Measurement Directiveness as a Cause of Response Bias: Evidence From Two Survey Experiments

Philip S. Brenner¹ and John DeLamater²

Abstract

Extant research comparing survey self-reports of normative behavior to direct observations and time diary data have yielded evidence of extensive measurement bias. However, most of this research program has relied on observational data, comparing independent samples from the same target population, rather than comparing survey self-reports to a criterion measure for individual respondents. This research addresses the next step using data from two studies. In each study, respondents completed a conventional survey questionnaire, including questions about frequency of religious behavior. Respondents were then asked to participate in a text messaging (short message service) data collection procedure, reporting either (1) participation in religious behavior specifically or (2) all changes in major activity without explicitly specifying religious behavior. Findings suggest that directive measurement, priming the respondent to consider the focal behavior, is a cause of measurement bias.

¹ University of Massachusetts Boston, Boston, MA, USA
² University of Wisconsin, Madison, WI, USA

Corresponding Author:
Philip S. Brenner, University of Massachusetts Boston, 100 William T. Morrissey Blvd, Boston, MA 02125, USA.
Email: philip.brenner@umb.edu
Keywords
survey research, response effects, measurement, social desirability, identity

Introduction
Social desirability bias—the phenomenon in which survey respondents report less of the bad stuff and more of the good than their actual behavior warrants (Tourangeau, Rips, and Rasinski 2000)—is one of the most investigated of survey artifacts. Survey respondents report less frequent drug and alcohol use (Aquilino 1994; Aquilino and LoSciuto 1990) and fewer illegal activities, arrests (Wyner 1980), and other embarrassing behaviors than their actual behavior suggests (Tourangeau and Smith 1996; Tourangeau and Yan 2007). They also report more frequent church attendance (Hadaway, Marler, and Chaves 1993, 1998; Presser and Stinson 1998) and exercise (Brenner and DeLamater 2014; Chase and Godbey 1984) than their actual behavior suggests, and that they voted in the last election at a higher rate than is reflected in actual turnout statistics (Andersson and Granberg 1997; Belli, Traugott, and Beckmann 2001; Bernstein, Chadha, and Montjoy 2001).

But measurement biases, like social desirability bias, do more than skew our estimates of the behavior of the population. Rather, they can create an illusion (or perhaps delusion) of who we are as a society. Of particular importance in the American case is the overreporting of religious behavior. Surveys tell us that 40 percent of Americans attend religious services frequently and regularly. This well-known and well-publicized fact (e.g., Newport 2006, 2010) contributes to the paradigm of American religious exceptionalism. Americans do not accidentally misreport religious behavior, like attending services. There is little random error to be found here. Rather, these are biases—systematic, unidirectional errors—that, if taken seriously, can yield as much insight as a valid survey report (Schuman 1982). However, without the ability to readily validate these claims, the biased survey estimate becomes a “truth” that informs and misinforms our understanding of a particular society. Thus, “everyone knows” the United States is the most religious of the advanced industrialized nations, and surveys confirm this conventional wisdom. Moreover, unlike election turnout figures, the true value of church attendance rates is rarely found in the pages of the New York Times, enabling the persistence of the myth of exceptional American religious behavior (Brenner 2011a).

Conventional wisdom aside, the United States has been the locus and focus of most of the research on the overreporting of church attendance.
Comparing survey estimates to counts from other sources (head counts from church visits, attendance figures from church records, and counts of automobiles in church parking lots), Hadaway et al. (1993) estimated an overreport of about 50 percent in a Midwestern county. Other research comparing survey estimates to head counts has replicated this finding in specific denominations and churches, including Presbyterians (Marcum 1999), Catholics (Chaves and Cavendish 1994), and an evangelical Protestant congregation (Marler and Hadaway 1999). Similar findings have emerged in Oxford County, Ontario, Canada, comparing survey estimates and a count of congregants (Hadaway and Marler 1997a) and the United Kingdom, comparing survey estimates and church records (Hadaway and Marler 1997b).

Criticisms of head counts, attendance records, and other criterion measures (Caplow 1998; Hout and Greeley 1998; Woodberry 1998) motivated research comparing survey estimates to those from other data collection methods. This research program compares conventional survey estimates to those of arguably better quality from time diaries. Direct survey questions, like those on most conventional surveys, prompt self-reflection on the part of the respondent, yielding an answer that subjective valuations of the focal identity may influence. In short, respondents’ sense of the importance of his or her religious identity may bias answers to questions about their religious behavior (Hadaway et al. 1998; Marler and Hadaway 1999). Conversely, the nondirective measurement provided by time diaries and other chronological data collection procedures more accurately measure normative behavior, like religious behavior, by avoiding the biases inherent to directive survey measurement (Bolger, Davis, and Rafaeli 2003; Niemi 1993; Stinson 1999; Zuzanek and Smale 1999). The respondent is never asked a direct question about religious behavior; consequently, he or she is not presented an opportunity to reflect on the importance of his or her religious identity as in a conventional survey question answering process.1

These studies compare estimates from time diaries to those from conventional surveys in the United States (Brenner 2011a; Presser and Stinson 1998) and Canada (Brenner 2011a, 2012a). They confirm the conclusion of the extant research that a substantively significant level of bias exists in conventional survey estimates of church attendance. This research program has been extended to Europe (Brenner 2011a) and West and South Asia (Brenner 2014) to test whether overreporting of religious behavior is idiosyncratic of North America or whether it represents a more widespread, or perhaps universal, phenomenon.2 Some evidence emerges for significant overreporting in Italy (Rossi and Scappini 2012) and the United Kingdom (Hadaway and Marler 1997b), but not necessarily to the extent and consistency of the American overreport.
Related research testing alternative causes has suggested that demographic correlates of religiosity (including social identities like gender, as well as marital and family status, age, education, and income) do not consistently predict the overreporting of religious behavior where and when it occurs (Brenner 2012b). However, the importance of the respondent’s religious identity emerges as a strong predictor of overreporting, suggesting it as a potential cause of bias in the measurement of the performance of the focal identity. Brenner demonstrated this effect of high religious identity importance in biasing self-reported church attendance in the United States (2011b), Canada (2012a), and in three predominantly Muslim countries (2014).

**Religious Identity as a Cause of Bias**

While much of our understanding of Americans’ religious behavior comes from sample surveys, the survey interview is not immune from the biasing effect of identity importance. According to Burke (1980:28), “the problem with most measurement situations is that without the normal situational constraints it becomes very easy for a respondent to give us that idealized identity picture which may only seldom be realized in normal interactional situations.” From this perspective, the directive survey question prompts the respondent to reflect on his or her self-concept—particularly on strongly valued identities—and answer questions accordingly. Hadaway et al. (1998:127, emphasis in original) argue that “over-reporting is generated by the combination of a respondent’s desire to report truthfully his or her identity as a religious, church-going person and the perception that the attendance question is really about this identity rather than about actual attendance.” The individual’s sense of religious identity noted by Hadaway et al. influences and is influenced by one’s social context (i.e., socialization). In essence, the answer to the survey question reflects the normative identities of the society in which the survey respondent lives. Thus, both the importance of religion in American society and individual Americans’ religious identities encourage bias in measures of religious behavior (Brenner 2011b; Hadaway et al. 1993, 1998).

Identity Theory (Stryker 1968, [1980] 2003) posits a set of interrelated concepts that allow us to understand this phenomenon as a function of religious identity vis-à-vis its relationship with other identities. The first concept, salience, is defined as the probability of an identity’s enactment, or the propensity to define a situation as one in which the focal identity is relevant. As this suggests, salience is a function of an individual’s corpus of identities, and the situations and interactions in which he or she finds himself or herself. Highly salient identities are those that are easily and often called

---

Sociological Methods & Research
forward and enacted. High identity salience encourages the individual to interpret situations as relevant for the performance of the focal identity (Stryker and Serpe 1982). As this suggests, some identities are likely to be enacted in situations in which the context of the interaction is irrelevant to the focal identity.

While the definition of salience implies a conscious choice, this implication is only an artifact of language. The individual is likely unaware of a given identity’s salience, in contrast with the second main concept—identity importance. Defined as the subjective value an individual places on an identity or its centrality in the individual’s self-concept, identity importance stands in contrast to salience as the individual is, by definition, aware of the value he or she places on an identity. Identity Theory posits a strong causal relationship between importance and salience; the higher the importance the individual places on an identity, the higher its salience (Brenner, Serpe, and Stryker 2014). Therefore, identities that are highly important to the individual are more likely to be called up and enacted.

Religious identity fits into the framework of Identity Theory well as it is performed in interaction with others. For example, religious identity is performed in attendance at religious services, in interaction with clergy and other coreligionists, and outside religious services when engaged in religious interaction (e.g., during Bible study groups, missionary or evangelical activity, or engaging in conversations about theology). These interactions do not necessitate a physically present other. The religious role identity can also be performed in interaction with the divine or an intermediary through prayer or other devotional activities, or with an earthly imagined or generalized other (e.g., practicing one’s “testimony;” Sharp 2010). As this suggests, individuals enact their religious identities in multifarious situations, given their centrality in many individuals’ self-concept. An individual with high religious identity importance may be motivated to call up his or her religious identity in a situation in which other individuals with relatively lower religious identity importance would not be religiously primed. As such, individuals can perform their religious identities during a religious service, on a prayer rug, or while reading scripture, as well as in situations like a classroom, a political meeting, the comments section of NYTimes.com, or even a survey interview.

Present Research

The research on overreporting has persuasively and robustly demonstrated the overreporting phenomenon. However, its ability to explain the cause of overreporting has been limited in two main ways. The first limitation is
founded in its reliance on observational data. With few exceptions (see Brenner 2012a), comparisons are made at the margins, computing differences between survey and diary estimates for different samples (see Brenner 2011a). Alternatively, a multiple imputation procedure akin to propensity matching is used to synthesize complete cases for respondent-level analysis (Brenner 2011b). Valuable as these analyses have been forwarding the research program, researchers should strive to collect both survey and diary data for the same respondents to be analyzed not only on the margins but also at the respondent level. Thus, this research collects both types of data for each respondent, allowing comparison of self-reported religious behavior from both directive (survey) and nondirective (diary) methods. Given the findings of prior research, bias is expected to emerge in the survey but not the diary reports of religious behavior.

**Hypothesis 1:** High levels of overreporting of religious behavior will emerge in answers to directive conventional survey questions when compared with a measure of the same religious behavior using a nondirective chronological measurement procedure.

The second limitation of this research program, related to the first, is in its reliance on correlational models. In each analysis, the dependent variable—overreporting of religious service attendance, defined as positive bias in survey reports of the religious behavior—correlates strongly with a number of different measures of religious identity importance. However, as both the independent and the dependent variables are measured observationally, the potential for an unobserved confounding factor cannot be ruled out. As this suggests, shifting toward an experimental approach would be a useful addition to the extant research program. This study does just this, extending the research program to include an experimental manipulation of the directiveness of the diary measure. In this experimental condition, a random subsample of respondents is explicitly asked to report religious behavior in the diary component of the data collection process. This manipulation is hypothesized to change respondent behavior, prompted by the directness of the prospective measurement. In short, telling respondents about the focal behavior before its measurement will alter their behavior in a socially desirable direction.

**Hypothesis 2:** Higher rates of overreporting of religious behavior will emerge when comparing an answer to a conventional survey question with a nondirective chronological measure compared to that from a directive chronological measurement procedure.
Respondents will also be asked to rate the importance of their religious identity. Following the findings of prior research, overreporting is hypothesized to be strongly associated with highly rated religious identity importance.

**Hypothesis 3:** Overreporting respondents will rate the importance of their religious identity higher than other nonattending respondents, but will not differ from validated attenders.

These three hypotheses are tested in two survey quasi-experiments. Respondents in both studies completed a brief web survey of daily life, including questions about frequency of religious behavior. In study 1, respondents were then asked to participate in a nondirective text messaging (short message service [SMS]) data collection procedure (described in the next section) similar to the diary method previously discussed, reporting all changes in major activity without explicitly specifying religious behavior. In study 2, respondents were randomly assigned to one of the two conditions: (1) nondirective, as in study 1 or (2) directive, in which they were asked to report participation in religious behavior specifically, in addition to all other behavior.

This design investigates the cause of misreporting by analyzing the difference between conventional survey reports and the text message reports both within and between studies. This comparison allows a focus on the nature of respondent priming; that is, does explicitly mentioning to the respondent the socially desirable focus of the study—even when using a chronological measurement procedure—alter respondent behavior in a socially desirable direction? This focus permits an investigation of the factors that encourage survey respondents to exaggerate their reports of the frequency of religious behavior.

**Data and Methods**

**Study 1**

A probability sample (random sample stratified by gender and year in school from a list of currently enrolled students) of 325 students from a large, public university in the Midwest was e-mailed an invitation to participate in a survey of daily life. Students were offered a 10-dollar incentive for completion of the conventional survey.

Of these respondents, 124 (38 percent response rate [RR]) completed a brief, 20-item web survey, including questions about use of university
facilities (e.g., libraries, student union, and recreation facilities), and other
daily activities on- and off campus. In this context, a standard question about
religious service attendance asked: “How often do you attend religious ser-
vice(s)? (1) Never, (2) about once or twice a year, (3) several times a year, (4) about once a month, (5) two to three times a month, (6) once a week, and (7) several times a week.” For comparison
with the chronological measure, this measure is dichotomized as regular and
frequent attendance: (1) two to three times a month or more and (0) about
once a month or less.

The final question asked respondents to enter their cell phone number if
they would be willing to participate in the second component of the
study—the chronological reporting procedure using SMS text messaging—and were offered an additional 30 dollars for its completion. Eighty-seven
respondents agreed and completed the texting component of the study (70
percent compliance rate; 27 percent final RR). Analysis of information (year
in school, gender) available on the sampling frame (record data from the bur-
sar) suggests that nonrespondents and respondents do not meaningfully differ
on these important demographic covariates.

Participants were assigned to one of the five 5-day field periods over two
weeks, each including a full weekend (including a Saturday and a Sunday)
during Lent, covering Passion Sunday and Palm Sunday in April 2011. Par-
ticipants were to report all major activities during the field period, as they
happened using SMS text messaging. A short training document (a two-
page pdf) was provided to participants, including examples of what to report
and an FAQ list (including instructions on how to report activities late, after
they occurred; see Brenner and DeLamater 2013, for details). No activities of
particular interest were highlighted, although a number of examples were
given (e.g., working out at campus recreation facilities, running errands, and
completing chores; studying, going to the library, and attending class; partici-
pat ing in or attending social or entertainment events). Participants were
sent brief reminder texts to encourage their participation; four on the first day
of their participation (10 a.m., 1 p.m., 5 p.m., and 8 p.m.), reduced to two in
the final days of the field period (10 a.m. and 8 p.m.).

While SMS has been used previously to collect in situ data (Alfvén 2010;
Anhøj and Møldrup 2004; Raento, Oulasvirta, and Eagle 2009), most of these
uses have used standard survey questions (Schober et al. 2013) or resembled
the experience sampling method (ESM). ESM measures at random or prese-
lected times of the day (Larson and Csikszentmihalyi 1983), rather than con-
tinuously as behavior occurs. Collecting diary-type data utilizes the strengths
of SMS by taking advantage of the idiomatic nature of SMS to avoid asking
direct, and directive, questions (Brenner and DeLamater 2013). Moreover, using SMS increases the appeal of the survey request by offering respondents in this hard-to-survey population a way to report using a technology relevant to the daily lives. Conventional survey modes commonly result in high rates of nonresponse among young adults (Groves and Couper 1998). Thus, adopting and adapting this frequently used technology increases interest to leverage participation (Groves, Singer, and Corning 2000).

The texting procedure resulted in nearly 2,000 observations. Each observation was coded for the activity reported. A dichotomous religious service attendance variable was created for each respondent, flagging respondents with any report of this activity during the entire five-day reference period.

Notably, these text message data are of high quality. An investigation of another normative behavior (physical exercise) compared these text reports to validation data from a reverse record check (admittance records to campus recreation facilities generated by the scanning of the students’ identification cards upon entrance). This investigation found the text message reports of exercise to be unbiased and with relatively little random error (Brenner and DeLamater 2013, 2014).

Study 2

A subject pool of 224 undergraduates enrolled in introductory sociology courses at a public university in the Northeast were asked to participate in a study of their daily activities. Students were offered extra credit as an incentive. Of these students, 75 completed a web survey nearly identical to that in study 1.

Following the web survey, respondents were randomly assigned to one of the two conditions. In the first condition, respondents were asked to text updates on their major activities without reference to the focal behavior of interest—religious service attendance. In the second condition, respondents were given identical instructions to respondents in the first condition, with one important change: They were also asked specifically to include reports of religious behavior (e.g., church, temple, synagogue, or mosque attendance) along with reports of other major daily activities. Reminders were sent to respondents at the same timing and frequency as in study 1. As in study 1, the texting portion of the study lasted for five days, including a weekend. Of the web survey respondents, 58 successfully completed the text message reports.

Both of the studies also included a conventional survey question measuring the importance of the respondent’s religious identity in the context of
other potentially important role identities: “Each of us fills a number of roles and participates in many different activities in our daily lives. How important to you is: participating in religious activities?” Other role identities measured included student, family, job, volunteering, student group membership, exercise and playing sports, and friend. In study 1, each identity was measured on a fully labeled five-point scale: (1) not at all important, (2) not very important, (3) somewhat important, (4) very important, and (5) extremely important. In study 2, this scale was increased to an 11-point scale (0–10) with the same end points. This change was made to allow respondents more choices at the well-used upper-end of the scale.

**Analysis Plan**

Three primary analyses test the three hypotheses listed previously. First, the conventional survey measures of religious behavior are compared to those from the text messaging procedures to estimate rates of overreporting in each study. Rates of attendance from the paired survey and chronological measures are compared using (1) McNemar’s $\chi^2$ to test for differences in reported attendance between the two data collection methods and (2) Cohen’s $d$ for proportions to assess the substantive size of the difference.

Second, overreporting is computed at the respondent level, as the positive difference between the conventional and chronological measures of attendance for each respondent $i$:

If $y_{s,i} - y_{t,i}$:

- $> 0$, the respondent is an overreporter.
- $< 0$, the respondent is an underreporter.
- $= 0$, and $y_i > 0$, the respondent is a validated attender.
- $= 0$, and $y_i = 0$, the respondent is an admitted nonattender.

where $s$ is the dichotomized survey measure of attendance and $t$ is the dichotomous text measure of attendance. Rates of overreporting are then compared between the two conditions in study 2.

Finally, rated levels of religious identity importance will be compared (where possible, given the outcome of the prior analyses) between overreporting respondents and those who accurately report their religious behavior—both validated attenders and admitted nonattenders—using the Mann–Whitney–Wilcoxon test.
Results

Study 1

The first study compares answers to a conventional survey item on religious service attendance to a measure of the same behavior from a non-directive chronological data collection procedure using text messaging. Over a quarter (28 percent) of study 1 respondents reported regular, frequent attendance, relatively evenly distributed between attending “two to three times a month” (15 percent) and “once a week or more” (13 percent; see Table 1). The remaining three-quarters of respondents (72 percent) reported on the survey that they attend about once a month or less, including over a quarter of respondents (28 percent) who reported that they never attend. This rate of attendance is comparable to that from the 2012 General Social Survey (GSS), a nationally representative omnibus survey, where a quarter of young adults, aged 18–25, reported regular, frequent attendance.

The SMS text message data collection procedure generated a much lower rate of religious service attendance. Approximately 10 percent of respondents texted an occasion of religious service attendance, all but one of whom reported frequent, regular attendance (two to three times a month or more) on the conventional survey questionnaire. The difference between these two rates ($\Delta = 17.2$ percentage points) is substantively ($d = .45$) and statistically significant, $\chi^2(1) = 13.2$, $p \leq .001$.

As this suggests, relatively few of the respondents reporting frequent and regular attendance on the survey actually attended services during the diary week. Only nine respondents attended—about a quarter (3 of the 13) of those who reported attending two to three times a month on the survey and nearly half (5 of the 11) who reported attending every week or more often on the survey (and the single respondent who reported never attending on the survey). This finding demonstrates a high rate of overreporting of religious behavior: two-thirds of respondents who claimed frequent and regular attendance on the survey did not attend during the five-day texting period which included a Saturday and Sunday during Lent.

As hypothesized, these overreporting respondents differ dramatically from other nonattenders. They rate their religious identity as much more important ($\eta = 4$, $\mu = 4.2$, $\sigma = .21$) than respondents who accurately report their infrequent or nonattendance ($\eta = 2$, $\mu = 1.9$, $\sigma = .14$; see Figure 1). This difference in the rated level of importance of the religious identity ($\Delta\eta = 2$, $\Delta\mu = 2.3$) is highly significant ($z = -5.5$, $p \leq .001$).
### Table 1. Comparison of Survey and Text Reports of Religious Service Attendance, Study 1.

<table>
<thead>
<tr>
<th>Survey report</th>
<th>Attended</th>
<th>Did not attend</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Percent</td>
<td>N</td>
</tr>
<tr>
<td>Never</td>
<td>1</td>
<td>1.1</td>
<td>23</td>
</tr>
<tr>
<td>About once or twice a year</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Several times a year</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>About once a month</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Two to three times a month</td>
<td>3</td>
<td>3.4</td>
<td>10</td>
</tr>
<tr>
<td>Once a week or more</td>
<td>5</td>
<td>5.7</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>10.3</td>
<td>78</td>
</tr>
</tbody>
</table>

**Figure 1.** Mean importance of religious identity by study and condition.
Also as hypothesized, no difference emerged in religious identity importance between the two types of self-reported attenders. Validated attenders ($\eta = 3.5, \mu = 3.5$) do not differ from overreporters in their ratings of religious identity importance ($\Delta \eta = 0, \Delta \mu = 0.7, z = 1.3, p = .200$).

**Study 2**

Study 2 explicitly tests the effect of making the chronological measurement procedure directive: If respondents are told directly what the focus of the research is, will it change the way they report and/or behave during the chronological measurement procedure?

The overall trend in self-reported and actual attendance in study 2 resembles that in study 1. Fewer than a quarter (22.3 percent) of the respondents reported regular, frequent attendance (two or three times a month or more often). The majority of respondents reported very infrequent attendance, with over half (57 percent) reporting attending about once or twice a year or less (see Table 2). The remaining 19 percent of respondents reported attending “several times a year.” While this is somewhat lower than the attendance rates in study 1, it is still comparable to GSS estimates for young adults of college-age.

Respondents were randomly assigned to two conditions: Condition 1 ($N = 27$) compared these conventional survey reports of religious service attendance to a nondirective texting procedure identical to that used in study 1. Condition 2 ($N = 31$) compared conventional survey reports to a directive texting procedure—respondents in this condition were given instructions that specified the focal behavior of interest but were otherwise identical to the instructions from condition 1.

This small change made a substantial difference in the reports of respondents. Five respondents texted reports of attendance, yielding a 9 percent attendance rate. Of these respondents who texted reports of attending religious services, four were in (directive) condition 2, a statistically significant difference, $\chi^2(1) = 6.4, p \leq .05$. Thus, texting respondents whose instructions told them specifically of the focal religious behavior were more likely to attend religious services than were respondents who were not told the focus of the study.

Comparing rates of overreporting between conditions further highlights this difference. Of the 13 respondents reporting regular, frequent attendance, only 5 attended, yielding a very high rate of overreporting (160 percent). However, overreporting respondents were not randomly distributed; more
Table 2. Comparison of Survey and Text Reports of Religious Service Attendance, Study 2.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Attended</th>
<th>Did not attend</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Percent</td>
<td>N</td>
</tr>
<tr>
<td>Survey report</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>About once or twice a year</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Several times a year</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>About once a month</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Two to three times a month</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Once a week or more</td>
<td>1</td>
<td>3.7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3.7</td>
<td>26</td>
</tr>
</tbody>
</table>
condition 1 respondents overreported than did condition 2 respondents. Only one (of seven) self-reported frequent attenders in condition 1 actually attended, compared to three (of six) self-reported frequent attenders in condition 2. This difference in overreporting is statistically significant, \( \chi^2(1) = 14.2, p \leq .001. \)

Study 2 respondents rated the importance of their religious identities somewhat lower than study 1 respondents (\( \eta = 3.5, \mu = 4.2, \sigma = 3.7 \)), with exactly half of the respondents rating their religious identity three or below on the scale from 0 to 10. This lower level of rated importance is likely a function of the reduced level of religiosity in the Northeastern United States relative to that in the Midwest. As hypothesized, overreporting respondents rate their religious identity as much more important (\( \eta = 9, \mu = 7.8, \sigma = 2.6 \)) than other nonattenders who report accurately (\( \eta = 2, \mu = 3.0, \sigma = 3.2 \)). This difference in the rated level of importance of the religious identity (\( \Delta \eta = 7, \Delta \mu = 4.8 \)) is highly significant (\( z = -3.4, p \leq .001 \)). Also as hypothesized, no difference emerged in religious identity importance between the two types of self-reported attenders. Validated attenders (\( \eta = 9.5, \mu = 8.5, \sigma = 2.4 \)) do not differ from overreporters in their ratings of religious identity importance (\( \Delta \eta = 0.5, \Delta \mu = 0.7, z = -0.33, p = .745 \)).

**Discussion**

As hypothesized, study 1 generated overreporting of church attendance on the conventional survey instrument. Two-thirds of the respondents reporting frequent and regular church attendance on the survey failed to attend during the five-day reference period, which included either Passion or Palm Sunday. Moreover, only one respondent reporting infrequent or nonattendance on the conventional survey (about once a month or less often) attended, as would be expected given that the reference period includes only one weekend. Were the surveys conducted a week or two later over the Easter weekend, or over multiple weekends, a different finding could potentially have resulted.

These results were replicated in study 2. Respondents in condition 1, the nondirective texting condition, looked similar to those from study 1, as overreporting was pronounced for self-reported frequent attenders in this condition. In condition 2, the directive texting condition, respondents changed their behavior to match their religious self-concepts as the instructions specifically mentioning the focal behavior invited reactivity. Moreover, one condition 2 respondent who reported attending “several times a year” texted attendance, adding credence to the notion that respondents changed their behavior to match the importance of their identities.
The divergence between respondents in the two conditions in study 2 is arguably attributable to the priming effect of directive measurement. Conventional surveys ask direct questions, initiating a process that can introduce measurement biases, as the respondent not only recalls instances of the focal behavior but also reflects on subjective judgments about the identity with which these behaviors are associated. For the example of religious behavior, like church attendance, the associated religious identity can be of high importance. The respondent who strongly values his or her religiosity may perceive questions about religious attendance as asking about this identity rather than actual behavior (Hadaway et al. 1998). As such, he or she may report more frequent attendance than is warranted by actual behavior, thereby introducing bias into the measurement process (Brenner 2011b).

The ability of the time diaries and similar procedures to avoid bias is arguably based on their nondirective measurement procedures. As they are typically used, time diaries utilize a chronological measurement procedure in which respondents are asked about occurrences and activities without mention of a focal behavior. Without a prime to focus on a particular normative identity, respondents are detoured around identity-based bias. In study 1 and condition 1 of study 2, respondents are not told about the focal behavior and, consequently, text reports of their religious behavior without bias. In comparison, the survey reports yield an estimate twice as high as the text reports, which is assumed to be an unbiased estimate of the actual rate of behavior. In condition 2 of study 2, the strength of the chronological measurement procedure is negated, its nondirectiveness forsaken by its alteration into a directive measurement procedure thereby causing reactivity that increased the attendance rate by nearly 350 percent.

Other potential causes of this difference are worth exploring. First, it is possible that condition 2 respondents have a higher rate of texted attendance not because they changed their behavior, but rather simply because they were prompted to remember to report by the instructions. While possible, we find explanation implausible. As previously discussed, the evidence for overreporting of church attendance on surveys is strong and consistent. This research, as well as related work on self-reports of other normative behaviors like voting and exercise, has suggested that lower estimates are more valid. Moreover, in prior research, we validated reports of a different normative behavior, exercise at campus recreation facilities, with record data from admission records to those facilities (Brenner and DeLamater 2014). We find that the nondirective texting procedure used in study 1 and condition 1 of study 2 yields unbiased estimates. There is no reason to believe that the procedure is working differently here.
Second, it is possible, although highly unlikely, that respondents faked their reports, texting religious behavior when none occurred. Again, evidence against falsified texts can be found in previous work validating texted reports of normative behavior. Little error was found in texted reports and, more importantly, that error was demonstrated no bias as it was equally distributed between over- and underreporting (Brenner and DeLamater 2014). There is no reason to believe the results here should differ dramatically.

Such claims of blatant falsification or differential memory are less likely than a second potential explanation: The priming effect of directive measurement pulled respondents’ behavior into alignment with the importance of their religious identities, yielding behavior that matched, or even exceeded, their survey reports. The higher rate of attendance in condition 2 of study 2 (compared to condition 1) demonstrates this effect. Explicitly telling respondents about the purpose of the diary procedure and the focal behavior modified respondent activity in a normative direction, much like the bright lights of the Hawthorne Works. A common cause, arguably based on religious identity importance, is generating bias in the survey report and causing identity-confirmatory behavior in the directive texting measure.

Limitations

Like more conventional chronologically based data collection procedures, text-based diaries have limitations. First, respondents may fail to report activities of very brief duration that happen frequently during the day. For example, trips down the hall to use the restroom or to the water fountain are likely to be omitted, as respondents tend to focus on longer activities (e.g., those that last for hours rather than minutes) and the sorts of activities around which the day is planned. Therefore, the focal activities of such a data collection procedure should be these sorts of major activities. Church attendance fits this qualification well.

Second, chronologically based data collection procedures place a heavy burden on respondents that can result in high rates of item and unit nonresponse, as respondents forget to report activities, choose to participate intermittently, or drop out early. Unfortunately, the SMS procedure may not relieve respondent burden relative to more conventional diary procedures; rather, it may lead to increased time spent on the data collection task, although this time may be more equally distributed throughout the diary day.

However, the SMS procedure does offer some promise, as it incorporates features that address these weaknesses and may lead to higher quality data. First, respondents are asked to report on behavior as it occurs, reducing recall
error. In order to reduce the burden of the data collection process, diaries can be, either by the researcher’s design or by the unilateral decision of the respondent, filled out at the end of the day or at the end of the reference period. However, shifting the timing of diary completion away from the time of occurrence of individual activities can result in poorer data quality, as respondents may introduce errors into the data collection procedure, like forgetting to include events or attributing them to incorrect times. Thus, creating an in situ measurement procedure like the SMS procedure used here can help to avoid retrospective reporting and its concomitant errors.

Second, the procedure avoids the editing and judging that can bias retrospective reports (see Tourangeau et al. 2000). Without the time to reflect and assess behavior, the SMS procedure avoids social desirability effect, reducing bias. While perhaps not true for all behaviors and activities, especially contranormative, illegal, or embarrassing activities (e.g., illicit drug use or sexual activity), or those of high frequency and brief duration (e.g., using the restroom or getting a drink of water), this procedure allows more accurate measurement of normative activity (Brenner and DeLamater 2014). While there is good evidence suggesting that the texting procedure produces unbiased, high-quality reports of socially desirable behaviors (Brenner and DeLamater 2013, 2014), we cannot guarantee that the text reports used here are unbiased as they cannot be validated.

Relatedly, small sample sizes and the nature of the population (university students) also limit the strength of inferences. While these samples are adequate to the task of testing theory, the results are statistically significant, and the findings suggestive, future research should investigate the veracity of the measurement of church attendance and other normative behaviors in large, nationally representative, probability sample surveys.

Conclusion

This research implicates identity as a cause of bias in the measurement of normative identity-related behavior. High religious identity importance encourages a pragmatic interpretation of the directive, conventional survey question about religious behavior to be about the respondent’s religious identity rather than about his or her actual behavior. These biased survey estimates are then reported as a social fact, creating, or at least contributing to, the illusion of a more religiously behaving population than is warranted. American religious exceptionalism, if defensible at all, is rooted not in actual behavior but in self-concept.
Yet, this research program is not only about improving survey measurement. The vexing problems of the measurement of religiousity are, counterintuitively, underappreciated. Survey artifacts offer sociologists of religion, survey methodologists, and other social scientists a fantastic opportunity. As Howard Schuman (1982) suggested, these types of measurement errors offer us an opportunity to understand culturally situated human behavior, once we accept that these “artifacts” are real and rooted in social processes. Recent work has approached the overreporting of religious behavior in just such a fashion, investigating the causes of bias in the measurement of religious behavior in order to understand Americans’ religious identities.

In conclusion, two survey experiments compared answers from conventional survey questions about religious behavior to measures of the same behavior from a chronological measurement procedure using SMS text messaging. The two chronological measurement procedures differed in their directiveness—the first did not mention the focal behavior and the second explicitly mentioned its focus on prayer. This manipulation changed the outcome, altering respondent behavior. Findings support the hypothesized primary role of identity as a cause of measurement error in survey measurement of normative behavior.

**Declaration of Conflicting Interests**
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Conway-Bascom Research funds provided to the second author by the University of Wisconsin-Madison Graduate School.

**Notes**

1. Notably, experimental survey questions aimed have had some success at reducing bias by, arguably, curtailing the self-reflection on the part of the respondent (Marler and Hadaway 1999; Presser and Stinson 1998). However, estimates of attendance from these experimental wording do not appear to completely eliminate bias and do not lessen the need for criterion measures (Marler and Hadaway 1999).

2. That is, this bias could be a consequence of survey measurement in which the source of bias pushes all estimates up (raising all boats), yielding little or no relative bias when comparing between countries.

3. Importance is often used synonymously with psychological centrality and prominence (see Brenner, Serpe, and Stryker 2014).
4. All response rates (RR) are computed as AAPOR RR. There were no ineligible cases or cases of unknown eligibility.

5. An alternative approach to this dichotomization between regular and frequent attendance (two to three times a month and more compared to less) applies a multiplier to each level of survey response based on the expected rate of attendance. For example, respondents answering “about once a month” should be expected to attend about one week in four, or 25 percent of the time. Respondents who report never attending and those reporting attending “once a week or more” should be expected to attend 0 and 100 percent of weeks, respectively. Other responses use the following multipliers: about once or twice a year (1.5/52 = 4 percent) and several times a year (6/52 = 11.5 percent). The products of these multipliers and the number of respondents selecting each survey response are then summed and divided by the total number of respondents to estimate a rate of weekly attendance.

This procedure yields estimates very similar or identical to those from the simple dichotomizations in Tables 1 and 2. In study 1, the multipliers yield an estimate of weekly attendance of 27.5 percent, identical to the dichotomization approach. In study 2, the attendance estimate for condition 1 based on the multipliers is 24.1 percent, compared to 25.9 percent using the dichotomization. For condition 2, the new estimate is 19.6 percent compared to 19.4 percent for the dichotomization approach. Notably, none of these adjusted percentages change the findings of the study.

6. The General Social Survey (GSS) rate includes the top four response categories for the variable ATTEND: “More than once a week,” “Every week,” “Nearly every week,” and “two to three times a month.” Analysis uses GSS weights and is the authors’ own.

7. Medians (\(\bar{y}\)) are included as many scales are ordinal, along with means (\(\mu\)) and standard deviations (\(\sigma\)).

8. While the validity of Hawthorne experiment has been challenged (Jones 1992), called “a glorified anecdote” by Richard Nisbett (Kolata 1998), the important point remains: Research participants may change their behavior when they know that they are being watched.

References


Anhøj, Jacob and Claus Møldrup. 2004. “Feasability of Collecting Diary Data from Asthma Patients through Mobile Phones and SMS (Short Message Service):
Response Rate Analysis and Focus Group Evaluation from a Pilot Study.” Journal of Medical Internet Research 6:e42.


Author Biographies

Philip S. Brenner is an assistant professor in the Department of Sociology and a senior research fellow in the Center for Survey Research at the University of Massachusetts Boston. His current research investigates identity processes in survey data collection. His work has recently appeared in Social Forces, Public Opinion Quarterly, Social Indicators Research, and the Social Psychology Quarterly.

John DeLamater is a Conway-Bascom Professor of Sociology at the University of Wisconsin, Madison. His research brings a broad, biopsychosocial perspective to the study of sexuality through the life course. His research has appeared in the Journal of Marriage and Family, the Journal of Sex Research, Journals of Gerontology Series B: Psychological Sciences and Social Sciences, and Social Indicators Research.