

## Manipulated vs. Measured: Using an Experimental Benchmark to Investigate the Performance of Self-Reported Media Exposure

Jennifer Jerit, Jason Barabas, William Pollock, Susan Banducci, Daniel Stevens & Martijn Schoonvelde

To cite this article: Jennifer Jerit, Jason Barabas, William Pollock, Susan Banducci, Daniel Stevens & Martijn Schoonvelde (2016) Manipulated vs. Measured: Using an Experimental Benchmark to Investigate the Performance of Self-Reported Media Exposure, *Communication Methods and Measures*, 10:2-3, 99-114, DOI: [10.1080/19312458.2016.1150444](https://doi.org/10.1080/19312458.2016.1150444)

To link to this article: <http://dx.doi.org/10.1080/19312458.2016.1150444>



Published online: 20 Apr 2016.



Submit your article to this journal [↗](#)



Article views: 38



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

## Manipulated vs. Measured: Using an Experimental Benchmark to Investigate the Performance of Self-Reported Media Exposure

Jennifer Jerit<sup>a</sup>, Jason Barabas<sup>a</sup>, William Pollock<sup>a</sup>, Susan Banducci<sup>b</sup>, Daniel Stevens<sup>b</sup>, and Martijn Schoonvelde<sup>c</sup>

<sup>a</sup>Department of Political Science, Stony Brook University, Stony Brook, New York; <sup>b</sup>Politics Department, University of Exeter, Exeter, UK; <sup>c</sup>Vrije Universiteit, Amsterdam, Netherlands

### ABSTRACT

Media exposure is one of the most important concepts in the social sciences, and yet scholars have struggled with how to operationalize it for decades. Some researchers have focused on the effects of variously worded self-report measures. Others advocate the use of aggregate and/or behavioral data that does not rely on a person's ability to accurately recall exposure. Our study illustrates how an experimental design can be used to improve measures of exposure. In particular, we show how an experimental benchmark can be employed to (1) compare actual (i.e., manipulated) and self-reported values of news exposure; (2) assess how closely the self-reported items approximate the performance of "true" exposure in an empirical application; and (3) investigate whether a variation in question wording improves the accuracy of self-reported exposure measures.

In order to study many contemporary political phenomena, it is crucial to know what political content a person has been exposed to in the mass media. Yet for decades, scholars have struggled to find the best way to operationalize media exposure. Concerns about measurement error and validity have even led some researchers to advocate abandoning self-reported media use measures altogether (e.g., Price & Zaller, 1993). However, important advances have been made in the measurement of media exposure (e.g., Althaus & Tewksbury, 2007; Dilliplane, Goldman, & Mutz, 2013; Guess, 2015; Prior, 2009a). Survey-based measures also are unique in their ability to capture people as they ordinarily encounter political information (Barabas & Jerit, 2009). In fact, one recent attempt to incorporate selection into media effects research relies upon a variant of a self-reported exposure question (Feldman, Stroud, Bimber, & Wojcieszak, 2013). Instead of rejecting the use of self-reported exposure measures, we must redouble our efforts to understand why people do not provide more accurate responses to such questions.

The purpose of this study is to explore the challenges of measuring media exposure through the use of an experimental benchmark. Scholars have used this approach in the areas of program evaluation and voter turnout (e.g., Arceneaux, Gerber, & Green, 2010; Green, Leong, Kern, Gerber, & Larimer, 2009; LaLonde, 1986), typically to compare the inferences from observational and experimental analyses. The random assignment of treatment and the resulting degree of internal validity is what makes the causal effect from an experiment the point of reference (i.e., the "benchmark") against which other methods are evaluated. In the present study, we employ an experimental benchmark to investigate the performance of self-reported exposure items.

**CONTACT** Jennifer Jerit  [jennifer.jerit@stonybrook.edu](mailto:jennifer.jerit@stonybrook.edu)  Department of Political Science, Stony Brook University, Social and Behavioral Sciences Building, Stony Brook, NY 11744.

Color versions of one or more of the figures in the article can be found online at [www.tandfonline.com/hcms](http://www.tandfonline.com/hcms).

We find that there is some imprecision in the self-reported exposure item, but the use of a randomized media treatment allows us to identify which respondents misreport and why. More specifically, our study provides an experimental confirmation of past work showing that people generalize from their interest in news to more specific questions about media exposure (Prior, 2009a). Additionally, whereas most of the emphasis in previous research has been on over reporting, we demonstrate that *under* reporting (e.g., forgetting that one was exposed) also is a problem. Finally, we leverage the experiment to assess the performance of two differently worded exposure measures, and show that an approach suggested by previous literature fails to increase the accuracy of self-reported exposure items.

## The existing literature

An experimental benchmark is a useful, but infrequently employed, method for investigating the problems of self-reported exposure measures (see Ansolabehere & Iyengar, 1995; or Vavreck, 2007; for notable exceptions). For the most part, scholars have either tried to “fix” the problems with self-reported exposure items (e.g., Bartels, 1993) or they have made theoretically-inspired modifications to the wording of existing self-report measures, using convergent and/or predictive validity to gauge the degree of improvement (e.g., Althaus & Tewksbury, 2007; Dilliplane et al., 2013; Prior, 2009a).<sup>1</sup> There is, however, a fundamental ambiguity with both approaches because the researcher remains ignorant of a person’s actual level of media exposure.<sup>2</sup>

To shed light on the problems associated with recall measures, it is crucial to establish a baseline against which self-reports can be compared. Vavreck (2007) does this by randomly assigning exposure to a mobilization message in a web-based survey at one point in time, and later asking people in the treatment and control groups to self-report their exposure to the images from the experimental stimuli. Thus, for each respondent, there is a true state of the world (i.e., their treatment assignment) as well as their *stated* level of exposure. In this way, the experimental benchmark gives the researcher the opportunity to “peak [sic] at the process that generates the data” (Vavreck, 2007, p. 333). In that particular study, people who over reported media exposure also overstated their political participation, and both forms of bias occurred among those who had high levels of civic duty and self-efficacy.

The Vavreck (2007) study identifies some of the motivational sources of misreporting. According to psychologists and survey researchers, however, the central problem with measuring media exposure is the *ability* of people to accurately recall the relevant behavior (e.g., Chang & Krosnick, 2003). One reason for the difficulty is the manner in which autobiographical memories of mundane behaviors are stored in the brain. Individual episodes of such behaviors are not stored in memory. “Instead,” write Schwarz and Oyserman, “the various instances of closely related behaviors blend into one global knowledge-like representation that lacks specific time or location markers” (2001, p. 137).

The recommendation for researchers seeking to recover accurate recall of a mundane behavior is to design survey questions in a way that facilitates memory. Markus Prior comes to a similar conclusion in a study that identifies ability, rather than motivation or socially desirable responding, as the cause of over reporting: “Most respondents are incapable of recalling most or all episodes of news exposure so they *estimate* their exposure,” (2009a, p. 904, emphasis added). This leads to two potential sources of error in self-reported items: the measure can miss people who were actually

<sup>1</sup>Convergent validity refers to the fit between independent measures of the same construct, while predictive validity refers to the ability of a construct to predict a criterion variable (e.g., political knowledge, in the case of media exposure; see Dilliplane et al., 2013).

<sup>2</sup>Nielsen ratings represent another alternative to self-reported exposure measures (Prior, 2009b), but they are an aggregate-level measure (i.e., the ratings represent the total audience for a particular program). Additionally, the ratings may contain error because (historically) they have been based on “people-meters,” which require viewers to push a button to indicate the beginning and ending of their viewing.

exposed but do not remember the treatment *and* it can include people who were never exposed but mistakenly report exposure (Vavreck, 2007, p. 332; also see Schwarz, 1999).

Notwithstanding these challenges, simple changes in questionnaire design may ameliorate the problem. Prior, for example, recommends that surveys “offer help with the estimation of news exposure” (2009a, p. 904). Likewise, Schwarz and Oyserman state that researchers can “improve the likelihood of accurate recall by restricting the recall task to a short and recent reference period and by providing recall cues...[such as] what happened, where it happened, and who was involved” (2001, p. 138). In the present study, we compare two versions of the self-reported exposure question: one that was modeled after the basic recall questions used by media researchers (the “general” version) and another that was designed to guide memory search and improve recall of the information treatment (the “specific” version).<sup>3</sup> In taking this approach, we follow the recommendation of Ansolabehere and Iyengar (1995), who state:

“The accuracy of new question wordings can only be gauged if researchers can measure actual exposure to a political measure, as well as the answer to the question. Experiments are ideally suited for this task. Survey researchers should make greater use of experiments as a means of testing and refining questions” (1995, pp. 23–24).

Taken together, the use of an experimental design will allow us to (1) compare actual (i.e., manipulated) and self-reported values of news exposure; (2) assess how closely the self-reported items approximate the performance of “true” exposure in an empirical application (e.g., as an independent variable predicting knowledge); and (3) investigate whether a subtle variation in question wording improves the accuracy of self-reported exposure measures. Although previous studies have explored changes in question wording (e.g., Prior, 2009a), to our knowledge none have evaluated alternative wordings against an experimental benchmark.

## Data and study design

We conducted a randomized laboratory experiment in which we controlled exposure to a news story. Because our goal was to compare actual (i.e., manipulated) and self-reported values of news exposure, we intentionally sought to implement our design in a controlled (i.e., laboratory) setting. We acknowledge that ordinary people consume news in settings that differ considerably from the one subjects experienced in our study. There have long been concerns about the realism of experimental research (e.g., Shadish, Cook, & Campbell, 2002). What our design may lack with respect to ecological validity (Morton & Williams, 2010) it compensates for in terms of the precision with which it can identify people who misreport their news exposure.

## Respondents

Participants were undergraduate students enrolled in political science classes at a large public university in the northeast region of the United States during the summer and fall of 2014. An initial study ( $n = 128$ ) was administered in May 2014. The study was replicated with different subjects later that same month ( $n = 25$ ) and again in November 2014 ( $n = 143$ ). In the analyses below, we combine data from all three studies (for a total  $n = 296$ ) and include indicators for the first and second experimental replications where appropriate.<sup>4</sup> The participants included 158 males and 135 females (with three people declining to indicate gender). The sample leaned slightly Democratic

<sup>3</sup>We view the general vs. specific distinction as one of many potential variations in question wording that may be used to facilitate recall (see Van Elsas, Lubbe, Van Der Meer, & Van Der Brug, 2014; or Timotijevic, Barnett, Shepherd, & Senior, 2009 for other applications).

<sup>4</sup>Subjects were recruited to participate in exchange for extra credit and instructed to sign up for the study through an online appointment system. The studies were approved by the Institutional Review Board at Stony Brook University (Application #: 580472). See Druckman and Kam (2011) on the use of student samples in experimental research.

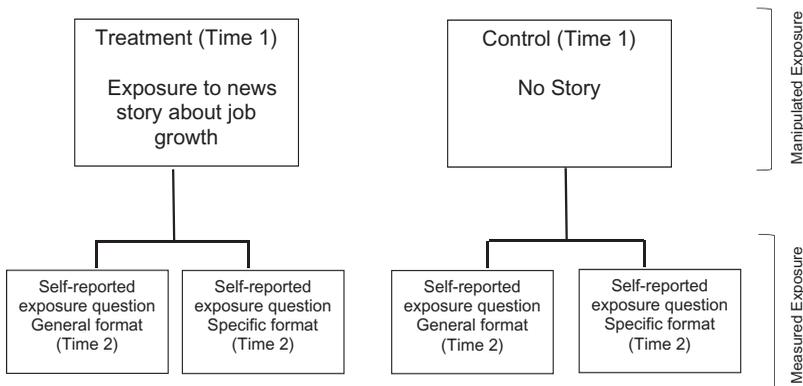
(on a 7-point partisan identification variable the mean was 5.1 with a standard deviation of 1.8). Table A1 provides additional details regarding sample characteristics.<sup>5</sup>

## Procedures

Our study is a 2 (Exposure: news story vs. no story) X 2 (Wording of self-reported exposure question: general vs. specific) between-subjects design that took place across two waves, approximately 2 days apart. Thus, participants were randomly assigned to receive a story or not, and then they were subsequently randomly assigned to receive one of the two question wordings of the media exposure measure.

The study occurred over two time points. At Time 1, participants completed a self-administered questionnaire on a computer in a private room. At the conclusion of the survey, subjects were informed about the opportunity to participate in a short follow-up study that would take place online at a time and place of their choosing (with participants entered into a drawing for a cash prize). Those who agreed to complete the follow-up questionnaire were sent a link to the study.<sup>6</sup>

Figure 1 shows some of the key design features of our study. At Time 1, participants were randomized to receive a story about job growth in the United States or no story (see Appendix for a screen shot of the stimulus).<sup>7</sup> The second experimental factor corresponds to a variation in question wording at Time 2, when participants answered a self-reported exposure item pertaining to the topic of job growth. There was a general version of the exposure item, which read, “Thinking about the past few days, do you recall seeing any information about job growth in the United States?” and had answer choices of “Yes” and “No.” The specific version was identical with the exception of a preamble that was intended to increase accuracy: “Thinking about the survey you completed a few days ago on campus, do you recall seeing any information about job growth in the United States?” The wording was intended to aid recall (Schwarz & Oyserman, 2001)



**Figure 1.** Design schematic.

Note: There was randomization into condition at Time 1 and Time 2. This design was implemented identically in each of the three replications noted in the text. The overall N for the study is 296.

<sup>5</sup>There were slight imbalances across conditions on partisanship, political knowledge, risk aversion, and presidential approval (see Table A1), however, a joint test reveals no significant difference in the overall composition of our experimental groups ( $p = .48$ ).

<sup>6</sup>In the original study, 66% of the respondents from Time 1 completed the follow-up. The re-contact rates for the first and second replications were 80% and 55%, respectively. Attrition was not significantly related to treatment assignment (see Table A1).

<sup>7</sup>The treatment was an edited version (338 words) of an actual *USA Today* story that appeared on the newspaper's website approximately 1 year before our study. We selected this article because it was geared toward young job seekers, and thus the topic should have had some appeal for our subjects.

by restricting the recall task to a short and recent reference period and by providing a cue regarding the location of the relevant behavior (e.g., a survey on campus).

## Measures

Most of the key outcome measures appear at Time 2, however, the original study and the first replication included a manipulation check immediately following the treatment at Time 1 (Mutz, 2011). The manipulation check consisted of three items asking respondents about details from the news story, and we will employ these questions to establish the effectiveness of the news treatment (i.e., that it had the intended effect on subjects).

As for the other outcomes, the Time 1 survey included measures of several individual difference variables that have been shown to be related to how people seek, acquire, and process information. There were six-items assessing Need for Cognition (NC) (Petty & Cacioppo, 1986;  $\alpha = .78$ ), four items on news elaboration (Eveland, 2002;  $\alpha = .93$ ), and a seven-item social desirability battery (Crowne & Marlowe, 1960;  $\alpha = .50$ ). Risk aversion was measured with a single question asking respondents to place themselves on a seven-point scale that ranged from “extremely comfortable taking risks” to “extremely uncomfortable taking risks” (with all response options labeled). Additionally, there was an item asking participants how often they pay attention to what is going on in government and politics, with five response options ranging from “Never” to “All the time.” All of the individual difference items were measured prior to the treatment at Time 1.<sup>8</sup>

The self-reported exposure question occurred at Time 2. As noted above, there was a general and specific version of the question, each with answer choices of “Yes” and “No.” After the self-reported recall item, all respondents received a follow-up asking how certain they were about seeing something about job growth (see Miller & Peterson, 2004 for more on meta-attitudinal measures). Response options ranged from “Absolutely certain” to “Not certain.” Respondents who recalled seeing something about job growth were asked an additional series of questions that probed the accuracy of their memories. The item (shown as a grid with “Yes” or “No” response options) asked whether the information pertained to “rates of job creation in different regions of the United States,” “a decline in the number of jobs in the restaurant industry,” or “a new federal program to train people receiving unemployment benefits.” Only the regional job creation item should have been answered affirmatively based upon the content of the news article treatment.

To investigate the possibility of acquiescence bias, we included a placebo item that asked respondents whether they recalled seeing anything about human cloning in the United States.<sup>10</sup> Finally, we repeated one of the knowledge questions from Time 1. The item asked, “According to economic forecasts, which region of the country has reported the fastest job growth in recent years? The answer choices were “East,” “West” (correct), “South,” and “Midwest.”

A novel feature of our study is that it both manipulates *and* measures exposure. As a result (and much like Vavreck, 2007), we can assess the degree to which a self-reported exposure measure identifies people who were exposed to the randomly-assigned treatment story. We also examine the performance of the manipulated and measured exposure items as independent variables in a model predicting treatment-relevant knowledge at Time 2. This is a useful exercise insofar as researchers often use self-reported media exposure variables in this capacity (e.g., Romantan, Hornik, Price, Cappella, & Viswanath, 2008). Finally, we compare two versions of the self-reported exposure (i.e.,

<sup>8</sup>To the extent that there are “errors” in self-reported measures of news exposure, these items will allow us to examine the characteristics of people who over or under report their exposure.

<sup>9</sup>When combined with the initial (dichotomous) recall item, this resulted in a six-point measure ranging from “Yes, absolutely certain” to “No, absolutely certain.”

<sup>10</sup>Respondents received a general or specific version of this question, depending on treatment assignment, along with the certainty follow-up. We confirmed via media content analysis that there had been no media coverage of human cloning around the time of the study.

recall) question. Based on work done by Schwarz and Oyserman (2001), we expected that the specific form of the recall question would guide memory search and help treated subjects remember elements of the treatment story.

## Empirical results

We begin the discussion of our results by demonstrating that exposure to the news story had the intended effect. As expected, treated subjects were significantly more likely to answer the three manipulation check questions correctly (2.5 vs. 1.2;  $|t| = 9.2$ ;  $p < .01$ ).<sup>11</sup> Even more impressive, the effect of the treatment persists over time, with treated respondents at Time 2 significantly more likely than control respondents to answer the knowledge question correctly (.51 vs. .27;  $|t| = 3.4$ ;  $p < .01$ ). Having shown that the treatment was effective, we next consider whether the Time 2 recall item recovered exposure among people who were in the treatment group at Time 1.

### *Does a self-reported measure identify who was (actually) exposed?*

There was a tendency for treated respondents to be more likely to answer “yes” to the self-reported exposure question at Time 2. This is reassuring, for one would expect there to be a relationship between what actually happened in the study (e.g., getting the job growth story or not) and subjects’ recollections. However, the correspondence is not particularly strong. In a probit model predicting self-reported exposure (i.e., answering “yes” to the recall question), the experimental treatment has positive and marginally significant effect (coeff = .30;  $p = .13$ ).<sup>12</sup> We observe a similar pattern with certainty of recall, with exposure to the treatment significantly increasing one’s certainty of recall on the self-reported exposure item (coeff = .36;  $p < .05$ ).<sup>13</sup> Based on the results so far, we have some confidence that the self-reported exposure measure is valid. In a design where we controlled who viewed the job growth story, the treatment has the expected (albeit weak) relationship to self-reported exposure and certainty of exposure.

At the same time, there is some imprecision in the self-reported media exposure measure. In our case, more than half (57%) of the respondents in control group report seeing something about job growth (i.e., they answer “Yes” to the self-reported exposure question at Time 2 even though they were not exposed to the story in the experiment). Additionally, a third of treated respondents say they do not recall hearing anything about job growth, even though we know they were treated (and that the treatment was effective).<sup>14</sup> Consistent with research on the difficulty of estimating mundane behaviors (e.g., Schwarz & Oyserman, 2001), many of the respondents in our study had difficulty figuring out whether they had been exposed to information about job growth.

### *Comparing measured and manipulated exposure as an independent variable*

Researchers often seek to measure what information people have been exposed to with the goal of relating self-reported exposure to other outcomes, such as learning and opinion change. Indeed, Dilliplane, Goldman, and Mutz write that “the gold standard for assessing the validity of media exposure is how well a measure predicts political knowledge *gain*” (2013, p. 238, emphasis original). Thus, we investigate the performance of the self-reported exposure measure in predicting knowledge

<sup>11</sup>All statistical tests are two-tailed unless otherwise noted.

<sup>12</sup>For the treatment indicator, the marginal effect is .11. This model includes indicators for the first and second replications, along with the pretreatment measure of respondent attention. Excluding the latter (only including controls for the replications) results in a  $p$ -value of .14 for the treatment indicator (marginal effect = .11).

<sup>13</sup>The sign and significance of the coefficient on the treatment indicator remains unchanged in a model that excludes the pretreatment measure of attention.

<sup>14</sup>67% of treated respondents accurately report exposure, which is slightly higher than the analogous figure reported in Ansolabehere and Iyengar (1995). Notably, there is no difference in the amount of time spent viewing the story according to self-reported recall status (i.e., answering “yes” or “no” to the recall question;  $p = .95$ ).

about job growth at Time 2. We expected that exposure to job growth information would be positively related to giving the correct answer to the knowledge question at Time 2. Tables 1 and 2 provide the relevant results.

Table 1 shows the results of a probit analysis in which the dependent variable is knowledge at Time 2 and the independent variable is the treatment indicator denoting whether the respondent received the job growth story at Time 1. Terms for the first and second replications are also included in the second column of results, but the conclusion is the same across both models. Being randomly assigned to view the story on job growth at Time 1 is positively and significantly ( $p < .01$ ) related to knowledge about this topic at Time 2. The marginal effect of the treatment indicator is substantively large (m.e. = .23), translating into a roughly 23 percentage point increase in knowledge for respondents exposed to the news story at Time 1.

A different conclusion emerges from Table 2, where we conduct a similar analysis but use *measured* exposure as the independent variable predicting Time 2 knowledge.<sup>15</sup> Because the analysis is now observational, we also included controls for gender and Need for Cognition, two individual-level characteristics that have been shown to be related to observed levels of knowledge (Bizer, Krosnick, Petty, Rucker, & Wheeler,

**Table 1.** Predicting knowledge (Time 2) with manipulated exposure.

	Coeff.	Coeff.
Manipulated Exposure (Treatment =1)	.63*** (.19)	.62*** (.19)
First Replication		.01 (.33)
Second Replication		.15 (.20)
Constant	-.62*** (.14)	-.68*** (.17)
Log-Likelihood	115.80	-115.53
$\chi^2$	10.89	11.46
N	182	182

Note: Cell entries denote probit coefficients with standard errors in parentheses.

\*  $p < .10$ ; \*\*  $p < .05$ ; \*\*\*  $p < .01$  (two-tailed)

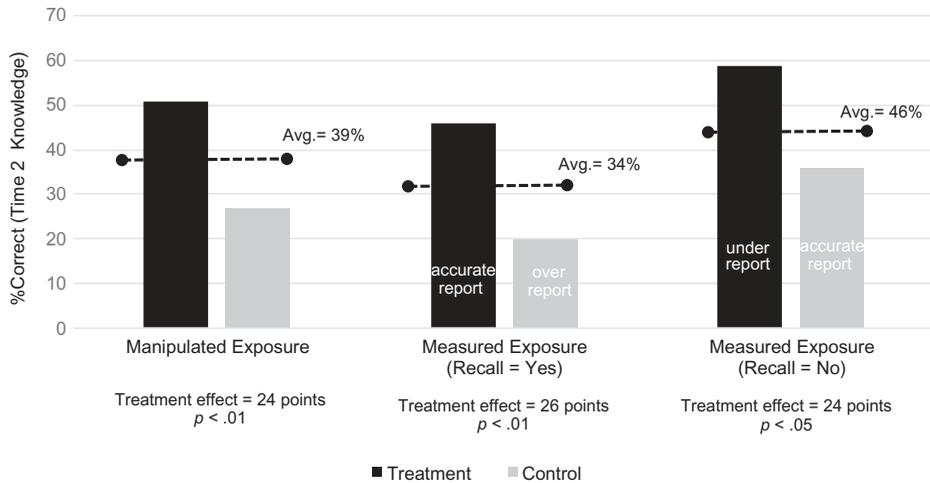
**Table 2.** Predicting knowledge (Time 2) with measured exposure.

	Coeff.	Coeff.	Coeff.
Measured Exposure (Recall=Yes)	-.33* (.19)	-.32 (.20)	-.42** (.20)
First Replication		.03 (.32)	-.01 (.34)
Second Replication		.14 (.20)	.05 (.21)
Male Respondent			.49** (.20)
Need for Cognition			.26* (.15)
Constant	-.09 (.15)	-.17 (.18)	-.34 (.20)
Log-Likelihood	-119.80	-119.55	-114.15
$\chi^2$	2.92	3.43	14.22
N	182	182	182

Note: Cell entries denote probit coefficients with standard errors in parentheses.

\*  $p < .10$ ; \*\*  $p < .05$ ; \*\*\*  $p < .01$  (two-tailed)

<sup>15</sup>We ignore the variation in question wording and collapse across the general and specific wording conditions (the effect of question wording is examined in the next subsection).



**Figure 2.** Percentage correct on wave 2 knowledge question, by treatment status and self-reported exposure.

Note: Columns represent the percentage giving the correct answer to the Wave 2 knowledge question. The average level of knowledge corresponds to the overall percent correct (i.e., across treatment and control conditions).

2000; Delli Carpini & Keeter, 1996). Irrespective of model specification, self-reported exposure is *negatively* related to knowledge at Time 2, and the variable is statistically significant in two of the three model specifications (at  $p \leq .10$  or better with the marginal effect ranging from  $-.12$  to  $-.16$ ).<sup>16</sup> The difference in results across Tables 1 and 2 is, on its own, quite striking. But consider the broader implications of the analyses thus far: A randomly assigned treatment *manipulating* news exposure increases the likelihood of saying “Yes” to self-reported recall item and knowledge related to the treatment. Yet the self-reported recall item—whose sole purpose is to *measure* exposure to the news treatment—is negatively related to knowledge.

To illustrate the source of this discrepancy, Figure 2 displays levels of knowledge across experimental groups at Time 2.

The two leftmost columns (“Manipulated Exposure”) show the percentage correct for treated and control subjects, and here the pattern mirrors previously-described findings. Treated subjects had significantly higher levels of knowledge than control subjects (51% vs. 27% resulting in a 24-point treatment effect;  $p < .01$ ). This difference represents the effect of manipulating exposure to job growth (i.e., randomly assigning some people to view the news story).

The remaining columns make analogous treatment vs. control comparisons, but now respondents are grouped according to whether they gave an affirmative response to the self-reported exposure question (Recall = Yes, middle set of columns) or a negative response (Recall = No, rightmost set of columns). Among those who said they had been exposed to information about job growth and who were in the treatment group, 46% give the correct answer to the knowledge question. In contrast, among those who said they had been exposed to information about job growth but were in the control group (i.e., no actual exposure), the corresponding figure is only 20% (resulting in a 26-point difference). The overall average percent correct among respondents who gave an affirmative response to the recall measure was just 34%, due in large part to the very low level of knowledge among people who over report their exposure.

A different pattern appears for respondents who indicated they did *not* recall seeing anything about job growth (the rightmost set of columns in Figure 2). People who said they did not recall anything but who

<sup>16</sup>The results remain unchanged when we include a measure of self-reported attention to politics, which itself is insignificant (coeff = .06; s.e. = .12).

were in fact treated in the experiment had high levels of knowledge (60% provide the correct answer). Among those who said they were not exposed and who were in the control group, approximately a third (36%) gave the correct answer (resulting in a 24-point difference). The overall percentage correct for respondents who said they were not exposed is 46%. Notice that this figure is higher than corresponding figure for people who reported exposure (34%, noted above). Thus, a combination of over reporting and under reporting *and* the differential levels of knowledge for both types of respondents is the source of the discrepancy across Tables 1 and 2.<sup>17</sup>

Insofar as there are errors in self-reported measures of news exposure, it behooves researchers to “[understand] the bias better” (Prior, 2012, p. 362). In our case, the randomized experiment permits an analysis of the individual-level correlates of misreporting. We are interested in those who falsely recall exposure (i.e., over reporting) and those who mistakenly forget it (i.e., under reporting). In separate models predicting over reporting and underreporting, the only individual-level factor that is statistically significant is self-reported attention (complete results shown in the Table A2). Respondents who report paying “attention to what’s going on in government in politics” are significantly more likely to over report exposure (i.e., say they recalled information even though they were in control group), while those who report lower levels of attention underreport (i.e., say they do not recall it even though they were in treatment group). In light of a difficult estimation task, respondents seem to use their habitual attention to news about government and politics as a cue when answering the self-reported recall question (see Prior, 2009a for a similar finding).<sup>18</sup> Consistent with the work of Prior (2009a), social desirability was not related to over reporting, nor was risk aversion, media elaboration, Need for Cognition, or gender.

### **Investigating the performance of differently worded exposure questions**

Having shown that the self-reported exposure measure contains some error, we now leverage the experiment to examine whether question format (general vs. specific wording) improves respondent accuracy. Because we know the “truth” (i.e., through a person’s treatment assignment), we can assess the performance of the general and specific wordings on the initial recall question, the follow-up item on certainty of recall, and the three-item battery that probed respondents for their memories about the treatment story.

Overall, question wording did not affect the accuracy with which people self-reported their exposure to news stories about job growth. Among treated respondents, receiving the specific language did not increase a person’s likelihood of giving an affirmative response to the self-reported exposure question (the mean value of *Recall* is .67 in both groups). Similarly, among the untreated, there was no difference in how respondents answered the self-reported exposure question across the specific and general conditions (.60 vs. .55;  $|t| = .60$ ;  $p = .55$ ). Question form also had no significant effects on certainty of self-reported recall ( $p > .50$  in analogous *t*-tests). Despite the addition of language that was intended to prompt memory of the “survey on campus,” there were no differences in accuracy among respondents receiving the general or specific wording either for the initial recall question or the certainty follow-up.

It is possible, however, that the specific language had more subtle effects when it comes to improving recall. Toward that end, we examined responses to the three-item battery probing what

<sup>17</sup>Table A3 provides a tabular presentation of the data from Figure 2. We estimated the model from Table 2 separately for respondents who accurately report their exposure and those who were inaccurate (over or under reporting). For accurate respondents, there is a positive but non-significant relationship between self-reported exposure and knowledge. For inaccurate respondents, that relationship is negative and significant ( $p < .001$ ). This pattern underscores the importance of devising self-reported exposure questions that elicit accurate responses.

<sup>18</sup>In explaining why more politically interested people over report media exposure, Prior states, they “rely too heavily on their generally high political involvement when estimating network news exposure without help. Even when they recall only a few episodes of news exposure, they may infer frequent exposure from their considerable interest in (and knowledge of) politics” (2009a, p. 901).

**Table 3.** The effect of question form on recall for elements of treatment story.

	Coeff.
Treatment	-0.14 (.28)
Specific	-0.14 (.32)
Specific X Treatment	0.89 * (.49)
$\mu_1$	-2.41 (.42)
$\mu_2$	-1.43 (.26)
$\mu_3$	-0.27 (.23)
Log-Likelihood	-99.83
$\chi^2$	4.59
N	114

Note: Coefficients are ordered probit estimates (standard errors in parentheses). Model N is lower because the analysis is run on Wave 2 respondents who indicated they saw a story about job growth.

\*  $p < .10$ ; \*\*  $p < .05$ ; \*\*\*  $p < .01$  (two-tailed)

respondents could remember about the treatment story.<sup>19</sup> In this analysis, the dependent variable is scored so that higher values denote level of accuracy (i.e., a score of three denotes respondents who said “yes” to the item on job growth and “no” to the restaurant and training program items). We predict performance on the scale with an ordered probit model including terms for *Treatment*, *Specific Question Wording*, and their interaction (*Treatment X Specific*). Table 3 presents the results.

The specific form of the self-recall question improves the ability of people to identify elements of treatment story, but only among people who were treated. The coefficient on *Treatment X Specific* is positively signed and statistically significant (coeff = .89;  $p = .07$ ). Translated into a more intuitive quantity, there is a 25% increase in the probability of obtaining the highest level of accuracy for people who were treated with the news story and received the specific form of the recall question, compared to those who were treated and received the general recall question. In an analysis of the placebo topic (not shown), the specific form also increased respondent accuracy. Few people said they had seen information about human cloning, but the specific question wording reduced the incidence even further (.06 vs. .01;  $|t| = 1.71$ ;  $p = .08$ ). The significant effect of question wording in these two analyses is noteworthy, both because the three-item battery was a difficult series of questions to answer correctly, and because a floor effect on the placebo question makes it that much more difficult for question wording to have a significant effect.

### Summary of results

This study employed experimental methods to better understand the challenges of using self-reported media exposure measures. Based upon the analyses reported here, there is evidence that the self-reported exposure variable contains error, but the experiment reveals that it is as much a product of under reporting as it is over reporting (see Figure 2). There also is suggestive evidence that misreporting occurs because respondents reason from their habitual news attention to answer a more specific question about news exposure. By contrast, other factors related to how people process information (e.g., Need for Cognition, news elaboration, risk aversion) were unrelated to

<sup>19</sup>As noted earlier, those who said they recalled seeing something about job growth were asked whether the information pertained to “rates of job creation in different regions of the United States,” “a decline in the number of jobs in the restaurant industry,” or “a new federal program to train people receiving unemployment benefits.” This question appeared in Wave 2 and it represents a more challenging recall task than the previous two items (overall recall and recall certainty). That said, we are not comparing groups formed by random assignment since we analyze a subset of people from the treatment group (i.e., those who recalled exposure).

misreporting among the subjects in this study. Finally, an attempt to facilitate recall with modified question wording was only partially successful: accuracy on the basic recall question did not improve, however, the specific wording did help some subjects better distinguish details from the news story they read at Time 1. It is important to emphasize that the insights from our analyses are derived from having an experimental benchmark—that is, *knowing* which individuals were treated with the news story. Such an approach provides unique leverage when it comes to uncovering the sources of bias and error in self-reported measures of media exposure.

## Conclusion

Media exposure has been described as “one of the most central concepts in all social science” (Prior, 2013, p. 632). Given its important place in research on public opinion and political behavior, there is room for methodological pluralism when it comes to understanding how best to measure this concept. Social science has benefited enormously from efforts to improve the wording of media use measures (e.g., Althaus & Tewksbury, 2007; Chang & Krosnick, 2003; Dilliplane et al., 2013), even if there is not complete consensus in that regard (e.g., Prior, 2013).<sup>20</sup> New technologies offer the promise of a behavioral measure by tracking media usage (e.g., Jackman, LaCour, Lewis, & Vavreck, 2012), but in many cases, people must agree to use the technology, which raises issues regarding obtrusiveness and non-compliance. Moreover, we are, in the view of some scholars still “years away” from reliable behavioral measures of the sort that academic researchers want (i.e., *individual-level* estimates of viewing behavior; see Goldman, Mutz, & Dilliplane, 2013, p. 640).<sup>21</sup> In the meantime, researchers must continue their efforts to develop better survey-based measures of media exposure.

Our contribution has been to develop a prototype of an experimental design that can be employed in a laboratory or online setting to explore a range of question wording variations and styles of media treatments. Much can be learned by comparing the performance of different exposure questions in a setting in which a news treatment has been delivered exogenously to some participants. While we did not find a significant difference in the accuracy of the general and specific versions of our self-report question, research on survey design (e.g., Schwarz & Oyserman, 2001; Tourangeau, Rips, & Rasinski, 2000) may suggest other question variations. As just one illustration, scholars might employ an experimental benchmark to demonstrate further the validity of the program list technique (Dilliplane et al., 2013). More generally, and as we documented here, an experimental benchmark also can be used to identify the individual-level correlates of misreporting. In our study, the errors in self-reports were as much a product of under reporting as over reporting. This is a novel finding since most of the emphasis in the literature has been on over reporting (e.g., see Prior, 2009b; or Price & Zaller, 1993).

That said, there are limitations with our study. While we were able to maintain a high degree of experimental control, the analytical leverage of a laboratory experiment comes at a price. In the natural world, citizens choose to attend to media sources, and modern democracies offer many possibilities in that regard (Prior, 2007; Stroud, 2011). Experiments that incorporate viewer choice have overturned some of the conventional wisdom regarding media effects (Arceneaux, Johnson, & Murphy 2012). Thus, it is possible that a person’s ability to recall exposure may differ according to how active he or she has been in selecting the information in question. Yet, there remains a place for self-reported measures even among research that involves viewer choice and selection effects. Indeed, one method for incorporating viewer choice entails the use of a self-reported exposure question (Feldman et al., 2013; also see Gaines & Kuklinski, 2011; fn 9). This particular approach

<sup>20</sup>Another challenge facing researchers who use survey-based measures of exposure is the steep decline in response rates (Kohut, Keeter, Doherty, Dimock, & Christian, 2012), and the potential for unobserved factors to be correlated with survey participation and self-reported media use (irrespective of how the exposure question is worded).

<sup>21</sup>Implicit techniques such as the Implicit Association Test (IAT) or lexical decision tasks are another class of tools for measuring media effects, however, many challenges remain in adapting these measures for communications studies (see Hefner, Rothmund, Klimmt, & Gollwitzer, 2011 for discussion).

involves post-hoc adjustment of experimental results based on a person's self-reported likelihood of selecting the source in "everyday life."<sup>22</sup> In our view, this example underscores how strong the tradition of using self-reported measures is in the social sciences and the importance of devising methods to improve their validity. We believe the analytical leverage provided by an experimental benchmark can be a powerful tool in that endeavor.

## Acknowledgments

We thank Yanna Krupnikov and Scott Clifford for helpful comments on previous versions of this article.

## References

- Althaus, S. L., & Tewksbury, D. H. (2007). Toward a new generation of media use measures for the ANES. *Report to the Board of Overseers, American National Election Studies*. Ann Arbor, MI.
- Ansolabehere, S. D., & Iyengar, S. (1995). *Messages forgotten: Misreporting in surveys and the bias toward minimal effects* (Unpublished manuscript). Cambridge, MA: Massachusetts Institute of Technology.
- Arceneaux, K., Gerber, A. S., & Green, D. P. (2010). A cautionary note on the use of matching to estimate causal effects: An empirical example comparing matching estimates to an experimental benchmark. *Sociological Methods and Research*, 39, 256–282. doi:10.1177/0049124110378098
- Arceneaux, K., Johnson, M., & Murphy, C. (2012). Polarized political communication, oppositional media hostility, and selective exposure. *Journal of Politics*, 74 (1), 174–186. doi:10.1017/S002238161100123X
- Barabas, J., & Jerit, J. (2009). Estimating the causal effects of media coverage on policy-specific knowledge. *American Journal of Political Science*, 53, 73–89. doi:10.1111/ajps.2008.53.issue-1
- Bartels, L. M. (1993). Messages received: The political impact of media exposure. *American Political Science Review*, 87, 267–285. doi:10.2307/2939040
- Bizer, G. Y., Krosnick, J. A., Petty, R. E., Rucker, D. D., & Wheeler, C. (2000). Need for cognition and need to evaluate in the 1998 national election survey pilot study. *Report to the Board of Overseers, American National Election Studies*. Ann Arbor, MI.
- Chang, L., & Krosnick, J. A. (2003). Measuring the frequency of regular behaviors: Comparing the "typical week" to the "past week". *Sociological Methodology*, 33, 55–80. doi:10.1111/some.2003.33.issue-1
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349–354. doi:10.1037/h0047358
- Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.
- Dilliplane, S., Goldman, S. K., & Mutz, D. C. (2013). Televised exposure to politics: New measures for a fragmented media environment. *American Journal of Political Science*, 57, 236–248. doi:10.1111/j.1540-5907.2012.00600.x
- Druckman, J. N., & Kam, C. D. (2011). Students as experimental participants: A defense of the 'Narrow Data Base'. In J. Druckman, D. Green, J. Kuklinski, & A. Lupia (Eds.), *Cambridge handbook of experimental political science* (pp. 41–57). New York, NY: Cambridge University Press.
- Eveland, W. P., Jr. (2002). News information processing as mediator of the relationship between motivations and political knowledge. *Journalism & Mass Communication Quarterly*, 79, 26–40. doi:10.1177/107769900207900103
- Feldman, L., Stroud, N. J., Bimber, B., & Wojcieszak, M. (2013). Assessing selective exposure in experiments: The implications of different methodological choices. *Communication Methods and Measures*, 7, 172–194. doi:10.1080/19312458.2013.813923
- Gaines, B. J., & Kuklinski, J. H. (2011). Experimental estimation of heterogeneous treatment effects related to self-selection. *American Journal of Political Science*, 55, 724–736. doi:10.1111/ajps.2011.55.issue-3
- Goldman, S. K., Mutz, D. C., & Dilliplane, S. (2013). All virtue is relative: A response to Prior. *Political Communication*, 30, 635–653. doi:10.1080/10584609.2013.819540
- Green, D. P., Leong, T. Y., Kern, H. L., Gerber, A. S., & Larimer, C. W. (2009). Testing the accuracy of regression discontinuity analysis using experimental benchmarks. *Political Analysis*, 1, 400–417. doi:10.1093/pan/mpp018
- Guess, A. M. (2015). Measure for measure: An experimental test of online political media exposure. *Political Analysis*, 23, 59–75. doi:10.1093/pan/mpu010
- Hefner, D., Rothmund, T., Klimmt, C., & Gollwitzer, M. (2011). Implicit measures and media effects research: Challenges and opportunities. *Communication Methods and Measures*, 5, 181–202. doi:10.1080/19312458.2011.597006

<sup>22</sup>Experimental subjects who have seen a stimulus are asked about the likelihood they would have ordinarily encountered that material in the real world.

- Jackman, S. D., LaCour, M. J., Lewis, J. B., & Vavreck, L. (2012, April). *Digital fingerprints: A new method for measuring political advertising*. Paper presented at the annual meeting of the Midwest Political Science Association, Chicago, IL.
- Kohut, A., Keeter, S., Doherty, C., Dimock, M., & Christian, L. (2012, May 15). *Assessing the representativeness of public opinion surveys*. Washington, DC: Pew Charitable Trusts Report.
- LaLonde, R. J. (1986). Evaluating the econometric evaluations of job training with experimental data. *American Economic Review*, 76, 604–620.
- Miller, J., & Peterson, D. (2004). Theoretical and empirical implications of attitude strength. *Journal of Politics*, 66, 847–867. doi:10.1111/j.1468-2508.2004.00279.x
- Morton, R., & Williams, K. (2010). *Experimental political science and the study of causality: From nature to the lab*. New York, NY: Cambridge University Press.
- Mutz, D. C. (2011). *Population-based survey experiments*. Princeton, NJ: Princeton University Press.
- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York, NY: Springer-Verlag.
- Price, V., & Zaller, J. (1993). Who gets the news? Alternative measures of news reception and their implications for research. *Public Opinion Quarterly*, 57, 133–164. doi:10.1086/269363
- Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. New York, NY: Cambridge University Press.
- Prior, M. (2009a). Improving media effects research through better measurement of news exposure. *Journal of Politics*, 71, 893–908. doi:10.1017/S0022381609090781
- Prior, M. (2009b). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, 73, 130–143. doi:10.1093/poq/nfp002
- Prior, M. (2012). Who watches presidential debates? Measurement problems in campaign effects research. *Public Opinion Quarterly*, 76, 350–363. doi:10.1093/poq/nfs019
- Prior, M. (2013). The challenge of measuring media exposure: Reply to Dilliplane, Goldman, and Mutz. *Political Communication*, 30, 620–634. doi:10.1080/10584609.2013.819539
- Romantan, A., Hornik, R., Price, V., Cappella, J., & Viswanath, K. (2008). A comparative analysis of the performance of alternative measures of exposure. *Communication Methods and Measures*, 2, 80–99. doi:10.1080/19312450802062539
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54, 93–105. doi:10.1037/0003-066X.54.2.93
- Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, 22, 127–160. doi:10.1177/109821400102200202
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-Experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Stroud, N. J. (2011). *Niche news: The politics of news choice*. New York: Oxford University Press.
- Timotijevic, L., Barnett, J., Shepherd, R., & Senior, V. (2009). Factors influencing self-report of mobile phone use: The role of response prompt, time reference and mobile phone use in recall. *Applied Cognitive Psychology*, 23, 664–683. doi:10.1002/acp.v23:5
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York, NY: Cambridge University Press.
- Van Elsas, E. J., Lubbe, R., Van Der Meer, T. W. G., & Van Der Brug, W. (2014). Vote recall: A panel study on the mechanisms that explain vote recall inconsistency. *International Journal of Public Opinion Research*, 26, 18–40. doi:10.1093/ijpor/edt031
- Vavreck, L. (2007). The exaggerated effects of advertising on turnout: The dangers of self-reports. *Quarterly Journal of Political Science*, 2, 325–343. doi:10.1561/100.00006005

## Appendix



### West leads in U.S. job growth

Seven of the 10 states with the fastest job growth this year are expected to be in the West.

Paul Davidson, USA TODAY

Go west, young job-seeker.



Seven of the 10 states with the fastest job growth this year will be in the West, as the region benefits from a stronger housing recovery and continued gains in its bread-and-butter energy, technology and tourism industries, according to forecasts by IHS Global Insight.

The states, which generally led the nation with rapid payroll increases last year, as well, are North Dakota, Texas, Arizona, Colorado, Utah, Idaho and Oregon.

Now that states such as Arizona and Nevada have worked through most of their home foreclosures, residential construction is rebounding sharply, spawning thousands of new jobs, economists say. "It was down so far, and the housing market has finally stabilized," says Richard Wobbekind, head of business research at University of Colorado, Boulder.

Other factors are also at work. North Dakota and Texas are riding an oil boom after largely avoiding the recession's most

punishing blows. Colorado and Utah, while enjoying a surge in oil and natural gas drilling, are also now high-tech centers helping satisfy Americans' appetite for mobile devices and applications.

Oregon is a semiconductor manufacturing hub. In Arizona, job growth is being fueled by a technology base that includes Apple's new 2,000-employee glass factory in Mesa, as well as surging tourism, now that rising household wealth is spurring more consumer spending.

As technology increasingly allows Americans to work remotely, the entire Western region is drawing more residents from other states who want to live amid scenic mountains and enjoy a better quality of life, says IHS economist Jim Diffley.

"They're just progressive, attractive places to live," Diffley says.

In the past, migration to the West was limited by the large number of Americans who couldn't move because they owed more on their mortgages than their homes were worth, says economist Chris Lafakis of Moody's Analytics. But the stock of so-called underwater homes has fallen dramatically.

"We expect 2014 to be the year when in-migration (to western states) picks up a lot," Lafakis says.

That will increase the need for local services and jobs.

**Table A1.** Balance of variables across conditions.

	Control	Treatment	Diff.	N
Male	.53 (.04)	.55 (.04)	-.01 (.06)	293
Party ID	.63 (.03)	.73 (.02)	-.10 *** (.04)	271
Foreign-born	.28 (.04)	.28 (.04)	.00 (.05)	293
Socially Desirable Responding	.48 (.02)	.51 (.02)	-.02 (.02)	293
Need for Cognition	.57 (.02)	.57 (.02)	.00 (.02)	294
News Elaboration	.68 (.02)	.67 (.02)	.00 (.02)	294
Knowledge (Wave 1)	.51 (.02)	.57 (.02)	-.06 * (.03)	290
Knowledge (Wave 2)	.72 (.03)	.71 (.03)	.01 (.04)	181
Internet consumption	.56 (.02)	.56 (.02)	.00 (.03)	293
Newspaper consumption	.12 (.02)	.04 (.10)	.08 (.10)	293
Radio consumption	.13 (.02)	.17 (.02)	-.03 (.03)	293
TV consumption	.28 (.02)	.32 (.02)	-.05 (.03)	293
Obama approval	.58 (.02)	.65 (.02)	-.07 ** (.03)	295
Attention to politics	.44 (.02)	.44 (.02)	.00 (.03)	295
Risk aversion	.40 (.02)	.35 (.02)	.05 * (.02)	295
Wave 2 Attrition	.35 (.04)	.41 (.04)	-.06 (.06)	296

\*Note: Cell entries are means (rounded to hundredths) with standard errors in parentheses. All variables scaled to [0,1]. In a model predicting treatment assignment at time 2 (specific versus general), the model chi square is insignificant ( $p = .48$ ).

\*\*  $p < .10$ ; \*\*  $p < .05$ ; \*\*\*  $p < .01$  (two-tailed)

**Table A2.** Factors that predict over and under reporting.

	Control/Recall	Treatment/No Recall
Attention	.25 *	-.41 ***
	(.14)	(.15)
Socially Desirable Responding	-.07	.30
	(.21)	(.26)
Risk Aversion	.03	-.03
	(.08)	(.10)
News Elaboration	-.04	.10
	(.15)	(.19)
Need for Cognition	.03	.02
	(.18)	(.22)
Male	.04	-.24
	(.21)	(.25)
Knowledge (Wave 1)	-.13	.17
	(.10)	(.12)
Constant	-1.16 ***	.13
	(.55)	(.60)
Log Likelihood	-107.47	-71.85
$\chi^2$	4.89	11.91
N	180	180

Note: Coefficients are probit estimates (standard errors in parentheses).

\* $p < .10$ ; \*\* $p < .05$ ; \*\*\* $p < .01$  (two-tailed)

**Table A3.** Percent correctly answering knowledge question at time 2, by treatment status and self-reported exposure

	Manipulated	Measured	
	Actual Exposure	Self-Reported Exposure (Recall = Yes)	Self-Reported Exposure (Recall = No)
Treatment	51% (n = 89)	46% (n = 59)	60% (n = 30)
Control	27% (n = 93)	20% (n = 54)	36% (n = 39)
Overall	39% (n = 182)	34% (n = 113)	46% (n = 69)

Note: Cell entries indicate percentage giving correct response to Wave 2 knowledge question, which asked: "According to economic forecasts, which region of the country has reported the fastest job growth in recent years?" Choices were "East," "West" (correct), "South," and "Midwest."