" or who refused to provide a response to the opinion question were not asked for their difficulty ratings, thus their data is coded as missing. It appears that most respondents did not find it very difficult to

---

[30]See Appendix A.3 for a further discussion of the 1998–1999 Multi-Investigator survey. This survey was designed such that each of the questions examined in this section were only administered to two-thirds of the sample.

answer these questions.[31] Even if respondents who said "don't know" on the opinion question have their difficulty ratings coded as very hard, the percent of very hard ratings would, on average, be less that ten percent. At the other end of the spectrum, approximately fifty percent of respondents thought the questions were not hard at all. In fact, assuming that missing difficulty ratings should be set at the maximum, the average difficulty ratings for the questions were all around 1.8 (Standard Deviation = 1.0) corresponding to a difficulty of just below not very hard.[32]

Table 6.11: Tabulation of Difficulty Ratings by Question

Question	Percent Not hard at all	Percent Not very hard	Percent Somewhat hard	Percent Very hard	Percent Missing	Sample Size
Set-Asides	49.51	20.17	22.99	4.65	2.68	709
Flexible Admit. Stand.	51.06	22.57	15.94	7.05	3.38	709
Workfare	53.97	17.56	18.41	8.36	1.70	706
Welfare Limits	50.99	18.84	22.24	7.22	0.71	706
Welfare for Legal Immig.	52.85	21.70	19.05	5.70	0.70	719
Edu. for Illegal Immig.	57.02	21.00	14.88	6.40	0.70	719

Although overall averages are very similar across questions, it is possible that individual respondents find some questions easier to answer than others. To determine if respondents' difficulty perceptions vary, difficulty ratings provided by each respondent were compared across the questions.[33] In total, 145 respondents provided the same difficulty for all four questions, 187 respondents provided ratings that were within one point of each other (i.e., rated all of them as not hard at all and not very

---

[31]It is important to note that this analysis assumes that the difficulty questions provide a good way of assessing how difficult respondents perceive each question to be. Evidence explored below concerning individual level variance in these difficulty ratings suggests that respondents do not usually provide the same difficulty rating for all six questions. While this is not concrete evidence that these difficulty measures are meaningful, it does indicate that respondents are considering these questions differently.

[32]For this analysis difficulty ratings are coded from 1, not hard at all, to 4, very hard. See Table A.5 for a summary of the difficulty ratings for these questions.

[33]As discussed in A.3, questions were randomized across the full sample, thus, each respondent was only asked four out of six of the questions.

hard), 303 respondents provided difficulty ratings that differed by two points (i.e., not hard at all and somewhat hard), and 432 respondents rated at least one of the questions as not hard at all and another as very hard. This suggests that most respondents' difficulty perceptions vary across questions even though average difficulty ratings were very similar.

Response Times in the 1998–1999 Multi-Investigator Study

The 1998–1999 Multi-Investigator Study also collected information about how long respondents took to answer the six policy questions about affirmative action, welfare, and immigrant rights mentioned above.[34] The active response times (length of time between when the interviewer finishes asking the question and the respondent begins answering) for these questions ranged between zero seconds and over two minutes.[35] Most respondents were able to answer the questions in less than twenty seconds but there are a few outliers who took significantly longer to answer questions.[36] Huckfeldt, Sprague & Levine (2000) suggest that many problems may occur in the coding of activated timers and recommend that response times of zero seconds and response times that were more than three standard deviations above the sample mean be coded as missing.[37]

---

[34]There are two main types of response timers: active and latent timers. Active response timers measure the time between when the interviewer finished asking a question and when the respondent began answering the question. Latent response timers measure the time between the beginning of one question and the beginning of the next. Thus latent response timers measures include not only how long the respondent takes to answer but also the length of time interviewers take to read the question and to provide explanations possibly introducing confounding effects from interviewers. However, as discussed in Mulligan et al. (2003), latent and active response timers are highly correlated.

[35]Response times of zero seconds suggest that participants provided responses before the interviewer finished asking the question.

[36]Over ninety-five percent of respondents answered each of these question in less than twenty seconds.

[37]Only a few response times of zero seconds were recorded for these questions: 4 for the question about set-asides, 2 for the flexible admission question, 2 for the workfare question, and 2 for the welfare limits questions; the response times for the welfare for legal immigrants and education for illegal immigrants were all positive. Slightly more response times were recorded that were more than three standard deviations above the sample mean: 18 for the question about set-asides, 21 for the flexible admission question, 12 for the workfare question, 6 for the welfare limits questions, and 15 for both questions related to immigrants. Trimming these respondents from the dataset amounts to losing less than three percent of the response times for each question. Analysis which included these response but re-coded zeros as 0.5. provided similar results.

Figure 6.6 contains histograms of the trimmed response times, in hundredths of seconds, for answers to each of these questions. The first thing to note is that, despite the trimming procedure, response times are highly skewed to the right. Having considered several different transformations of response times, it appears that the distribution of the log of response times most closely resembles a normal distribution.[38] Thus, subsequent analysis will incorporate the log of the trimmed response times.
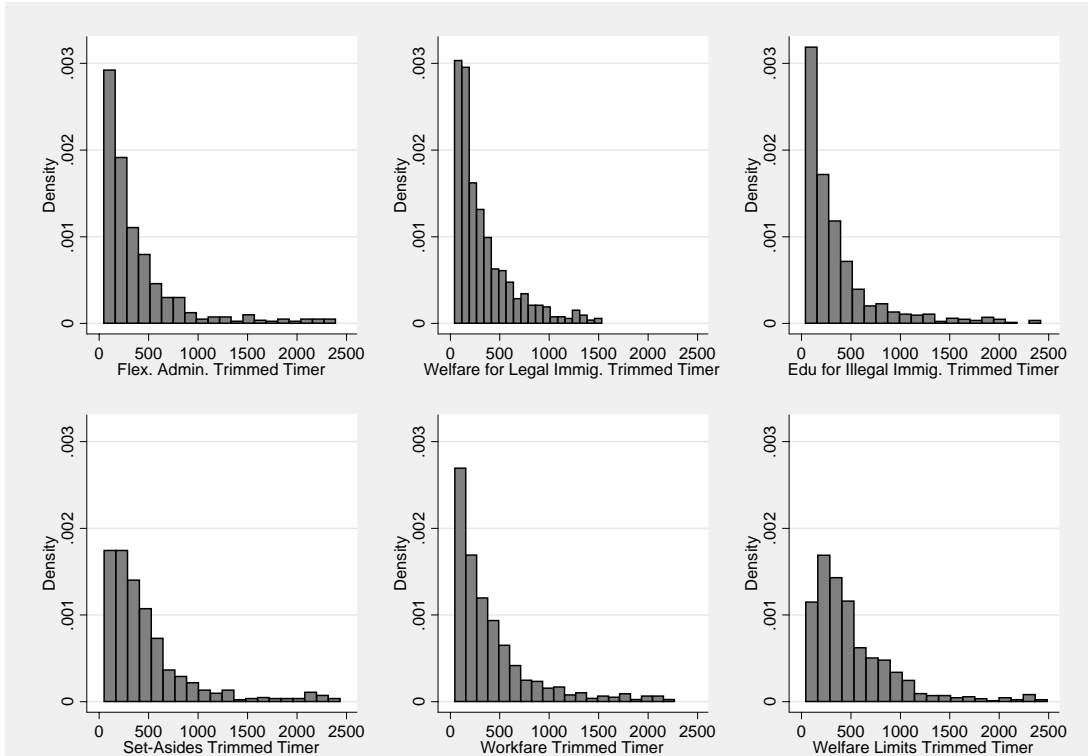
Figure 6.6: Histogram of Response Times for Questions in MI1998

Since in some of the following analysis, response times will be used as a measure of how difficult respondents found the question to answer, Figure 6.7 contains the distribution of response times by respondent-reported difficulty measures for the question about government set-asides for minority owned businesses.[39] This figure suggests

---

[38]Transformation considered include: the cubic, identify, square, square-root, natural log, inverse, inverse of the square, inverse of the cubic, inverse of the square-root.

[39]Respondents who take longer to answer the question are assumed to perceive the question as more difficult. For a similar interpretation of response timers, see Albertson, Brehm & Alvarez (Forthcoming, 2004).
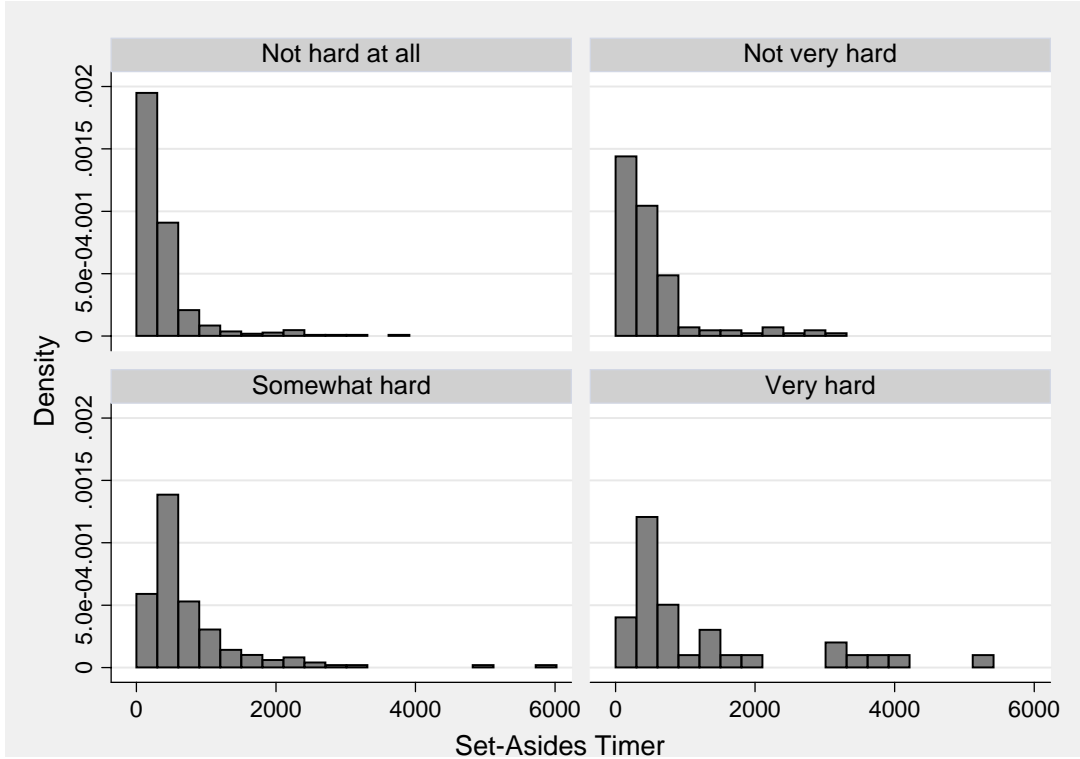
Figure 6.7: Histogram of Response Times by Difficulty

that response times and reported question difficulties are positively correlated. The same relationship is evident for each of the other five public policy questions under consideration. In fact, simple correlations between response times and reported question difficulty ratings were between 0.22 and 0.39. Hence, it appears reasonable to include response times together with the difficulty measures in the analysis of shirking as identified by "no opinion" response behavior.

6.2.6 Incorporating New Difficulty Measures into the Analysis of No Opinion Responding

Maintaining the assumption that shirking is indicated by "no opinion" responding, this chapter extends the analysis by incorporating both respondent-reported difficulty measures and response times in an attempt to better understand the factors that

influence survey response behavior. For the six policy questions in the MI1998 survey, respondents were not explicitly given a "no opinion" response option but a small fraction of respondents still indicated that they did not have an opinion on each question. Given the randomization design implemented in MI1998, each respondent was only asked four of the six policy questions. Thus the maximum number of times that they could express a "no opinion" response is four. Table 6.12 indicates how frequently respondents provided "no opinion" responses—almost ninety-five percent of respondents always provided an opinion in response to these policy questions. This low "no opinion" response rate is undoubtedly partly attributable to the lack of an explicit "no opinion" response option.[40]

Table 6.12: No Opinion Response Rates for the Six MI1998 Policy Questions

Count Non-Responses	Frequency	Percent
Zero	1010	94.66
One	47	4.40
Two	9	0.84
Four	1	0.09

Since this data has such a small fraction of "no opinion" responses and because those who indicated they did not have an opinion on a question were not asked the follow-up question about difficulty, it is not feasible to examine shirking behavior separately for each question.[41] Instead, the dependent variable in the subsequent analysis is the number of times a respondent provided a "no opinion" response on the policy questions that they were asked. Respondents who did not provide a difficulty rating for the questions they were asked are re-coded as having rated them as very

---

[40] For a discussion of how the inclusion of "no opinion" response options affects response selection see, for example, Krosnick, Holbrook, Berent, Carson, Hanemann, Kopp, Mitchell, Presser, Ruud, Smith, Moody, Green & Conaway (2002).

[41] Attempts to estimate a logit model of shirking using the stacked "no opinion" response data (one observation for each respondent-question pair) was unsuccessful because there was not enough variation in the "no opinion" response indicator and the difficulty measures.

hard. As suggested in Section 5.4, this type of ordinal data may be viewed as indicating different levels of shirking. Thus a simple extension to the probability model expressed in Equation (5.2) implies that an ordered logit is appropriate for testing what factors affect "no opinion" response rates.

Table 6.13 contains the results from several ordered logit regressions of the number of times each respondent indicated that they did not have an opinion. The first model in this table, contained in columns two and three, reports the results from running an ordered logit regression of the number of times respondents indicated "no opinion"on a variety of demographic characteristics. These results indicate that the only variables that have a significant effect on "no opinion" responding behavior are dummy variables for ages over sixty-five and residence in the northeast or south. Notice, however, that this model does not do very well at explaining the variance in the "no opinion"count data—the pseudo R-squared is only 0.07. The second model in this table augments the first by adding a subjective difficulty measure. Since this analysis is being done across questions, the difficulty measure included here is the sum of the difficulty measures provided by respondents for each of the questions they were asked.[42] This measure of difficulty has a statistically significant positive effect on "no opinion" response behavior across these questions. In other words, the more difficult the respondent finds the questions the more likely he is to provide "no opinion" responses. Also, the measure of political informedness (developed from responses to seven factual questions) becomes marginally significant in the second model. The final columns in this table contain the results from further expanding the model to include a measure of how long it took respondents to answer these questions. Since this analysis looks at "no opinion" response behavior across all the policy questions, the measure of response time is the log of the sum of the active timers for these questions. This variable also has a significant positive effect on the probability of providing "no opinion" responses. None of the models in Table 6.13 explain the

---

[42]Recall that respondents were only asked four out of six of the questions and corresponding difficulty ratings. Including indicators for which policy questions respondents received did not have a significant effect on the analysis.

variance in "no opinion" responding very well, although the pseudo R-squared of the third model is almost three times that of the first model. This may occur because there are relatively few cases in which respondents provided "no opinion" responses.[43]

Table 6.13: Ordered Logit Analysis of No Opinion Response in the MI1998 Policy Questions

Count	Model 1		Model 2		Model 3	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Over 65 Years Old	1.22*	0.32	1.25*	0.33	1.24*	0.33
Some High School	1.20	1.12	0.79	1.13	0.85	1.14
High school Grad	0.64	1.10	0.29	1.11	0.30	1.12
Some College	0.26	1.14	0.08	1.15	0.08	1.16
College Graduate	0.38	1.17	0.12	1.19	0.16	1.20
Graduate Work	1.32	1.17	1.09	1.19	1.16	1.19
Political Informedness	-0.14	0.09	-0.15$^+$	0.09	-0.15	0.09
Northeast	-0.99*	0.49	-1.00*	0.49	-1.03*	0.50
Midwest	-0.36	0.36	-0.30	0.37	-0.37	0.37
South	-0.64$^+$	0.38	-0.64$^+$	0.38	-0.69$^+$	0.38
Democrat	-0.89	0.35	-0.94	0.36	-0.95	0.36
Republican	-0.43	0.36	-0.36	0.36	-0.39	0.36
Sum of Difficulty Measures	·	·	0.13*	0.03	0.13*	0.03
Log Trimmed Timer	·	·	·	·	0.52*	0.26
Cut 1	2.43	1.11	3.61	1.17	7.40	2.25
Cut 2	4.24	1.15	5.46	1.21	9.25	2.27
Cut 3	6.57	1.49	7.82	1.54	11.61	2.46
Number of Observations	1056		1056		1056	
Pseudo R-squared	0.07		0.11		0.12	

* Statistically significantly different from zero at the $p \leq 0.05$ level.
$^+$ Statistically significantly different from zero at the $p \leq 0.10$ level.

Unfortunately, if "no opinion" responding is used to identify shirking, there is no way to extend the analysis beyond tests for the impact of respondent and question characteristics on this form of shirking behavior. Also, because "no opinion" respond-ing behavior is so rare in MI1998 policy questions, it is possible that respondents who did not want to exert much effort used some other response strategy when they were

---

[43]The infrequency of "no opinion" responding in this data suggest perhaps a more appropriate model for this analysis would be a rare-events logit as suggested by King & Zeng (2001).

shirking. It has been suggested that behaviors such as choosing the first response or randomly selecting response options might be indicative of cognitive shortcutting (Krosnick 1991). Although it may be fruitful to consider how these different behaviors are manifest in surveys, the MI1998 policy questions do not provide the ideal situation for examining these behaviors since they only offer two response options: for or against. Instead, the next section turns attention to identifying shirking behavior through direct measures of response times.

Chapter 7   Response Times as Indicators of Shirking

In the middle of difficulty lies opportunity (Einstein 1989).[1]

Since the results from in-depth interviews in Chapter 2 suggest that not all shirkers will choose the "no opinion" response option (and vica versa), this chapter explores the use of response times to identify shirking behavior. As discussed in Section 5.5.3, response may provide an alternative means of identifying shirking behavior. Assuming that respondents who carefully contemplate their answers take relatively longer to answer survey questions, it is possible to establish criteria to identify shirking based on responses times. As with the "quick-wrong" shirking behavior described above, this response time measure of shirking can be incorporated into substantive analysis to account for differences between shirkers and their non-shirking counterparts.

In this chapter, two different approaches to identifying shirkers based on their response times. The first approach identifies shirkers using outlier analysis; respondents who answer surveys in significantly less time than is predicted by a model of response time are determined to be shirking. The second measure of shirking based on response times classifies respondents based on demographic characteristics and difficulty reports and identifies shirking behavior within these group classification. Using group level estimates of the amount of time that should be sufficient to ensure reliable responses, respondents are identified as shirking if they take less time than is sufficient for they type. But before undertaking this analysis, however, it is important to consider the relationship between response times, effort, and response reliability.

---

[1] Authorities from the Albert Einstein Archives at the Hebrew University of Jerusalem were unable to authenticate the attribution of this quotation and believe the original source might be one of the Kalendersprueche which is famous only for its association with Einstein.

7.1 Response Times, Effort, and Accuracy

To relate response times, effort, and accuracy, it is useful to revise Equation 5.1 which specifies the probability of shirking as a function of respondent optimal effort, e_{iq}^{r*}, and surveyor optimal effort, e_{iq}^{s*}. To convert this equation into a function of response times, the relationship between effort and time must be explored.

As discussed in Section 5.5.3, studies indicate that response times are positively correlated with effort (Bassili & Fletcher 1991, Fazio & Williams 1986, Fletcher & Chalmers 1991, Fletcher & Chalmers 1991). Thus, response times may viewed as proxy for e^{r*} with

$$e_{iq}^{r*} = \theta(T_{iq}), \tag{7.1}$$

where T_{iq} is the time respondent i takes to answer question q and θ is the appropriate positive functional relation between time and effort.[2] Given this relationship between respondents' optimal effort and time, it only remains to determine when respondents have exerted enough effort to provide reliable responses.

If the surveyor's optimal effort level, e_{iq}^{s*}, were constant across respondents, i.e., $e_{iq}^{s*} = e_q^{s*} \forall i$, then shirkers could be identified any respondent, i, whose response times on question q were such that $\theta(T_{iq}) <= e_q^{s*}$.[3]

The above discussions, however, suggest that e_{iq}^{s*} varies with respondent characteristics and the perceived question difficulty. To help establish when respondents are exerting sufficient effort to provide reliable responses, it is useful to return to the examination of factual opinion responses. Because the reliability or accuracy of factual responses can be established from auxiliary data, it is possible to analyze the relationship between response times and accuracy in this data. Assuming that the relationship between accuracy or reliability and response times is the same across

---

[2]Having compared several transformations of response times (including the cubic, identify, square, square-root, natural log, inverse, inverse of the square, inverse of the cubic, inverse of the square-root) the log-linear transformation of time was found to provide a sample distribution that most closely resembled a continuous normal distribution. Furthermore the residuals from the log-linear regression appeared to satisfy the Gauss-Markov assumptions.

[3]Error will be introduced in this model because e_q^{s*} will have to be estimated.

both opinion and factual information questions, results obtained from the relationship between factual response accuracy and time can be generalized to opinion questions.

To examine all the factors that might affect the time needed to provide an accurate factual response, this work again uses the Indianapolis 1996 data to test which demographic characteristics affect the response times required to provide accurate answers. Results regressing the log of response times for accurate answers on a variety of individual characteristics are contained in Table 7.1.[4] According to these results, the amount of time respondents take to answer factual questions accurately is related to overall political knowledge (measured by correct responses on other factual questions), age, and education. Baseline speed, measured by the time respondents took to answer the demographic questions, also has a positive effect on the amount of time respondents take to correctly answer questions. Because the time data examined in this analysis is from latent response times, average interviewer times were also included in this analysis as suggested by Mulligan et al. (2003). The results suggest that, when using latent response times, interviewer behavior has a significant effect on how much time respondents take to provide reliable answers.[5] However, Mulligan et al. (2003) found that there are no interviewer effects evident in active response timers.

---

[4]In addition to the log-linear regressions considered here, OLS regression and Poisson regressions were estimated. Plots of the residuals from the OLS regression versus the fitted value suggest that the linear model violated the Gauss-Markov assumptions. Plots of the residuals from the log-linear model, on the other hand, do not display any pattern. Comparisons of first differences between the log-linear and poisson models suggest the two models produce similar results thus only the log-linear model is considered in this analysis.

[5]Analysis that included interviewer dummies effects only indicated that a few of the interviewers has significantly different latent timers.

Table 7.1: Analysis of Response Times for Accurate Factual Responses

	Const. Question		B of Rights Question		Veto Question	
	Coef.	Std. Err.	Coef.	Std. Err	Coef.	Std. Err.
Other Fact Q's Correct	-0.07*	0.01	-0.06*	0.02	-0.06*	0.02
Married	0.00	0.02	-0.01	0.03	0.02	0.03
Female	0.05*	0.02	0.00	0.02	0.05*	0.02
Working	-0.02	0.02	-0.02	0.03	-0.03	0.03
Age 18 to 25	-0.15	0.11	-0.18	0.11	-0.28*	0.13
Age 25 to 35	-0.08$^+$	0.04	0.00	0.05	-0.09$^+$	0.05
Age 35 to 45	-0.05	0.04	-0.06	0.05	-0.10*	0.05
Age 45 to 55	-0.09*	0.04	-0.03	0.05	-0.08	0.05
Age 55 to 65	-0.10*	0.04	-0.08$^+$	0.05	-0.12*	0.05
Age 65 to 75	-0.04	0.03	-0.03	0.04	-0.09*	0.04
Grades 1 to 11	0.07	0.05	0.27*	0.10	0.06	0.08
High School	0.10*	0.03	0.18*	0.04	0.12*	0.04
Some College	0.10*	0.03	0.12*	0.03	0.07*	0.03
Vocational. or Tech. Degree	0.01	0.05	0.09	0.07	0.09	0.07
College Degree	0.05*	0.02	0.12*	0.03	0.08*	0.03
White	-0.11*	0.03	-0.06$^+$	0.03	0.01	0.03
Income less than \$15K	0.09	0.05	0.01	0.06	0.07	0.06
Income \$15-25K	0.06$^+$	0.04	0.01	0.05	0.00	0.05
Income \$25-35K	0.03	0.03	0.01	0.04	0.00	0.04
Income \$35-50K	0.01	0.03	-0.01	0.03	0.05	0.03
Income \$50-75K	0.03	0.02	-0.01	0.03	-0.01	0.03
Log Resp Time for Demog Q's†	0.05*	0.02	0.09*	0.03	0.10*	0.03
Log Intrv. Av. Fact Q Times‡	0.00$^\diamond$	0.00	0.00$^\diamond$	0.00	0.00$^\diamond$	0.00
Interested in Politics	-0.01	0.01	0.00	0.01	0.01	0.01
Party Membership	0.00	0.01	0.01	0.01	0.00	0.01
Constant	6.66*	0.19	5.64*	0.24	5.69*	0.24
Number of Observations	1512		1168		1010	
Adj. R-squared	0.13		0.09		0.14	

† The time to answer demographic questions is measured as the sum of the amount of time respondents took to answer questions about marrital status, employment status, education level, race and birth year. These questions were chosen as a bseline since the are straightforward to answer and this information should be easy to recall.

‡ This is the average response times by interviewer. This variable has been included to control for possible interviewere effects.

* Statistically significantly different from zero at the $p \leq 0.05$ level.

$^+$ Statistically significantly different from zero at the $p \leq 0.10$ level.

$^\diamond$ These values are statistically significant but very small. The estimated coefficients (standard errors) for each model are 0.00007 (0.00002), 0.00009 (0.00002), and 0.0001 (0.00002).

These regression results suggest that relationship between reliability and response time appears to be mediated by demographic characteristics such as age, education

and political informedness. This supports the conclusion that e_{iq}^{s*} is not constant, thus, identification of shirking is only feasible if we can establish when respondents are exerting sufficient effort to provide reliable responses. In order to achieve this goal, response times will be analyzed using both the outlier and classification approaches described above.

The first approach identifies shirkers using outlier analysis; respondents who answer surveys in significantly less time than is predicted by a model of response time are determined to be shirking. The second measure of shirking based on response times classifies respondents based on demographic characteristics and difficulty reports and identifies shirking behavior within these group classifications. Using group level estimates of the amount of time that should be sufficient to ensure reliable responses, respondents are identified as shirking if they take less time than is sufficient for they type. But before undertaking this analysis, however, it is important to consider the relationship between response times, effort, and response reliability.

7.2 Response Times and Shirking—Outliers Analysis

The outlier approach to identifying shirkers involves establishing which respondents provided answers in significantly less time than would be predicted by a model of response time. Since the above analysis of accuracy in factual response questions suggests that demographic characteristics such as age, education, and political informedness affect the amount of time respondents take to correctly answer factual question, these same variables deserve to be included in the analysis of response times for opinion questions.

To undertake an analysis of shirking as identified by response times, data from the six MI1998 policy questions will be used again. Table A.6 contains the results from regressing the log of response times for the question about government set-asides

Table 7.2: Shirking as Determined by Outlier Measure of Response Times

Question	Not Shirking		Shirking	
	Frequency	Percent	Frequency	Percent
Set-Asides	670	95.77	26	4.23
Flex. Admit	605	98.09	12	1.94
Workfare	598	96.92	19	3.08
Welfare Limits	600	96.00	25	4.00
Welfare for Legal Immigrants	609	97.13	18	2.87
Education for Illegal Immigrants	612	97.30	17	2.70

for minority businesses on several respondent and survey characteristics.[6] Using the results from this analysis, it is possible to identify outliers. Table 7.2 contains the number and proportion of respondents who were determined to be outliers on each question. For the purpose of this analysis respondents were determined to be outliers if their response time residuals were more than 1.65 standard deviations below the average residual.

These results suggest that between 1.94 and 4.23 percent of the respondents are identified as shirkers when shirking is defined based on the outliers measure described above. These shirking rates are lower than those indicated by "no opinion" responding in the 1980 and 1984 NES Pre-Election Surveys but this may result because these questions offered only a two-point response scale.[7]

As with the "quick-wrong" shirking behavior described above, it is possible to incorporate this outlier measure of shirking into substantive analyses of survey data. Since the highest level of shirking was observed in the question about government set-asides, a variety of analyses were undertaken on this variable. For example, a logit analysis of this is contained in Table A.7. In this model, measure of ability, motivation, and difficulty do not appear to have any affect on this measure of shirking. Since this may indicate that this measure of shirking is not ideal, this work turns next

---

[6]Similar results were obtained for all six of the MI1998 policy questions.

[7]The analysis of difficulty ratings indicated that questions with only two response options were rated as significantly easier to answer compared to other types of questions.

Table 7.3: Shirking as Determined by Group Classification Measure of Response Times

	Not Shirking		Shirking	
Question	Frequency	Percent	Frequency	Percent
Set-Asides	670	97.53	17	2.47
Flex. Admit	612	89.21	74	10.79
Workfare	680	98.27	12	1.73
Welfare Limits	673	96.42	25	3.58
Welfare for Legal Immigrants	694	98.58	10	1.42
Education for Illegal Immigrants	701	99.57	3	0.43

to consider an alternate method for identifying shirking based on response times—the group classification method.

7.3 Response Times and Shirking—Analysis Based on Group Classification

To account for the fact that the surveyor's optimal effort level, e_{iq}^{s*}, is not constant across respondents, respondents are classified based on their demographic characteristics and their reported difficulty ratings. Assuming that within each group, g, the surveyor's optimal effort is constant, i.e., assuming that $e_{iq}^{s*} = e_{gq}^{s*} \forall i \in g$, respondents will be determined to be shirking if their response times less than the appropriate group level cut-point, e_{gq}^{s*}. In the subsequent analysis, group level cut-points, e_{gq}^{s*}, were set at 1.75 standard deviations below the mean of the log of response times for all the respondents in group g.[8] Table 7.3 contains the number and proportion of respondents who were determined to be shirking based on this measure.

Tables 7.3 shows that between 0.4 and 10.79 percent of the respondents are identified as shirkers when shirking is defined by the group classification method for

---

[8]This cut-point is used because it represents the one-tailed ninety-five percent confidence interval. Other cut-points including two standard deviations below the mean of the group level log response times was also considered but at this level no respondents were identified as shirking.

response times; between three and seventy-four of the response times were indicative of shirking. Again, these shirking rates are lower than those indicated by "no opinion" responding in the 1980 and 1984 NES Pre-Election Surveys but this may result because these questions offered only a two-point response scale.[9] Another notable fact in this data is that shirking appears to vary significantly across questions.

The question asking respondents how they feel about flexible admissions standards at universities appears to have received the highest percent of shirking responses.[10] One possible reason for these differences is that these questions are more or less likely to trigger conflicting predispositions. Analyzing the randomized question frames used in this survey, Alvarez & Brehm (2000) tested the hypothesis that conflicting predispositions, activated by the different question frames, affected response reliability. Unfortunately, the tests of this hypothesis were inconclusive. The alternative inferential measure of ambivalence, the average difficulty measure, does not provide much more explanatory power since it does not vary significantly across questions—the average difficulty rating for the flexible admission standards question was 1.79 (Standard Deviation = 1.01) compared to 1.88 (Standard Deviation = 1.00) on the set-asides question and 1.72 (Standard Deviation = 0.96) on the question about education for illegal immigrants.

Before proceeding with the analysis of shirking as identified by response times, it is important to recall that the relationship between effort and time is assumed to be positive and mediated only by demographic characteristics (age, education and political informedness levels) and question difficulty perceptions. Since the relationship between reliability and response times was established by examining factual questions administered in the Indianapolis 1996 survey, it is appropriate to verify the relationship between shirking behavior and other variables that we expect to be correlated with shirking.

---

[9]The analysis of difficulty ratings indicated that questions with only two response options were rated as significantly easier to answer compared to other types of questions.

[10]It has been suggested that this may result from the fact that this question is vague (Kiewiet N.d.).

Table 7.4 contains the results from a logit analysis of shirking on the question about flexible university admission standards as measured by group level cut-points.[11] The dependent variable is an indicator of whether a response was determined to be associated with shirking. According to these results the only variable that is strongly correlated with shirking behavior is the reported question difficulty ratings— respondents who indicated that questions were difficult have response times that were indicative of shirking.[12] Although only marginally significant, respondents who self-identify as being Democrats are slightly less likely to shirk while respondents who took longer to answer a question about world affairs are slightly more likely to be identified as shirkers. Variables such as education and political informedness, however, do not appear to have a significant effect on shirking behavior.

---

[11]This question is examined because it has the highest level of shirking as identified from the response times.

[12]Recall that respondents are identified as shirking if their response times were less than 1.75 standard deviations below their group mean response time where groups were determined based on groups demographic characteristics and respondent-reported difficulty.

Table 7.4: Logit Analysis of Shirking Measured from Response Times for the Flexible Admission Question

	Coef.	Std. Err.	F. Diff[†]	F.D. Std. Err.
Political Informedness	0.03	0.22	0.03	0.12
Female	0.42	0.71	0.01	0.05
Democrat	-1.49[+]	0.85	-0.03	0.06
Republican	-1.57	1.14	-0.03	0.07
Education - Categorical	-0.24	0.33	-0.02	0.10
Working	1.81	1.36	0.06	0.10
Married	-1.17	0.76	-0.03	0.06
Age	0.00	0.03	-0.01	0.11
Difficulty Rating for Flex. Admin Q	0.88[*]	0.29	0.20[*]	0.19
Time to answer Gov't Trust Q	0.00	0.00	0.01	0.25
Time to answer World Affairs Q	0.00[◇]	0.00	0.46	0.33
Time to answer Trustworthiness Q	0.00	0.00	0.00	0.21
Constant	-4.68	2.45	·	·
Number of Observations	675			
Pseudo R-squared	0.24			

[†] First differences are calculated as the difference between the predicted probability if all the explanatory variables set at their minimum and the predicted probability if the variable under consideration is increased to its maximum all else the same. Standard errors of the first differences are estimated using Monte Carlo simulation techniques similar to those suggested by (King, Tomz & Wittenberg 1999) and (King, Tomz & Wittenberg 2000).

[*] Statistically significantly different from zero at the $p \leq 0.05$ level.

[+] Statistically significantly different from zero at the $p \leq 0.10$ level.

[◇] The actual value of this coefficient is 0.009 (Standard Deviation = 0.0005), which is statistically significant at the $p \leq 0.10$ level.

Despite the fact that most of the dependent variables on the logit analysis of shirking behavior identified from response times are not significant, this model is capable of accounting for almost 25 percent of the variation in shirking which is striking considering that in previous models of shirking in the form of "no opinion" responding, presented in Table 6.13, only about ten percent of the variation was accounted for by the exogenous variables. It is also possible that the lack of statistical significance for most of the coefficients in this regression indicates that these variables have been accounted for appropriately in the establishment of the shirking indicator. Since shirking was determined for groups identified by their age, education, political

Table 7.5: Comparing Frequency of Shirking from Different Measures

Question	NR Shirking	Timer Shirking	Both	Neither	Total
Set-Asides	19	17	0	651	687
Flex. Admit	21	74	1	591	686
Workfare	11	12	0	669	692
Welfare Limits	4	25	1	669	698
Welfare for Legal Immigrants	5	10	0	689	704
Education for Illegal Immigrants	5	3	0	696	704

informedness, and difficulty ratings, it is possible that the effect of these variables has been completely accounted for in the development of the shirking estimates.

An alternative way of examining shirking behavior as identified by response times is to compare this measure of shirking with shirking as identified by respondents who provided "no opinion" responses. Table 7.5 contains a comparison of the frequency of shirking as defined by these two measures. The first thing to note is that, as mentioned previously, shirking rates on these questions are relatively low, independent of which measure is considered. On the questions about set-asides for minority businesses, workfare, and education for illegal immigrants, the shirking rates are comparable across the two different measures while, for the remaining questions, the rate of shirking identified from response times is several orders of magnitude higher than that identified by "no opinion" responding. In fact, for the question about welfare limits the shirking rate as identified by response times is six times that identified from "no opinion" responses. Still the response timer measure only identifies 24 respondents, or 3.5 percent of the respondents, as shirking on the welfare limit question. As mentioned previously the highest shirking rate is identified by response timers on the flexible admission standards question.

In addition to comparing the shirking rates defined by these two measure, it is possible to compare how many respondents are identified as shirking by both measures. This data is contained in the third column of Table 7.5. These results suggest

that these two measures of shirking are orthogonal; across all six questions, only two cases are identified as shirking according to both measures. Response time measures actually suggest that respondents who indicate that they do not have an opinion are doing so after careful contemplation.[13] Assuming that the response times measure accurately captures shirking, these results suggest that "no opinion" responding may not be a good measure of shirking. This may result because respondents are ambivalent concerning this issue. Unfortunately, determining what predispositions are activated when respondents are asked to answer these questions is difficult. As mentioned previously, Alvarez & Brehm (2000) were unable to uncover any evidence that conflicting predispositions were triggered in these questions through the use of randomized question frames.

Of course the results concerning shirking from timer measures depend on the assumed relationship between effort and response times. Assuming that respondents who take less time relative to other respondents within the same group implies that, once different demographic and difficulty characteristics that are used to define the groups are controlled for, the relationship between response times and effort is positive.[14]

One advantage to using response times as indicators of shirking is that it is possible to compare the response behavior of shirkers with that of non-shirkers.[15] Assuming that shirkers have been correctly identified, it is possible to consider how the existence of this behavior affects substantive measures of interest. Of course the effect of shirking on analytical predictions will depend on the percent of respondents in the sample who are identified as shirking. As Table 7.5 suggests, the level of shirking as measured by response times is likely to vary across questions. Since the highest level

---

[13]This may partly be attributed to the fact that these questions did not explicitly offer a "no opinion" response. Thus the "no opinion" response may only have been provided after serious contemplation.

[14]Identified difficulty measures attempts to control for this possibility but it is not entirely clear that this measure accurately captures how hard respondents found the question. A significant problem is the fact that respondents who provided "no opinion" were not asked the difficulty question and thus have been assumed to have the highest difficulty rating.

[15]Shirking as identified by "no opinion" responding depends on the response outcome so there is no variation in response behavior for these shirkers.

of shirking was observed in the question concerning flexible admission standards, the following analysis will focus on how analytic results concerning support for flexible admission standards are affected by shirking behavior.

Table 7.6 contains simple tabulations of the responses to the flexible admissions standard question. From these results it appears that shirkers are less likely to support flexible admission standards than their non-shirking counterparts. Although at first there is no reason to expect shirkers not to support flexible admission standards, it is important to note that this question was worded such that the "for" response option was provided first followed by the "against" option. Thus, it is possible that the higher fraction of shirkers indicating lack of support for flexible admissions standards is simply an artifact of the order in which these response options were provided.[16] It is possible that respondents who did not wish to exert much effort in answering a question opted for the latter response option.[17]

Table 7.6: Tabulations of Support for Flexible Admissions Standards and Shirking

	Non-Shirkers	Shirkers
Percent For Flex. Admit.	48.73	38.36
Percent Against Flex. Admit.	51.27	61.64
Pearson Chi-Squared	2.80	
P-Value	0.09	

The Pearson Chi-Squared results in Table 7.6 indicate that the distribution of responses provided by shirkers is statistically different from that of their non-shirking counterparts. However, the question remains as to how this difference in response selection translates into behavioral parameters of interest. To examine this issue, Table 7.7 contains the results from a regression of support for flexible admission

---

[16]For a discussion of these response types of response order effects, see, for example, (Becker 1954, Krosnick & Alwin 1987, Dillman et al. 1995, Narayan & Krosnick 1996).

[17]Studies of short-term memory suggest that when presented arguments orally, objects presented at the end are more likely to be active in memory (Henson 1998, Baddeley 1994, Lewandowsky & Murdock 1989). Since this survey was administered over the telephone, it is possible that many of the shirking respondents simply chose the "against" option because this response was still active in their memory.

standards on several independent variables that are likely to affect this outcome. Since the dependent variable is a binary outcome (for or against), a probit analysis was implemented.

Table 7.7: Probit Analysis of Support for Flexible Admission Standards

	Coef.	Std. Err.	F. Diff†	F.D. Std. Err.
Political Informedness	-0.03	0.03	-0.08	0.08
Female	0.18+	0.11	0.07	0.04
White	-0.16	0.14	-0.06	0.05
Democrat	0.19	0.12	0.07	0.05
Republican	-0.18	0.14	-0.07	0.05
Have Children	0.05	0.12	0.02	0.04
Support Women's Rights	0.07	0.07	0.05	0.05
Support Reducing Wealth Gap	0.07	0.06	0.05	0.04
Support Gov't Help for Blacks	0.30*	0.06	0.24*	0.05
Indicator of Shirking	-0.32+	0.17	-0.11	0.06
Constant	-0.27	0.22	·	·
Number of Observations	647			
Pseudo R-squared	0.09			

† First differences are calculated as the difference between the predicted probability if all the explanatory variables set at their minimum and the predicted probability if the variable under consideration is increased to its maximum all else the same. Standard errors of the first differences are estimated using Monte Carlo simulation techniques similar to those suggested by (King, Tomz & Wittenberg 1999) and (King, Tomz & Wittenberg 2000).
* Statistically significantly different from zero at the $p \leq 0.05$ level.
+ Statistically significantly different from zero at the $p \leq 0.10$ level.

Table 7.7 suggests that only gender and a measure of whether respondents agree that the government in Washington should make a major effort to improve the social and economic position of blacks are significant determinants of support for flexible admission standards; women are more likely to support flexible admissions standards as are those who agree that government should help blacks. Also notable in this analysis is the fact that the indicator for shirking is statistically significantly different from zero at the $p \leq 0.10$ level. This confirms the finding in Table 7.6 that shirkers are less likely to support flexible admission standards even after controlling for other factors that might influence attitudes towards flexible admissions standards. In

other words, the behavior of shirking respondents appears to be shifted towards not supporting flexible admissions standards, ceteris paribus.

In addition to affecting the intercept, shirkers might also be influenced differently by exogenous variables. This would translate into parameters that differ across shirkers and non-shirkers. To verify this hypothesis, a probit analysis, like that in Table 7.7, that included terms interacting the indicator of shirking with each of the dependent variables was performed. The results from this analysis are contained in Table A.8. This nested model provides the basis for testing the joint significance of the parameters associated with shirkers. The likelihood-ratio Chi-Squared statistic for this test is 11.04 (DF=10) and the associated p-value was 0.35. Thus at standard significance levels we can not reject the null hypothesis that shirkers do not have statistically significantly different parameters than their non-shirking counterparts.

An alternative way that shirking might affect responses is by affecting the estimated response variability. This implies that a heteroskedastic probit analysis might be appropriate. Results from this analysis are contained in Table 7.8. In this analysis the variance was assumed to depend on both an indicator of shirking and the level of respondent political informedness. Looking at the log of the estimated variance, more politically informed respondents appear to have significantly less variance as do respondents who are shirking. Again, this supports the hypothesis that shirkers are displaying response behavior that is different from their non-shirking counterparts. The lower variance suggest that perhaps many of the shirkers are applying the same rule for choosing responses. As suggested above, one possibility is that most shirkers are simply choosing the response option that is most active in their mind.

Table 7.8: Heteroskedastic Probit Estimates of Support for Flexible Admissions Standards

	Coef.	Std. Err.	Mean First Diff.[†]	First Diff. Std. Err.
Political Informedness	-0.01	0.01	-0.015	0.028
Female	0.05	0.03	0.018	0.022
White	-0.10[+]	0.06	-0.037	0.022
Democrat	0.05	0.04	0.018	0.014
Republican	-0.03	0.04	-0.010	0.014
Have Children	0.01	0.03	0.003	0.011
Support Women's Rights	0.02	0.02	0.016	0.015
Support Reducing Wealth Gap	0.02	0.02	0.020	0.013
Support Gov't Help for Blacks	0.06[+]	0.04	0.050	0.030
Constant	-0.03	0.07	.	
Log of Variance				
Indicator of Shirking	-0.68[*]			
Political Informedness	-0.27[*]			
hline Number of Observations	647			
Wald Chi-Squared	4.34			

[†] First differences are calculated as the difference between the predicted probability if all the explanatory variables set at their minimum and the predicted probability if the variable under consideration is increased to its maximum all else the same. Standard errors of the first differences are estimated using Monte Carlo simulation techniques similar to those suggested by King, Tomz & Wittenberg (1999) and King, Tomz & Wittenberg (2000).

[*] Statistically significantly different from zero at the $p \leq 0.05$ level.

[+] Statistically significantly different from zero at the $p \leq 0.10$ level.

Another factor to note about the heteroskedastic probit results is that support for government assistance to blacks remains a marginally significant determinant of support for flexible admissions standards.[18] However, this effect is slightly smaller given that the variance introduced by shirking has been controlled for by using a heteroskedastic estimator. In these results, the coefficient on the indicator for gender, female, is no longer statistically significant. This undoubtedly arises because women were more likely to have low political informedness measures. Also, the indicator for whether respondents are white is now statistically significant at conventional levels. The negative coefficient on white indicates that respondents who are white are less likely to support flexible admissions standards. This result is reasonable since it is

[18] Notice though, that none of the first differences in this table are statistically significantly different from zero.

possible that white respondents would be negatively impacted by such admissions policies.

Since respondents identified as shirkers are defined to be those who do not exert enough effort to provide reliable responses, it is appropriate to perceive of these responses as unreliable. Given that these responses are questionable, several different approaches might be appropriate: ignore the problem, throw out shirkers, or impute behavior for shirkers. The choice between these three approaches will depend on the severity of the shirking problem and the behavior of shirkers. For example, if shirkers randomly select a response option then excluding them from the analysis will affect the efficiency of estimates but not the consistency. However, if shirkers are displaying some systematic behavior, such as a recency effect, if these responses are not omitted or corrected, the estimates will be inconsistent.

Table 7.9 examines how support for flexible admissions standards varies as the procedure to account for shirking is altered. The first two columns indicate the support for flexible admissions standards if we assume that shirkers are providing legitimate responses. As a comparison the next few columns indicate the support for flexible admissions policies changing how the shirkers are measured. For example, if we assume that all the shirkers really do not support this policy then only 47.02 percent of the sample appear to support flexible admissions standards. On the other hand, if we assume that respondents who are shirking actually should have indicated that they support flexible admissions standards, then this percentage increases to 49.49 percent—this is almost two percent more support for flexible admissions standards that indicated if shirking behavior is ignored. The final pair of columns indicates the support for flexible admissions standards if shirking respondents are omitted from the analysis. According to the results, the support for flexible admissions standards is higher than would be predicted if we simply ignored shirking as suggested by the results in Table 7.6. However, it is important to recall that these estimates will be inefficient and may be biased if shirkers are inherently different from their non-shirking counterparts.

Table 7.9: Different Approaches to Accounting for Shirking

| | Ignore Shirking | | Set Shirkers=0 | | Set Shirkers=1 | | Omit Shirkers | |
	Freq.	Percent	Freq.	Percent	Freq.	Percent	Freq.	Percent
For	329	47.81	323	47.02	340	49.49	300	48.94
Against	358	52.19	364	52.98	347	50.51	313	51.06
Total	687	100.00	687	100.00	687	100.00	613	100.00

7.4 Asking Questions about Difficulty and Timing Answers

The new data on difficulty and response times explored in this chapter improves the ability to explain "no opinion" response behavior. But the analysis of "no opinion" responding is of limited use because, other than identifying that perceptive measures of question difficulty are positively correlated with this behavior, this analysis does not contribute much more than standard unit non-response analysis. Response times provide an alternative means of identifying shirking but, in the MI1998, these measures appear to be orthogonal to measures of shirking based on "no opinion" response. This suggests that "no opinion" response behavior may not be a perfect indicator of shirking as assumed in Chapter 6.[19] The results from shirking as identified by response times indicate that shirking behavior can have a significant impact on the substantive conclusions drawn from survey data. In this situation, analytical methods need to be established to account for respondents who display behavior indicative of shirking. One possibility, explored above is the use of estimators that explicitly model the heteroskedastic as a function of whether respondents are shirking or not. Before further exploring the suggestions from this analysis, it is useful to establish whether the prevalence of shirking varies across survey modes. With this in mind,

---

[19]Of course it is also possible that the approach to using response times established in this paper does not fully capture shirking.

the next chapter compares shirking behavior in telephone and Internet surveys to determine whether one mode is less likely to produce shirking behavior.

Chapter 8   Shirking in Answers to Telephone and Internet Surveys

> We are now witnessing another development in survey methodology, the consequences of which may prove to be even more profound [than the introduction of random sampling and interviewing by telephone]. It is the collection of survey data through self-administered electronic survey by email [and] the World Wide Web (Dillman 2000, p. 352).

The Internet has recently been recognized as a valuable instrument for the collection of opinion data (Schillewaert, Langerak & Duhamel 1998). Collection of survey data over the Internet has also attracted a lot of attention because telephone and face-to-face surveys are becoming increasingly difficult to conduct—Americans are increasingly difficult to contact at home or on the phone and increasingly are unwilling to participate in traditional surveys (Remington 1992).[1] Low costs, rapid turn around, and the ability to present complex graphical material are well recognized strengths of this new survey mode (Dillman 2000, Krosnick & Chang 2000, Couper, Traugott & Lamias 2001, Couper 2000, Alvarez, Sherman & VanBeselaere 2003, Berrens, Bahora, Jenkins-Smith & Weimer 2003). However, skepticism remains concerning the ability to obtain useful information through web-based surveys.

Although the primary concern with Internet surveys remains the issue of sampling, it is also important to determine whether reliable responses can be obtained through Internet surveys.[2] The shirking model described above provides a new way

---

[1] Available data suggests that response rates for both phone and face-to-face surveys are declining. For example, responses rates for the NES face-to-face surveys have declined from approximately 77 percent in 1952 to 64.3 percent in 2000. The 2002 NES was administered over the telephone and had a response rate of 66 percent; the NES telephone surveys have tended to have lower response rates – the 2000 Pilot Pre-Election telephone survey response rate was only 56.5 percent (National Election Studies (NES) 1995). The current Population Survey (CPS)—conducted by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics—shows non-interview rates of between 4-5 percent from 1964 through 1994, but an increasing non-interview rate thereafter, reaching almost eight percent in 2003; the refusal rate in 1964 was barely over one percent in the CPS monthly surveys, but stands around four percent in 2003 (Bureau of Labor Statistics 2002).

[2] For other recent comparisons of Internet and Phone surveys see (Krosnick & Chang 2000) and (Berrens et al. 2003).

of comparing telephone and Internet survey responses. Using extensive matched telephone and Internet survey data available from the Internet survey project at Caltech, this chapter compares shirking behavior between Internet and telephone survey respondents.[3]

In March 2001, matched telephone and Internet surveys were administered as part of the Caltech Internet Survey Project (CISP).[4] Despite slight differences in administration, these surveys provide an excellent opportunity to compare shirking behavior across the survey modes since they contained identical opinion questions with the same response options.[5]

8.1 Comparing Caltech Internet Survey Respondents

8.1.1 Demographics Characteristics

Table 8.1 shows the distribution of several demographic characteristics of the Caltech Internet and phone survey respondents. As a comparison, demographic data from the August 2000 Current Population Survey (CPS) is also included.[6] From this table it appears that the distribution of demographic characteristics varies quite dramatically across the samples. Tests comparing the distribution of the respondent characteristics indicate that, relative to the telephone survey, Internet respondents were less likely to be married and have children. Internet respondents were more likely to have some

---

[3]See Alvarez, Sherman & VanBeselaere (2003) for a thorough discussion of this Internet survey project.

[4]See Appendix A.1 for a more discussion about this particular survey.

[5]Given the differences in the telephone and Internet mediums, it is impossible to implement identical survey administration procedures. However, there were a few unforseen differences, such as the ability to skip questions on the Internet survey, that are likely to impact the comparison of respondent behavior across the two survey modes.

[6]The CPS is a household survey conducted monthly by the U.S. Census Bureau (Bureau of Labor Statistics 2002). Because this is a household survey the data is not directly comparable to the individual level data collected by the Caltech surveys. In order to account for this difference, only the primary family members are considered and this data is weighted using the household weight provided with the CPS data. The Aug. 2000 CPS data is used because this survey included the computer and internet usage supplement.

college eduction or a vocational degree. The apparent difference in age of the two groups, however, disappears when attention is restricted to phone survey respondents who have Internet access.[7] Other comparisons indicate that the Internet respondents differ from the CPS respondents on all the demographic characteristics while the telephone respondents have significantly different race, education, region and number of children. These results suggest that the Internet and telephone samples differ on several dimensions and neither of them appear to be drawn from the same distribution as the CPS.[8]

In addition to the above mentioned demographic characteristics, the Caltech Internet and telephone survey respondents were also asked about political party identification and religious beliefs. Table 8.2 contains a tabulation of these characteristics for the Internet respondents, the phone respondents, and the phone respondents with Internet access. It is clear that the Internet sample differs significantly from both telephone survey samples in terms of these characteristics. Even when restricting the comparison to phone survey respondents with Internet access, the Internet respondents are more likely to be Democrats or independents and they tend to identify less with political parties. Internet respondents also tend to have weaker religious affiliations.

---

[7]Pearson Chi-Squared tests for differences between the telephone and Internet respondents are contained in the Table A.1 and the same statistics for a comparison of the Internet and CPS respondents is contained in Table A.2.

[8]This raises the question of whether the Caltech survey samples are representative. This is a serious concern if the analysis is to be generalized to the American population as a whole but by focusing only on internal comparisons we can put aside the discussion of sample representativeness.

Table 8.1: Demographic Comparison—Internet, Telephone, and CPS

Characteristic	Internet Survey Respondents	Phone Survey Respondents	Phone Respondents w/ Net	Weighted CPS	Weighed CPS w/ Net
Gender					
Male	31.60	47.00	51.64	47.44	48.28
Female	68.40	53.00	48.36	52.56	51.75
Sample Size	*2332*	*1500*	*978*	*89888*	*48829*
Age					
18 to 29	26.76	20.93	25.15	21.42	24.57
30 to 39	21.61	19.13	22.60	20.70	23.93
40 to 49	25.13	21.00	24.85	24.07	24.90
50 to 59	17.84	17.40	16.67	15.22	15.82
60 to 69	5.87	10.67	7.57	9.91	6.77
Over 69	2.79	10.87	3.17	11.68	4.01
Sample Size	*2332*	*1500*	*978*	*89888*	*48829*
Race					
Non-white	12.65	21.07	20.35	26.80	19.65
White	87.35	78.93	79.65	73.20	80.35
Sample Size	*2332*	*1500*	*978*	*89888*	*48829*
Education					
No High School	5.27	7.42	2.20	16.38	5.68
High School	23.71	32.36	26.13	32.84	25.89
Some Coll.-Voc.	37.05	23.01	26.23	26.81	32.29
Coll. Graduate	20.84	23.35	28.54	16.27	25.04
Post-Graduate	13.12	13.86	16.89	7.76	12.10
Sample Size	*2332*	*1443*	*953*	*89888*	*48829*
Marital Status					
Other	50.81	43.10	40.08	43.34	37.96
Married	49.19	56.90	69.92	56.66	62.04
Sample Size	*2340*	*1443*	*953*	*89888*	*48829*
Number of Children					
Zero	37.14	27.07	30.64	68.45	62.19
One	15.27	16.34	16.06	13.25	15.71
Two	22.36	25.81	26.06	12.14	15.25
Three or More	25.34	30.79	27.23	6.16	6.85
Sample Size	*2332*	*1426*	*940*	*89888*	*48829*
Region					
North East	18.01	16.49	17.09	19.10	19.61
Midwest	23.93	24.88	23.52	22.87	23.30
South	34.95	34.03	32.81	35.53	32.90
West	23.11	24.60	26.58	22.50	24.19
Sample Size	*2332*	*1443*	*948*	*89888*	*48829*

Table 8.2: Other Respondent Demographic Characteristics

Characteristic	Internet Survey Respondents	Phone Survey Respondents	Phone Respondents w/ Net
Party Identification			
Republican	24.4 0	32.30	32.98
Democrat	35.42	34.27	32.56
Other	3.30	2.25	2.76
Independent or no party reg	23.89	21.28	22.80
Not Reg to Vote	12.99	9.90	8.91
N	2332	1424	943
$\chi^{2\dagger}$		33.66^{*}	30.44^{*}
Strength of Partisan identification			
Strong	17.41	31.67	28.69
Moderately strong	19.98	15.95	18.42
Moderate	21.10	26.35	28.22
Weak	5.45	6.43	6.38
Not at all or Independent	28.17	17.86	16.53
Not Reg to Vote	7.89	1.75	1.77
N	2332	1260	847
$\chi^{2\dagger}$		181.84^{*}	124.91^{*}
Religion			
Protestant	28.90	33.40	32.92
Catholic	17.92	23.93	24.34
Jewish	2.87	1.80	2.35
Islamic	0.43	0.47	0.61
Other	24.44	20.13	19.63
None	16.21	12.47	14.01
Don't Care to Respond	9.22	7.80	6.13
N	2332	1500	978
$\chi^{2\dagger}$		44.84^{*}	36.13^{*}
Strength of religious affiliation			
Strong	36.11	50.00	47.85
Moderately strong	19.13	14.67	15.64
Moderate	21.87	16.07	17.69
Weak	7.80	5.53	6.34
Not Applicable	9.22	5.87	6.65
Don't Care to Respond	5.87	7.87	5.83
N	2332	1500	978
$\chi^{2\dagger}$		93.77^{*}	41.87^{*}

[†] χ^{2} statistics are from the comparison of the different phone samples with the Internet sample.

[*] Denotes a χ^{2} statistic that is statistically significant at the 0.05 level.

From the above results it is clear that these two samples differ on many dimensions that will likely affect shirking behavior. Thus differences will be considered in the analysis of shirking behaviors across survey modes.

8.1.2 Knowledge and Behavior

In addition to comparing demographic characteristics, it is possible to compare responses to "core" knowledge and behavioral questions.[9] Table 8.3 presents data on post-election reports of turnout for the 2000 election and a measure of political informedness.[10] As often noted in political opinion surveys, reported turn-out rates are higher than actual voter turnout.[11] However, the Internet respondents were less likely to report that they voted in the 2000 election than the phone survey respondents.[12] It is also interesting to note that the Internet survey respondents are more politically informed when political information is measured by an ability to recognize four political figures.[13] This difference in political informedness is particularly important since this attribute has been shown to be correlated with shirking behavior. As a result, it will be important to consider the different levels of political informedness between the two samples when analyzing shirking behavior.

---

[9]Core questions were determined to be those that were similar to those asked every year by the National Election Study (NES).

[10]Political informedness is calculated from respondents' ability to identify the jobs or political offices held by Dick Cheney, William Rehnquist, Trent Lott, and Colin Powell. Respondents able to correctly identify more political figures received a higher political informedness score.

[11]The reported turn-out rate for Internet survey respondents was 68 percent while 75 percent of phone survey respondents indicated that they voted. According to the Federal Election Commission (www.fec.gov/pages/2000turnout/reg&to00.htm), the actual turnout rate for the 2000 election was 51.3 percent of the voting age population.

[12]This may be evidence that Internet respondents are more likely to answer sensitive questions honestly. However, there is insufficient data in this survey to conclusively test this hypothesis.

[13]It has been suggested that this discrepancy may occur because Internet survey respondents can easily obtain answers from the Internet. Anecdotal evidence, such as survey completion times, email feedback from respondents, and differential political informedness levels by recruitment method suggests that respondents are not in fact searching for information about these people while simultaneously completing the survey.

Table 8.3: Behavior and Knowledge Questions

Item	Internet Survey Respondents	Phone Survey Respondents	Phone Respondents w/ Net
Voted in 2000 Election			
Yes	68.43	73.00	74.95
No	30.71	27.00	25.05
Don't Know	0.86	0.00	0.00
N	*2327*	*1500*	*978*
$\chi^{2\ddagger}$		19.90*	20.17*
Political Informedness[†]			
Identify 0 of 4 positions	12.14	15.73	12.47
Identify 1 of 4 positions	9.73	15.80	14.83
Identify 2 of 4 positions	26.33	27.93	27.81
Identify 3 of 4 positions	4.29	9.53	9.10
Identify 4 of 4 positions	47.51	31.00	35.79
N	*2327*	*1500*	*978*
$\chi^{2\ddagger}$		137.57*	65.98*

[†] Political informedness is generated from respondents' ability to identify the jobs or political offices held by Dick Cheney, William Rehnquist, Trent Lott, and Colin Powell.

[‡] The Pearson statistic is $\chi_p^2 = n \sum_r \sum_c \frac{(p_{rc} - p_{0rc}^2)}{p_{0rc}}$ where n is the total number of observations, p_{rc} is the estimated proportion for the cell in the rth row and c^{th} column of the table, and p_{0rc} is the estimated proportion under the null hypothesis of independence. χ^2 statistics are from the comparison of the different phone samples with the Internet sample.

[*] Denotes a χ^2 statistic that is statistically significant at the 0.05 level.

8.1.3   Feelings and Opinions

Another interesting set of questions frequently included in political opinion surveys asks respondents to provide "feeling thermometers" for the different presidential candidates; these questions require respondents to indicate how they feel about presidential candidates on a "thermometer" scale from zero (cold/negative) to one-hundred (warm/positive). These questions, included in the telephone and Internet surveys administered in March 2001, provide a succinct summary measure of respondents feelings toward the 2000 presidential candidates. Figure 8.1 contains graphical presentations of Bush's and Gore's thermometer scores for Internet and phone survey
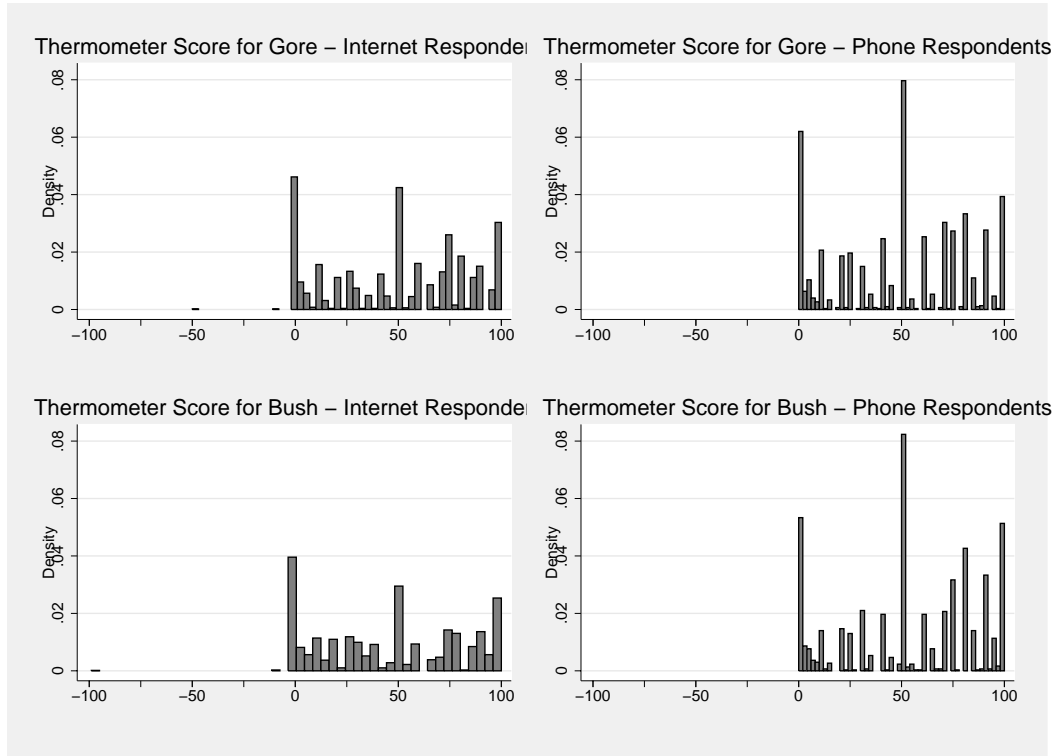
Figure 8.1: Histograms of Thermometer Scores for Bush and Gore

respondents.[14] The first obvious fact is that all four figures indicate a distribution of thermometer scores that are tri-modal with modes at zero, fifty and one-hundred. In comparing the Internet and phone respondents, it appears that Internet respondents are "colder" towards both Bush and Gore.

Another interesting feature of these questions is that the Internet respondents were very likely to skip this question—only seventy-three percent of Internet respondents provided a thermometer score for Bush and Gore. In comparison, the telephone survey obtained responses from all 1,500 respondents. This might suggest that these questions were more difficult for Internet respondents to answer—Internet respondents had to type in a numerical response for these questions while other questions only required respondents to select among the available response options. However, low response rates to thermometer questions are also evident in the NES surveys.

[14]An unforseen design issue also allowed Internet respondents to provide numbers between negative ninety-nine and zero.

In 1980 the "no opinion" response rate to the thermometer ratings for Carter and Reagan were 19.0 and 16.5 percent, respectively. In 1984, 13.8 and 21.5 percent of respondents refused to provide thermometer rating for Carter and Reagan. This suggest that perhaps the telephone survey obtained unusually high response rates. Since this question did not provide a "don't know" response option, interviewers were required to obtain a numerical response to this question and thus the high response rates may be an artifact of interviewer probing.[15]

Among the other "core" questions in most political opinion surveys are questions concerning institutional approval and vote choice. Included in the March 2001 Caltech Internet and telephone surveys were questions asking respondents whether they approved of George Bush, the Congress, and the Supreme Court. Table 8.4 contains the χ^2 results from tests comparing the distributions of these question across the two different samples. It is clear from this data that the telephone and Internet respondents differ in terms of their sentiments concerning how politicians are handling their jobs.

---

[15]Interviewers are usually paid for each completed survey they obtain. Since a "don't know" response option was not explicitly provided for this question, interviewers had incentives to probe respondents to provide a numerical rating in response to the thermometer questions. It is not clear what behavior to expect from respondents who are obliged to provide a response but do not know how to rate their feelings about candidates on this thermometer score. Random selection of responses or selection of the "first" response option are both feasible strategies for respondents who do not know how to express their opinions. However, to analyze this behavior, it would be necessary to make assumptions concerning the indicators of these different behavior choices.

Table 8.4: Political Opinions Questions—Chi-Squared Tests

Question	Net vs. Phone	Net vs. Phone w/Net
Approve of George Bush	84.08[*]	881.49[*]
Approve of Congress	137.00[*]	137.18[*]
Approve of Supreme Court	108.41[*]	135.39[*]

[*] The $\chi^2_p = n \sum_r \sum_c \frac{(p_{rc} - p^2_{0rc})}{p_{0rc}}$ where n is the total number of observations, p_{rc} is the estimated proportion for the cell in the rth row and c^{th} column of the table, and p_{0rc} is the estimated proportion under the null hypothesis of independence. All of these χ^2 statistic are statistically significant at the 0.05 level.

8.2 Shirking—No Opinion Responding Across Modes

Given the model of shirking as described in Chapter 5, it is possible to compare shirking behavior across both telephone and Internet survey modes. Assuming that "no opinion" responding is indicative of shirking, a comparative analysis is possible using the matched telephone and Internet surveys; examining "no opinion" response behavior provides a useful test of the hypothesis that Internet respondents exert less effort than their telephone counterparts because the relationship between Internet respondents and the surveyor is weaker (Green, Krosnick & Holbrook 2002).

In comparing respondent behavior across identically worded telephone and Internet surveys, the analysis need not worry about question characteristics because the questions were identical in both survey modes. Individual level characteristics, however, are still likely to be important.

The March 2001 survey contained several questions asking respondents for their opinion on political issues varying from election reform to congressional approval. This section examines responses to the eighteen opinion questions in this survey that explicitly allowed a "don't know" option.[16] Using the responses to these eighteen

---

[16]Attention is restricted to these questions because phone survey respondents were only allowed to choose answers that were among the provided response options. The Internet survey respondents on the other hand were able to skip a question resulting in a missing response. The questions

questions, a "no opinion" response rate can be established for each individual. Unfortunately the fact that Internet respondents were able to skip questions complicates the analysis of "no opinion" responding. In the phone survey, we can only measure "no opinion" responding when respondents chose the "don't know" option. In the Internet survey data, "no opinion"behavior may be displayed either by the choice of the "don't know" option or by the submission of an empty response.[17] In order to account for these discrepancies, the subsequent analysis considers two different approaches to measuring "no opinion" responses. The first compares the choice of the "don't know" response option by telephone respondents with the same behavior for the Internet respondents who completed all survey questions.[18] The second comparison includes all phone and Internet respondents but counts empty responses as a "no opinion" responses.

Table 8.5 contains the mean, standard deviation, and skewness of "no opinion"behavior for the two samples. The mean number of "don't know" responses for the phone survey was 1.49 (Standard Deviation = 2.06, skewness=2.34); for the Internet respondents who completed all questions the mean number of "don't know" responses was 1.36 (Standard Deviation = 1.86, skewness=2.51); if a missing answer is counted as a "no opinion" response, the mean number of "no opinion" responses for the Internet respondents is 2.28 (Standard Deviation=2.25, skewness=2.27). Simple t-test comparing these means indicate that there is no statistically significant difference in "no opinion" responding behavior across the survey modes. However, since, as suggested by the theory outlined in Chapter 5, respondent characteristics affect shirking behavior, it is useful to test how different demographic sub-sets of respondents behaved. From Table 8.5 it appears that less educated and less politically

---

considered correspond to the phone survey questions 2, 4, 5, 6, 14, 15, 23, 24 26, 27a-i. These questions represent just under half the total questions in this survey; the Internet survey had a total of 40 questions and the phone survey contained 44 questions in addition to the eleven demographic questions.

[17]It is not entirely clear that providing an empty response is the same as choosing the "don't know" option; it is possible that respondents unintentionally skipped questions and thus this behavior would not correspond to "no opinion" responding.

[18]Only the 408 Internet respondents who completed all of these questions are included in this sample.

informed respondents are slightly more likely to provide "no opinion" responses but again these differences are not statistically significant. However, these demographic characteristics alone may not fully capture respondent motivation and ability. In order to control for all the factors that will affect shirking behavior, it is useful undertake a multivariate analysis. As discussed in Section 3.1.1, if we assume that the proportion of "no opinion" responses has a logistic distribution, it is appropriate to run weighted least-squares regressions on the log-odds ratio.

Statistic	Phone Resp. DK	Net - Excl. Resp w/ Missing	Net Resp. DK & Missing
	Full Sample		
Mean	1.49	1.36	2.28
Standard Deviation	2.06	1.80	2.25
Skewness	2.34	2.63	2.27
N	1500	408	2332
	Low Education[†]		
Mean	1.84	2.15	3.11
Standard Deviation	2.41	2.64	2.78
Skewness	1.92	2.56	1.96
N	574	80	676
	Low Education[‡]		
Mean	1.16	1.16	1.94
Standard Deviation	1.52	1.47	1.88
Skewness	1.88	1.55	2.19
N	869	328	1650
	Low Political Informedness[*]		
Mean	0.98	0.91	1.53
Standard Deviation	1.36	1.20	1.41
Skewness	2.04	1.47	1.70
N	508	244	1208
	Low Political Informedness[◊]		
Mean	1.84	2.02	3.09
Standard Deviation	2.360	2.28	2.66
Skewness	2.05	2.25	1.91
N	892	164	1124

[†] No High school or High school.

[‡] Some College, College, Vocational, Graduate or Professional.

[*] ID two or less politicians or not answer these questions.

[◊] ID three or more politicians.

Table 8.6 contains the results from these regressions. When the sample includes only the 408 Internet respondents who completed all questions, the coefficient on the dummy for the Internet survey is not statistically significant. This suggests that Internet respondents who answered all questions are not more likely to provide "don't know" responses than phone survey respondents. However, if the measure of "no opinion" responding includes missing responses, then this coefficient is positive and significant—Internet respondents are more likely to give "no opinion"when skipped

questions are included. It is also interesting to note that increased political informed-ness results in reduced levels of "no opinion" responding as expected. The same is true for education and strength of party identification.[19] Being white, having children, and supporting either Bush or Gore also decrease "no opinion" responding. On the other hand, being female is associated with an increase in "no opinion" responding.

Table 8.6: No Opinion Responding—WLS Regression on the Log-Odds[†]

Variable	Net & Phone Excl.Resp. w/Miss.		Net & Phone DK and Missing	
	Coefficient	Std. Err	Coefficient	Standard Error
Internet Survey	-0.02	0.05	0.27[*]	0.03
Political Informedness	-0.54[*]	0.06	-0.56[*]	0.04
Age	0.16[*]	0.07	-0.03	0.05
Education	-0.44[*]	0.07	-0.29[*]	0.05
White	-0.11[*]	0.05	-0.07[*]	0.03
Female	0.14[*]	0.04	0.15[*]	0.03
Have children	-0.18[*]	0.05	-0.06[*]	0.03
Married	0.03	0.04	-0.07[*]	0.03
Republican	-0.16	0.10	-0.11[*]	0.05
Democrat	-0.09	0.09	-0.05	0.05
Independent	-0.13	0.09	-0.06[+]	0.04
Strength of Party Ident.	0.28[*]	0.07	0.17[*]	0.05
Protestant	-0.05	0.05	0.02	0.03
Catholic	-0.06	0.05	0.00	0.03
Jewish	0.06	0.14	0.21[*]	0.09
No Religion	-0.16[*]	0.06	-0.07[+]	0.04
Strength of Religious Ident.	0.00	0.06	0.02	0.04
Support Bush	-0.26[*]	0.06	-0.23[*]	0.04
Support Gore	-0.17[*]	0.05	-0.33[*]	0.04
Support Nader	0.01	0.11	-0.27[*]	0.06
Intercept	-1.10[*]	0.13	-1.16[*]	0.08
N Internet	408		2327	
N Phone	1232		1232	
Adjusted R-squared	0.18		0.22	

[†] Since the log of zero is undefined, $P_i = 0$ has been replaced by $P_i = 0.0001$.
[*] Statistically significantly different from zero at the $p \leq 0.05$ level.
[+] Statistically significantly different from zero at the $p \leq 0.10$ level.

[19]Strength of party identification is measured from zero to one where zero corresponds to strong partisan identification and one to no partisan identification.

8.3　Stacking up Internet Survey Responses

Taken together these results provide somewhat mixed support for the hypothesis that Internet survey respondents exert less effort than their telephone counterparts. While it does appear that when missing responses are considered "no opinion" responses, Internet respondents are more likely to display behavior suggestive of shirking, this may say more about how surveys are administered than the effort levels exerted by respondents since it is quite possible that some of the skipped questions resulted from unintentional mistakes. Also, requiring respondents to answer every questions may encourage random selection of responses or, worse yet, encourage respondents to break-off from answering the survey. The lack of data from the partially completed telephone surveys may be confounding these results. As in the analysis contained in Chapters 6 and 7, demographic characteristics, such as education and political informedness, appear to affect the "no opinion" response measure of shirking in a predictable fashion.

It does appear that characteristics such as education, political knowledge or informedness, gender, and race have predictable affects on the use of the "no opinion" response strategy.

Chapter 9 Conclusions about Shirking on Survey Answers

> *USA Today* has come out with a new survey: Apparently three out of four people make up seventy-five percent of the population (Letterman September 27, 2002).

Since survey data is only as meaningful as the response provided by survey participants, everyone who uses survey data should be concerned about how respondents answer survey questions. By proposing a rational choice model of respondent behavior, this work attempts to provide a new framework within which to examine survey response behavior. Unlike previous literature, the shirking model has clearly testable empirical implications concerning the determinants of respondent behavior.

Before analyzing shirking behavior among survey respondents, the issue of identification must be addressed. Unfortunately, most readily available opinion data cannot easily distinguish respondents who are shortcutting their cognitive processes from those who provide reliable responses. Results from in-depth interviews and evidence from existing research in this area suggest that "no opinion" responding and answering survey questions relatively quickly might be indicative of shirking. However, much work remains to be done to establish a reliable means of identifying shirkers. By obtaining more interview data and by asking respondents to discuss how they behaved when answering survey questions, it should be feasible to better identify responses that will be unreliable.

Given that shirking can be identified, the next issue addressed in this work is the determination of factors which affect shirking behavior. Assuming that it takes effort to provide reliable responses and that rational agents optimally select their effort levels, it is possible to establish clear tests for how ability, motivation, and difficulty affect respondent behavior. Since there are no direct measures of shirking, several

different indicators are explored in this work. From this analysis many different conclusions about shirking are derived.

First, this work suggests that the inclusion of subjective measures of question difficulty improves our ability to explain "no opinion" response behavior. Supplementing existing data with measures of task difficulty, this work also concludes that respondents are more likely to shirk when they perceive questions to be difficult. Results from the 1980 and 1984 NES surveys also suggest that "no opinion" responding may be more prevalent towards the end of long surveys. A clear recommendation from this observation is that long surveys should be avoided. If long surveys are unavoidable, randomizing the order in which questions are presented should minimize "no opinion" response rates for questions that would otherwise have appeared at the end of the survey.

Assuming that time and effort are positively correlated, response times offer another means of identifying respondents who are not are shirking. After having established factors that determine how much effort respondents would have to exert in order to provide reliable responses, response times are analyzed to establish which respondents are shirking. Again, the results indicate that respondents who are more able, motivated, who find questions easy to answer, and are less conflicted are less likely to display shirking behavior.

If shirking can be identified based on response times, it is possible to measure the frequency oconsider the impact that this behavior has on analysis of substantive survey data. The first thing to note is that the impact of shirking will be related to the prevalence of this behavior. If a significant fraction of respondents display shirking behavior, evidence explored in Chapters 6 and 7 suggests that shirking might impact the reliability of even the simple summary statistics derived from survey data. Preliminary analysis does, however, indicates that it may be possible to account for shirking behavior in multivariate analysis through the use of empirical tools such as heteroskedastic estimators.

Of course the discussion of shirking is far from over. With better data, shirking behavior could be more reliably established and a better insight into the determinants and effects of this behavior could be established. As mentioned above, by collecting more interview data in which respondents have the opportunity to discuss how they behaved when answering survey questions, will provide better means of identify shirking. It may even be expedient to simply ask respondents to indicate how much effort they exerted in answering a question.

As suggested by the MI1998 analysis which includes subjective difficulty measures, the collection of more extensive data concerning perceptions of question difficulty are likely to provide researchers with a better understanding of respondent behavior. Given that difficulty may be introduced at any of the four cognitive stages of survey response, it may be expedient to ask respondents to identify the different components or sources that make a question difficult. Third-party difficulty reports might also prove more useful if they are collected concurrently with survey responses. Alternatively, interviewers might be able provide insight from their experiences into when responses are less likely to be reliable.[1] Since interviewers are familiar with the survey questions and have information about how respondents perceive questions, they may also be able to report question difficulty measures. A comparison of these different measures of difficulty deserves attention.

Another approach to establishing whether respondents provide reliable responses might be to obtain repeated measures from the same respondents. Using more modern survey approaches than those implemented by Converse (1964), it may be possible to identify when respondents are providing legitimate responses. This type of analysis might also attempt to obtain better measure of predispositions and predisposition activation to help distinguish respondents who are shirking from those who are ambivalent.

---

[1]At one point the National Elections Study obtained debriefings from interviews. This data or other similar data might provide new insights into respondent behavior.

By collecting and analyzing data about respondent behavior, survey researchers will be better equipped to ensure that the conclusions drawn from survey data are meaningful and reliable.

Bibliography

Achen, C. 1975. "Mass Political Attitudes and the Survey Response." *American Political Science Review* 69:1218–1223.

Albertson, B., J. Brehm & R. M. Alvarez. Forthcoming, 2004. "Ambivalence as Experienced Conflict." In *Ambivalence and the Structure of Public Opinion*, ed. S. C. Craig & M. D. Martinez. Palgrave Macmillan p. Chapter 3.

Aldrich, J. H. & R. D. McKelvey. 1977. "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections." *American Political Science Review* 71:111–130.

Alvarez, R. M. 1998. *Information and Elections.* Ann Arbor, MI: University of Michigan Press.

Alvarez, R. M. & J. Brehm. 2000. "Binding the Frame: How Important Are Frames for Survey Response." *Annual Meetings of the American Political Science Association* Sept. 2-5, 1999.

Alvarez, R. M. & J. Brehm. 2002. *Hard Choices, Easy Answers.* Princeton, NJ: Princeton University Press.

Alvarez, R. M., R. P. Sherman & C. VanBeselaere. 2003. "Subject Acquisition for Web-based Surveys." *Political Analysis* 11:23–43.

Baddeley, A. D. 1994. "Working Memory: The Interface between Memory and Cognition." In *Memory Systems*, ed. D. L. Schacter & E. Tulving. Cambridge, MA: MIT Press.

Bassili, J. N. & J. Fletcher. 1991. "Response-Time Measurement in Survey Research: A Method for CATI and a New Look at Non-Attitudes." *Public Opinion Quarterly* 60:390–399.

Becker, S. L. 1954. "Why an Order Effect." *Public Opinion Quarterly* 18:271–278.

Berrens, R., A. Bahora, H. Jenkins-Smith & D. Weimer. 2003. "The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples." *Political Analysis* 11:1–22.

Blau, P. M. 1964. *Exchange and Power in Social Life.* New York, NY: Wiley.

Bradburn, N. M., L. Rips & S. Shevell. 1987. "Answering Autobiographical Questions: The Impact of Memory and Inference on Survey." *Science* 236:157–161.

Brehm, J. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation.* Ann Arbor, MI: The University of Michigan Press.

Bureau of Labor Statistics. 2002. "Current Population Survey: Design and Methodology." *U.S. Department of Labor,* http://www.bls.census.gov/cps/tp/tp63.htm.

Cacioppo, J. T. & W. L. Gardner. 1999. "Emotion." *Annual Review of Psychology* 30:191–214.

Cannell, C., P. Miller & L. Oksenberg. 1981. "Research on Interviewing Techniques." In *Sociological Methodology*, ed. S. Leinhard. San Francisco, CA: Jossey-Bass pp. 389–437.

Church, A. H. 1992. "Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis." *Public Opinion Quarterly* 57:62–79.

Clark, H. H. 1985. "Language Use and Language Users." In *The Handbook of Social Psychology: Volume 2. Special Fields and Applications*, ed. G. Lindzey & E. Aronson. New York, NY: Random House pp. 179–231.

Clark, H. H. & M. F. Schober. 1992. "Asking Questions and Influencing Answers." In *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, ed. J. M. Tanur. New York, NY: Wiley pp. 15–48.

Cochran, W. 1953. *Sampling Techniques*. Third ed. New York, NY: John Wiley and Sons.

Converse, P. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. D. Apter. New York, NY: Free Press pp. 206–261.

Conway, M. A. 1996. "Autobiographical Knowledge and Autobiographical Memories." In *Remembering our Past*, ed. C. C. Rubin. New York, NY: Cambridge University Press pp. 67–93.

Couper, M. P. 2000. "Web Surveys: A Review of Issues and Approaches." *Public Opinion Quarterly* 64:464–494.

Couper, M. P., M. W. Traugott & M. J. Lamias. 2001. "Web Survey Design and Administration." *Public Opinion Quarterly* 65:230–253.

Dickinson, J. R. & E. Kirzner. 1985. "Questionnaire Item Omission as a Function of Within-Group Question Position." *Journal of Business Research* 13:71–75.

Dillman, D. A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York, NY: Wiley-Interscience.

Dillman, D. A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. Second ed. New York, NY: John Wiley and Sons.

Dillman, D. A., T. L. Brown, J. E. Carlson, E. H. Carpenter, F. O. Lorenz, Robert Mason, John Saltiel & Roberta L. Sangster. 1995. "Effects of Category Order on Answers in Mail and Telephone Surveys." *Rural Sociology* 60:674–687.

Einstein, A. 1989. "Three Rules of Work." In *Words of Wisdom*, ed. William Safire & Leonard Safir. New York, NY: Fireside.

Fazio, R. H. & C. Williams. 1986. "Attitude Accessibility as a Moderator of the Attitude-Perception and Attitude-Behavior Relations: An Investigation of the 1984 Presidential Election." *Journal of Personality and Social Psychology* 51:505–514.

Fletcher, J. F. & M. C. Chalmers. 1991. "Attitudes of Canadians toward Affirmative Action: Opposition, Value Pluralism and Nonattitudes." *Political Behavior* 12:67–96.

Friedman, W.J. & A.J. Wilkins. 1985. "Scale Effects in Memory for the Time of Events." *Memory And Cognition* 13:168–175.

Gallegos, J. G. 1974. "An Experiment in Maximizing Response to Telephone Interviews Through Use of a Preliminary Letter, Based on the Principles of Exchange." *Department of Sociology,* Washington State University, Working Paper.

Gamson, W. A. & A. Modigliani. 1987. "The Changing Culture of Affirmative Action." In *Research in Political Sociology, Volume 3*, ed. Richard D. Braungart. Greenwich, CT: JAI Press.

Gosnell, H. F. 1927. *Getting-Out-The-Vote: An Experiment in the Stimulation of Voting.* Chicago, IL: University of Chicago Press.

Goyder, J. C. 1987. *The Silent Majority: Nonrespondents on Sample Surveys.* Boulder, CO: Westview Press.

Graesser, A. C., M. Singer & T. Trabsso. 1994. "Constructing Inferences During Narrative Comprehension." *Psychological Review* 101:1055–1124.

Green, M. C., J. A. Krosnick & A. L. Holbrook. 2002. "The Survey Response Process in Telephone and Face-to-Face Surveys: Differences in Respondent Satisficing and Social Desirability Response Bias." *National Elections Studies,* 1982 Method Comparison Project, Working Paper.

Greene, W. H. 2000. *Econometric Analysis.* New York, NY: Prentice Hall.

Grice, H. P. 1989. *Studies in the Way of Words.* Cambridge, MA: Harvard University Press.

Groenendijk, J. & M. Stokhof. 1997. "Questions." In *Handbook of Logic and Language*, ed. J. van Bengthem & A. ter Meulen. Amsterdam: Elsevier pp. 1055–1124.

Groves, R. M. 1989. *Survey Errors and Survey Costs.* New York, NY: John Wiley and Sons.

Henson, R. N. A. 1998. "Short-Term Memory for Serial Order: The Start-End Model." *Cognitive Psychology* 36:73–137.

Higginbotham, J. 1996. "The Semantics of Questions." In *The Handbook of Contemporary Semantics Theory*, ed. S. Lappin. Oxford: Blackwell pp. 361–384.

Hippler, H., N. Schwarz & S. Sudman, eds. 1987. *Social Information Processing and Survey Methodology.* New York, NY: Springer-Verlag.

Holbrook, A., J. A. Krosnick, P. S. Visser, W. L. Gardner & J. T. Cacioppo. 2001. "Attitudes Toward Presidential Candidates and Political Parties: Initial Optimism, Inertial First Impressions, and a Focus on Flaws." *American Journal of Political Science* 45:930–950.

Huckfeldt, R. & J. Sprague. 2000. *Indianapolis-St. Louis Election Study, 1996–1997* [computer file]. First ICPSR ed. Ann Arbor, MI: Conducted by the Indiana University Center for Survey Reseach, Bloomington, IN [producer]. Inter-university Consortium for Political and Social Research [distributor].

Huckfeldt, R., J. Sprague & J. Levine. 2000. "Dynamics of Collective Deliberation in the 1996 Election: Campaign Effects on Access, Certainty and Accuracy." *American Political Science Review* 94:641–652.

Jabine, T., M. Straf, J. Tanur & R. Tourangeau, eds. 1984. *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines.* Washington, DC: National Academy Press.

James, J. M. & R. Bolstein. 1992. "Large Monetary Incentives and Their Effects on Mail Survey Response Rates." *Public Opinion Quarterly* 54:346–361.

Jobe, J. B. & E. F. Loftus. 1991. "Cognitive Aspects of Survey Methodology." *Applied Cognitive Psychology, Special issue* 91.

Johnson, T. & S. McLaughlin. 1990. "GMAT Registrant Survey Design Report." *Los Angeles, CA* Los Angeles Graduate Admission Council.

Kelley, S. Jr. & T. W. Mirer. 1974. "The Simple Act of Voting." *American Political Science Review* 61:572–591.

Key, V.O. Jr. 1966. *The Responsible Electorate.* New York, NY: Vintage Books.

Kiewiet, D. R. N.d. *Personal Conversation with Author.* May 10, 2004.

Kinder, D. R. & L. M. Sanders. 1990. *Divided by Color: Racial Politics and Democratic Ideals.* Chicago, IL: University of Chicago Press.

King, G. & L. Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis* 9:137–163.

King, G., M. Tomz & J. Wittenberg. 1999. "CLARIFY: Software For Interpreting and Presenting Statistical Results. Version 1.2.1." *Harvard University.* http://gking.harvard.edu/.

King, G., M. Tomz & J. Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 33:347–361.

Kolodner, J. 1985. "Memory for Experience." In *The Psychology of Learning and Motivation*, ed. G. H. Bower. Vol. 19 Orlando, FL: Academic Press pp. 1–57.

Krosnick, J. A. 1991. "Response Strategies for Coping with the Cognitive Demands Of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213–236.

Krosnick, J. A., A. L. Holbrook, M. K. Berent, R. T. Carson, W. M. Hanemann, R. J. Kopp, R. C. Mitchell, S. Presser, P. A. Ruud, V. K. Smith, W. R. Moody, M. C. Green & M. Conaway. 2002. "The Impact of "No Opinion" Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice?" *Public Opinion Quarterly* 66:371–403.

Krosnick, J. A. & D. F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement." *Public Opinion Quarterly* 51:201–219.

Krosnick, J. A. & D. F. Alwin. 1991. "The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Characteristics." *Sociological Methods and Research* 20:139–181.

Krosnick, J. A. & L. Chang. 2000. "A Comparison of the RDD Telephone Survey Methodology With Internet Survey Methodology." *Ohio State University,* http://www.psy.ohio-state.edu/social/krosnick.htm.

Kurbat, M., S. Shevell & L. Rips. 1994. "A Year's Memories: Effects of the Calendar on Recalled Autobiographical Events." *Memory and Cognition* 26:532–552.

Lehnert, W. 1978. *The Process of Question Answering: A Computer Simulation of Cognition.* Hillsdale, NJ: Erlbaum.

Letterman, D. September 27, 2002. "The David Letterman Show." *CBS* Show 1877.

Lewandowsky, S. & B. Murdock. 1989. "Memory for Serial Order." *Psychological Review* 96:25–57.

Lewis, D. K. 1975. "Adverbs of Quantification." In *Formal Semantics of Natural Language*, ed. E. L. Keenan. New York, NY: Cambridge University Press pp. 3–15.

Lippmann, W. 1922. *Public Opinion.* New York, NY: Macmillan.

Lodge, M., K. McGraw & P. Stroh. 1989. "An Impression-Driven Model of Candidate Evaluation." *American Political Science Review* 83:399–419.

Mas-Colell, A., M. D. Whinston & J. R. Green. 1995. *Microeconomic Theory.* Oxford: Oxford University Press.

Mayo, E. 1933. *The Human Problems of an Industrial Civilization.* New York, NY: MacMillan.

McFadden, D. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, ed. P. Zarembka. New York, NY: Academic Press chapter 4, pp. 105–142.

Miller, W. E. & The National Election Studies/Center for Political Studies. 1999. *American National Election Study, 1980* [computer file]. Second ICPSR ed. Ann Arbor, MI: Conducted by University of Michigan, Center for Political Studies. Inter-university Consortium for Political and Social Research [producer and distributor].

Miller, W. E. & The National Election Studies/Center for Political Studies. 2000. *American National Election Study, 1984* [computer file]. Third ICPSR ed. Ann Arbor, MI: Conducted by University of Michigan, Center for Political Studies. Inter-university Consortium for Political and Social Research [producer and distributor].

Mueller, J. 1973. *War, Presidents and Public Opinion.* New York, NY: Wiley.

Mulligan, K., J. T. Grant, S. T. Mockabee & J. Q. Monson. 2003. "Response Latency Methodology for Survey Research: Measurement and Modeling Strategies." *Political Analysis* 11:289–301.

Narayan, S. & J. A. Krosnick. 1996. "Education Moderates Some Response Effect in Attitude Measurement." *Public Opinion Quarterly* 60:58–88.

National Election Studies (NES). 1995. "Data Quality." *Center for Political Studies,* http://www.umich.edu/ nes/studyres/datainfo/dataqual.htm.

Nietzsche, F. 1883. *Thus Spoke Zarathustra Part 2,* 1920. translated by Thomas Common. 1920. New York, NY: Vikinig Penguin.

Pillemer, D. B., L. R. Goldsmith, L. Krensky, S. N. Kleinman & S. H. White. 1991. "Chapters in Narratives: Evidence from Oral Histories of the First Year in College." *Journal of Narrative and Life History* 1:3–14.

Radford, A. 1997. *Syntactic Theory and the Structure of English: A Minimalist Approach.* New York, NY: Cambridge University Press.

Remington, T. D. 1992. "Telemarketing and Declining Survey Response Rates." *Journal of Advertising Research* pp. 6–7.

Riker, W. H. & P. C. Ordeshook. 1968. "A Theory of the Calculus of Voting." *The American Political Science Review* 62:25–42.

Roper, B. 1986. "Evaluating Polls with Poll Data." *Public Opinion Quarterly* 50:10–16.

Rosch, E. H. & B. B. Lloyd, eds. 1978. *Principles of Categorization.* Hillsdale, NJ: Erlbaum.

Rosenstone, S. J. 1997. *Detroit Area Study, 1989: Political Participation in the Detroit Area* [computer file]. Second ICPSR ed. Ann Arbor, MI: University of Michigan, Department of Sociology, Detroit Area Studies [producer], 1989, Inter-university Consortium for Political and Social Research [distributor].

Rush. 1990. "Freewill." In *Chronicles.* Island Mercury.

Schank, R. C. 1982. *Dynamic Memory.* New York, NY: Cambridge University Press.

Schank, R. C. & R. P. Abelson. 1977. *Scripts, Plans, Goals and Understanding.* Hillsdale, NJ: Erlbaum.

Schillewaert, N., F. Langerak & T. Duhamel. 1998. "Non-probability Sampling for WWW Surveys: A Comparison of Methods." *Journal of the Market Research Society* 40:307–322.

Schuman, H. 1992. "Context Effects: State of the Art/State of the Past." In *Context Effects in Social and Psychological Research*, ed. N. Schwarz & S. Sudman. New York, NY: Springer-Verlag pp. 35–47.

Schuman, H. & S. Presser. 1981. *Questions and Answers in Attitude Surveys.* New York, NY: Academic Press.

Schwarz, N. & S. Sudman, eds. 1996. *Answering Questions.* San Francisco, CA: Jossey-Bass.

Simon, H. A. 1957. *Models of Man.* New York, NY: Wiley.

Smith, A. F. 1991. "Cognitive Processes in Long-Term Dietary Recall." *Vital and Health Statistics, U.S. Government Printing Office* 6.

Smith, A. F., J. B. Jobe & D. Mingay. 1991. "Retrieval from Memory of Dietary Information." *Applied Cognitive Psychology* 5:269–296.

Sniderman, P. M. & H. Brady. 1999. *1998–1999 Multi-Investigator Study* [computer file]. Berkeley, CA: The Survey Research Center, University of California, Berkeley [producer and distributor].

Sperber, D. & D. Wilson. 1986. *Relevance: Communication and Cognition.* Cambridge, MA: Harvard University Press.

Stack, F. & L. Martin. 1987. "Thinking, Judging and Communicating: A Process Account of Context Effects in Attitude Surveys." In *Social Information Processing and Survey Methodology*, ed. H. Hippler, N. Schwarz & S. Sudman. New York, NY: Springer-Verlag pp. 123–148.

Stein, J. Nov. 4, 2002. "The New Politics of Pot." *Time Magazine* pp. 57–66.

Sudman, S., N. M. Bradburn & N. Schwarz. 1995. *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology.* San Francisco, CA: Jossey-Bass.

Tourangeau, R. & K. Rasinski. 1988*a.* "Cognitive Precesses Underlying Context Effects in Attitude Measurement." *Psychological Bulletin* pp. 299–314.

Tourangeau, R. & K. Rasinski. 1988*b*. "Cognitive Processes Underlying Context Affects in Attitude Measurement." *Psychological Bulletin* 103:299–314.

Tourangeau, R., K. Rasinski, N. Bradburn & R. D'Andrade. 1989. "Belief Accessibility and Context Effects in Attitude Measurement." *Journal of Experimental Social Psychology* 25:401–421.

Tourangeau, Roger, Lance J. Rips & Kenneth Rasinski. 2000. *The Psychology of Survey Response.* New York, NY: Cambridge University Press.

Tulving, E. 1983. *Elements of Episodic Memory.* Oxford: Oxford University Press.

VanBeselaere, C. 2001. "Sample Selection Issues in Internet Surveys." *California Institute of Technology.* http://www.hss.caltech.edu/ carla/.

Washington, G. April 21, 1778. "Letter to John Banister." In *George Washington Papers at the Library of Congress, 1741–1799.* Libary of Congress.

Wilson, T. D., D. Kraft & D. Dunn. 1989. "The Disruptive Effects of Explaining Attitudes: The Moderating Effects of Knowledge about the Attitude Object." *Journal of Experimental Psychology: General* 125:249–263.

Wilson, T. D. & S. Hodges. 1992. "Attitudes as Temporary Constructions." In *The Construction of Social Judgments*, ed. L. Martin & A. Tesser. New York, NY: Springer-Verlag pp. 37–66.

Zaller, J. R. 1992. *The Nature and Origins of Mass Opinion.* New York, NY: Cambridge University Press.

Zaller, J. R. & S. Feldman. 1992. "A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences." *American Journal of Political Science* 36:579–616.

Appendix

Appendix A   Data Appendix

A.1   The Caltech Internet Survey Project (CISP)

In January 2000, the California Institute of Technology (Caltech) Internet survey project, a project designed to study the use of the Internet for public opinion surveying, began recruiting respondents for an Internet survey panel.[1] After completing a demographic survey, registered panel members were notified by email about new surveys. Between January 2000 and April 2003, seven different surveys were implemented with respondents drawn from the existing pool of registered subjects and newly recruited subjects. See Alvarez, Sherman & VanBeselaere (2003), for a complete discussion of this survey project and the process by which respondents were recruited to participate in this survey.

A.1.1   March 2001 Matched Internet and Telephone Surveys

In March 2001, matched Internet and telephone surveys were conducted as part of the Caltech Internet Survey Project.[2] The Internet survey began on March 22, 2001, when registered respondents were sent an email asking them to visit the survey website and complete a new survey.[3] At the same time, new respondents were recruited using a subscription method.[4] The telephone survey was conducted by Interviewing Services of America (ISA).[5]

---

[1]The Principal Investigators of the project are R. Michael Alvarez and Robert Sherman. The survey website is available at http://survey.caltech.edu/.

[2]A codebook for this Internet and matched telephone survey can be obtained from http://survey.caltech.edu/research.html under *2001 Winter Telephone Survey* and *2001 Winter Internet Survey.*

[3]Initial emails were sent starting March 22, 2001. Registered participants who had not respondent by April 2, 2001, were sent a reminder email.

[4]The subscription recruitment was provided by ValueClicks' *UltraLeads^{tm}* model. See (Alvarez, Sherman & VanBeselaere 2003), for a detailed discussion of this recruitment method.

[5]The ISA used Random Digit Dialling (RDD) to obtain their survey sample. While this should help ensure that the initial sampling frame is representative, selection bias may be introduced when

The March 2001 Internet survey received 2,815 unique responses. Of these respondents 2,444 could be matched with demographic information.[6] Since this study focused on survey questions related to voting behavior and preferences, we will restrict attention only to the 2,340 respondents who were eligible to vote.[7] Omitting those who did not answer any questions or stopped answering half-way through, the sample size reduces to 2,332. Only 408 of the Internet survey respondents provided responses to ALL the questions in the March 2001 survey. The ISA provided 1,500 completed telephone survey responses.[8] All 1,500 phone survey respondents answered every opinion question but some refused to respond to demographic questions and will thus be excluded from the analysis.[9]

Table A.1 contains the Pearson-χ^2 statistics comparing the distributions of demographic characteristics between the phone and Internet respondents.[10] These results, together with those in Table 8.1, suggest that the Internet sample contains more females and whites. The Internet respondents were less likely to be married and had fewer children. Internet respondents were more likely to have some college eduction or vocational degree. Comparing phone and Internet respondents directly, it appears

---

potential respondents refuse to participate. It took approximately two weeks for ISA to obtain 1500 completed survey responses.

[6]Email addresses were used to match survey responses with demographic data provided during registration. If respondents did not provide identical email addresses on both the registration form and the March 2001 survey they could not be matched. In addition, there appear to be several survey respondents who never registered with the Caltech Internet survey panel.

[7]For the purpose of this study, respondents are considered to be eligible to vote if they are over 18.

[8]For a discussion of the response rate for this survey see (Alvarez, Sherman & VanBeselaere 2003). Also, in the administration of the phone survey, there were 199 suspended interviews for which no data was maintained. In an attempt to match the telephone and Internet survey, the Internet respondents who stopped (or never started) answering the opinion questions are omitted from the analysis.

[9]Over 200 respondents refused to answer a question asking them to indicate how strongly they identified with the party with which they are registered. Refusal rates were also high for questions about educational attainment, marital status, region of residency, and number of children. It is also important to note that Internet respondents were asked to provide demographic information prior to answering the opinion questions while phone survey respondents received the demographic questions at the end of the survey.

[10]The Pearson statistic is $\chi_p^2 = n \sum_r \sum_c \frac{(p_{rc} - p_{0rc}^2}{p_{0rc}}$ where n is the total number of observations, p_{rc} is the estimated proportion for the cell in the rth row and c^{th} column of the table, and p_{0rc} is the estimated proportion under the null hypothesis of independence.

that Internet respondents are younger, but when we restrict attention to phone survey respondents with Internet access this difference disappears.

Table A.1: Compare Distribution of Demographic Variables - Phone vs. Net

Category	Phone vs. Net. χ^2-value	Phone vs. Net p-value	Phone w/Net vs. Net. χ^2-value	Phone w/Net vs. Net p-value
Gender	93.20	0.00	117.95	0.00
Age	149.28	0.00	4.98	0.42
White	48.25	0.00	32.15	0.00
Education	89.81	0.00	63.67	0.00
Marital Status	21.23	0.00	31.23	0.00
Kids	41.86	0.00	13.38	0.00
Region	2.56	0.46	4.63	0.20

Table A.2 contains similar test statistics comparing the telephone and Internet respondents with the CPS respondents.[11] The phone survey respondents appear to differ from the CPS in terms of race, education, number of children, and region of residency. The phone survey consisted of a larger proportion of white and more highly educated respondents. These respondents also had more children and were less likely to reside in the Northeast. The Internet sample has a significantly different distribution for all the demographic characteristics even if we restrict attention to CPS respondents with Internet access.[12] The Caltech Internet sample has more female, young, white and unmarried respondents than the CPS sample. The fact that the Caltech Internet sample is not representative suggests that it is not a random sample from the population of interest.[13] The same, however, may be said about the Caltech telephone survey.

---

[11]The CPS data is weighted by probability of sample inclusion. To account for the survey design, the Pearson-χ^2 statistic is converted to an F statistic with non-integer degrees of freedom.

[12]The p-value for region is only marginally significant at seven percent.

[13]These results suggest that it would be inappropriate to use the Caltech Internet survey data to make substantive statements about the population of interest without accounting for selection bias. See (VanBeselaere 2001) for an analysis of sample selection in the Caltech Internet surveys and methods for correcting these biases. For a general discussion of the sampling issues related to Internet surveys, see (Alvarez, Sherman & VanBeselaere 2003) and (Couper 2000).

Table A.2: Compare Distribution of Demographic Variables - CISP vs. CPS

CISP Category	Phone vs. CPS F-value	Phone vs. CPS p-value	Net vs. CPS w/Net F-value	Net vs. CPS w/Net p-value
Gender	0.12	0.73	245.86	0.00
Age	1.64	0.15	5.17	0.00
White	24.66	0.00	69.28	0.00
Education	47.28	0.00	7.85	0.00
Marital Status	0.03	0.86	153.64	0.00
Kids	636.62	0.00	446.29	0.00
Region	3.74	0.01	2.37	0.07

A.1.2 2003 Attitudes Survey

In January 2003, the Internet Survey Project began the collection of responses to the 2003 Attitudes Survey. This survey focused primarily on collecting attitudes towards marijuana and euthanasia policy. In an attempt to understand the determinants of opinion expressions, additional questions that tried to measure respondents' values and predispositions were incorporated.

To supplement the survey data, Caltech students were selected to participate in in-depth interviews after having completed the survey. In total, eleven students were interviewed during this process and their survey data was supplemented with the interview results.

A.2   1980 and 1984 NES Variables

A.2.1   The 1980 NES Pre-Election Survey Variables

The following demographic variables were used form the 1980 NES Pre-Election Survey:

- Census region (V800023)
- Education (V800436)

- Respondent's gender (V800720)

- Respondent's race (V800720)

- Respondent's age (V800408)

- Respondent's party identification (V800225)

- Social class (V800691)

- Parental status (does respondent have children) (V800410)

- Strength of party affiliation (v800226)

Several of the respondent characteristics included in the analysis were created based on the above demographic information. Below is an outline of how some of these variables were created.

- High Education is a dummy variable indicating that respondents's education level was either Some College, Junior College, B.A., or Advanced Degree;

- High Poll Attention is a dummy indicating whether a respondent has read or heard results from any national elections polls (v800219);

- Candidate Awareness was derived from questions asking respondents to identify congressional candidates in their district (V800826). The following values were assigned to this variable:

    1 if the respondent was able to correctly identify at least one congressional candidate and this candidate's party affiliation;

    $\frac{1}{2}$ if respondent was able to correctly identify at least one congressional candidate but unable to determine this candidate's party affiliation;

    0 if the respondent was unable to correctly identify even one congressional candidate

- TV Attention to Gov't is a dummy variable indicating that respondents self-reported paying a great deal of attention to television news about government or politics (V800218)

- Interested in Politics is a dummy variable capturing respondents who self-reported that they were very much interested in the campaign (V800053)

Table A.3 contains some summary statistics for the demographic variables from the 1980 NES Pre-Election Survey used in the analysis contained in Chapter 6.

Table A.3: Summary Statistics for Variables in the NES 1980 NR Analysis

Variable Name	Sample Size	Mean	Standard Deviation	Min	Max
Region	1614	2.48	1.01	1	4
Northeast	1614	0.21	0.41	0	1
North Central	1614	0.27	0.44	0	1
South	1614	0.35	0.48	0	1
West	1614	0.17	0.38	0	1
Education	1610	5.59	2.49	1	10
Higher Education	1610	0.37	0.48	0	1
Female	1614	0.57	0.50	0	1
White	1611	0.87	0.33	0	1
Age	1612	44.54	18.06	18	96
Over 75	1612	0.06	0.24	0	1
Have kids	1614	0.71	0.45	0	1
Party Identification	1613	3.13	1.66	1	5
Republican	1613	0.22	0.42	0	1
Independent	1613	0.24	0.43	0	1
No Party Affiliation	1613	0.12	0.32	0	1
Other Party	1613	0.00	0.07	0	1
Democrat	1613	0.41	0.49	0	1
Candidate Awareness	1614	0.24	0.43	0	1
TV Attention to Gov't	1466	0.46	0.50	0	1
Interested in Politics	1565	0.30	0.46	0	1

A.2.2 The 1984 NES Pre-Election Survey Variables

The following demographic variables were used form the 1984 NES Pre-Election Survey:

- Census region (V840009)
- Education (V840438)
- Respondent's gender (V840707)
- Respondent's race (V840708)
- Respondent's age (V840429)
- Respondent's party identification (V840315)
- Social class (V840685)
- Strength of party affiliation (V840316)

Several of the respondent characteristics included in the analysis were created based on the above demographic information. Below is an outline of how some of these variables were created.

- High Education is a dummy variable indicating that respondents's education level was either Some College, Junior College, B.A., or Advanced Degree;

- Child(ren) in house is an indicator of the presence of children in the house derived from questions asking respondents to identify the number of children in the house within a variety of different age ranges (V840071, V840072, v840073, V840074);

- Candidate Awareness was derived from questions asking respondents to identify congressional candidates in their district (V840741). The following values were assigned to this variable:

  1 if the respondent was able to correctly identify at least one congressional candidate and this candidate's party affiliation;

  $\frac{1}{2}$ if respondent was able to correctly identify at least one congressional candidate but unable to determine this candidate's party affiliation;

  0 if the respondent was unable to correctly identify even one congressional candidate;

- Discuss Politics Frequently is a dummy variable indicating that the respondent self-reports discussing politics more than once a week (v840106)

- TV Attention to Gov't is a dummy variable indicating that respondents self-reports paying a great deal of attention to television news about government or politics (V840109);

- Interested in Politics is a dummy variable capturing respondents who self-reported that they were very much interested in the campaign (V840075)

Table A.4 contains some summary statistics for the demographic variables from the 1984 NES Pre-Election Survey used in the analysis contained in Chapter 6.

Table A.4: Summary Statistics for Variables in NES 1984 NR Analysis

Variable Name	Sample Size	Mean	Standard Deviation	Min	Max
Region	2257	2.56	1.02	1	4
Northeast	2257	0.18	0.39	0	1
North Central	2257	0.27	0.45	0	1
South	2257	0.33	0.47	0	1
West	2257	0.21	0.41	0	1
Education	2243	5.78	2.44	1	10
Higher Education	2243	0.41	0.49	0	1
Female	2257	0.56	0.50	0	1
White	2245	0.87	0.34	0	1
Age	2237	44.41	18.08	17	94
Over 75	2237	0.07	0.26	0	1
Child in House	2257	0.42	0.49	0	1
Party	2236	2.95	1.68	1	5
Republican	2236	0.27	0.44	0	1
Independent	2236	0.25	0.43	0	1
No Party Affiliation	2236	0.10	0.31	0	1
Other Party	2236	0.00	0.04	0	1
Democratic	2236	0.37	0.48	0	1
Strength of Partisanship	2257	0.29	0.45	0	1
Social Class	2191	2.87	1.76	0	8
TV Attention to Gov't	2164	0.35	0.48	0	1
Political Informedness	2257	0.24	0.41	0	1
Interested in Politics	2250	0.28	0.45	0	1

A.3 1998–1999 Multi-Investigator Data

The 1998–1999 Multi-Investigator Study was a national random-digit telephone survey administered by the Survey Research Center of the University of California, Berkeley that attempted to collect data about how people form their opinions. As part of this effort, a series of six policy questions concerning affirmative action, welfare, and immigrant rights were followed by questions asking respondents to indicate how difficult these questions were to answer. Active response times were also collected with each of these questions.

Variables used from the 1998–1999 Multi-Investigator Study:

al1b When it comes to setting aside a certain number of government construction contracts for businesses that are owned and operated by minorities, are you for or against this?

al1d How do you feel about allowing state universities to have flexible admissions standards in order to promote racial diversity? Are you for or against this?

al2b How do you feel about requiring women with children to work at a job in order to stay on welfare? Are you for or against this?

al2d Are you for or against limiting the number of months a woman with children can stay on welfare?

al3b How do you feel about allowing legal immigrants from other countries, who are here legally, to receive welfare in the U.S.? Are you for or against this?

al3d How do you feel about providing public education to the children of illegal immigrants who are in this country illegally? Are you for or against this?

In the MI1998 survey respondents were randomly assigned to one of three groups. Group one was asked the questions about set-asides, flexible admission standards, workfare, and welfare limits; group two: set-asides, flexible admission standards, welfare for legal immigrants, and education for illegal immigrants; group three: workfare, welfare limits, welfare for legal immigrants, and education for illegal immigrants. Table A.5 contains summary measures of the difficulty ratings provided for these questions.

Table A.5: Summary of Difficulty Ratings by Question

Question	Observations	Mean	Standard Deviation	Min	Max
Set-Asides	709	1.88	1.00	1	4
Flex. Admit	709	1.86	1.03	1	4
Workfare	706	1.85	1.05	1	4
Welfare Limits	706	1.87	1.02	1	4
Welfare for Legal Immigrants	719	1.80	0.97	1	4
Education for Illegal Immigrants	719	1.72	0.96	1	4

Table A.6 contains the results from regressing the log of response times for the question about minority set-asides on a variety of variables that might affect response times (Number of observations = 614 and Adjusted R-squared = 0.22).

Table A.6: Regression on Log of Response Times for the Minority Set-Asides Question

	Coef.	Std.Err.	P-Value
time to answer all other issue questions	0.19	0.05	0.00
time to answer question about trusting gov't	0.00	0.00	0.58
time to answer question about global politics	0.00	0.00	0.56
time to answer trust question	0.00	0.00	0.36
versional1b	0.25	0.11	0.03
versional1c	0.02	0.11	0.88
versional1d	-0.05	0.11	0.64
versional1e	0.01	0.11	0.91
versional1f	0.03	0.11	0.80
versional1g	0.18	0.11	0.10
question difficulty	0.30	0.03	0.00
political informedness	-0.02	0.02	0.32
education - categorical	-0.10	0.03	0.00
working	-0.03	0.08	0.72
ideology	0.01	0.01	0.45
party identification	0.00	0.02	0.85
strength of party identification	0.03	0.07	0.69
age	0.00	0.00	0.98
female	0.07	0.06	0.25
white	0.01	0.08	0.88

Table A.6 (Continued)

	Coef.	Std.Err.	P-Value
married	0.01	0.07	0.91
have kids	0.09	0.08	0.30
income summary	0.00	0.00	0.61
urban	0.01	0.02	0.52
religion	0.02	0.01	0.09
northeast	0.13	0.09	0.17
south	0.09	0.08	0.28
midwest	0.11	0.09	0.18
authoritarian measure	0.03	0.08	0.70
support women's rights	-0.04	0.04	0.41
fear of god measure	-0.03	0.04	0.47
support reducing wealth gap	-0.05	0.03	0.12
support gov't assistance to blacks	-0.02	0.03	0.56
interviewer2	-1.02	0.88	0.25
interviewer3	-0.20	0.54	0.71
interviewer4	-0.13	0.56	0.82
interviewer5	-0.20	0.60	0.74
interviewer6	0.06	0.55	0.91
interviewer7	-0.36	0.54	0.51
interviewer8	0.00	0.61	1.00
interviewer9	-0.35	0.54	0.52
interviewer10	-0.21	0.53	0.70
interviewer11	-0.34	0.55	0.54
interviewer12	-0.39	0.54	0.46
interviewer13	-0.42	0.55	0.45
interviewer report—respondent could hear	0.01	0.09	0.91
interviewer report—respondent was able	0.02	0.06	0.72
interviewer report—confidence in report	0.08	0.15	0.60
interviewer report—number of refusal	0.04	0.05	0.35
Constant	4.17	0.80	0.00

Table A.7 contains the results from a logit analysis of shirking on the question about minority set-asides as measured by the outlier response timer measure.[14] The dependent variable is an indicator of whether a response was determined to be associated with shirking.

---

[14]This question is examined because it has the highest level of shirking as identified by this measure of shirking.

Table A.7: Logit Analysis of Outlier Shirking on the Question about Government Set-Asides for Minorities

	Coef.	Std. Err.
High Political Informedness	0.56	0.52
Female	0.42	0.71
Education - Categorical	-0.22	0.19
Age	-0.000.01	
Difficulty Rating for Set-Asides Q	0.04	0.21
Constant	-2.68*	1.05
Number of Observations	607	
Pseudo R-squared	0.01	

* Statistically significantly different from zero at the $p \leq 0.05$ level.

Table A.8 contains the estimates from a nested probit analysis of support for flexible university admissions standards which includes the group level measure of shirking.

Table A.8: Nested Probit Analysis of Support for Flexible Admissions

	Coef.	Std. Err.	P-Value
Political Informedness	-0.04	0.03	0.29
Female	0.19	0.11	0.10
White	-0.10	0.15	0.51
Democrat	0.23	0.13	0.08
Republican	-0.10	0.14	0.50
Have Children	0.09	0.13	0.50
Support Women's Rights	0.07	0.08	0.37
Support Reducing Wealth Gap	0.07	0.06	0.26
Support Gov't Help for Blacks	0.30	0.06	0.00
Shirking*Political Informedness	-0.02	0.10	0.85
Shirking*Female	-0.03	0.41	0.94
Shirking*White	-0.51	0.55	0.35
Shirking*Democrat	-0.44	0.42	0.29
Shirking*Republican	-1.06	0.57	0.06
Shirking*Have Children	-0.39	0.41	0.33
Shirking*Support Women's Rights	-0.01	0.25	0.98
Shirking*Support Reducing Wealth Gap	-0.03	0.20	0.89
Shirking*Support Gov't Help for Blacks	0.10	0.23	0.66
Indicator of Shirking	0.81	0.78	0.30
Constant	-0.36	0.23	0.12
Number of Observations	647		
Pseudo R-Squared	0.10		

A.4 Indianapolis–St. Louis Election Study, 1996–1997

The Indianapolis–St. Louis Election Study, 1996–1997, was conducted by the Indiana University Center for Survey Research and is primarily concerned with patterns of social communication and influence over the course of the 1996 United States presidential election campaign. This research, which began in March of 1996 and ended in early January of 1997, involved computer-assisted telephone interviewing (CATI). In addition to collecting demographic and political opinion data, latent response times were collected for each of the following questions:

- Whose responsibility is it to determine if a law is constitutional or not?

- What are the first ten amendments to the constitution called?

- How much of a majority is required for the U.S. Senate and House to override a presidential veto?