

**Measurement Error as a Threat to Causal Inference:
Acquiescence Bias and Deliberative Polling**

G. Michael Weiksner, Stanford University*

Draft Paper prepared for the
Political Methodology Conference in Ann Arbor, 2008

** DO NOT CITE OR DISTRIBUTE WITHOUT PERMISSION **

Abstract

Experiments, unlike observational studies, are rarely criticized for yielding invalid causal inferences. However, I identify measurement error as a threat to causal inference of an experiment. In particular, acquiescence bias, a common and substantial source of measurement error within surveys, may be correlated with experimental manipulations. Using data from a survey experiment embedded in a Deliberative Poll, I find that acquiescence bias causes significant measurement error and that the bias differs before and after deliberation. I conclude that even experimental researchers should heed the recommendation by questionnaire design researchers to refrain from asking agree/disagree questions completely and instead ask only construct-specific questions to avoid this threat to validity.

* A sincere debt of gratitude goes to Jim Fishkin, Bob Luskin and Jon Krosnick. Jim and Bob approved of the inclusion of the survey experiment in their Deliberative Poll and have been supportive in everyway of this research. The inspiration for this research, and indeed the story line of the literature review, comes from Jon and his course of survey design. I also thank Daniel Schneider, Neil Malhotra, and Josh Pasek. Of course, I alone bear responsibility any remaining errors.

Introduction

The well-known strength of experiments is their internal validity gained from random assignment of subjects to treatment groups. There are only two common and related criticisms of experiments. First, critics may raise procedural concerns, such as haphazard versus true random assignment. Second, critics may question the interpretation of the experimental manipulation. Rarely, if ever, do critics call the internal validity (that is, the basis of a causal inference) of a well-administered experiment into question. In this paper, I identify that measurement error as a real threat to internal validity.

One important type of measurement error is acquiescence bias, the tendency of some respondents to agree with a statement independent of its content. The harm of acquiescence bias has been fairly well documented by researchers studying questionnaire design. Three findings are relevant. First, agree/disagree questions are highly susceptible to acquiescence bias. Second, the bias is substantial. Third, acquiescence increases when respondents are less motivated to optimize their answers.

Until now, acquiescence bias has been ignored by experimental researchers. Many experiments use surveys to gather data. In these surveys, agree/disagree questions are quite common. The potential threat is that a common secondary effect of experiment treatment may be to change how motivated a respondent is. Hence, experiments that both use agree/disagree questions and explore treatments that affect the motivation of respondents may have measurement error that is correlated with treatment. When measurement error is correlated with treatment, then causal inference may not be valid.

This paper proceeds as follows: in the first section, I review the acquiescence bias literature: what acquiescence bias is, why it occurs, and how much contamination it causes. In

the second section, I apply the theory to a political example, Deliberative Polling. I define Deliberative Polling, identify how acquiescence bias threatens the validity of its conclusions and review the procedure and results of a survey experiment embedded in a Deliberative Poll. In the third section, I explore the significance for experimental research in political science and propose a strategy to avoid the threat of contamination.

What Acquiescence Bias Is

Concerns about acquiescence bias can be traced back to Cronbach and his pioneering work on measurement error. In his works, he developed two different concepts of acquiescence. Since he does not explicitly distinguish these two different concepts of concepts, I will label them as “nonrandom guessers” and “logically inconsistent guessers” and define them below.

“Nonrandom guessers” always pick the same response in true/false items when they are uncertain of the correct answer. Cronbach (1941, 1942) showed that poor students do better on questions when the correct answer is true because poor students were more likely to guess true than false. In addition, he found that tests with questions where the correct answers are false yield more reliable and valid measures of knowledge.

In contrast, “logically inconsistent guessers” are respondents who demonstrate “any tendency causing a person consistently to make different responses to test items than he would have made if the same content had been presented in a different form.” (Cronbach, 1946) In other words, a logically inconsistent guesser might agree with a statement when prompted “do you agree or disagree” but disagree with the same statement if the prompt is reversed to “do you disagree or agree”.

Rorer (1965) points out that these two definitions are technically independent. For example, the logically inconsistent guesser in the example above is not a nonrandom guesser.

Conversely, a nonrandom guesser may answer “true” on every question in a particular survey without logically contradicting herself.

Modern references to acquiescence bias seems to have settled on Cronbach’s first definition, non-random guessing. Schuman and Scott (1989) say “acquiescence or agreeing-response bias refers to a presumed tendency for respondents to agree with attitude statements presented to them.” (p.203) By this definition, we can measure acquiescence bias simply by looking at the likelihood that a respondent will agree with an arbitrary survey question.

But I think this definition is incomplete. Imagine a particular survey that includes a battery of positively worded statements about the United States. Might a patriot not truly agree with every one of these pro-American statements? I think a complete definition of acquiescence bias must include the counterfactual nature of Cronbach’s second definition, logically inconsistent guessing.

Indeed, Schuman and Scott’s review focuses on experiments where participants seem to agree with both a statement and its reverse. This approach to studying acquiescence bias suggests a better way to define acquiescence bias: the tendency of respondents to agree with a statement independent of the content of the statement. (As a side note, the other independent dimension that Chronbach identified can then be called response form bias: the tendency to respond the same way based on the form of the question only.) As I shall elaborate later, Schuman and Scott’s review fits this more precise definition quite well.

Why Acquiescence Bias Occurs

It takes two factors for acquiescence bias to occur. Researchers must ask a question that is susceptible to acquiescence bias and then respondents must acquiescence. I analyze each of these factors in turn.

First, why do researchers ask questions prone to acquiescence bias? Krosnick (2007) argues that certain types of questions, agree/disagree, true/false and yes/no questions, are chosen by researcher because they are easy to administer:

1. An agree/disagree question is answered in two thirds the time that it takes to answer a question with construct-specific response choices. (Krosnick, 2007)
2. There are fewer “don’t know” responses to yes/no questions than questions with construct-specific response choices. (Krosnick, 2007)
3. People prefer agree/disagree questions over questions with construct-specific response choices. (Krosnick, 2007)

In addition to Krosnick’s list of reasons, I would add that: researchers find it easier to create agree/disagree, true/false and yes/no questions. These questions are easier to employ in surveys because researchers may avoid making choices about what concepts they really want to measure.

Agree/disagree questions are susceptible to acquiescence bias, yet are highly prevalent in surveys. To understand why they are used so frequently by researchers, we must understand what the alternative is. The alternative to an agree/disagree question is a construct-specific question. Designing construct-specific questions requires the researcher to find the “hidden variable.” For example, a typical hospital satisfaction survey might ask whether a respondent agrees with the following statement: “The nurses treated me with respect.” In contrast, a researcher could ask a similar construct-specific question: “How respectfully did the nurses treat you?”

What is the theoretical difference between the two versions of the question? In the agree/disagree version, the respondent has to make and integrate two separate judgments. First, she must decide how respectfully she thinks the nurses treated her. Second, she must decide how sure she is in her judgment. If the respondent “strongly agrees” with that statement, does it mean that she thinks the nurses were exceptionally respectful or that he or she is quite certain that the nurses were indeed respectful? It is unclear. In most cases, it is more likely the researcher wants just the answer to the first question so Krosnick and others advocate merely asking the construct-specific question to begin with.

Second why do respondents acquiesce in the first place? Three explanations have been given: norms of conduct (Schlenker, 1980; Leary & Kowalski, 1990), status differential (Lanski & Leggett, 1960), and satisficing theory (Krosnick & Alwin, 1987). First, norms of conduct suggest that respondents want to be polite and agreeable. Hence, they want to avoid disagreeable conflict with the questioner and tend to agree. Second, respondents may feel that they have lower social status than the questioner, and, as such, that they tend to agree with the higher-status questioner. Third, satisficing theory predicts that respondents may pick an acceptable but non-optimal answer to speed to complete the survey more quickly.

In theoretical discussions, analysis of acquiescence bias has become increasingly sophisticated. In particular, early studies (e.g., Couch & Keniston, 1960 and Hamilton, 1968) tend to focus exclusively on dispositional factors, e.g., acquiescence is as a personality trait of certain respondents. More recent studies of acquiescence bias, however, have also included situational factors like the order and wording of questions in their accounts (see Krosnick, 1999; e.g., McClendon, 1989 and Moum, 1988). For example, items that appear later in the survey

and with vague wording are more likely to induce acquiescent behavior than items earlier in the survey with clear wording.

The Harm Caused By Acquiescence Bias

There is substantial evidence that a portion of respondents answering questions are acquiescent on occasion. First, respondents are 10-20% more likely to agree than disagree with contentless statements (Krosnick, 2007). Second, item reversals (e.g., asking “Pepsi or Coke?” versus “Coke or Pepsi?”) have only weak negative (and sometimes even positive) correlations across items. Moreover, 10-15% of respondents will agree with both an item and its opposite but few people disagree with both. Third, surveys often exhibit a 15-20% bias in comparison with construct-specific questions. Respondents correctly answer more construct-specific questions than agree/disagree questions on factual matters. Construct-specific questions also show more agreement with informants than agree/disagree questions. (Krosnick, 2007)

Taken individually, the evidence in different studies on acquiescence bias may be consistent with alternative explanations. For example, can we be sure that respondents use the same strategy to respond to both contentless and meaningful questions? Item reversals are notorious for having “knife-edge” possibilities where a person might truly agree with a statement and its reverse. (Consider the pair of statements: “Freer trade leads to lower prices” and “Freer trade leads to higher prices.” You could legitimately disagree with both statements if you believe that freer trade has no impact on prices.) Some researchers may object that agree/disagree questions are actually the right measurement, and the construct-specific questions are the inaccurate ones.

The cognitive model of satisficing is the only theory that is consistent with all the evidence across all the relevant studies. Individually the evidence is largely consistent with acquiescence

bias theory, and alternative explanations are ruled out. For example, the fact that double disagreement is very unlikely but double agreement is common helps rule out the “knife-edge” objection. Also, construct-specific questions are more likely than agree/disagree questions to be consistent on factual questions and with informants.

Acquiescence bias harms data quality in two ways: measurement of constructs may be biased and correlations among questions and constructs may be overstated. Biased measures of constructs means that self-reported agreement with a statement may overstate real agreement. This bias is substantial: experimental and observational research studies find that measures of attitudes may be inflated by 15% to 25%. Distorted correlations may yield spurious results when analyzing relationships across individuals. Unlike the measurement bias at the aggregate level, analyses that require high-quality, individual-level data like correlations and regressions are impossible to correct.

How often do respondents agree with a statement when, if they would consider it more fully, they would actually disagree? Schuman, Scott (1989) designed a clever experiment to answer this question. Over two periods, randomly assigned subsamples were asked three versions of questions addressing the same construct. In the first period, each subsample was asked one of the following questions:

1. “Individuals are more to blame than social conditions for crime and lawlessness in this country. How strongly do you agree or disagree with this statement?”
2. “Social conditions are more to blame than individuals for crime and lawlessness in this country. How strongly do you agree or disagree with this statement?”

3. “Which in your opinion is more to blame for crime and lawlessness in this country— individuals or social conditions?”

In the second period a few months later, participants were again asked a question about this construct. If participants received the first version of the question above in the first period, they instead received question two in the second period. Conversely, participants who received the second version in first period were instead the second version in the second period. Finally, participants who were asked third version in the first period received that same in the second period too.

Notice that it is inconsistent to repeat the same answer (e.g., double agreement or double disagreement) in the first two subsamples when the question changes. Conversely, it is inconsistent to switch answers in the third group where the questions remain the same. Interestingly, double agreement (on average, 25.6% of participants) is more than twice as likely as double disagreement (on average, 9.9%) in each of the first two groups. Meanwhile, there is no statistically significant pattern for inconsistency in the third group.

Schuman and Scott further argue that we should break the 25.6% of participants that exhibit double agreement into two components: real acquiescence bias and random error. Fortunately, the double disagreement provides an upper bound on the random error. If everyone who disagreed at both times did so randomly, we know that random error is at most 9.9%. However, some or even all of the double disagreement may be real naysaying. Hence, they conclude that the real acquiescence, net of the random error, falls in the range of 16-25%.

It is also possible to use statistical analyses to estimate acquiescence bias and correct for it. Krosnick (2007) summarizes how magnitudes and even directions of responses can change

when you correct for acquiescence bias when surveying political attitudes. Krosnick finds that population estimates may increase in some cases or decrease in other cases when you control for acquiescence bias. In some cases, statistical corrections can even reverse the substantive inferences you would draw from a population estimate without including for such a factor (e.g., Jackman 1973; Winkler, Kanouse & Ware, 1982). For example, early studies that found that people were susceptible to authoritarianism may actually only show that people are susceptible to acquiescence bias.

Yet, statistical controls are not a panacea. Although statistical analyses (such as including a method factor) can help correct incorrect population estimates, they cannot correct individual-level responses. Even with the most sophisticated controls, we cannot reliably perform regression and other analyses that relate individuals or variables because the measurement has been potentially systematically biased. The same flaw holds even when we modify the survey design to include a balanced set of item reversals. (Krosnick, 2007)

Deliberative Polling

Deliberative Polls are an exercise in counterfactual public opinion: what would the public think, given the chance to think? They are *deliberative* in the sense that participants are provided carefully balanced briefing materials laying out the major arguments for and against given policy proposals or electoral choices and discuss these materials in small groups and pose questions to experts and leaders in plenary sessions over an entire weekend. They are *polls* in the sense that participants are drawn from a random sampled. In addition, Deliberative Polls are a *quasi-experiment* since participants are surveyed before and after deliberation.

Fishkin, Luskin and colleagues have studied extensively how deliberation affects public opinion (Fishkin, 1991; Luskin, Fishkin, n.d.; Luskin, Fishkin, et al., 2002). In over two dozen

Deliberative Polls, they have found that political attitudes measured in polls conducted post-deliberation often differ significantly from pre-deliberation polls. Moreover, changes in knowledge can be used to predict changes in attitude. This pattern of attitude change is offered as evidence that post-deliberation opinions are more informed than pre-deliberation opinions.

However, satisficing theory (Simon, 1957; Krosnick & Alwin, 1987) also suggests that deliberation may cause opinion shifts. Satisficing theory predicts that people must be motivated to choose an optimal rather than merely acceptable response. Agreeing to a survey question is an acceptable (although not necessarily optimal) response. Moreover, evidence from Deliberative Polls indeed shows that participants are more motivated after spending a weekend of intense deliberation. Hence, changes in the level of agreement may be affected by both changes in attitude and changes in motivation.

Does acquiescence bias distort the measurement of opinion in Deliberative Polls? Can we disentangle the effects of information and other intended treatments from potentially distorting measurement effects?

Two Threats to the Interpretation of the Results of Deliberative Polls

Given the tight fit between explanation and evidence and the documented harm of acquiescence bias in traditional surveys, how might acquiescence bias affect the analysis of Deliberative Polls? I have identified two plausible consequences: (1) a threat to the **construct validity**¹ and (2) a threat to the **internal validity**² of Deliberative Polls.

The threat to the construct validity is that acquiescence bias may distort the measurement of considered opinions after deliberation. According to Luskin, Fishkin and Jowell (2002), the

¹ Construct validity is whether you are justified making the theoretical conclusions based on your operationalization. In layman's terms, construct validity is whether you are measuring what think and say that you are measuring.

² Internal validity is whether you are justified in making conclusions about causal relationship based on the evidence that you have collected.

goal of the Deliberative Polling enterprise is to narrow the gap between off-the-top-of-the-head opinions and well-informed opinions. They claim, and provide evidence that, the polls conducted after the deliberation process do indeed close this gap. For present purposes, the issue is: does acquiescence bias threaten the validity of this claim? And even if the results of a Deliberative Poll after deliberation represent more informed opinion than traditional polling techniques (as I believe), does acquiescence bias introduce meaningful measurement bias?

The threat to internal validity is that acquiescence bias may distort the quasi-experimental evidence that deliberation makes better citizens. A secondary goal of Deliberative Polls is to examine the claims that deliberation makes better citizens (Luskin, Fishkin, 2002). The primary analytical technique in this stream of research is to compare participants on a variety of measures before and after deliberation. Like other survey research, acquiescence bias may introduce measurement error. Specific to deliberation as a quasi-experiment, acquiescence bias may pose an additional threat to the research if it has differential effects before and after deliberation.

For example, consider a hypothetical example based on the agree/disagree question from Schuman and Scott: “Individuals are more to blame than social conditions for crime and lawlessness in this country.” Imagine that participants agreeing with the previous statement decreased from 45% before deliberation to 37% after deliberation. At first blush, we would conclude participants were 8% less likely to blame individuals for crime after deliberation than before. However, imagine further that 20% of the respondents acquiesced before deliberation but only 3% acquiesced after deliberation. If we correct for acquiescence bias, we now see only 25% of participants blamed individuals before deliberation and 34% after deliberation. In other words, participants were actually 9% more likely to blame individuals after deliberation. (It is important to note that opinion measured after deliberation is still a better measure than opinion

measured by traditional measures, because the final measure is both informed and free from acquiescence bias. This thought experiment only demonstrates that the effect of deliberation may be incorrectly attributed to deliberation rather than acquiescence reduction.)

However, satisficing theory does not always predict overstating or reversing results. In fact, in many cases satisficing theory predicts that the effects of acquiescence bias may *attenuate* the effects of deliberation.

Consider: what if the percentage who agree with the statement "Convicted pedophiles have a right to live in our neighborhood" rises from 35% before deliberation to 38% after deliberation. The 3% change actually understates the actual persuasion, because some respondents who changed their minds falsely reported agreeing with the statement before deliberation due to acquiescence bias. Of course, there are also likely to be cases where we indeed do overstate persuasion by lumping in acquiescence bias reduction together with persuasion effects.

How plausible are these two threats, really?

The threat to construct validity seems less worrying. The theory of Deliberative Polls states and the evidence shows that deliberation increases respondent motivation. Moreover, the theory of satisficing and the evidence show that increasing motivation reduces acquiescence bias. At a minimum, Deliberative Polling does indeed provide a better reflection of true public opinion than traditional polls.

In contrast, the threat to internal validity is more plausible. Satisficing theory predicts that unmotivated respondents acquiesce more than motivated respondents do. Based on satisficing theory, opinion changes may result from reducing acquiescence bias after deliberation rather than real persuasion. In particular, satisficing theory predicts that respondents would be

less likely to acquiescence after deliberation than before deliberation and hence become less likely to agree with statements regardless of their true beliefs.

So, I predict that acquiescence bias is not likely to reduce the construct validity of Deliberative Polls. However, I do worry that the internal validity of Deliberative Polls is threatened by measurement error.

Procedure

In early May, residents of San Mateo County were randomly sampled, surveyed by phone and offered \$200 to attend a weekend deliberation on housing issues in the county. Those who agreed to participate were provided carefully balanced briefing materials presenting the major arguments for and against given policy proposals or electoral choices and discussed these materials in small groups and posed questions to experts and leaders in plenary sessions over a weekend. In a sense, this intervention is a quasi-experiment to explore counterfactual public opinion: what would the public think, given the chance to think?

In addition to this quasi-experimental set up, both surveys before and after deliberation included a manipulation of the wording of three agree/disagree questions. The manipulation consisted of randomly assigning one of two versions of logically opposite statements (e.g., a between subjects design). These measures were repeated before and after deliberation. For example, the original wording of one question was “San Mateo County can have job growth without housing growth.” The construct of this question is whether job growth requires housing growth, and the original statement is affirmative. Hence, the logical opposite asks “San Mateo

County cannot accommodate job growth without housing growth.” The original statements and their logical opposites are show in Table 1.³

[Table 1 about here]

Results & Discussion

To prepare the data, the survey data were recoded. For the questions in the original wording, they rescaled between 0 and 1 where 0 represents “strongly disagree” and 1 represents “strongly agree.” For the questions with the logically opposite wording, the items were reverse coded so that the scale taps agreement with construct in the original wording. So, the logically opposite worded statements were recoded between 0 and 1 where 0 represents “strongly agree” and 1 represents “strongly disagree.”

[Table 2 about here]

Table 2 shows the results of the experiment. Row A shows the agreement with the item as originally worded. Before deliberation (T1), the mean of Q1 was 0.542 for participants and 0.544 for non-participants. From these results, one might conclude that a majority of both participants and non-participants *supports* the notion that “San Mateo County can have job growth with housing growth” before deliberation.

However, Row B shows the agreement with the logically opposite item, reverse coded. Before deliberation, the mean of Q1 was 0.416 for participants and 0.340 for non-participants.

³ Please note that there exists a remote but logical possibility of a “knife edge” attitude. Also, these questions are highly technical and it is possible that questions on more common political subjects like abortion would have lower levels of acquiescence bias. That said, these effects sizes are quite large and these many research projects ask respondents to evaluate complex statements like the ones in this study.

Using the logically opposite wording, one might conclude that a majority of both groups *opposes* the notion about job and housing growth. In other words, the substantive conclusion about the majority opinion on this empirical premise differs depending on the question wording.

Hence, Row A – Row B represents the level of agreement with a statement and its logical opposite, e.g. acquiescence bias. Continuing the above example, the level of acquiescence bias for participants and non-participants for Q1 is 0.126 and 0.204, respectively. Across all groups and questions, acquiescence bias ranges from 0.096 to 0.256. These levels demonstrate that a large share of measured agreement on these three items is actually due to acquiescence.

Finally, the bottom row shows the average acquiescence bias across the three items. Interestingly, participants are more acquiescent before deliberation than non-participants (0.183 versus 0.161). However, acquiescence bias among participants is reduced after deliberation (0.183 versus 0.142).

To test the significance of these differences, I ran a regression to predict measured agreement of an item based on the time of the question (e.g., before or after deliberation), participation in the Deliberative Poll, the item and the wording of the item. The regression was of the following form:

$$y_{i,j,k,l} = C + T_i + W_j + Q_k + P_l + 2 \text{ way interactions} + 3 \text{ way interactions} + 4 \text{ way interactions}$$

Where $y_{i,j,k,l}$ is the response at time i , with wording j , on question k , by participant/non-participant l , recoded from 0 to 1 where 1 indicates strong agreement with the original wording of the statement (and hence strong disagreement with the logically opposite wording) and

T_i, W_j, Q_k, P_l are the time, wording, question and participant/nonparticipant for the response $y_{i,j,k,l}$.

C is the constant.

The coefficients have straightforward interpretations: they represent levels of agreement with the base question in the original wording. Hence, the constant means that we predict that non-participants answering the original wording of Q1 before deliberation would be placed at 0.542, or between “neither agree nor disagree” and “somewhat agree.”

Given the data, certain coefficients are zero by definition. The dummies for questions are independent, so that all interactions that include $q2 * q3$ must be set to zero, (e.g., Opposite * $q2 * q3$, Time * $q2 * q3 * participant$, Time * Opposite * $q2 * participant$, Time * Opposite * $q3 * participant$, Time * Opposite * $q2 * q3$ and Opposite * $q2 * q3 * participant$.) In addition, non-participants were not surveyed after deliberation so the interactions between the dummy for participants and time must also be zero.

Table 3 shows the parameter estimates of the regression model described above.

[Table 3 about here]

The coefficients to examine are: (1) Opposite (is there evidence of acquiescence bias in general?), (2) Opposite * participants (do participants differ from non-participants?) and (3) Opposite*time (is acquiescence bias reduced after deliberation?). We also have to examine the interactions of these coefficients with the question dummies, since these interactions tell us whether there is different behavior for each of the questions. We can consider the fixed effects for questions, time and participants as controls that do not require interpretation. In addition, we can consider the interactions of questions and time as controls.

We have three questions to examine: acquiescence bias in general, differences between participants and non-participants and differences between participants before and after deliberation.

To the first question: is there evidence of acquiescence bias in general? The answer is clearly yes. Holding all else constant, respondents score between -0.126 and -0.230 points ($p < .001$ for all three questions) lower when given the logically opposite wording of the question than in the original version depending on the question.

To the second question: is there evidence that participants differ from non-participants in terms of acquiescence bias? Here the answer seems to be maybe. On question 1 before deliberation, participants score -.078 points lower ($p = .019$) when given the original wording of the question than logically opposite wording as compared to the non-participants. Note that this difference is in the opposite direction than what we would predict. On questions 2 and 3, however, participants are not statistically different from non-participants).

To the third question: is there evidence that deliberation changes acquiescence bias? The answer seems to be probably yes. The interaction Opposite*time is significant ($p = 0.042$) and holding all else equal, participants are less likely to acquiesce by 0.105 points after deliberation than before deliberation on question 1. The interaction Opposite*time*q3 is not significant, so we can conclude that that acquiescence bias reduction in the third item is similar to the first. However, the interaction Opposite*time*q2 is indeed significant and negatively signed. In fact, the reduction in acquiescence bias on question 2 is insignificantly different from zero when analyzed directly. The three way interaction of Opposite*time*question is insignificant, making this interpretation general across the three questions.

Conclusion

The results from this experiment embedded in a Deliberative Poll are quite striking. For experimental researchers, the results demonstrate that the internal validity of an experiment may be threatened by measurement error. On at least one of three agree/disagree questions, there was a substantial difference in the amount acquiescence bias before and after deliberation. Moreover, the magnitude of measurement error is very large. On each of the three questions studied in the poll, I find substantial differences in reported attitudes between the original statement and its logical opposite. Indeed, this difference is to reverse the “majority” opinion on two of three questions.

The news for Deliberative Polling is mixed. These results are mostly good news for policy makers but not quite as good for researchers. For policy makers, there are two positive results. First, the poll after deliberation yields a clearer picture of informed opinion because acquiescence bias is reduced after deliberation compared to before deliberation. Second, participants and non-participants may be more similar attitudinally than previous thought. Although non-participants and participants may not differ attitudinally, they appear to differ due to acquiescence bias. For researchers, however, comparison of how opinions change before and after deliberation may be distorted. In particular, there appears to be a bias towards less agreement after deliberation than before deliberation.

Before I give my final conclusion, I will anticipate and address two common rebuttals by defenders of agree/disagree question. First, researchers in experimental settings may concede that acquiescence may bias some measures yet claim that findings are valid so long as the bias is equivalent across treatment groups. Yet, the results of this research show that some

experimental manipulations may motivate respondents to optimize their responses and hence the assumption of equivalent bias across groups is violated. In these cases, the internal validity of the experiment may be threatened by acquiescence bias.

Second, researchers may revise their data collection to include balanced batteries of statements. (Balanced batteries have an equal number of positive and negatively valenced questions.) But this approach does not solve the problem. Consider some preliminary results gathered by Jackman (2008) on racial resentment and the Democratic primary. He finds a relationship between racial resentment and support for Hillary Clinton. However, I would argue that this result is seriously attenuated if not spurious because of acquiescence bias. Why? Clinton supporters are older, female and less educated: precisely the profile of respondents who are likely to acquiesce. Respondents who acquiesce would agree with each of the six questions and hence may have a score near zero. But a zero score places a respondent in the top quintile of racial resentment even though this score may merely be an artifact of response bias and not racial resentment.

So, what can be done? From research on questionnaires, we know the right answer: never ask agree/disagree questions and only ask construct-specific questions. For example, the first item from the Deliberative Poll (“San Mateo County can have job growth without housing growth”) might be reworded to ask “How important is housing growth to promoting job growth?” This reworded question has no easy answer for a respondent, so she will be more likely to optimize her response. If a researcher reflects carefully on what construct she wants to measure, it is always possible (if perhaps a little awkward) to convert an agree/disagree question into a construct-specific one.

Of course, this introspection requires more work on the part of the researcher. It may also take longer to administer because it forces respondents to optimize their answer. (Although it cost less time than is often assumed, because construct-specific questions save respondents from having to guess what the researcher really wants.) But the payoff is significantly higher quality data. And when the data quality may be so bad that the validity of an expensive experiment is threatened, I would argue that construct-specific questions are a necessity, not a cost.

Table 1. Question wordings*Question Group 1*

Original wording	San Mateo County can have job growth without housing growth
Logically opposite wording	San Mateo County cannot accommodate job growth without housing growth.

Question Group 2

Original wording	Public funding for housing developments will bring the price down to what most people can afford
Logically opposite wording	Public funding for housing developments will not be enough to bring the price down to what most people can afford

Question Group 3

Original wording	Public funding for housing developments can be allocated fairly
Logically opposite wording	It is impossible to allocate public funding for housing developments fairly

Table 2. Agree/disagree items recoded 0 to 1

	T1						T2		
	Non-Participants			Participants			Participants		
	q1	q2	q3	q1	q2	q3	q1	q2	q3
A. Original wording	0.54 2	0.48 7	0.51 8	0.54 4	0.52 6	0.55 6	0.33 3	0.43 8	0.57 1
B. Logically opposite wording, reverse coded	0.41 6	0.24 7	0.39 9	0.34 0	0.27 0	0.46 7	0.23 4	0.20 8	0.47 5
Acquiescence Bias (A-B)	0.12 6	0.24 0	0.11 9	0.20 4	0.25 6	0.08 9	0.09 9	0.23 0	0.09 6
Average Acquiescence Bias across all items		0.16 1			0.18 3			0.14 2	

Table 3. Parameter Estimates

Dependent Variable: Agreement with original statements

Parameter	B	Std. Error	t	Sig.	Partial Eta Squared
Intercept	.542	.015	36.793	.000	.193
Time	-.210	.036	-5.927	.000	.006
Opposite	-.126	.021	-5.918	.000	.006
q2	-.055	.021	-2.635	.008	.001
q3	-.024	.021	-1.136	.256	.000
participant	.002	.023	.086	.931	.000
Time * q2	.122	.050	2.420	.016	.001
Time * q3	.226	.050	4.476	.000	.004
Time * Opposite	.105	.052	2.036	.042	.001
q2 * participant	.038	.033	1.129	.259	.000
q3 * participant	.036	.033	1.081	.280	.000
Opposite * participant	-.078	.034	-2.338	.019	.001
Opposite * q2	-.114	.030	-3.770	.000	.003
Opposite * q3	.007	.030	.217	.828	.000
Time * Opposite * q2	-.078	.073	-1.072	.284	.000
Time * Opposite * q3	-.112	.073	-1.532	.126	.000
Opposite * q2 * participant	.062	.048	1.281	.200	.000
Opposite * q3 * participant	.109	.048	2.253	.024	.001

*Note: The following interactions are set to zero because they are redundant: q2 * q3, Time * q2 * participant, Time * q3 * participant, Time * Opposite * participant, Time * q2 * q3, q2 * q3 * participant, Opposite * q2 * q3, Time * q2 * q3 * participant, Time * Opposite * q2 * participant, Time * Opposite * q3 * participant, Time * Opposite * q2 * q3, Opposite * q2 * q3 * participant.*

References

- Ackerman, B. and Fishkin, JS. Deliberation day. In *Debating Deliberative Democracy..* Fishkin JF, Laslet P (eds). Malden, MA: Blackwell Publishing Ltd, 2003
- Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. Journal of Abnormal and Social Psychology, 60, 151-174.
- Cronbach, L. J. (1941). An experimental comparison of the multiple true-false and multiple multiple-choice tests. Journal of Educational Psychology, 32, 533-543.
- Cronbach, L. J. (1942). Studies of acquiescence as a factor in the true-false test. Journal of Educational Psychology, 33, 401-415.
- Cronbach, L. J. (1946). Response sets and test validity. Educational and Psychological Measurement, 6, 475-494.
- Fishkin JF, Laslet P (eds). *Debating Deliberative Democracy*. Malden, MA: Blackwell Publishing Ltd, 2003.
- Fishkin JS. (1991) *Democracy and Deliberation*. New Haven, CT: Yale University Press.,
- Hamilton, D. L. (1968). Personality attributes associated with extreme response style. Psychological Bulletin, 69, 192-203.
- Jackman, M. R. (1973). Education and prejudice or education and response-set? American Sociological Review, 38, 327-339.
- Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. Public Opinion Quarterly, 51, 201-219.
- Krosnick, J. A., and M. A. Milburn. 1990. "Psychological Determinants of Political Opinionation." Social Cognition 8:49-72.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied Cognitive Psychology, 5, 213-236.
- Krosnick, J.A. (1999). Survey research. Annual Review of Psychology, 50, 537-567.
- Leary, M.R. & Kowalski, R.M. (1990). Impression management: A literature review and two-component model. Psychological Bulletin, 107, 34-47.
- Lenski, G. E., & Leggett, J. C. (1960). Caste, class, and deference in the research interview. American Journal of Sociology, 65, 463-467.

- Luskin, R. (2002). From denial to extenuation (and finally beyond): Political sophistication and citizen performance. In *Thinking about Political Psychology*, ed. Jame Kulinski, New York: Cambridge University Press
- Luskin, R. and Fishkin, J.(n.d) Deliberation and “better citizens.”
- Luskin R. and Fishkin, J (2005). *Deliberative Polling, Public Opinion, and Democracy: The Case of the National Issues Convention.*
- Luskin R, Fishkin JS, Jowell R. (2002). Considered opinions: Deliberative Polling in the U.K. *British Journal of Political Science.* 32:671-686.
- McClendon, M. J. (1989). Acquiescence: A test of the cognitive limitations and question ambiguity hypotheses. Paper presented at the 1989 meetings of the American Association for Public Opinion Research, St. Petersburg, Florida.
- Moum, T. (1988). Yea-saying and mood-of-the-day effects in self-reported quality of life. *Social Indicators Research*, 20, 117-139.
- Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, 63, 129-156.
- Schlenker, B.R. (1980). *Impression management: The self-concept, social identity, and interpersonal relations.* Monterey, CA: Brooks/Cole
- Schuman, H., & Presser, S. (1981). Questions and answers in attitude surveys. New York: Academic Press. Chapter 8, "The acquiescence quagmire."
- Schuman, H., & Scott, J. (1989). Response effects over time: Two experiments. *Sociological Methods and Research*, 17, 398-408.
- Toner, B. (1987). The impact of agreement bias on the ranking of questionnaire response. *Journal of Social Psychology*, 127, 221-222.
- Young IM. (2003) *Activist Challenges to Deliberation Democracy.* In: Fishkin JF, Laslet P (eds). *Debating Deliberative Democracy.* Malden, MA: Blackwell Publishing Ltd.