The analysis of survey data with framing effects

The Analysis of Survey Data with Framing Effects

Jacob Goldin Daniel Reck*

September 11, 2017

Abstract

A well-known difficulty in survey research is that respondents' answers to questions can depend on arbitrary features of a survey's design, such as the wording of questions or the ordering of answer choices. In this paper we describe a novel set of tools for analyzing survey data characterized by such *framing effects*. We show that the conventional approach to analyzing data with framing effects – randomizing survey-takers across frames and pooling the responses – generally does not identify a useful parameter. In its place, we propose an alternative approach and provide conditions under which it identifies the responses that are unaffected by framing. We also present several results for shedding light on the population distribution of the individual characteristic the survey is designed to measure.

*Goldin: Stanford Law School, email:jsgoldin@law.stanford.edu. Reck: Department of Economics, University of California Berkeley, email: dreck@berkeley.edu. For valuable comments, we thank Angus Deaton, Edward Freeland, Michael Gideon, Allyson Holbrook, Michael Frakes, Daniel Ho, Bo Honore, David Lee, Yair Listokin, Charles Manski, Jeffrey Rachlinski, Alex Rees-Jones, Maya Sen, Joel Slemrod, Martin Wells, Justin Wolfers, and participants in seminars at the University of Michigan, Princeton University, and the Junior Empirical Legal Studies workshop at Cornell Law School. All errors are our own.

Introduction

A well-known difficulty in survey research is that how survey-takers respond to a question may depend on seemingly arbitrary details about how the question is asked. For example, respondents may answer differently depending on the order of questions (Moore, 2002; Deaton, 2012), the order in which answer choices are listed (Holbrook et al., 2007), the grouping of responses into categories (Schwarz, 1990), or any number of minor variations in the manner in which questions or responses are worded (Schuman and Presser, 1981; Chong and Druckman, 2007). Such *framing effects* arise in many contexts; large literatures in psychology, political science, communications, and marketing are devoted to documenting and explaining their presence.

Despite the attention paid to framing effects in recent decades, the range of practical solutions available to survey researchers remains limited. The conventional wisdom is that when framing effects are unavoidable, researchers should balance them by randomly assigning an equal number of survey-takers to each version of the survey questionnaire and then analyze the pooled responses.[1] Some researchers acknowledge problems with this approach but note the lack of better alternatives (e.g., Schwarz and Oyserman, 2001).

In this paper we present a simple empirical approach for the analysis of survey data characterized by framing effects.[2] Our results highlight important shortcomings with the conventional pooling approach. In its place, we propose several alternatives and argue they are more likely to shed light on questions of interest to the researcher.

Our formal analysis focuses on binary-response questions in which an arbitrary feature of the survey – the *frame* – affects the responses of a subset of survey-takers. We assume that each respondent is observed answering a given survey question only once, under one of two possible frames. We label respondents as *consistent* if they would

[1] The following statements are typical of the literature: "Randomization ... does not reduce the impact of context at the level of individual respondents. It simply ensures that these influences result in random noise rather than systematic bias in the sample as a whole." (Sudman, Bradburn and Schwarz, 1995). "Our findings suggest that survey organizations should routinely rotate the order of response choices to guard against creating bias in results." (Holbrook et al., 2007). "Acquiescence bias can be reduced by balancing scales so that the affirming response half the time is in the direction of the construct and half the time is in the opposite direction (e.g. six agree/disagree items on national pride, with the patriotic response matching three agree and three disagree responses)." (Presser et al., 2004).

[2] In other work, we derive analogous results to study the identification of decision-makers' preferences in settings characterized by inconsistent choice data (Goldin and Reck, 2015). The goal of the current paper is to apply this approach to survey data in which framing effects are present and to use these results to assess the conventional approach for dealing with framing effects in survey research.

select the same response under both frames and *inconsistent* if their response would vary depending on the frame under which the question is asked. This potential outcomes framework has been widely applied in causal inference analysis (e.g., Holland, 1986; Imbens and Angrist, 1994). In contrast to that literature, our goal is not to estimate the *effect* of the treatment (i.e., the frame) on a variable but rather to recover the distribution of a variable after removing the treatment's effect.

Considering the problem from this light, we show that the pooling approach does not itself eliminate the bias induced by framing effects. Rather, that approach yields a response share that is a weighted sum of (1) the response share of the consistent respondents, and (2) the fraction of respondents assigned to each frame, typically 0.5, where the weights depend on the fraction of respondents that are affected by the frame. Outside a narrow set of applications, this weighted sum is unlikely to be the parameter of interest to the researcher. In its place, we propose several alternatives and describe the conditions under which they are valid.

First, we consider identification of the consistent response share – the response share for the subset of respondents who are unaffected by the frame. Consistent responses are most likely to be of primary interest to the researcher in settings where the survey is designed to measure some attitude or belief and where frame-varying responses indicate an incoherent or unformed opinion about the question being asked. For example, a researcher conducting an opinion survey may wish to exclude the responses of a survey-taker who simply states that she prefers whichever answer is presented first. We show that under the assumption that the frame affects respondents in a uniform direction, an assumption we label *frame monotonicity*, it is straightforward to estimate the consistent response share from the data. Intuitively, only consistent respondents ever select an answer that is "against the frame" – for example, choosing the second option when the answer order favors the first option. Consequently, examining the share of respondents who respond against the frame sheds light on the share of consistent respondents selecting each response.

The plausibility of frame monotonicity varies by setting. When the assumption fails, the consistent response share is only partially identified. Importantly, however, we show that when frame monotonicity is assumed erroneously, the approach we propose for identifying the consistent response share will be biased, but will be *less* biased than the conventional pooling approach.

In other applications, even the respondents who answer inconsistently will have well-defined (but unobservable) answers to the question being asked that the researcher

hopes to recover. Plainly, in at least some cases, such as when the researcher is seeking to learn about respondents' past behavior, a "true" answer to the survey question exists for even those individuals whose responses depend on the frame. For example, suppose a researcher samples respondents from a population and asks them how much television they watch per week, but that the wording of the answer choices affects responses (Schwarz, 1990). If the goal of the survey is to learn the television habits of the population, the researcher will be interested in the behavior of both the consistent and inconsistent respondents.

We provide additional results for applications like this – where the goal is to recover the distribution of a characteristic among all respondents (consistent and inconsistent alike). First, we describe worst-case bounds for this parameter. Second, we describe a new re-weighting technique that exploits respondents' observable covariates to extrapolate from the consistent respondents to the full population. Intuitively, the answers for the inconsistent respondents are treated as missing data, which can be imputed using the answers of the consistent respondents with similar observable covariates.

The remainder of the paper proceeds as follows. Section I describes our basic assumptions and provides notation to study framing effects. Section II analyzes the conventional pooling approach from this perspective and highlights its shortcomings. Section III derives results for estimating the consistent response share. Section IV provides results for identifying demographic characteristics of the consistent and inconsistent respondents. Section V considers settings in which even the inconsistent respondents are assumed to have some value of the variable of interest to the survey researcher. Section VI provides an extended illustration of our proposed approach. Section VII concludes.

I. Model and Notation

Consider a set of individuals, denoted by i. Each individual answers a binary question, with response indicated by $y_i \in \{0, 1\}$. Many survey questions have this binary form. For example, y_i might denote whether one agrees with a particular statement or supports a particular candidate for office.

The purpose of the survey question is to elicit information about some individual characteristic, which we refer to as the *characteristic of interest,* y_i^*. For each individual, y_i^* is either 0, 1, or undefined. For example, if a survey question asks whether a respondent believes a particular statement to be true, y_i^* would indicate whether or

not the respondent actually believes the statement in question. In the case that the respondent lacks a belief about the statement, y_i^* would be undefined.

Each individual answers the survey question once, under one of two possible *frames*. Let d_i denote the frame to which individual i is assigned, $d_i \in \{d_0, d_1\}$. Let $y_i(d)$ denote how i would answer if asked the survey question under frame d. That is, $y_i(d) = y_i$ when $d_i = d$. A framing effect occurs when a change in the presentation of a question is associated with a change in at least some respondents' answers, and where that change in presentation is, by assumption, unrelated to the information the survey researcher seeks to measure. More abstractly, a survey question is designed to elicit information about a characteristic of a respondent (y_i^*), and the defining feature of a frame is that the characteristic of interest does not depend on the frame, i.e., y_i^* does not depend on d. For example, a researcher seeking to learn whether an individual supports a proposed policy would typically hope to measure an attitude that is unrelated to whether the survey question is phrased positively or negatively, and to the order in which the answer choices are presented. Similarly, the presence of interviewer effects – in which interviewer characteristics affect respondents' answers – or survey method effects – such as differences in the answers of respondents taking the survey in-person versus online – will typically satisfy the definition of a frame. This definition of a frame is similar to the definition proposed by Salant and Rubinstein (2008).

Let $c_i \in \{0, 1\}$ indicate whether an individual's response would be consistent across frames: $c_i \equiv \mathbf{1}\{y_i(d_0) = y_i(d_1)\}$. The fraction of respondents affected by the frame is given by $P(c_i = 0) = 1 - E[c_i]$. Let Y_0 denote the mean response among individuals observed under frame d_0, $Y_0 = E[y_i|d_i = d_0]$, and let $Y_1 = E[y_i|d_i = d_1]$.

Throughout the paper, we assume that population moments such as Y_0 and Y_1 are directly observable to the researcher, setting aside issues of sampling error. This simplifying assumption allows us to abstract from issues of statistical inference. In practice, researchers wishing to apply our results would replace population moments with their finite-sample analogues. For example, one would replace Y_0 with the mean response observed among survey-takers in d_0 and Y_1 with the mean response observed among survey-takers in d_1. It is straightforward to calculate the associated standard errors for these estimates using the delta method (see Supplementary Appendix B for derivations). Section VI illustrates how one would apply our results to observed survey data exhibiting framing effects.

Finally, we assume that framing effects are observed, so $Y_0 \neq Y_1$. Without loss of generality, we assume $Y_1 > Y_0$.

Examples

The above notation accommodates a number of types of survey questions and framing effects. We provide several examples here.

Answer Order In a Gallup Organization (2003) telephone poll on opinions concerning the Iraq war, certain respondents were asked the following question:

> Do you think the Bush administration (1) Provided information about Iraq's weapons of mass destruction that was accurate, or (2) Provided information about Iraq's weapons of mass destruction it thought was accurate, but turned out to be inaccurate?

The order of answers (1) and (2) was reversed for half of the respondents. The response share was observed to vary depending on which answer choice was presented first. When (1) was presented first, 51% of respondents reported believing that the Bush administration provided accurate information. When (2) was presented first, only 41% reported believing this statement.

Question Order Moore (2002) documents question-order effects in a 1997 Gallup survey. Respondents were asked the following question about both Bill Clinton and Al Gore:

> Do you generally think Bill Clinton [Al Gore] is honest and trustworthy?

Respondents' answers varied depending on which politician was asked about first. When the Clinton question was asked first, 50% of respondents reported thinking Clinton was trustworthy, whereas 57% reported thinking so when the Gore question was asked first. Conversely, 68% reported thinking Gore to be trustworthy when the Gore question was asked first, but only 60% did so when the Clinton question was first.

Question Wording (Acquiescence Bias) Schuman and Presser (1981) document the presence of numerous survey framing effects, including one that they refer to as "acquiescence bias" – the tendency of respondents to agree with the question being asked, regardless of the content. In one of their studies, individuals were randomly assigned to one of two versions of the same question. The fraction agreeing with the stated proposition is provided in brackets.

> Individuals are more to blame than social conditions for crime and lawlessness in this country [60%]
>
> Social conditions are more to blame than individuals for crime and lawlessness in this country [57%]

Question Wording (Gain/Loss Framing) Tversky and Kahneman (1981) asked experimental participants about their willingness to accept risky policies that have the potential to save large numbers of lives. Respondents were asked two versions of a question after being randomly divided into the gain frame and the loss frame:

> Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:
>
> [Gain Frame] If Program A is adopted, 200 people will be saved. If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. Which of the two programs would you favor? [Program A: 72%, Program B: 28%]
>
> [Loss Frame] If Program C is adopted 400 people will die. If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.Which of the two programs would you favor? [Program C: 22%, Program D: 78%]
>
> Note that Programs A and C are identical, as are Programs B and D.

These examples illustrate the range of framing effects that have been documented in survey research. In many cases, framing effects are difficult to avoid: questions and responses must be provided to respondents in some order, and they must be worded in some way. The remainder of the paper investigates possible approaches for dealing with framing effects when they arise.

II. The Conventional Pooling Approach

As described above, the conventional approach to analyzing survey data that exhibit framing effects is to randomize respondents evenly across frames and then to proceed using the responses pooled across the frames. For example, political polling surveys, recognizing the potential for response-order effects, typically rotate which major candidate is listed first and which is listed second; however, after this initial randomization, candidate order usually plays no further role in the analysis. This section shows that,

contrary to the conventional wisdom, simply randomizing survey-takers across frames and pooling the responses does not eliminate the bias associated with framing effects.

Suppose that respondents are randomly assigned to each frame, with equal probability of being selected for each. The pooled response share, Y_p, is defined as:

$$Y_p \equiv \frac{Y_1 + Y_0}{2}$$

We will assume that the frame to which an individual is assigned is unrelated to how the individual would respond under either frame:

Assumption 1 (*Frame Exogeneity)* Frame assignment d_i is independent of the vector $(y_i(d_0), y_i(d_1))$.

This assumption is satisfied, for example, if respondents are randomly assigned across frames.

We next assume that but for the observed framing effect, survey responses accurately reflect the characteristic of interest:

Assumption 2 (*Consistent Responses Reflect the Characteristic of Interest)* For each i, $c_i = 1 \implies y_i = y_i^*$.

Assumption 2 implies that the answers to the survey question obtained from the consistent respondents – those whose answers are unaffected by the frame – reflect the characteristic of interest for members of that group. In addition, the assumption implies that if a respondent is consistent, then y_i^* is defined.[3] The assumption represents a weakening of the baseline assumption typically made in survey analysis, namely that every respondent's answer corresponds to the characteristic of interest – i.e. that $y_i = y_i^*$ always. In practice, of course, survey responses may fail to accurately reflect the characteristic of interest for a number of reasons apart from framing effects, such as forgetfulness, deception, or inattentiveness by the survey-taker. We restrict our analysis here to errors caused by framing effects to focus on the key features of our approach.

Finally, Assumption 2 implicitly defines the types of framing effects to which our approach is meant to be applied: those in which framing effects are a "nuisance," and prevent the researcher from recovering the characteristic of interest from the observed survey responses. In other settings, framing effects can actually be desirable, such as

[3]We consider the question of whether y_i^* is defined for inconsistent respondents in Section V, below.

when a researcher employs the Random-Response-Technique (RRT) to obtain higher-quality responses to sensitive questions (Jann, Jerke and Krumpal, 2011; Wolter and Preisendörfer, 2013). In practice, virtually all instances in which the pooling approach is applied are cases in which framing effects are viewed as undesirable.

A useful quantity for understanding the pooling approach is the *response share among consistent respondents*, $E[y_i|c_i = 1]$. Under Assumption 2, this parameter is equal to the mean of the characteristic of interest among the consistent respondents, which we denote by π_c. Formally, $\pi_c = E[y_i^*|c_i = 1] = E[y_i|c_i = 1]$. Although the consistent response share cannot be directly observed from the data, the following proposition provides conditions under which it can be identified.

Proposition 1: The Pooled Response Share

Under Assumptions 1 and 2, the pooled response share identifies the following weighted sum: $Y_p = E[c_i]\pi_c + (1 - E[c_i])\left(\frac{1}{2}\right)$.

Proof By the law of iterated expectations

$$Y_0 = E[y_i(d_0)|c_i = 1, d_i = d_0]\, P(c_i = 1|d_i = d_0) + E[y_i(d_0)|c_i = 0, d_i = d_0]\, P(c_i = 0|d_i = d_0).$$

Applying Assumption 1 yields

$$Y_0 = E[y_i(d_0)|c_i = 1]\, P(c_i = 1) + E[y_i(d_0)|c_i = 0]\, P(c_i = 0).$$

Under Assumption 2, $y_i(d_0) = y_i(d_1) = y_i^*$ when $c_i = 1$. Hence we can use the definition of π_c to write

$$Y_0 = \pi_c\, P(c_i = 1) + E[y_i(d_0)|c_i = 0]\, P(c_i = 0). \tag{1}$$

Similarly, one can show that

$$Y_1 = \pi_c\, P(c_i = 1) + E[y_i(d_1)|c_i = 0]\, P(c_i = 0). \tag{2}$$

Finally, note that when $c_i = 0$, $(y_i(d_0), y_i(d_1)) \in \{(0, 1), (1, 0)\}$, and thus $y_i(d_0) + y_i(d_1) = 1\ \forall i$. Hence,

$$E[y_i(d_0)|c_i = 0] + E[y_i(d_1)|c_i = 0] = 1. \tag{3}$$

Substituting (1) and (2) into the definition of Y_p and applying (3) yields the desired result. ■

When survey respondents are evenly divided between frames, Proposition 1 shows that the pooled response share is a weighted average of the consistent response share and 0.5.[4] To understand the intuition, consider two extreme examples. First suppose that all respondents are consistent, so $E[c_i] = 1$. In that case the pooled response share would yield the share of the population that (consistently) selects $y_i = 1$, $Y_p = \pi_c = E[y_i] = E[y_i^*]$. Next, consider the opposite extreme, in which all respondents are sensitive to the frame, so that $E[c_i] = 0$. In that case, assigning half of the respondents to d_0 and half to d_1 would result in half of the respondents selecting each answer, yielding $Y_p = 0.5$. Finally, when there are both consistent and inconsistent respondents in the population, the pooled response share will simply be the weighted average of these two extremes, where the weights depend on the fraction of respondents in each group.

Proposition 1 has a straightforward interpretation without Assumption 2. The proposition still holds as stated, but with π_c redefined as $E[y_i|c_i = 1]$. That is, the pooled response share is still a weighted sum of the consistent response share and 0.5, but the consistent response share is not necessarily informative about the consistent respondents' characteristic of interest, y_i^*. Thus, the violation of Assumption 2 would not typically support the use of the pooling approach – it just makes it more difficult to interpret the parameter that approach identifies.

We can think of few practical settings in which a researcher would want to identify the parameter identified by the pooled response share. In particular, the pooled response share will depend on the magnitude of the framing effect. For example, a pooled response share close to 0.5 could indicate either that many respondents are affected by the frame or that an equal number of respondents would consistently select each of the available answer options.[5] In the next sections, we describe alternative approaches for dealing with framing effects in survey analysis.

[4]Typically the pooling approach is implemented using equal assignment to both frames. Under frame monotonicity (Assumption 3, below), one can show that assigning some arbitrary fraction θ to d_1 and the rest to d_0 implies $Y_p = E[c_i]\pi_c + (1 - E[c_i])\theta$.

[5]One narrow set of cases in which the pooled response share does identify a useful parameter is when the goal of a survey is to predict a future choice that is characterized by the same framing effect as the survey itself. For example, suppose election polling data exhibits candidate order effects. If actual voting behavior will exhibit the same candidate order effects as the survey, and if the order of candidate names on the actual ballot will itself be randomized, then the pooled response share will estimate the weighted average associated with the actual election results.

III. Identifying the Consistent Response Share

This section considers how researchers might attempt to recover information about π_c – the response share for the subset of survey-takers whose responses are consistent across frames. In contrast to the parameter identified by Y_p, π_c may reflect information that is of primary interest to the researcher. First, π_c provides information about the quantity the survey was designed to measure, y_i^*, in a way that is not mechanically related to the frame in which respondents are observed. Additionally, in the context of survey questions designed to measure respondents' attitudes or beliefs, for example, the inconsistent respondents may lack a meaningful opinion about the question being asked, so that y_i^* is undefined. In such cases, researchers may wish to isolate the responses of the consistent individuals from the responses of those whose answers depend on the frame. Conversely, in situations in which the researcher assumes that all inconsistent respondents do have a well-defined y_i^*, such as questions about past behavior, one may wish to extrapolate from the consistent responses to recover the characteristic of interest for the inconsistent respondents. Identifying π_c is the first step in solving this problem, the remainder of which is considered in Section V.

Identifying π_c would be trivial if the researcher could observe individual respondents under both frames; the inconsistent respondents could be individually identified and their responses discarded.[6] In contrast, when each respondent is only observed under one frame, the degree to which π_c can be identified depends on whether the frame in question affects all respondents in a uniform direction. In particular, consider the following assumption:

Assumption 3 (*Frame Monotonicity*) For each i, $y_i(d_1) \geq y_i(d_0)$.

For example, if one observes that a higher fraction of respondents answer "Yes" when asked version A of a question as compared to version B, frame monotonicity implies that there are no respondents who answer "No" to version A but "Yes" to version B. Clearly this assumption will be more plausible in some settings, such as with default effects (where it is difficult to imagine that a non-trivial number of respondents always select

[6] We assume that choices are observable in just one frame because the vast majority of data on framing effects are collected using between-subjects designs, especially in recent years. This tendency stems from concerns that, when respondents are asked the same question under multiple frames, the order in which respondents answer questions from multiple frames can affect the responses (LeBoeuf and Shafir, 2003), or that respondents might guess the hypothesis under study (Tversky and Kahneman, 1983). There has nevertheless been some debate among psychologists about the use of within- versus between-subject designs in the study of judgment, and some studies on framing effects do use within-subjects designs (see Lambdin and Shaffer, 2009 for a review).

whichever response is not marked as the default), and less likely to hold in other settings, such as with response-order effects (where one might imagine that some respondents always select the first option while others always select the most recent option). The following proposition describes the role of frame monotonicity in identifying π_c.

Proposition 2: The Consistent Response Share

Let $Y_c = \frac{Y_0}{Y_0+1-Y_1}$, and maintain Assumptions 1 and 2.

2.1 Under frame monotonicity, $E[c_i] = Y_0 + 1 - Y_1$ and $Y_c = \pi_c$.

2.2 Without frame monotonicity, $E[c_i]$ and π_c are bounded as follows:

2.2.1 $|1 - Y_1 - Y_0| \leq E[c_i] \leq Y_0 + 1 - Y_1$, and

2.2.2 $Y_c \leq \frac{1}{2} \implies \pi_c \in [0, Y_c]$ and $Y_c \geq \frac{1}{2} \implies \pi_c \in [Y_c, 1]$

Proof of 2.1 Applying frame monotonicity to (1) and (2) yields $Y_0 = \pi_c\, E[c_i]$ and $Y_1 = \pi_c\, E[c_i] + 1 - E[c_i]$, so that $Y_0 + 1 - Y_1 = E[c_i]$. Substituting these into the definition of Y_c yields the result. ■

The proof of Proposition 2.2, and of all further results, is contained in the supplementary appendix.

Proposition 2.1 establishes that the consistent response share is point-identified when frame monotonicity holds. Intuitively, frame monotonicity implies that only consistent individuals respond against the frame, so that any respondent selecting $y_i = 1$ under d_0 would also select $y_i = 1$ under d_1. Consequently, we know that Y_0 respondents are consistent *and* select $y_i = 1$. Similar logic implies that $1 - Y_1$ of respondents are consistent and select $y_i = 0$. Scaling the former by the total fraction of respondents that are consistent yields the share of consistent respondents who select $y_i = 1$, or π_c.

Proposition 2.2 establishes that without frame monotonicity, the consistent response share is partially identified. Borrowing terminology from Imbens and Angrist (1994), we can divide the set of inconsistent respondents into two groups, those who are affected by the frame in the same direction as the majority of inconsistent respondents (the *frame-compliers*) and those who are affected by the frame in the opposite direction (the *frame-defiers*). That is, a frame-defier selects $y_i = 1$ if and only if the frame is d_0. Intuitively, the presence of frame-defiers means that some respondents who are inconsistent will be misclassified as consistent in the computation of Y_c; the misclassified group will contain the frame-defiers, plus an offsetting number of frame-compliers. The

frame-defiers will respond with $y_i = 1$ if and only if they are assigned to d_0; the frame-compliers if and only if they are assigned to d_1. Since there are an equal number of frame-defiers and frame-compliers in the group of misclassified respondents, the group, on average, answers $y_i = 1$ half of the time under both frames. Because the misclassified group's behavior – in the aggregate – is the same under each frame, the group as a whole appears to be consistent (even though, in reality, each individual member of the group is actually inconsistent). And because the behavior of this group is attributed to the consistent respondents, the misclassification will bias Y_c upwards from π_c when π_c is in fact below 0.5 and downwards when the opposite is true. Thus the failure of frame monotonicity implies that Y_c is biased away from π_c towards 0.5.

The following corollary highlights an important practical implication of this result for survey researchers. Recall that Y_p denotes the pooled response share obtained from randomly assigning half of the respondents to each frame, $Y_p = \frac{1}{2}(Y_0 + Y_1)$.

Corollary to Proposition 2.2: Relative Bias in Y_c and Y_p

Under Assumptions 1 and 2, $|Y_c - \pi_c| \leq |Y_p - \pi_c|$, where the inequality is strict whenever $Y_c \neq Y_p$. In addition, $Y_c - \pi_c > 0 \iff Y_p - \pi_c > 0$.

Although neither Y_c nor Y_p will generally be equal to the consistent response share in the absence of frame monotonicity, this corollary states that the bias in the latter will be at least as large as the bias in the former. Moreover, the two quantities will be biased in the same direction, suggesting that moving from Y_p to Y_c will always be a conservative adjustment.

Examples

Table 1 illustrates the estimation of the consistent response share, assuming frame monotonicity as in Proposition 2.1, for the examples described in Section I. In the first row, for example, we compute the fraction of consistent respondents stating a belief that the information on Iraq was accurate. In this example, 41 percent of respondents indicated the information was accurate when this answer choice was second (d_0) and 51 percent did so when this answer choice was first (d_1). As such, we calculate that $P(c_i = 1) = 0.41 + 1 - 0.51 = 0.90$, and $Y_c = \frac{0.406}{0.90} \approx 0.45$. In words, 90 percent of respondents were consistent and of that group, 45 percent believed the information was accurate. Similar calculations are provided for each of the examples.

Table 1: Estimating the Consistent Response Share: Illustrations

	Answer Corresponding to $y_i = 1$	Response Share under d_0	Response Share under d_1	Fraction Consistent	Consistent Response Share
Survey		Y_0	Y_1	$P(c_i = 1)$	Y_c
Response Order	Information Provided Was Accurate	41	51	90	45
Question Order (Clinton)	Clinton Trustworthy	50	57	93	54
Question Order (Gore)	Gore Trustworthy	60	68	92	65
Question Wording (Acquiescence Bias)	Individuals More to Blame	43	60	84	52
Question Wording (Gain/Loss Framing)	Prefer Program A or C	22	72	50	44

Note: This table reports estimates of the consistent response share, as in Proposition 2, for the examples described in Section I. Missing and "don't know" responses are discarded where applicable. All quantities represent percentages.

IV. Which Respondents are Consistent?

Thus far our focus has been on the distribution of answers to the survey question being asked among the population of respondents. However, in some applications, the consistency (or lack thereof) of the respondents will itself be an issue of primary interest to the researcher. For example, a political campaign may be quite interested in likely voters whose stated preferences between two candidates depend on the order in which the candidates are listed, or upon which features of the candidates are made salient in the wording of the survey question. Understanding the distribution of such voters could help a campaign better target political messages, for example.

Without observing a single respondent under multiple frames, it is impossible to identify precisely which individuals are consistent and which are not. Under the assumption of frame monotonicity, however, researchers can identify the aggregate distribution of observable covariates of the consistent and inconsistent decision-makers. The approach in this section is analogous to Abadie (2003), who, in a causal inference setting, shows how to identify the aggregate characteristics of the "compliers" despite the fact that individual members of that group cannot be identified. Formally, we suppose that individuals have observable covariates $g_i \in G$. We denote the response shares under each frame, conditional on respondents' observable covariates, by $Y_0(g) \equiv E[y_i|d_i = d_0; g_i = g]$ and $Y_1(g) \equiv E[y_i|d_i = d_1; g_i = g]$. Finally, we will replace our (unconditional) frame exogeneity assumption with the following conditional frame exogeneity assumption:

Assumption 1' (*Conditional Frame Exogeneity*) For each i and for all $g \in G$,
$(d_i \perp (y_i(d_0),\, y_i(d_1)) \mid g_i = g)$

Like frame exogeneity, conditional frame exogeneity is guaranteed when respondents are assigned to frames at random.

Proposition 3: Relating Consistency to Observable Covariates Under Assumptions 1' and 3:

3.1 The distribution of g among the consistent respondents is given by $P(g_i = g \,|\, c_i = 1) = \frac{Y_0(g)+1-Y_1(g)}{E_g[Y_0(g)+1-Y_1(g)]}\, P(g_i = g)$.

3.2 The distribution of g among the inconsistent respondents is given by $P(g_i = g \,|\, c_i = 0) = \frac{Y_1(g)-Y_0(g)}{E_g[Y_1(g)-Y_0(g)]}\, P(g_i = g)$.

Intuitively, the distribution of g among the consistent (or inconsistent) respondents can be found by adjusting the overall distribution of g by the relative propensity of type-g respondents to be consistent (or inconsistent), relative to the overall population. An illustration of Proposition 3 can be found in Section VI.

V. Framing Effects When the Characteristic of Interest is Defined For All Respondents

In some settings, the individual characteristic of interest, y_i^*, will be well-defined for all respondents, even those whose response to the survey depends on the frame. For example, framing effects have been documented in surveys that solicit self-reported behavioral frequency data, such as the frequency with which respondents watch television or engage in risky health behaviors (see Schwarz and Oyserman (2001) for a number of examples). When the question is about past behavior, it is clear that a "true answer" exists for all individuals, even if the presence of a framing effect prevents that answer from being revealed by the survey question. More controversially, respondents may have a well-defined true answer even when the question being asked is about an attitude or belief (see Fischhoff, 1991). This section provides tools for settings in which researchers seek to identify the distribution of characteristics among the population of respondents, including those whose responses depend on the frame.

Assumption 4 (Characteristic of Interest is Well-Defined For All Respondents) For each i, $y_i^* \in \{0, 1\}$.

The parameter of interest in this problem is the *characteristic share* in the full population, $E[y_i^*]$.

We begin by providing worst-case bounds for the characteristic share (in the spirit of Manski, 1989).

Proposition 4: Partially Identifying the Distribution of the Characteristic of Interest Suppose Assumptions 1, 2, and 4 are satisfied.

4.1 Under frame monotonicity, $Y_0 \leq E[y_i^*] \leq Y_1$.

4.2 Without frame monotonicity, $Y_c \geq \frac{1}{2} \implies E[y_i^*] \in [Y_0 - (1 - Y_1), 1]$ and $Y_c \leq \frac{1}{2} \implies E[y_i^*] \in [0, Y_0 + Y_1]$.

Given response shares under two frames, Y_0 and Y_1, it might be tempting to conclude that the characteristic share, $E[y_i^*]$, lies somewhere between these values. Proposition 4 reveals, perhaps surprisingly, that such an interpretation is only valid when the researcher assumes not only that the consistent responses reveal the characteristic of interest (Assumption 2), but *also* assumes frame monotonicity (Assumption 3). Intuitively, when there are a large number of inconsistent respondents who respond against the frame, the magnitude of the framing effect can appear much smaller than it really is. Proposition 4.2 shows that without frame monotonicity, the response shares under the two frames still provides some information – in the form of a one-directional bound – about the characteristic of interest. These bounds will be more informative when Y_0 and $1 - Y_1$ are very different, which tends to occur when the observed framing effect, $Y_1 - Y_0$, is small and response shares are far from 0.5.

Some researchers may wish to go further, making additional assumptions to point-identify the distribution of the characteristic of interest in the population. Under Assumptions 1 and 2, Proposition 2.1 identifies the distribution of the characteristic of interest among the consistent respondents. The problem of recovering this distribution for the full population then parallels the well-studied problem of accounting for selection into a particular survey sample. Here, the goal is to account for selection into the population of consistent respondents (e.g., Manski, 2003). Note though that one important difference from the standard setting is that researchers cannot directly observe whether a particular respondent is consistent, whereas one can typically observe whether any given respondent is included in the sample.

The approach we develop below parallels the use of post-stratification weights to address the presence of missing data. The key assumption is that conditional on observable covariates, the consistent and inconsistent respondents have the same distribution of the characteristic of interest:

Assumption 5: Conditional Consistency Independence

$$\forall g \in G,\ cov(c_i, y_i^* | g_i = g) = 0,$$

where g denotes observable covariates, as in the previous section.

Proposition 5: The Characteristic Share Under Conditional Consistency Independence Let $Y_c(g) = \frac{Y_0(g)}{Y_0(g)+1-Y_1(g)}$. Under Assumptions 1-5, $E[y_i^*] = E_g[Y_c(g)]$.

To implement the approach suggested by Proposition 5, one first divides the respondents by observable group $g \in G$, then calculate $Y_c(g)$ for each group. One then obtains $E[y_i^*]$ from a weighted sum of each $Y_c(g)$, with weights based on the distribution of g in the full population. The technique is analogous to post-stratification weights frequently employed in survey analysis (Holt and Smith, 1979), which correct for the fact that some respondents are more likely to select into the sample than others. In our setting, *consistency weights* correct for the fact that some respondents are more likely to select into the consistent subgroup of the population – the subgroup whose characteristic of interest is revealed by their responses to the survey question.

Carrying the analogy further, for conventional post-survey non-response weights to eliminate selection bias, it must be the case that respondents' propensity to participate in the survey is uncorrelated with unobservable correlates of the variable being investigated. Our conditional consistency independence assumption guarantees exactly this; it will fail when respondents' consistency is related to the distribution of y in unobservable ways. As such, the more individual covariates the researcher can observe that are potentially correlated with a respondent's consistency, the more confident the researcher can be that using consistency weights will recover the consistency share for the population.

VI. Illustration

In this section, we illustrate our proposed approach for dealing with survey framing effects. The survey we focus on is the 2003 Gallup telephone poll that solicited beliefs about the Bush administration provision of information leading up to the Iraq war, and that was described in the first example in Section I. For ease of illustration, we drop 31 respondents from the raw data who did not respond to the survey question as well as 6 respondents who did not classify themselves as Democrats, Republicans, or Independents.

A natural first step for a survey researcher concerned about a framing effect is to investigate whether such an effect is actually present. This is plain from the responses summarized in Table 1. When the "accurate information" answer was presented first, 51% of respondents reported believing that the Bush administration provided accurate information ($Y_1 = 0.51$). In contrast, when the "inaccurate information" answer was provided first, only 41% of respondents reported believing that the information provided

by the Bush administration was accurate ($Y_0 = 0.41$). A simple t-test confirms the difference in responses between the two frames is statistically significant ($p = 0.016$).

Having confirmed that a framing effect exists, the researcher next confronts the question of what can be learned about the consistent respondents, those who report having the same belief regardless of which answer choice is presented first. Proposition 2 provides the mean response of this group if the assumptions underlying the proposition are satisfied. Frame exogeneity (Assumption 1) is satisfied with this data because the survey's answer order was randomized across respondents. In addition, it seems likely that consistent responses would reflect the characteristic of interest (Assumption 2), since most respondents would have lacked incentives to falsely report their beliefs to the survey administrators (the survey was anonymous and not conducted in person). Consequently, Proposition 2 tells us that how much can be learned about the consistent respondents depends on whether frame monotonicity (Assumption 3) holds – i.e., whether every inconsistent respondent selects whichever answer choice is listed first. If one is willing to make this assumption, Proposition 2.1 implies that 89.8% of respondents are consistent (standard error 4.2%), and that of that group, 45.3% (standard error 2.4%) believe that the information provided by the Bush administration was accurate.[7] Without assuming frame monotonicity, the survey data provide much less information: applying Proposition 2.2, the fraction of consistent respondents believing the information was accurate could be anywhere between 0 and 45.3%.

The researcher can next use Proposition 3 to investigate which types of survey-takers are most sensitive to the framing effect. Table 2 illustrates this result, focusing on the survey-taker's party affiliation. For example, the share of Democrats among the inconsistent respondents is calculated as $P(g_i = DEM\,|c_i = 0) = \frac{Y_1(DEM)-Y_0(DEM)}{E_g[Y_1(g)-Y_0(g)]}\ P(g_i = DEM) = \left(\frac{0.449-0.310}{0.104}\right)\ (0.193) \approx 0.256$.[8] The main substantive finding in Table 2 is that Democrats are over-represented among the inconsistent respondents – they constitute just 19 percent of all respondents but 26 percent of those respondents whose answers depend on the answer order.

[7] The standard errors are derived using the delta method, as described in Appendix B.

[8] The denominator of this expression, 0.104, estimates the fraction of inconsistent respondents in the population, $E[c_i]$. To estimate this, we use a weighted sum over the three groups, $E_g[Y_1(g) - Y_0(g)] = \sum_g [Y_1(g) - Y_0(g)]P(g)$ and the numbers in the first three columns of Table 2. This estimator identifies $E[c_i]$ under Assumption 1' (Conditional Frame Exogeneity). If frames are randomly assigned, both this expression and the one in Proposition 2 will identify $E[c_i]$. With random assignment in finite samples, however, it is preferable to use the estimator employed here, since doing so ensures that the shares in 3.1 and 3.2 will sum to one.

Table 2: Illustration of Proposition 3: Respondent Characteristics by Consistency

	Response Share under d_0	Response Share under d_1	Fraction of All Respondents	Fraction of Consistent Respondents	Fraction of Inconsistent Respondents
	Y_0	Y_1	$P(g_i = g)$	$P(g_i = g \mid c_i = 1)$	$P(g_i = g \mid c_i = 0)$
Party					
Democrat	31.0 (6.1)	44.9 (7.2)	19.3 (1.7)	18.5 (1.8)	25.6 (16.8)
Republican	49.2 (4.4)	59.1 (4.2)	48.5 (2.1)	48.8 (2.3)	45.9 (20.0)
Independent	33.3 (5.4)	42.6 (4.9)	32.2 (2.0)	32.7 (2.2)	28.5 (18.6)
Total	50.9 (3.0)	40.7 (3.0)	1	1	1

Note: $y_i = 1$ indicates respondent answered that the Bush Administration provided accurate information about Iraqi weapons of mass destruction. All quantities represent percentages. Standard errors are provided in parentheses beside point estimates. Standard errors were obtained via the delta method, see Appendix B. Source: Gallup Organization (2003).

Whether there is more to glean from the survey data depends on whether the researcher believes that the characteristic of interest is defined for those respondents who are subject to framing effects. In the context of this survey, there are two possibilities. First, it may be that the survey-takers whose responses depend on the question order simply lack any well-defined belief about whether the Bush administration provided accurate information. In that case, the researcher should limit his or her attention to the consistent respondents. Alternatively, it may be that at least some of the inconsistent respondents do have an opinion about the survey question, notwithstanding the fact that the answer order interfered with that opinion being expressed. In that case, our proposed approach is for the researcher to use the information recovered about the consistent respondents to estimate the (unobservable) opinions of the inconsistent respondents, and thereby recover the distribution of opinions for the entire population of survey-takers.

We have provided two approaches for proceeding when the characteristic of interest is defined for all respondents, including those who are inconsistent. First, Proposition 4 allows one to bound the fraction of all survey-takers believing the Bush administration's information was accurate. If one assumes frame monotonicity, Proposition 4.1 implies that this quantity is between 40.7% and 50.9% of all survey-takers. Without frame monotonicity, Proposition 4.2 implies that we may conclude only that fraction of

survey-takers believing the information to be accurate is below 91.6%. Without frame monotonicity, therefore, very little can be gleaned about the distribution of survey-takers' beliefs from these data.

The second approach we developed allows one to obtain a precise estimate for the distribution of opinions in the population of survey-takers, but requires imposing conditional consistency indepndence (Assumption 5) in addition to our other assumptions. In the context of our running example, this means that we must assume the opinions of the inconsistent respondents are the same, on average, as the opinions of the consistent respondents who share their observable characteristics. For the sake of illustration, suppose respondents' only observable characteristic was their party affiliation. Our approach would then be to separately estimate the distribution of opinions among the consistent respondents of each political party and then to re-weight those estimates based on the prevalence of each party in the overall population of survey-takers, as illustrated in Table 3.

The results of this exercise suggest that 45.3 percent of all survey-takers believe the information provided by the Bush administration's was accurate. Intuitively, Democrats are over-represented in the inconsistent group and have a lower tendency to believe that the Bush administration information was accurate, so the fraction of inconsistent respondents believing the information was accurate is less than the corresponding fraction among consistent respondents. Note that because a relatively small fraction of all survey-takers are inconsistent, the discrepancy between the estimates of $E[y_i^* | \, c_i = 1]$ and $E[y_i^*]$ is quite small in this example.

VII. Conclusion

This paper presents a new set of empirical tools to study a ubiquitous problem in survey research: the sensitivity of responses to seemingly arbitrary features of survey design. As in other settings, the degree to which the parameters of interest can be identified from the data depend on the assumptions the researcher is willing to impose.

Two limitations of our work deserve particular consideration in future research. First, we have focused on the relatively simple setting of binary response survey questions with two frames. Analyzing settings with additional frames or answer choices requires further assumptions. One straightforward generalization obtains when responses are sensitive to multiple binary frames, such as answer choice order and posi-

Table 3: Illustration of Proposition 5: Consistency Weights

	Democrat	Republican	Independent
Fraction of consistent believing information was accurate, $E[y_i^*\|c_i = 1, g_i = g]$	36.0(5.5)	54.6(3.4)	36.7(4.2)
Fraction of all respondents, $P(g_i = g)$	19.3(1.7)	48.5(2.1)	32.2(2.0)
Fraction of inconsistent respondents, $P(g_i = g\|c_i = 0)$	25.6(16.8)	45.9(20.0)	28.5(18.6)
Fraction of population believing information was accurate, $E[y_i^*]$		45.3 (2.0)	
Fraction of inconsistent believing information was accurate, $E[y_i^*\|c_i = 0]$		44.8 (12.9)	
Fraction of consistent believing information was accurate, $E[y_i^*\|c_i = 1]$		45.3 (2.4)	

Note: All quantities represent percentages. Standard errors are provided in parentheses below point estimates. Standard errors were obtained via the delta method, see Appendix B. Source: Gallup Organization (2003).

tive/negative question wording. In such cases, under an appropriately modified monotonicity assumption, one can focus on the two most extreme frames – the two frames making respondents most likely to select one or the other answer – as the two frames in the binary framework considered here. Second, our approach is aimed at eliminating the bias induced by framing effects, but other sources of bias could still be a problem. Generalizing the approach proposed here to non-binary survey questions and to settings characterized by other types of bias – such as random choice, forgetfulness, or selection effects – are important directions for future research.

References

Abadie, Alberto. 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics*, 113(2): 231–263.

Chong, Dennis, and James Druckman. 2007. "Framing Theory." *Annual Review of Political Science*, 10: 103–106.

Deaton, Angus. 2012. "The Financial Crisis and the Well-Being of Americans." *Oxford Economic Papers*, 64(1): 1–26.

Fischhoff, Baruch. 1991. "Value Elicitation: Is There Anything in There?" *American Psychologist*, 46(8): 835.

Goldin, Jacob, and Daniel Reck. 2015. "Preference Identification Under Inconsistent Choice."

Holbrook, Allyson, Jon Krosnick, David Moore, and Roger Tourangeau. 2007. "Response Order Effects in Dichotomous Categorical Questions Presented Orally." *Public Opinion Quarterly*, 71(3): 325–348.

Holland, Paul. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association*, 81(396): 945–960.

Holt, D. Tim, and T.M. Fred Smith. 1979. "Post Stratification." *Journal of the Royal Statistical Society*, 142(1): 33–46.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–475.

Jann, Ben, Julia Jerke, and Ivar Krumpal. 2011. "Asking sensitive questions using the crosswise model an experimental survey measuring plagiarism." *Public Opinion Quarterly*, nfr036.

Lambdin, Charles, and Victoria A Shaffer. 2009. "Are Within-Subjects Designs Transparent?" *Judgment and Decision Making*, 4(7): 554–566.

LeBoeuf, Robyn, and Eldar Shafir. 2003. "Deep Thoughts and Shallow Frames: On the Susceptibility to Framing Effects." *Journal of Behavioral Decision Making*, 16(2): 77–92.

Manski, Charles. 1989. "Anatomy of the Selection Problem." *The Journal of Human Resources*, 24(3).

Manski, Charles. 2003. *Partial Identification of Probability Distributions (Springer Series in Statistics*. Springer.

Moore, David W. 2002. "Measuring New Types of Question-Order Effects: Additive and Subtractive." *Public Opinion Quarterly*, 66(1): 80–91.

Organization, Gallup. 2003. "Gallup News Service Poll No. 2003-37: Terrorism/Homosexual Civil Unions/Iraq/Children/College/Dangerous Drivers." Roper Center Public Opinion Archives Survey Dataset.

Presser, Stanley, Mick Couper, Judith Lessler, and Elizabeth Martin. 2004. *Methods for Testing and Evaluating Survey Questions.* Wiley.

Salant, Yuval, and Ariel Rubinstein. 2008. "(A, f): Choice with Frames." *The Review of Economic Studies*, 75(4): 1287–1296.

Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context.* SAGE.

Schwarz, Norbert. 1990. "Assessing Frequency Reports of Mundane Behaviors: Contributions of Cognitive Psychology to Questionnaire Construction." *Review of Personality and Social Psychology*, 11: 98–119.

Schwarz, Norbert, and Daphna Oyserman. 2001. "Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction." *American Journal of Evaluation*, 22(2): 127–160.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1995. *Thinking about Answers : The application of Cognitive Processes to Survey Methodology.* Jossey-Bass.

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science*, 211(4481): 453–458.

Tversky, Amos, and Daniel Kahneman. 1983. "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological review*, 90(4): 293.

Wolter, Felix, and Peter Preisendörfer. 2013. "Asking sensitive questions: An evaluation of the randomized response technique versus direct questioning using individual validation data." *Sociological Methods & Research*, 42(3): 321–353.