Framing Effects in Survey Research: Consistency-Adjusted Estimators*

Jacob Goldin† Daniel Reck‡

February 25, 2015

Abstract

A well-known difficulty in survey research is that respondents may answer survey questions in ways that are systematically related to arbitrary features of the survey's design, such as the way questions are worded or the order in which answers are presented. We develop a simple framework for analyzing such framing effects for binary survey questions in which individual survey-takers are observed under one of two frames. We first show that the conventional approach to analyzing data with framing effects – randomly assigning respondents across frames and then pooling the data – yields a biased estimate for the mean of the variable being measured. We then propose an alternative estimator, the consistency-adjusted mean, and provide conditions under which it identifies the mean of the survey variable for the subset of respondents whose answer is unaffected by the frame. When framing effects push respondents in heterogeneous directions, the consistency-adjusted mean will be biased, but less biased than the conventional approach. We also show how to estimate the distribution of covariates among the consistent and inconsistent respondents, and provide techniques for identifying the mean of the survey variable among the full population. We illustrate our proposed techniques by applying them to data from previous surveys characterized by framing effects.

*For valuable comments, we thank Angus Deaton, Edward Freeland, Michael Gideon, Allyson Holbrook, Bo Honore, David Lee, Yair Listokin, Charles Manski, Alex Rees-Jones, Joel Slemrod, Justin Wolfers, and participants in the Princeton Industrial Relations Lunch and the Michigan Summer Seminar. We are also grateful to Jonathan Schuldt and Paul Sniderman for making their data available to us. All errors are our own.

†Department of Economics, Princeton University, email: jgoldin@princeton.edu

‡Department of Economics, University of Michigan, email: dreck@umich.edu

Introduction

A well-known difficulty in survey research is that respondents' answers to survey questions may vary systematically according to how the survey is designed. For example, respondents may answer differently depending on the order in which questions are asked (Moore, 2002; Deaton, 2012), the order in which answers are listed (Holbrook et al., 2007), the grouping of responses into categories (Schwarz, 1990), or any number of variations in the manner in which questions or responses are worded (Schuman and Presser, 1981; Krosnick and Fabrigar, Forthcoming; Chong and Druckman, 2007). Such *framing effects* arise in many contexts; large literatures in psychology, political science, behavioral economics, 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 researchers should design their surveys to avoid potential framing effects when possible, and, when framing effects are unavoidable, that randomly assigning respondents across frames and then applying standard statistical techniques to the pooled data eliminates framing effects as a source of bias.[1] Others recognize the unsatisfactory nature of the conventional approach but note the lack of better alternatives.[2] The goal of this paper is to provide researchers with new tools for analyzing survey data that exhibit framing effects. Indeed, we show that the conventional approach – randomizing across frames and then working with the pooled data – yields a biased estimate for the mean of the survey variable (the variable about which the survey respondents are being asked). In its place, we suggest a new estimator and show that it yields useful information under plausible assumptions.

To study framing effects in survey research, we apply a potential outcomes frame-

[1]For example, 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 and et al, 2004) :

[2]For example, Schwarz and Oyserman (2001) write, "[R]esearchers ... may reverse the order in which the items are listed for half of the respondents. Although this ensures that the researcher becomes aware of possible response order effects, it remains unclear what to do with the results aside from the less than satisfying solution of averaging over both sets of answers."

work (Neyman, 1923).[3] We consider binary response questions in which an arbitrary feature of the survey – the *frame* – affects the responses of a subset of survey-takers. Each individual respondent answers a given survey question only once, under one of two possible frames. Initially, we also assume that when the frame affects a respondent's answers, it does so in a uniform direction for all respondents. The parameter of interest is the mean of the binary outcome, i.e. the fraction of respondents for whom the answer is one of the two options.

Using this model, we show that randomizing the assignment of respondents in equal proportions across the two frames does not eliminate the bias induced by framing effects; rather, it biases the estimated mean towards 0.5. To understand the source of this bias, consider two extreme cases. First, suppose that all respondents are completely immune to framing, so that they answer consistently across frames. In this case, the unadjusted mean will be an unbiased estimator of the population mean. For the second case, suppose that *all* respondents are influenced by framing: everyone's response is determined by the frame alone. In this case, randomly assigning half of the respondents to one frame and half to the other frame causes the unadjusted sample mean to equal 0.5. Finally, when some but not all respondents are subject to framing effects, the unadjusted sample mean is a mixture of the true mean for the consistent sub-population of respondents and 0.5 for the sub-population of inconsistent respondents.

To overcome this bias, we show how our assumptions permit point identification of the *consistent-subgroup mean*, the mean of the survey variable for the subset of respondents whose answers are unaffected by the frame. Hence we refer to this new estimator as the *consistency-adjusted mean*. In many applications, the consistent-subgroup mean will be the quantity of primary interest to the researcher; those respondents whose answers depend on the frame may have only poorly-formed opinions or beliefs about the subject being investigated.

The intuition behind our proposed estimator is as follows. First consider the set of individuals who select a given response to the survey question "against the frame" – that is, when their response is *not* favored by the frame to which they are assigned. We

[3] The potential outcomes framework has been widely applied in causal inference analysis (e.g., Holland, 1986; Imbens and Angrist, 1994). In those studies, researchers utilize experimental or non-experimental data to estimate the causal effect of some treatment on an observed outcome variable. To our knowledge, we are the first to apply the approach to study framing effects, where the goal is not to estimate the *effect* of the frame on a variable but rather to estimate the mean of a variable for those whom the frame does not affect. The framework we develop is an application of the theory in Goldin and Reck (2014) to survey research.

know that these respondents are consistent; otherwise, they would have selected the response favored by their assigned frame. Second, the average difference in responses between individuals assigned to the two frames represents the fraction of respondents that are subject to framing effects – when everyone responds consistently, the difference in responses between the frames will be zero. This insight permits us to obtain the fraction of respondents who do choose consistently. Scaling the first quantity (the fraction selecting a given response against the frame) by the second (the fraction that are consistent) yields the fraction selecting that response among all those who are consistent, i.e., the consistent-subgroup mean.

Next, we consider the implications of relaxing the assumption that frames affect survey-takers' responses in a uniform direction (an assumption we label *frame-monotonicity*). Although the consistency-adjusted mean no longer identifies the consistent-subgroup mean when frame-monotonicity fails, it nonetheless remains preferable to the conventional, unadjusted mean. The reason is that when frame-monotonicity fails, both the unadjusted sample mean and the consistency-adjusted sample mean will be biased towards 0.5, but the former will be more biased than the latter.

Apart from estimating the mean of the survey variable, researchers may wish to understand the characteristics associated with consistent and inconsistent respondents. For example, suppose a political campaign asks likely voters to select which of two candidates they plan to vote for, and that the order in which the candidates are listed to respondents affects the fraction of respondents selecting each candidate. Instead of simply discarding the responses of the inconsistent respondents, the researchers working for the campaign may wish to understand which voters are susceptible to such response-ordering effects, which may indicate that the respondent does not yet have a fully-formed opinion of the candidates. Similarly, knowing which parts of the population are likely to be influenced by a particular message can be useful for targeting advertising or other campaigns. Our approach provides a method for recovering the distribution of observable characteristics for the consistent and inconsistent respondents, which can then be used to predict whether or not a particular respondent is likely to be sensitive to the frame.

Finally, in certain applications, such as survey questions about behavior (rather than opinions), the researcher will be interested in the distribution of the survey variable in the full population – that is, the quantity of interest will be the distribution of the survey variable over both consistent and inconsistent respondents. For example, suppose a survey asks respondents how much television they watch per week, and the wording of the answer choices affects the amount of television respondents report watching (Schwarz, 1990). If the goal of the study is to learn about the television viewing habits of the population, the researcher will be interested in the viewing behavior of the inconsistent respondents as well as those whose answers are unaffected by the answer wording. Motivated by such applications, we develop techniques for estimating the mean of the survey variable in the full population. In particular, we derive worst-case bounds and a new weighting estimator, and describe the conditions under which these techniques shed light on the parameter of interest.

To illustrate our results in practice, we apply them to data from previous studies of survey framing effects. The results highlight that our proposed estimators can provide useful information across a range of settings, using only the type of data that is routinely collected.

The remainder of the paper proceeds as follows. Section I introduces our notation and assumptions. Section II utilizes that framework to show that the conventional approach to analyzing surveys with framing effects is biased. Section III develops our proposed alternative, the consistency-adjusted sample mean. Section IV relaxes the frame-monotonicity assumption. We discuss how to recover observable characteristics of the consistent and inconsistent populations in Section V. Section VI proposes two techniques for recovering information about the survey variable meant in the full population. Section VII concludes.

I. Notation and Assumptions

We presume to have data on N individuals from a random sample of the population, denoted $i = 1,...,N$. Each individual answers a survey question attempting to measure the distribution in the population of some binary variable, denoted by $y_i \in \{0, 1\}$. For example, y_i could denote whether an individual agrees with a particular statement or supports a particular candidate for office. By assumption, the true value of the survey variable, y_i, does not depend on the frame.

Each individual responds to the survey a single time, under one of the two possible frames, which are denoted by $d_i \in \{d_l, d_h\}$. For example, d might denote whether the survey question is worded positively or negatively, or it might denote the order in which two answer choices are presented to the respondent. Let d_i denote the *frame* to which individual i is assigned. The value of the survey variable reported by individual i is denoted by $\hat{y}_i \in \{0, 1\}$. Unlike the true value (y_i) the value that an individual reports

($\hat{y}_i$) may depend on the frame. We use $E[\hat{y}|d]$ to denote the mean reported value among individuals observed under frame d, $E[\hat{y}|d] \equiv E[\hat{y}_i|d_i = d]$. We restrict our attention to situations in which framing effects exist, i.e., in which $E[\hat{y}|d_l] \neq E[\hat{y}|d_h]$.

Although each individual is observed under one frame only, it will be useful to denote an individual's potential response under each frame by $\hat{y}_i(d)$, for $d \in \{d_l, d_h\}$. That is, $\hat{y}_i(d) = \hat{y}_i$ when $d_i = d$.

Let $\psi_i \in \{0, 1\}$ indicate whether an individual responds consistently across the two frames, $\psi_i = 1 \iff \hat{y}_i(d_h) = \hat{y}_i(d_l)$. Because respondents are only observed under a single frame, note that one cannot observe ψ_i for any particular individual.

We now introduce the three assumptions that form the basis for our framework.

Assumption 1: The Consistency Principle

$$\hat{y}_i(d_h) = \hat{y}_i(d_l) \Longrightarrow y_i = \hat{y}_i \tag{1}$$

The Consistency Principle states that individuals whose responses do not vary across frames are reporting their true value of the survey variable. To the extent that individuals systematically mis-report the survey variable for reasons unrelated to framing effects, our approach will yield biased estimates.

Assumption 2: Unconfoundedness

$$d_i \perp (\hat{y}_i(d_h),\, \hat{y}_i(d_l)) \tag{2}$$

Unconfoundedness requires that respondents are not systematically assigned to frames in ways that are correlated with their potential responses. This assumption is satisfied, for example, if respondents are randomly assigned across frames.

Assumption 3: Frame-Monotonicity

$$\forall i,\ \hat{y}_i(d_h) \geq \hat{y}_i(d_l) \tag{3}$$

Frame-Monotonicity states that framing effects are uni-directional; to the extent that a survey's design affects individuals' responses, it does so in a uniform way for all individuals.[4] For example, if some respondents are more likely to select whichever

[4]Note that frame-monotonicity is an example of a monotone treatment response assumption of the

answer is marked as the default, there must not be any individuals who do the opposite, i.e., select $y = 0$ if and only if $y = 1$ is the default. We consider the implications of relaxing this assumption in Section IV.

Throughout the paper, we maintain the assumption that the sample of observed respondents are drawn independently and identically-distributed (i.i.d.) from the underlying population, which guarantees that sample means converge to the corresponding population moments. This assumption rules out survey non-response bias.

II. Bias of the Conventional Approach

As described in the introduction, the conventional approach to analyzing survey data that exhibit framing effects is to randomize respondents across frames and then to proceed using the pooled data. 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 plays no further role in the analysis. This section shows that simply randomizing respondents across frames does not eliminate the bias associated with framing effects.

Suppose that respondents are randomly assigned to two groups, one for each of the frames. Let ρ_h denote the fraction of the population observed under frame d_h. In surveys with two frames, ρ_h is typically set equal to 0.5. Because our focus will be on identification rather than inference, we will state our results in terms of population moments; the relevant population moments can be estimated from their corresponding sample means. Let μ_c denote the mean of y among the consistent respondents, $\mu_c = E[y_i|\psi_i = 1]$. The following proposition demonstrates that the unadjusted reported mean, $E[\hat{y}]$, is equal to a weighted average of (1) the mean of the survey variable among the consistent respondents, and (2) the fraction of the population assigned to the high frame, ρ_h.

Proposition 1: The Unadjusted Mean

Under the Consistency Principle (1), Unconfoundedness (2), and Frame-Monotonicity (3),

$$E[\hat{y}] = \mu_c\, P(\psi_i = 1) + \rho_h\, P(\psi_i = 0)$$

type studied in Manski (1997).

Proof Applying the law of iterated expectations,

$$E[\hat{y}] = E[\hat{y}_i|\psi_i = 0]\, P(\psi_i = 0) + E[\hat{y}_i|\psi_i = 1]\, P(\psi_i = 1) \tag{4}$$

Begin with the inconsistent respondents (the first term in 4). The law of iterated expectations allows us to write

$$E[\hat{y}_i|\psi_i = 0] = E[\hat{y}_i|\psi_i = 0,\ d_i = d_l]\, P(d_i = d_l|\psi_i = 0) + E[\hat{y}_i|\psi_i = 0,\ d_i = d_h]\, P(d_i = d_h|\psi_i = 0) \tag{5}$$

By Frame-Monotonicity and the definition of ψ_i, the conditions $d_i = d_l$ and $\psi_i = 0$ jointly imply $\hat{y}_i = 0$, so that $E[\hat{y}_i|\psi_i = 0, d_i = d_l] = 0$. Similarly, the same two assumptions imply $E[\hat{y}_i|\psi_i = 0,\ d_i = d_h] = 1$. Finally, Unconfoundedness guarantees $P(d_i = d_h|\psi_i = 0) = P(d_i = d_h) = \rho_h$. Substituting these results into (5) yields

$$E[\hat{y}_i|\psi_i = 0] = \rho_h \tag{6}$$

Turning to the consistent respondents (the second term in 4), recall that by definition, $\psi_i = 1 \iff \hat{y}_i(d_l) = \hat{y}_i(d_h)$. Consequently, the Consistency Principle implies that $\psi_i = 1 \iff \hat{y}_i = y_i$. Thus we can write

$$E[\hat{y}_i|\psi_i = 1] = E[y_i|\psi_i = 1] \equiv \mu_c \tag{7}$$

Substituting (6) and (7) into (4) yields the desired result. ■

When survey respondents are evenly divided between the two frames, Proposition 1 shows that the unadjusted mean is biased towards 0.5. To understand why, consider two extreme examples. First suppose that all respondents are consistent, $P(\psi_i = 1) = 1$. Then the unadjusted mean estimates the average value of y in the population, $E[\hat{y}] = E[y_i|\psi_i = 1] = E[y_i]$. Now consider the other extreme, when $P(\psi_i = 0) = 1$ and all respondents are sensitive to the frame. In this case, respondents select $\hat{y}_i = 1$ if and only if they are assigned to the frame d_h, so the unadjusted mean identifies the fraction of individuals assigned to d_h, $E[\hat{y}] = E[\hat{y}_i|\psi_i = 0] = \rho_h$. Finally, when there are both consistent and inconsistent respondents in the population, the parameter estimated by the unadjusted mean will simply be the weighted average of these two extremes.

Proposition 1 shows that under the conventional approach to analyzing surveys with framing effects, the decisions of the researcher (i.e., the choice of ρ_h) affect the information obtained by the survey. Simply randomizing across frames does not solve the problem; it simply shapes the bias. The more respondents assigned to d_h, the greater the unadjusted mean will be. In the next section, we propose an alternative approach for dealing with framing effects in survey analysis.

III. The Consistency-Adjusted Mean

Instead of simply randomizing the population between frames and working with the pooled data, Assumptions 1-3 permit the researcher to recover the *consistent-subgroup mean*, the average value of y for the subset of survey-takers whose responses are unaffected by the frame, $\mu_c = E[y_i|\psi_i = 1]$. Obtaining this statistic would be trivial if the researcher could observe individual respondents under both frames; the inconsistent respondents could be individually identified and their responses discarded. However, the following proposition shows that this parameter is point-identified under our assumptions even when each respondent is only observed under a single frame.

Proposition 2: The Consistency-Adjusted Mean

Let $Y_c = \frac{E[\hat{y}|d_l]}{E[\hat{y}|d_l]+1-E[\hat{y}|d_h]}$ *and let* μ_c *denote the consistent-subgroup mean,* $\mu_c = E[y_i|\psi_i = 1]$. *Under the Consistency Principle (1), Unconfoundedness (2), and Frame-Monotonicity (3),* $Y_c = \mu_c$.

Proof *Lemma 1:* $E[\hat{y}|d_l] = \mu_c\, P(\psi_i = 1)$

Proof of Lemma 1:

By the law of iterated expectations,

$$E[\hat{y}|d_l] = E[\hat{y}_i|d_l, \psi_i = 0]\, P(\psi_i = 0|d_l) + E[\hat{y}_i|d_l, \psi_i = 1]\, P(\psi_i = 1|d_l) \tag{8}$$

Note that by Frame Monotonicity and the definition of ψ_i, the conditions $\psi_i = 0$ and $d_i = d_l$ jointly imply $\hat{y}_i = 0$. Thus $E[\hat{y}_i|d_l, \psi_i = 0] = 0$. Additionally, Unconfoundedness implies $P(\psi_i = 1|d_l) = P(\psi_i = 1)$ and $\mathrm{E}[\hat{y}_i|d_l, \psi_i = 1] = E[\hat{y}_i|\psi_i = 1]$. Finally, note that the Consistency Principle implies $E[\hat{y}_i|\psi_i = 1] = E[y_i|\psi_i = 1] \equiv \mu_c$. Substituting these results into (8) yields Lemma 1.

Lemma 2: $E[\hat{y}_i|d_h] = \mu_c\, P(\psi_i = 1) + P(\psi_i = 0)$

The proof of Lemma 2 is analogous to Lemma 1. Applying Lemmas 1 and 2 to the terms in the denominator of Y_c yields

$$E[\hat{y}_i|d_l] + 1 - E[\hat{y}_i|d_h] = P(\psi_i = 1) \tag{9}$$

Proposition 2 follows from applying (9) and Lemma 1 to the definition of Y_c. ■

Corollary 2.1: The Fraction of Consistent Respondents

Under the assumptions for Proposition 2, the fraction of consistent respondents is given by $P(\psi_i = 1) = E[\hat{y}|d_l] + 1 - E[\hat{y}|d_h]$[5]

Proof The result is identical to line (9) in the proof to Proposition 2. ■

An important benefit of the consistency-adjusted mean Y_c is that it provides information about the survey variable in a way that is not mechanically related to the design of the survey, even in the presence of framing effects. For example, when a survey of likely voters yields different aggregate responses depending on which candidate is listed first, the consistency-adjusted mean estimates the fraction planning to vote for each candidate from the sub-population of respondents who select the same candidate regardless of the order in which they are listed. Intuitively, $E[\hat{y}|d_l]$ and $E[\hat{y}|d_h]$ estimate the fraction of consistent respondents with $y_i = 1$ and $y_i = 0$ (respectively). Scaling $E[\hat{y}|d_l]$ by the total fraction of consistent respondents yields the fraction of that population with $y_i = 1$.

Illustrations of the Consistency-Adjusted Mean

1. Acquiescence Bias: Attitudes Toward Crime In their seminal book, 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. For example, in one study individuals were randomly assigned one of the following two versions of the same question:

> Version A: "Individuals are more to blame than social conditions for crime and lawlessness in this country"
>
> Version B: "Social conditions are more to blame than individuals for crime and lawlessness in this country"

[5]This corollary is the analogue within our framework of previous results for identifying the fraction of a population that responds to a treatment (see, e.g., Section 5 of Zhang and Rubin (2003)).

The researchers found that 60 percent of respondents assigned to Version A agreed with their statement, but that 57 percent of respondents assigned to Version B agreed with their statement as well.

To apply our framework to this setting, let y_i indicate whether a respondent believes that social conditions are more to blame for crime and lawlessness, and let d_h indicate that the respondent was assigned to Version B of the question.[6] Thus $E[y|d_h] = 0.57$ and $E[y|d_l] = 0.40$. Our maintained assumptions are satisfied if (1) individuals who would report that social conditions are more to blame under both question versions actually hold this belief (Consistency Principle); (2) no respondent would answer "No" to both versions of the question (Frame Monotonicity); and (3) the randomization of respondents between question versions was successful (Unconfoundedness). When these assumptions are satisfied, Proposition 2 implies that the difference between the two numbers above, $0.57 - 0.40 = 0.17$ is the fraction of the population that is is inconsistent, while $1 - 0.17 = 0.83$ is the fraction that are consistent. Since we know from the original data that 40 percent of respondents are consistent *and* they believe that social conditions are more to blame for crime, we can conclude that among those who are consistent, the fraction believing that social conditions are more to blame than crime is $Y_c = \frac{0.40}{0.83} = 0.48$.

2. Question-Order: Politician Trustworthiness Moore (2002) documents question-order effects in a 1997 Gallup survey. Respondents were asked, "Do you generally think [Bill Clinton / Al Gore] is honest and trustworthy?" Moore (2002) shows that respondents' answers varied depending on which politician they were asked about first. When the Clinton question was asked first, 50 percent of respondents reported thinking that Clinton was trustworthy, whereas 57 percent reported thinking so when they were first asked about Al Gore. Conversely, 68 percent reported believing Al Gore to be trustworthy when the Gore question was first, but only 60 percent did so when the Gore question was second.

Let y_i^k indicate whether i believes politician $k \in$ {Clinton, Gore} to be trustworthy and let d_h indicate that the Gore question was asked first. When Proposition 2 is satisfied, we may conclude that 93 percent of the respondents were unaffected by ques-

[6] One shortcoming of this particular test for acquiescence bias is the possibility that the findings could reflect the fact that some respondents simply believed that individuals and social conditions were equally to blame for crime and lawlessness. In that case, the consistent subgroup mean would identify the fraction of respondents who believed social conditions were more to blame, out of all respondents who believed that one of the causes was more to blame than the other.

tion order with respect to their assessment of Clinton's trustworthiness, and 92 percent with respect to their opinion about Gore. Among the consistent respondents, 54 percent believed Clinton to be trustworthy and 65 percent believed Gore to be trustworthy.

3. Question Wording: Beliefs about Climate Change Schuldt, Konrath and Schwarz (2011) investigate the presence of framing effects on Americans' reported beliefs about climate change. The authors document large framing effects among respondents who self-identify as Republican. Specifically, they find that such respondents are more likely to agree that the Earth's temperature has been changing when the phenomenon is referred to as "climate change" (60 percent) as opposed to "global warming" (44 percent).

Let y_i indicate whether a respondent agrees that the Earth's temperature has been changing[7] and let d_h indicate that the phenomenon is referred to as "climate change." When Proposition 2 is satisfied, we can conclude that the question wording affects the beliefs of 16 percent of Republicans and that among those who are consistent, 52 percent believe that the phenomenon is occurring.

4. Answer Order: Political Polling A large literature suggests that polls about candidate preferences can be influenced by the order in which candidates are listed. The data in this example come from a Gallup telephone survey conducted on October 15 and October 16, 2012, and relate to the 2012 presidential election (Gallup Organization, 2012). The question asked respondents (while randomizing the order in which candidates were listed):[8]

> Suppose that the presidential election were being held today, and it included Barack Obama and Joe Biden as the Democratic candidates and Mitt Romney and Paul Ryan as the Republican candidates. Would you vote for Mitt Romney and Paul Ryan, the Republicans [Barack Obama. and Joe Biden, the Democrats] or Barack Obama and Joe Biden [Mitt Romney and Paul Ryan, the Republicans]?

[7]Technically, "global warming" and "climate change" may refer to distinct phenomena (such as changing variability in temperature without an increasing trend). But as the authors point out, the two terms are used interchangeably in public discourse and opinion polling on this issue.

[8]Individuals who indicated they were unsure were asked "As of today, do you lean more toward...", followed by the candidates, presented in the same order as in the first question. Respondents indicating a preference in the follow-up question are coded according to their stated preference in this question. We discard observations where the individual indicated a preference for a third candidate, or did not know or refused the follow-up question.

The data indicate a strong *recency* effect (recency effects are commonly observed on oral surveys, see Holbrook et al. (2007)). When the Republican candidates were listed second, the average level of support for Romney was 55 percent. In contrast, when the Democrats were listed second, Romney's support fell to 48 percent.

Let y_i indicate support for the Republican candidates, and let d_h indicate whether the Republicans were listed second. From Proposition 2, we can conclude that 7 percent of survey respondents were affected by candidate order and that 52 percent of the consistent respondents expressed support for Romney.

5. Issue Framing: Losses versus Gains In a classic study Tversky and Kahneman (1981) asked experimental participants about their willingness to accept risky policies that have the potential to save large numbers of lives.[9] 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?
>
> [**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?

Note that Programs A and C are identical, as are Programs B and D. Yet under the Gain Frame, 72 percent of respondents selected Program A, whereas under the Loss Frame, only 22 percent of respondents selected Program C.

Let y_i indicate whether a respondent supports the risky option (Programs B or D) and let $d_h$indicate that the respondent was assigned to the Loss Frame. The estimated population moments are thus $E[y|d_l] = 0.28$ and $E[y|d_h] = 0.78$. When Proposition 2 is

[9]Although this particular experiment was conducted in a lab, similar issues can arise when survey-takers gauge support for real policies in the field.

satisfied, we can conclude that 50 percent of respondents consistently prefer one policy or the other, and that of those who are consistent, 56 percent prefer the riskier policy.

In many settings, the consistent-subgroup mean will be the parameter of primary interest to the researcher. For example, politicians surveying their constituents about their support for a policy may be most interested in the opinions of those who have well-formed and sufficiently stable opinions on the issue. Similarly, political pollsters surveying potential voters about which candidate they support may wish to throw out those responses that are sensitive to something as arbitrary as the order in which candidates are listed, perhaps assuming that those respondents haven't yet settled on a preferred candidate. In these settings, the consistent subgroup mean is the parameter of interest. In other applications, the researcher may be more interested in the distribution of the survey variable in the full population – i.e., among both consistent and inconsistent respondents.[10] Identifying the full population mean will be the focus of Section VI.

IV. Relaxing Frame-Monotonicity

Thus far we have maintained the assumption that any framing effects that occur affect all respondents in a uniform direction. For example, if one observes that a higher fraction of respondents answers "Yes" in response to version A of a question as opposed 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 default effects (where it is difficult to imagine that a non-trivial number of respondents always select whichever response is not marked as the default), and less likely to hold in other settings, such as response-order effects (where one might imagine that some respondents always select the first option while others always select the most recent option).

Without frame-monotonicity, the consistency-adjusted mean Y_c does not identify the consistent-subgroup mean μ_c. The bounds derived in this section show that partial identification of μ_c remains possible even when frame-monotonicity does not hold.

[10] In still other settings, the parameter of interest is the frame-sensitive choice under a particular frame or distribution of frames. To predict election results with ballot-order effects, for example, we should estimate the distribution of votes under the ballot orders that will occur in the election. In this case, the survey frame(s) should be made as similar as possible to the frame(s) under which the variable of interest, such as actual votes, is determined. This setting differs from other questions in survey research because, in this case, researchers essentially use one opinion poll to predict the distribution of answers to another, more important opinion poll (the election itself).

Proposition 3 : Partial Identification of the Consistent Subgroup Mean Without Frame-Monotonicity

Let $Y_c = \frac{E[\hat{y}|d_l]}{E[\hat{y}|d_l]+1-E[\hat{y}|d_h]}$ *and let* μ_c *denote the consistent-subgroup mean,* $\mu_c = E[y_i|\psi_i = 1]$*. Under the Consistency Principle (1) and Unconfoundedness (2):*

(3.1) *If* $Y_c > \frac{1}{2}$*, then* $Y_c \leq \mu_c \leq 1$

(3.2) *If* $Y_c < \frac{1}{2}$*, then* $0 \leq \mu_c \leq Y_c$

(3.3) *If* $Y_c = \frac{1}{2}$*, then* $Y_c = \mu_c = \frac{1}{2}$

Proof Borrowing terminology from Imbens and Angrist (1994), we can divide the set of inconsistent respondents into two groups, those who are effected 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 responds $\hat{y}_i = 1$ if and only if the frame is d_l. Let f_i indicate whether respondent i is a frame-defier and let α denote the fraction of inconsistent respondents who are of this type, $\alpha = P(f_i = 1 \,|\, \psi_i = 0)$.[11]

Next, using the law of iterated expectations, write

$$\begin{aligned} E[\hat{y}|d_l] =& E[\hat{y}_i|d_l, \psi_i = 1]\, P(\psi_i = 1|d_l) \\ &+ E[\hat{y}_i|d_l, \psi_i = 0,\ f_i = 0]\, P(\psi_i = 0|d_l)\, P(f_i = 0|\psi_i = 0, d_l) \\ &+ E[\hat{y}_i|d_l, \psi_i = 0,\ f_i = 1]\, P(\psi_i = 0|d_l)\, P(f_i = 1|\psi_i = 0, d_l) \end{aligned} \tag{10}$$

By Unconfoundedness and the definition of f_i, (10) simplifies to

$$E[\hat{y}_i|d_l] = \mu_c\, P(\psi_i = 1) + \alpha P(\psi_i = 0) \tag{11}$$

Similarly, it is straightforward to show

$$E[\hat{y}_i|d_h] = \mu_c\, P(\psi_i = 1) + (1-\alpha) P(\psi_i = 0) \tag{12}$$

[11] To illustrate, suppose that 60 percent of respondents consistently answer "Yes" to a "Yes/No" question, 35 percent answer "Yes" only under Version A of the question, and 5 percent answer "Yes" only under Version B. In this example, $\alpha = \frac{0.05}{0.35+0.05} = 0.125$. Because d_h and d_l are labeled based on the net effect of the frame, note that it must always be the case that $\alpha < 0.5$. Note also that $\alpha = 0.5$ is ruled out whenever the researcher observes some framing effect.

Substituting (11) and (12) into the definition of Y_c yields

$$Y_c = \frac{P(\psi_i = 1)\,\mu_c + \alpha\,P(\psi_i = 0)}{2\,\alpha\,P(\psi_i = 0) + P(\psi_i = 1)} \tag{13}$$

Subtracting μ_c from both sides of (13) yields

$$Y_c - \mu_c = \frac{P(\psi_i = 0)\alpha\,(1 - 2\mu_c)}{2\,\alpha\,P(\psi_i = 0) + P(\psi_i = 1)} \tag{14}$$

To prove the proposition, note that $P(\psi_i = 0) > 0$, $P(\psi_i = 1) > 0$, and $\alpha \geq 0$. Thus $\mu_c > \frac{1}{2}$ implies $Y_c < \mu_c$. This, along with the fact that μ_c is by definition less than 1, proves (3.1). The derivations of (3.2) and (3.3) follow from (14) as well. ■

Proposition 3 shows that μ_c may be partially identified even in settings where frame-monotonicity does not hold. The following corollary highlights an important practical implication of this result for survey researchers.

Corollary 3.1: The Advantage of the Consistency-Adjusted Mean to the Unadjusted Mean

Let Y_u denote the (unadjusted) mean response obtained by randomly assigning respondents to d_l and d_h with equal probabilities, $Y_u = \frac{1}{2}\,E[\hat{y}|d_l] + \frac{1}{2}\,E[\hat{y}|d_h]$. Under the Consistency Principle (1) and Unconfoundedness (2):

(3.1') *If $Y_c > Y_u$, then $Y_u < Y_c < \mu_c$*

(3.2') *If $Y_c < Y_u$, then $Y_u > Y_c > \mu_c$*

(3.3') *If $Y_c = Y_u$, then $Y_u = Y_c = \mu_c$*

Proof Substituting (11) and (12) into the definition of Y_u yields

$$Y_u = \mu_c\,P(\psi_i = 1) + \frac{1}{2}\,P(\psi_i = 0) \tag{15}$$

Using (13) and (15), we can write

$$Y_u - Y_c = \frac{\left(\frac{1}{2} - \alpha\right)(1 - 2\,\mu_c)\,P(\psi_i = 1)P(\psi_i = 0)}{2\,\alpha\,P(\psi_i = 0) + P(\psi_i = 1)} \tag{16}$$

To prove (3.1'), suppose that $Y_c > Y_u$. From (16), this implies that $\mu_c > \frac{1}{2}$ (since $\alpha < \frac{1}{2}$ by definition). But from (3.1), this only occurs when $\mu_c > Y_c$. Thus we can conclude that $\mu_c > Y_c > Y_u$, proving the result. The proofs of (3.2') and (3.3') are analogous. ■

Three lessons can be taken away from this corollary. First, when respondents are evenly assigned between frames, frame-monotonicity is not required for Proposition 1 to hold; that is, Y_u is still biased towards 0.5 relative to the consistent-subgroup mean. Intuitively, half of the inconsistent respondents, either the frame-compliers under d_h or the frame-defiers under d_l, will still report $\hat{y}_i = 1$ and the other half will still report $\hat{y}_i = 0$.

Second, when frame monotonicity does not hold, Y_c is also biased away from μ_c toward 0.5. 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 $\hat{y}_i = 1$ if and only if they are assigned to d_l; the frame-compliers if and only if they are assigned to d_h. Since there are an equal number of frame-defiers and frame-compliers in the group of misclassified respondents, the group, on average, answers $\hat{y}_i = 1$ 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 when μ_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 towards 0.5.

Finally, despite the fact that Y_c is biased when frame-monotonicity fails, the misclassified group causing the bias in Y_c is necessarily smaller than the group of inconsistent respondents causing the bias in Y_u. As a result, the consistency-adjusted mean is less biased, in absolute terms, than the unadjusted mean Y_u.

V. Which Respondents are Consistent?

Thus far our focus has been on eliminating the bias caused by frame-sensitive respondents in order to learn about the distribution of the survey variable. 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 pollster working for an electoral 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 survey. Understanding who such voters are could be quite useful for better targeting political messages. Similarly, both advertisers and advocacy groups may wish to know which types of consumers are most responsive to particular types of messaging.

Without exposing a single respondent to multiple frames, it is impossible to identify precisely which individuals are consistent and which are not.[12] However, this section proposes tools that allow the researcher to estimate the aggregate distribution of characteristics of the consistent and inconsistent decision-makers.[13]

Suppose that the sample is partioned into K groups, based on observable characteristics of the respondents, $G = \{g_1, g_2, ...g_K\}$. Let $g_i \in G$ denote the group of individual i and let P_g denote the fraction of the population in group g.[14] Let $E[\hat{y}|d_j, g]$ denote the mean response of individuals in group g observed in frame $j \in \{l, h\}$. For the results in this section, we can modify the Unconfoundedness assumption so that it need only hold after conditioning on one's group:

Assumption 3': Conditional Unconfoundedness

$$d_i \perp (\hat{y}_i(d_h),\, \hat{y}_i(d_l)) \mid g_i \tag{17}$$

Proposition 4: Relating Consistency and Observable Characteristics *Define* $\bar{\psi} = \sum_{g \in G} \left[P_g \left(E[\hat{y}|d_l, g] + 1 - E[\hat{y}|d_h, g]\right)\right]$. *Under Assumptions 2 and 3':*[15]

(4.1) *The distribution of g among the consistent respondents, $P(g_i = g \mid \psi_i = 1)$, is equal to* $(P_g)\, \frac{E[\hat{y}|d_l,g]+1-E[\hat{y}|d_h,g]}{\bar{\psi}}$.

[12]Even if obtaining this information were possible, in many contexts it may be that surveying a respondent under one frame would bias their subsequent answers under alternate frames (LeBoeuf and Shafir, 2003).

[13]The results in this section are analogous to Abadie (2003), who, in the context of instrumental variable estimation, shows how to identify the aggregate characteristics of the "compliers" in instrumental variable estimation, despite the fact that individual members of that population cannot be identified.

[14]Like the other population moments described here, P_g can be consistently estimated from its sample analogue.

[15]Note that the Consistency Principle is not required for this result to hold; that is, the question of which respondents are consistent is distinct from the question of whether the responses of those who are consistent reveal the true value of y.

(4.2) The distribution of g among the inconsistent respondents, $P(g_i = g \,|\, \psi_i = 0)$, is equal to $(P_g)\,\frac{E[\hat{y}|d_h,g]-E[\hat{y}|d_l,g]}{1-\bar{\psi}}$.

(4.3) The fraction of consistent respondents, $P(\psi_i = 1)$ is equal to $\overline{\psi}$.[16]

Proof *Lemma 1: The fraction of group g that is consistent, $P(\psi_i = 1 \,|\, g_i = g)$, is equal to $E[\hat{y}|d_l, g] + 1 - E[\hat{y}|d_h, g]$.*

Proof of Lemma 1

By the same logic as in Proposition 2, Conditional Unconfoundedness and Frame Monotonicity allow us to obtain

$$E[\hat{y}_i|d_l,\, g_i = g] + 1 - E[\hat{y}_i|d_h,\, g_i = g] = P(\psi_i = 1 \,|\, g_i = g)$$

Proof of (4.1)

Note that by Bayes rule, $P(g_i = g|\psi_i = 1) = \frac{P(\psi_i=1|g_i=g)\,P(g_i=g)}{P(\psi_i=1)}$. Now consider the terms in (4.1). First, note that $\bar{\psi}$ is equal to $P(\psi_i = 1)$ by the law of iterated expectations and Lemma 1. Additionally, from Lemma 1 we know that $E[\hat{y}|d_l, g] + 1 - E[\hat{y}|d_h, g]$ is equal to $P(\psi_i = 1 \,|\, g_i = g)$. Combining these results yields (4.1).

Proof of (4.2)

The proof is analogous to (4.1).

Proof of (4.3)

Because ψ_i is binary, $P(\psi_i = 1) = E[\psi_i]$, which equals $\sum_g E[\psi_i \,|\, g_i = g]\, P(g_i = g)$ by the law of iterated expectations. The result then follows from the definition of $\overline{\psi}$ and Lemma 1. ■

Illustration of Proposition 4: Beliefs About Climate Change

Schuldt, Konrath and Schwarz (2011), discussed above, documented that substantially more Republicans reported believing in "climate change" than in "global warming." Although the authors of that paper were able to investigate the intensity of framing effects among various subgroups of respondents by repeating their analysis with various subgroup, Proposition 4 allows us to go further. Specifically, Proposition 4 allows us to identify and compare the aggregate characteristics of the consistent and inconsistent

[16] Under assumptions (2) and (3'), $\bar{\psi}$ is a better estimate for $P(\psi_i = 1)$ than the one derived in Corollary 2.1 because even when (unconditional) Unconfoundedness (3) holds, $\bar{\psi}$ has desirable finite-sample properties in the presence of spurious correlation between frame assignment and observable characteristics.

groups of respondents – that is, to identify which types of respondents are most prone to reporting differing beliefs when presented with the "climate change" and "global warming" frames. Table 1 displays the results of this analysis using the data from Schuldt, Konrath and Schwarz (2011).[17]

Table 1: Climate Change Beliefs - Characteristics of Consistent and Inconsistent Respondents

	"Global Warming" Frame $E[y\|d_x, g]$	"Climate Change" Frame $E[y\|d_y, g]$	Fraction Consistent $E[\psi\|g]$	Among All Respondents $P(g)$	Among Inconsistent $P(g \mid \psi = 0)$
Gender					
Female	0.54	0.68	0.86	0.57	0.46
Male	0.43	0.64	0.79	0.43	0.54
Education					
No College	0.48	0.65	0.83	0.47	0.50
College	0.51	0.67	0.84	0.53	0.50
Age					
18 - 44	0.51	0.73	0.78	0.35	0.45
45 - 57	0.52	0.69	0.83	0.33	0.33
58 -	0.45	0.56	0.89	0.32	0.21

Source: Authors' calculations from data collected by Schuldt, Konrath and Schwarz (2011). Analysis restricted to self-identified Republicans.

The results suggest several interesting patterns. First, women were less affected than men by the framing; men represented only 43 percent of all respondents but 54 made up percent of the inconsistent group. Second, older respondents tended to be more consistent. The three age groups we examined each represented approximately one-third of the population, but the youngest group was over-represented among the inconsistent respondents whereas the oldest group was under-represented. Finally, college graduates appear no more likely to make consistent responses.

[17]Standard errors may be obtained by bootstrapping or using the delta method, see Appendix D to Goldin and Reck (2014) for a discussion.

Note that it is also possible to apply Proposition 2 to the data of a particular subset of respondents to identify the average responses of the consistent respondents within that group. For example, doing so yields that 65 percent of the consistent respondents below age 44 believed that climate change / global warming was occurring whereas that fraction was only 51 percent for the consistent respondents aged 58 and older.

VI. Identifying the Full Population Mean

This section considers the problem of estimating the mean of the survey variable among the full population of respondents in the presence of framing effects. Although in many applications the researcher's focus will be on the subset of consistent survey-takers, there will also be cases in which the distribution of the survey variable among the full population of respondents – both consistent and inconsistent – will be of primary interest. 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). Another example is exit polling data for an election, where answer order may bias which candidate a voter recalls voting for (especially in more obscure contests such as judicial elections), but pollsters nevertheless want to know how the population voted. In such cases, there exists a well-defined "correct" answer to the survey variable even for those respondents whose reported answers vary by frame; their susceptibility to framing effects simply prevents that true value from being revealed to the researcher.[18]

Given the consistent-subgroup mean, the problem of recovering the full population mean parallels the well-studied question of inference under sample selection.[19] As discussed below, when selection into the consistent subgroup is uncorrelated with the sample variable, the consistency-adjusted mean will provide an unbiased estimate for the full population. Otherwise, obtaining an unbiased estimate of the full population requires additional techniques. An important difference from the standard setting is that researchers can typically observe whether individual respondents are included in the sample or not (e.g., whether they respond to the survey); in contrast, the researcher cannot directly observe whether an individual respondent is consistent. As described below, this difference complicates the task of adapting the standard sample selection techniques to this setting.

[18]Understood this way, the problem parallels the one addressed in the "Contaminated Outcomes" literature, in that the goal is to learn about a parameter of interest in settings where the observed data contain some fraction of erroneous observations (Joel L Horowitz and Charles F Manski, 1995) (here, the erroneous observation occurs when a survey-taker responds according to the frame). And as in that literature, the answer here will also turn on whether the contaminated outcomes are generated in a way that is systematically related to the underlying variable of interest.

[19]See Manski (2003).

VI.1 Relationship Between the Consistent-Subgroup and Full Population Means

Proposition 2 permitted researchers to estimate the consistent-subgroup mean, i.e., the mean of the survey variable for the subset of the population whose responses are unaffected by the frame, $\mu_c = E[y_i|\psi_i = 1]$. The next proposition clarifies the relationship between this quantity and the full population mean, $E[y]$.

Proposition 5: The Consistent-Subgroup and Full Population Means *Under Assumption 1-3,* $Y_c = E[y] + \frac{cov(y_i,\psi_i)}{P(\psi_i=1)}$.

Proof From Proposition 2, $Y_c = E[y_i|\psi_i = 1]$. We will show $E[y_i|\psi_i = 1] = E[y] + \frac{cov(y_i,\psi_i)}{P(\psi_i=1)}$.

By the law of iterated expectations,

$$E[y_i|\psi_i = 1] = \frac{E[y_i] - E[y_i|\psi_i = 0]\, P(\psi_i = 0)}{P(\psi_i = 1)}$$

$$= \frac{E[y_i\, \psi_i]}{P(\psi_i = 1)}$$

where the second equality follows from the binary nature of the variables. The result follows from applying the identity $E[y_i\, \psi_i] = cov(y_i, \psi_i) + E[y_i]\, E[\psi_i]$. ■

Proposition 5 shows that the extent to which the consistent-subgroup mean departs from the full population mean depends on the empirical relationship between respondents' susceptibility to framing effects and their value of y_i. Understanding this relationship is essential for recovering the distribution of y for the full population. However, in the special case that frame-sensitivity is uncorrelated with respondents' true answers, the consistent-subgroup mean will be identical to the full population mean. We label this special condition "consistency independence."

Assumption: Consistency Independence

$$cov(y_i, \psi_i) = 0 \tag{18}$$

With consistency independence, the parameter identified by the consistency-adjusted mean is equal to the mean value of y over the full population.

Corollary 5.1: Identifying the Population Mean Under Consistency Independence *Under Assumptions 1-3 and Consistency Independence, $Y_c = E[y_i]$.*

Proof Follows immediately from Proposition 5. ■

When consistency independence holds, estimating the consistent subgroup mean, $E[y_i|\psi_i = 1]$, is equivalent to estimating the full population mean, $E[y_i]$.[20] Frequently, however, the plausibility of this assumption will be suspect. For example, suppose that a survey is investigating support for a political candidate. Consistency independence will fail, for example, if more highly educated respondents are more likely to choose consistently, and education is empirically correlated with respondents' support for the candidate. That is, if more highly educated people are more likely to support the candidate, Proposition 5 shows that the consistency-adjusted mean will overestimate the full population mean.

VI.2 Bounds for the Full Population Means

Absent any information on the empirical relationship between respondents' frame-sensitivity and the distribution of y, Assumptions 1-3 are still sufficient to permit partial identification of $E[y_i]$. This section derives worst-case bounds (Manski, 1989) by first identifying the consistent-subgroup mean and then making the most extreme assumptions possible regarding the distribution of y_i among the inconsistent respondents.

Proposition 6: Partial Identification of the Full Population Mean with Frame Monotonicity *Under Assumptions 1-3, $E[y] \in [E[\hat{y}|d_l], E[\hat{y}|d_h]]$.*

Proof By the law of iterated expectations, $E[y_i] = E[y_i|\psi_i = 1]\, P(\psi_i = 1) + E[y_i|\psi_i = 0]\, P(\psi_i = 0)$. Because $P(\psi_i = 0) \geq 0$ and $E[y_i|\psi_i = 0] \in [0, 1]$, we know

$$E[y_i|\psi_i = 1]\, P(\psi_i = 1) \leq E[y_i] \leq E[y_i|\psi_i = 1]\, P(\psi_i = 1) + P(\psi_i = 0) \tag{19}$$

[20] Consistency independence thus parallels the "missing-at-random" assumption in the literature on inference under sample selection.

Additionally, by Lemma 1 to Proposition 2, we know that $E[\hat{y}|d_l]$ is equal to the left-most quantity in (19). Similarly, the right-most quantity in (19) is given by $E[\hat{y}_i|d_h]$, proving the result. ■

The bounds derived in Proposition 6 are quite intuitive. Note however that the result requires frame monotonicity in order to hold. The greater the fraction of consistent respondents, the more informative the bounds will be. The following Proposition demonstrates that even without frame monotonicity, the observed data partially identies the full population mean, but depending on the data the bounds might be quite wide.

Proposition 7: Partial Identification of the Full Population Mean without Frame Monotonicity *Under Assumptions 1 and 2:*

(7.1) *If* $Y_c > \frac{1}{2}$*, then* $E[y_i] \in [E[\hat{y}_i|d_l] - (1 - E[\hat{y}_i|d_h]),\ 1]$

(7.2) *If* $Y_c < \frac{1}{2}$*, then* $E[y_i] \in [0,\ E[\hat{y}_i|d_l] + E[\hat{y}_i|d_h]]$

Proof By re-arranging (11) and (12), we obtain that

$$P(\psi_i = 1) = \frac{E[\hat{y}_i|d_l] + 1 - E[\hat{y}_i|d_h] - 2\alpha}{1 - 2\alpha} = 1 - \frac{E[\hat{y}_i|d_h] - E[\hat{y}_i|d_l]}{1 - 2\alpha} \tag{20}$$

$$\mu_c = \frac{E[\hat{y}_i|d_l] - \alpha\left(E[\hat{y}_i|d_h] + E[\hat{y}_i|d_l]\right)}{E[\hat{y}_i|d_l] + 1 - E[\hat{y}_i|d_h] - 2\alpha} \tag{21}$$

Substituting (20) and (21) into our expression for $E[y_i]$, (4), and simplifying, we obtain that

$$E[y_i] = \frac{E[\hat{y}_i|d_l] - \alpha\left(E[\hat{y}_i|d_h] + E[\hat{y}_i|d_l]\right)}{1 - 2\alpha} + \mu_n \frac{E[\hat{y}_i|d_h] - E[\hat{y}_i|d_l]}{1 - 2\alpha} \tag{22}$$

where $\mu_n = E[y_i|\psi_i = 0]$.

Proof of (7.1): Suppose $Y_c > \frac{1}{2}$. To obtain a lower bound on $E[\phi_i]$, set $\mu_n = 0$, so (22) implies that

$$E[y_i] \geq \frac{E[\hat{y}_i|d_l] - \alpha\left(E[\hat{y}_i|d_h] + E[\hat{y}_i|d_l]\right)}{1 - 2\alpha}. \tag{23}$$

The right-hand side is decreasing in α under our assumptions, so we can obtain a lower-bound that only depends on observable moments using the maximum feasible α. Because $\mu_c \leq 1$, we know from (21) that $\alpha \leq \frac{1 - E[\hat{y}_i|d_h]}{2 - E[\hat{y}_i|d_h] - E[\hat{y}_i|d_l]}$. Using the highest possible value of α in (23) yields that

$$E[y_i] \geq E[\hat{y}_i|d_l] - (1 - E[\hat{y}_i|d_h]) \tag{24}$$

It is straghtforward to verify that $Y_C > \frac{1}{2} \implies E[\hat{y}_i|d_l] > 1 - E[\hat{y}_i|d_h]$, implying this bound will be informative. The upper bound of $E[y_i] \leq 1$ is trivial.

Proof of (7.2): Suppose $\underline{Y_c < \frac{1}{2}}$. To obtain an upper bound on $E[\phi_i]$, set $\mu_n = 1$, so (22) implies that

$$E[y_i] \leq \frac{E[\hat{y}_i|d_h] - \alpha \left(E[\hat{y}_i|d_h] + E[\hat{y}_i|d_l]\right)}{1 - 2\alpha}. \tag{25}$$

The right-hand side in this case is increasing in α, so we should find a maximum feasible α for an upper bound. Because $\mu_c \geq 0$, we know from (21) that $\alpha \leq \frac{E[\hat{y}_i|d_l]}{E[\hat{y}_i|d_h]+E[\hat{y}_i|d_l]}$. Combining this with (25) yields

$$E[y_i] \leq E[\hat{y}_i|d_l] + E[\hat{y}_i|d_h].$$

Similarly to before, $Y_C < \frac{1}{2} \implies E[\hat{y}_i|d_l] + E[\hat{y}_i|d_h] < 1$, so the bound is non-trivial. The lower bound in this case is trivial. ■

Even without the assumption that frames monotonically affect survey responses, calculating the mean of the survey variable within each frame still conveys some information on the population mean. In this case the researcher can only obtain a non-trivial bound in one direction, depending on the relative share of respondents choosing against the frame (represented by Y_c). Intuitively, the bounds are calculated using the most extreme values possible of the outcome for the inconsistent respondents (μ_n), and the share of inconsistent respondents who are frame-defiers, α that are consistent with what we know about μ_c and the observed population moments. The degree to which these bounds place restrictions on the population mean depends on how far Y_c is from $\frac{1}{2}$. For example, $E[\hat{y}_i|d_l] = 0.05$ and $E[\hat{y}_i|d_h] = 0.20$ imply $E[y_i] \leq \frac{1}{4}$. In the case in which Y_C is exactly equal to $\frac{1}{2}$, the bounds provide no additional information at all because in that case the observed population moments do not constrain the feasible values of α.

VI.3 Consistency Weights

Although Assumptions 1-3 are sufficient to recover the subgroup-consistent mean of y, Proposition 5 highlights that this parameter diverges from the full population mean of y, $E[y_i]$, except in special cases (i.e., consistency independence). The result in this section takes advantage of survey-takers' observable characteristics to recover the full population mean under a relaxed version of the consistency independence assumption.

As in Section V, suppose that the sample is partioned into K groups, based on observable characteristics of the respondents, $G = \{g_1, g_2, ...g_K\}$. Let $g_i \in G$ denote the group of individual i. We now assume that consistency independence holds, conditional on one's group.

Assumption: Conditional Consistency Independence

$$cov(y_i, \psi_i \mid g_i) = 0 \tag{26}$$

In words, this assumption states that among members of the same group, a respondent's susceptibility to framing effects is uncorrelated with his or her value of y. For example, support for a candidate may be correlated with both consistency and education, so long as education is observable. As in other matching-on-observables techniques, the more individual characteristics associated with consistency that are observable, the more likely the assumption is to hold.

As in Section V, let $E[\hat{y}|d_j, g]$ denote the mean response of individuals in group g observed in frame $j \in \{l, h\}$, $E[\hat{y}|d_j, g] = E[\hat{y}_i|d_i = d_j,\, g_i = g]$. Let $w(g)$ denote the *consistency-weight* for group $g \in G$, $w(g) \equiv \left(\frac{1}{E[\hat{y}_i|d_l]+1-E[\hat{y}_i|d_h]}\right)(P_g)$, where as above P_g denotes the fraction of the population in group g. Finally, define the *consistency-weighted mean*: $Y_w = \sum_g E[\hat{y}|d_l, g]\, w(g)$

Proposition 8: The Consistency-Weighted Mean *Under Assumptions 1, 2, 3', and Conditional Consistency Independence 26,* $Y_w = E[y]$.

Proof Using the definition of $w(g)$, re-write Y_w as

$$Y_w = \sum_{g \in G} (P_g) \left(\frac{E[\hat{y}_i|d_l]}{E[\hat{y}_i|d_l] + 1 - E[\hat{y}_i|d_h]} \right) \tag{27}$$

By the same logic as Proposition 2 (but with the Conditional Unconfoundedness assumption replacing the unconditional version), it is straightforward to show that $\frac{E[\hat{y}_i|d_l]}{E[\hat{y}_i|d_l]+1-E[\hat{y}_i|d_h]}$ is equal to $E[y_i|\psi_i = 1,\, g_i = g]$. But by Conditional Consistency Independence, we have $E[y_i|\psi_i = 1,\, g_i = g] = E[y_i|g_i = g]$.

Thus

$$Y_w = \sum_{g \in G} E[y_i | g_i = g] \, P(g_i = g)$$

But by the law of iterated expectations, $\sum_{g \in G} E[y_i|g_i = g] \, P(g_i = g) = E[y_i]$, yielding the result. ■

The consistency weights proposed here are 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 responses to the survey can be observed. The weights themselves can be understood in familiar terms: an individual's weight represents the inverse of the probability that the individual is consistent, i.e. that she falls into the consistent subgroup, along with an adjustment for the fact that certain groups may be disproportionately observed under a particular frame.[21]

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. Similarly, 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 characteristics 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 full population average.

[21] In particular, it is straightforward to show that $\frac{1}{\overline{Y}_{l,g} + 1 - \overline{Y}_{h,g}}$ converges to the inverse of $P(\psi_i = 1 | g_i = g)$ and that $\frac{N_g}{N_{j,g}}$ converges to the inverse of $P(d_i = d_j \mid g_i = g)$. Even under random assignment of frames, the adjustment for the fact that certain groups may be disproportionately observed under a particular frame is desirable because it ensures that the results are unaffected by spurious correlation between observables and frame assignment in small samples.

VII. Conclusion

In this paper we have developed a straightforward framework to understand a ubiquitous problem in survey research: the sensitivity of responses to seemingly-arbitrary features of the survey's design. We showed how the conventional approach to dealing with this problem results in biased estimates and proposed an alternative technique. Even when our proposed estimator fails due to frame non-monotonicity, the parameter it identifies is closer to the consistent subgroup mean than is the conventional approach. We also adapted identification techniques from other settings to provide tools for identifying the characteristics of consistent and inconsistent respondents and for identifying the mean of the survey variable for the full population.

The degree to which the parameters discussed here can be identified from the data depend on the assumptions the researcher is willing to impose. When only the consistency principle is imposed, the data permit partial identification of μ_C and $E[y_i]$, where the bounds on the latter (Proposition 7) are wider than those on the former (Proposition 3). Adding frame monotonicity permits point identification of μ_C (Proposition 2) and narrows the bounds on $E[y_i]$ (Proposition 6). Finally, assuming either Consistency Independence (Corollary 5.1) or Conditional Consistency Independence (Proposition 8) allows $E[y_i]$ to be point identified as well.

In applying our results, two limitations are important to keep in mind. First, we have focused on the relatively simple setting of binary survey variables with two frames; applying the approaches in settings with additional frames or answer choices requires further assumptions. 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 Financil Crisis and the Well-Being of Americans."

Gallup Organization. 2012. "Gallup News Service Poll: Monthly Indicators Update." Roper Center Public Opinion Archives Survey Dataset.

Goldin, Jacob, and Daniel Reck. 2014. "Preference Identification Under Inconsistent Choice: A Reduced-Form Approach."

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

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

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

Horowitz, Joel L, and Charles F Manski. 1995. "Identification and robustness with contaminated and corrupted data." *Econometrica: Journal of the Econometric Society*, 281–302.

Imbens, Guido W, and Joshua D Angrist. 1994. "Identification and estimation of local average treatment effects." *Econometrica: Journal of the Econometric Society*, 467–475.

Krosnick, J, and L Fabrigar. Forthcoming. *The Handbook of Questionnaire Design.* Oxford University Press.

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

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

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

Manski, Charles F. 1997. "Monotone treatment response." *Econometrica: Journal of the Econometric Society*, 1311–1334.

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

Neyman, J. 1923. "On the Application of Probability Theory to Agricultural Experiments." *Statistical Science*, 5.

Presser, Stanley, and et al. 2004. *Methods for Testing and Evaluating Survey Questions.* Wiley.

Schuldt, Jonathon, Sara Konrath, and Norbert Schwarz. 2011. ""Global Warming" or "Climate Change"?" *Public Opinion Quarterly.*

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

Schwarz, N. 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: 1270169.

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.

Zhang, J, and D Rubin. 2003. "Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by "Death"." *Journal of Educational and Behavioral Statistics*, 28: 353–368.