

CHEATING ON POLITICAL KNOWLEDGE QUESTIONS IN ONLINE SURVEYS

AN ASSESSMENT OF THE PROBLEM AND SOLUTIONS

SCOTT CLIFFORD*

JENNIFER JERIT

Abstract Survey researchers worry about the quality of data produced by online surveys. One concern is that respondents might cheat on performance questions, such as political knowledge, invalidating their responses. Yet, existing evidence is unclear about the prevalence of cheating, and scholars lack a validated method for coping with the problem. In this paper, we demonstrate that such cheating behavior varies considerably by sample and provide some evidence that it is motivated by self-deceptive enhancement. We experimentally test a variety of methods for reducing cheating and find that common methods, such as timers, are not the most effective approach. By contrast, a commitment mechanism, in which respondents affirm their choice not to cheat, is more efficacious. Although cheating in online surveys can distort estimates of knowledge and decrease the validity of the measure, there are methods for coping with this problem.

Across the social sciences, survey research is increasingly being conducted online because this mode is cost effective and convenient. Many prestigious organizations are adding online components (e.g., the American National Election Studies), and some companies operate completely online (e.g., YouGov/Polimetrix, Knowledge Networks/GfK). Yet, important issues regarding data quality arise when data are collected remotely. Measures of political

SCOTT CLIFFORD is an assistant professor in the Department of Political Science at the University of Houston, Houston, TX, USA. JENNIFER JERIT is a professor in the Department of Political Science at Stony Brook University, Stony Brook, NY, USA. The authors thank the following people for helpful comments and suggestions on earlier versions of this paper: Craig Burnett, John Bullock, Mona Kleinberg, Yanna Krupnikov, Thomas Leeper, Adam Seth Levine, and Hillary Shulman. This work was supported by the College of Arts and Sciences at Stony Brook University and the Division of Research at the University of Houston. *Address correspondence to Scott Clifford, University of Houston, Department of Political Science, 3551 Cullen Boulevard Room 447, Houston, TX 77204-3011, USA; e-mail: sclifford@uh.edu.

doi:10.1093/poq/nfw030

Advance Access publication August 12, 2016

© The Author 2016. Published by Oxford University Press on behalf of the American Association for Public Opinion Research.

All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

knowledge obtained from online surveys may be compromised if respondents use an external source to answer these items (e.g., look up the answers on the internet). This is an important concern because political knowledge is a central construct in political science, communications, and related fields.¹ As a crucial individual difference variable, political awareness influences attention to and reception of elite discourse (Zaller 1992), the use of heuristics (Lau and Redlawsk 2001), and other aspects of information processing such as effortful thinking, motivated reasoning, and value trade-offs (e.g., Kam [2005]; Jacoby [2006]; Taber and Lodge [2006]).

In this paper, we document how often respondents seek outside assistance while answering political knowledge questions online. Rates of self-reported cheating vary considerably across commonly used populations, from single digits in MTurk samples to over 40 percent in some student samples.² We also examine several approaches to reducing cheating in experiments across a variety of samples, and find a substantial variation in the effectiveness of different approaches. A simple request not to cheat has only modest effects, while the use of commitment language (Krosnick 1991; Clifford and Jerit 2015) is more useful. Additionally, the commitment mechanism leads to the largest reduction in knowledge scores among respondents who are the most likely to cheat. More generally, our analyses indicate that the manipulations are most effective among student populations, who are most inclined to cheat, and least effective among MTurk respondents, who are least inclined to cheat. Overall, cheating can distort estimates of political knowledge and decrease the validity of the measure, but there are methods for decreasing the incidence of this behavior.

The Debate over Cheating in Online Questionnaires

There is a lively debate about how often respondents seek outside assistance on knowledge questions in online surveys. Several scholars question the frequency of cheating (e.g., noting the tendency of respondents to satisfice), and some evidence seems to support this belief (Gooch and Vavreck 2015). In one study, for example, difficult questions were not answered correctly at rates better than chance, which suggests that people are not looking up answers (Berinsky, Huber, and Lenz 2012, 359; also see Munzert and

1. We focus on survey-based measures of knowledge because, as Mondak observes, “there is compelling evidence that political awareness is best represented with data from survey batteries that measure factual knowledge” (2001, 224; also see Delli Carpini and Keeter [1996]).

2. We use the word “cheating” to be consistent with past research (e.g., Jensen and Thomsen 2014; Shulman et al. 2014; Burnett 2015), and employ the term to describe a specific kind of respondent behavior—namely, consulting an external source when answering web-based knowledge questions. Our discussion focuses on students, workers on Amazon’s Mechanical Turk (MTurk) platform, and panelists in online surveys such as YouGov because these are commonly used subject populations in social science research (Krupnikov and Levine 2014).

Selb [forthcoming]).³ Other researchers acknowledge that political knowledge scores are higher in online samples, but they maintain that the difference is due to sample composition rather than cheating (Ansolabehere and Schaffner 2014). Collectively, these studies imply that cheating on web-based political knowledge questions is uncommon.

However, a different conclusion emerges from a mode experiment in which student subjects were randomized to take a questionnaire online or in a lab (Clifford and Jerit 2014). In that study, online participants had significantly higher political knowledge scores than those completing the survey in a lab. Due to the random assignment of survey mode, the Clifford and Jerit (2014) study provides some of the clearest evidence of cheating in online surveys. Additionally, the analysis revealed that cheating harms the descriptive validity of political knowledge scales (Luskin and Bullock 2011), as well as the criterion validity of the measure. Both findings support our assumption that cheating can impair the quality of political knowledge scales. Other studies using student (Shulman and Boster 2014; Burnett 2015) and adult subjects (Fricker et al. 2005; Strabac and Aalberg 2011) also find higher political knowledge scores online compared to telephone or paper-and-pencil surveys. Cheating appears to be prevalent in nationally representative samples as well. When respondents are directly asked whether they looked up answers, cheating rates among national samples range from 13 percent (Shulman et al. 2014) to 22 percent (Jensen and Thomsen 2014). Taken together, the evidence regarding cheating suggests enough of a problem to warrant further examination of its prevalence and ways to reduce this behavior.

The Motivations for Cheating in an Anonymous Online Survey

We contribute to this literature by developing an account of why a person might cheat on knowledge questions in the first place. It is often argued that online surveys produce low levels of social desirability bias (e.g., Holbrook and Krosnick 2010), resulting in the conclusion that respondents have little incentive to consult external sources on web-based knowledge questions. However, there is an important distinction between two forms of socially desirable responding: impression management (IM), which is a “deliberate tendency to over report desirable behaviors and under report undesirable ones” and self-deceptive enhancement (SDE), which refers to the “tendency to give honestly believed but overly positive reports about oneself” (Booth-Kewley, Larson, and Miyoshi 2007, 464, emphasis in original). Notably, SDE

3. Gooch and Vavreck (2015) report low rates of cheating, but their study took place in a lab-like setting, making it different than the typical online survey in which respondents participate at a time and place of their own choosing.

occurs on attributes involving social and intellectual competence (Paulhus and John 1998) and leads to over-claiming of knowledge across a wide variety of domains (Paulhus et al. 2003). Thus, while IM pressures may indeed be low in an online survey (though see Krupnikov, Piston, and Bauer [2016]), SDE is *higher* due to the greater opportunity to self-enhance in an online setting (Booth-Kewley, Larson, and Miyoshi 2007).

But what does it mean for respondents to self-enhance in the area of political knowledge? A recent line of research demonstrates that the internet allows people to “offload” the burden of remembering facts and that offloading takes place without much (if any) conscious awareness (Sparrow, Liu, and Wegner 2011; Ward 2013). Consequently, people feel as if they “know” the information simply because they know where to find it. This research has shown, for example, that subjects who were instructed to access the internet while answering trivia questions took credit for their higher scores, rating themselves as more knowledgeable and having higher cognitive self-esteem than subjects who were not instructed to access the internet (Ward 2013; Fisher, Goddu, and Keil 2015). The implication is that as the internet becomes the primary way people acquire information about the world, they fail to differentiate data that are personally known from that obtained online.

These various lines of research imply that online surveys are ripe for distortions in political knowledge. The internet provides the opportunity for people to deceive themselves into believing they are knowledgeable, while self-deceptive enhancement provides the motive (Shulman and Boster 2014). Yet, the effects of SDE might differ by subject population due to variation in intrinsic interests and material incentives. In particular, research has shown that the drive for self-enhancement is strongest when an individual values the topic (Brown 2012). If student subjects, who are typically recruited from social science courses, place greater value on being politically knowledgeable than other respondents, they may have greater desire to self-enhance on knowledge questions.⁴ The opportunity cost of looking up answers may also affect the incidence of cheating. Compared to students and online panelists, MTurk respondents face the greatest opportunity cost from cheating, due to the unique structure of that platform (Chandler, Mueller, and Paolacci 2014). Time spent looking up answers to knowledge questions represents foregone earnings from completing other tasks. In contrast, students and online panelists typically participate at their leisure, are limited in the number of surveys they can take, and do not forfeit potential earnings if they spend a moment looking up answers to questions.

The preceding discussion suggests that cheating rates vary considerably by sample, just as response styles such as survey satisficing (Hauser and Schwarz forthcoming) and social desirability (Krupnikov, Piston, and Bauer 2016) vary

4. Indeed, in the data we report below, student subjects had a tendency to place more importance on being politically knowledgeable than campus staff ($p < .10$).

across populations. This variation may explain why cheating on web-based knowledge questions appears common in some studies (e.g., Clifford and Jerit 2014), but less common in others (Shulman et al. 2014; Ansolabehere and Schaffner 2015; Gooch and Vavreck 2015). Looking up answers should be the most prevalent among student subjects because they face low opportunity costs and a stronger motivation to self-enhance on studies about politics.⁵ In contrast, cheating should be rare among MTurk respondents because they face high opportunity costs for doing so. Adult subjects and members of online panels fall between students and MTurk respondents on both dimensions, and thus should also fall between the two in terms of their tendency to consult external sources on web-based political knowledge questions.

Given the nascent status of the literature, another purpose of our study is to examine the effectiveness of various instruction sets in reducing cheating behaviors. The treatments are described in the next section, but in brief, we used instruction sets that were suggested by the existing literature or were being employed by major survey organizations. Remarkably, certain practices, such as the inclusion of timers on political knowledge questions, have never been systematically evaluated despite their use by many researchers and polling firms.⁶ Based on the preceding discussion, we expect that the treatments will be most efficacious among respondents who are inclined to cheat. Consequently, the treatments should vary in effectiveness both within and across samples: they should be most effective among people who are high in SDE and student subjects, and least effective within MTurk samples.

Data and Measures

To test our expectations, we conducted 10 studies across populations commonly used in social science research. Table 1 summarizes the key features of each study. Four samples were drawn from undergraduate subject pools at two universities in different regions of the country, with sample sizes of 84, 845, 271, and 66. Student 1 and Student 4 were conducted at Stony Brook University, while Student 2 and Student 3 were run at the University of Houston.

Additionally, we collected data from several adult samples, including a study of campus staff (Kam, Wilking, and Zechmeister 2007) that was administered in parallel with Student 4 at Stony Brook University ($N = 59$), an original survey conducted by YouGov ($N = 1,000$), and a team module on the 2014 Cooperative Congressional Election Study (CCES; $N = 1,000$). The CCES data piggybacked

5. Additionally, student subjects may be likely to offload the burden of remembering facts because many of them came of age in the internet era (Kleinberg and Lau 2014).

6. Shulman and Boster (2014, 187) and Jensen and Thomsen (2014, 3353) both encourage future researchers to examine how to reduce cheating on knowledge questions in online surveys.

Table 1. Summary of Data Sources

	Sample size (control group)	Approximate date of study	Included experimental manipulation(s)	Number of knowledge Qs
Student 1	84 (46)	May 2014	Direct request	5
Student 2	845 (203)	November 2014	Forgiving, Timer, Commitment	8
Student 3	271 (135)	March 2015	Commitment	8
Student 4	66 (35)	April 2015	Commitment	4
Campus staff	59 (30)	May 2015	Commitment	4
YouGov	1,000 (354)	December 2015	Timer, Commitment	8
CCES	1,000 (NA)	November 2014	None	5
MTurk 1	500 (NA)	July 2014	None	6
MTurk 2	505 (NA)	June 2015	None	12
MTurk 3	300 (167)	April 2015	Commitment	4

NOTE.—The data reported in Student 1, Student 4, and Campus Staff were from the second wave of two-wave studies. See [appendix A](#) for details on the administration of all studies.

on an unrelated study, and there we examine cheating on a five-item Wordsum battery, which measures vocabulary rather than political knowledge. The Wordsum measure is often used as a proxy for intelligence (e.g., Gooch and Vavreck 2015) and should create the same self-enhancement dynamics as political knowledge (Shulman et al. 2014; Burnett 2015). Finally, we have data from three separate MTurk samples that were part of unrelated studies (MTurk 1–3; $N = 500, 505, 300$). Several of the studies included experimental manipulations designed to reduce cheating, which we discuss in detail below.

In each study, political knowledge was measured with batteries of 5–12 questions, all but one of which were in multiple choice format.⁷ Details on question wording are provided in appendices B and C, but all of the scales included questions about institutions, politicians, and policy-specific facts (Barabas et al. 2015), and they are representative of knowledge scales used by survey researchers.⁸ Immediately following the knowledge items, respondents were asked whether they had looked up any answers. The question, shown below, was designed to be as forgiving as possible to decrease misreporting (e.g., Peter and Valkenburg [2011]):

Many people struggle to remember facts, even when they know them, and so they get help remembering. When you were answering the factual knowledge questions, did you get help from any other source, such as the internet or another person? (Please be honest, this is for statistical purposes only.)

Naturally, our estimate of cheating may be conservative (i.e., a *low* estimate) if some subjects underreport cheating behavior. However, we show that multiple pieces of evidence, including knowledge scores and time spent on the questions, converge with this measure.

Prevalence of Cheating across Samples

Figure 1 displays the rates of self-reported cheating by sample, excluding data from experimental treatment conditions.⁹ Among the student samples, rates of

7. With the exception of MTurk 2 ($N = 505$), there was no explicit “Don’t Know” (DK) option (see Miller and Orr [2008] for discussion).

8. The scales vary in terms of length and apparent difficulty, as judged by percent correct on the scale (which ranges from 44 to 77 percent correct). There also were differences in the topics of the surveys in which the knowledge scales were embedded. Any of these factors might influence the motivation to cheat, but as we report below, we observe the expected patterns despite this variation. Moreover, data from Study 4 suggest that scale differences do not seem to affect rates of self-reported cheating. In that study, students were randomized to receive one of two knowledge batteries. One knowledge scale was substantially more difficult, as evidenced by lower scores ($p < .01, d = .34$). Yet, respondents receiving the more difficult scale were no more likely to report cheating ($p = .56$).

9. Sample sizes may be smaller than those reported in table 1, due to item non-response on the cheating question.

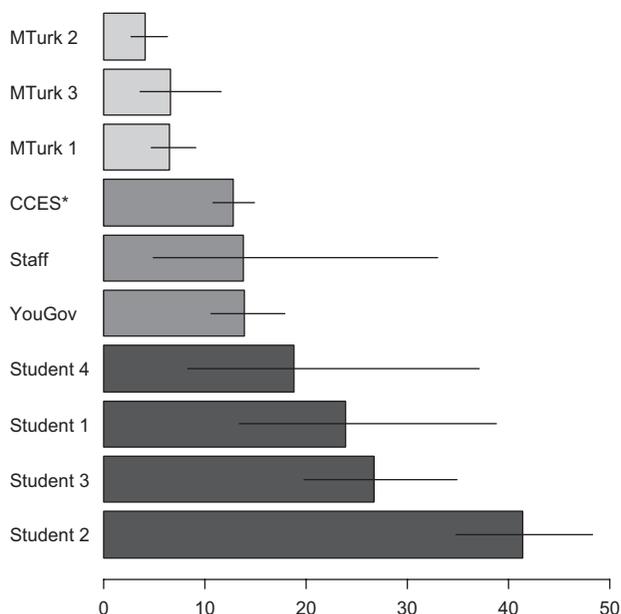


Figure 1. Rates of Self-Reported Cheating on Knowledge Questions in Online Surveys. Bar length represents the percentage of self-reported cheating on knowledge questions among student, staff, MTurk, and national samples (lines represent 95 percent confidence intervals). Studies arranged in ascending order by self-reported cheating rates. See table A1 for details. CCES cheating rates come from five-item Wordsum measure. Data from control conditions only. Sample size may be smaller than reported in table 1 due to item non-response.

self-reported cheating are relatively high, ranging from 24 to 41 percent. In contrast, self-reported cheating is much lower in the MTurk samples, ranging from 4 to 7 percent. The disparity between students and MTurk subjects is consistent with the notion that the financial structure of MTurk provides an incentive not to look up answers.¹⁰ As expected, the reported cheating rates in our adult samples are between the rates for students and MTurk subjects. In both the campus staff and the YouGov samples, 14 percent of respondents reported cheating. Similarly, 13 percent reported cheating on the five-item Wordsum battery in the 2014 Cooperative Congressional Election Study (CCES).

Other metrics (time spent on the question, knowledge score; shown in table A1 of online appendix B) support our claim about variable motives and opportunities for cheating across samples. Self-reported cheaters spent

10. In later analyses, we examine whether MTurk subjects are more likely to underreport cheating for fear of not being compensated.

significantly longer answering the knowledge questions, with standardized effect sizes (Cohen's d) ranging from medium to large, $d = .63$ to 1.27 . This is the expected pattern if respondents are seeking outside assistance on knowledge questions (cf. Jensen and Thomsen [2014]). At the same time, there was considerable variation in the median time per question across sample types, with students taking more time than YouGov respondents and MTurk workers. This difference is consistent with the claim that student subjects face both lower opportunity costs and a stronger motivation to cheat on political knowledge questions than other types of respondents. Notably, even among students, the average time per question was low, with the median student who did not report cheating averaging 13 seconds per question and the median student who did report cheating averaging 27 seconds. Thus, our data cast doubt on the efficacy of timers for reducing cheating since the time limits are often longer than 30 seconds (see also Jensen and Thomsen [2014]).

When it comes to knowledge scores among respondents who self-report cheating, there is a tendency for cheaters to have higher knowledge scores, but this difference is statistically significant only in four larger studies (see table A1 in online appendix B for details).¹¹ Among respondents in the YouGov sample, people who reported cheating scored significantly *lower* than those who did not report cheating ($p < .05$). This suggests that the relationship between cheating and knowledge scores may be complex. Cheating seems most likely to take place among respondents who do not know (or who are not certain of) the correct answer. The key issue is whether these respondents overcome their knowledge deficit by consulting the internet. Data across our 10 studies indicate that while cheaters frequently score higher than non-cheaters, sometimes they do not.

The Effect of Interventions Designed to Reduce Cheating

In this section, we examine whether changes in survey design reduce cheating, beginning with a series of experiments conducted on student samples. We focus on students because this is the subject population that is most likely to cheat. Upon identifying the most effective treatments, we examine the effects of these manipulations among a large national sample and an MTurk sample.

In our first experiment (Student 1), all subjects received a standard set of instructions to the knowledge section, but half were randomly assigned to also receive a statement asking them not to look up answers during the survey ("Direct Request"), an approach used in the 2012 ANES Time Series (see also Berinsky, Huber, and Lenz [2012]).¹² In the control condition, 23.9 percent

11. Those studies include students, online panelists from the CCES, and MTurk workers.

12. The language was as follows: "Now we have a set of questions concerning various public figures. We want to see how much information about them gets out to the public from television, newspapers, and the like. [Please give your best guess and do NOT use outside sources like the Internet to search for the correct answer.]"

reported cheating, while 18.4 percent reported cheating in the treatment condition, a difference that is not statistically significant ($p = .54$). Knowledge in the treatment group is lower than in the control (3.7 vs. 3.9, respectively), but the difference is not statistically significant or substantively large ($p = .44$, $d = .17$).

In our second student study (Student 2), we explored three different manipulations: Timer, Commitment, and Forgiving Request. All respondents (including those in the control group) received the standard introduction (see note 12), while those in the treatment conditions also saw additional text described below. The Timer treatment placed a 30-second timer on each question and explained to subjects that the screen would automatically advance after 30 seconds (Ansolabehere and Schaffner 2014; Prior, Sood, and Khanna 2015; Bullock et al. 2015).

Timer: “Please do NOT use outside sources like the Internet to search for the correct answer. You will have 30 seconds to answer each question.”

The second treatment asked respondents whether they were willing to answer the knowledge questions without getting help and requested a yes or no response (“Commitment”; see Krosnick [1991] or Clifford and Jerit [2015]). We expect this treatment to be effective because few people will reject a request from an interviewer, and subsequently they will be motivated to maintain consistency with their previous commitment (Cialdini et al. 1978; Cannell, Miller, and Oskenbeurg 1981).

Commitment: “It is important to us that you do NOT use outside sources like the Internet to search for the correct answer. Will you answer the following questions without help from outside sources?” (Yes, No)¹³

The third and final treatment was a stronger version of the Direct Request from the Student 1 study. Like the Direct Request, the treatment instructed subjects not to look up answers, but it also included additional forgiving language intended to reduce social desirability pressures (“Forgiving Request”; e.g., Duff et al. [2007]).

Forgiving Request: “Many people will know the answers but won’t be able to remember them at the moment. However, it is essential that we only measure what people are able to recall on their own. Thus, it is important to us that you do NOT use outside sources like the Internet to search for the correct answer.”

The comparisons from the Student 2 sample are shown in figure 2. On average, subjects in the control condition answered 5.1 questions out of 8 correctly,

13. Only 5 out of 222 subjects (2.3 percent) answered “no” to the question posed by the treatment. All subjects are retained for analysis regardless of their response.

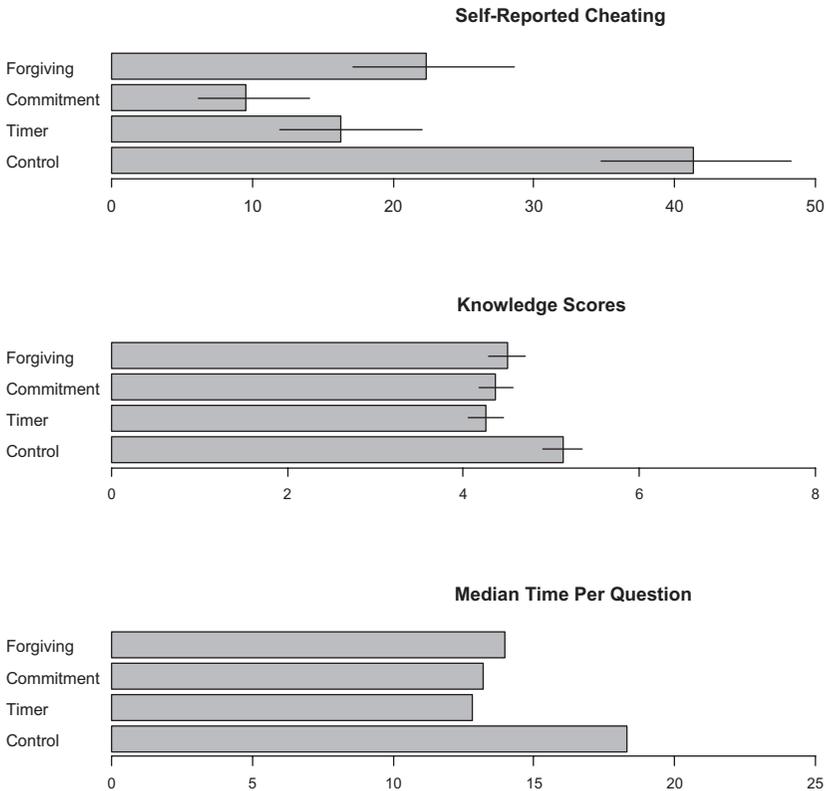


Figure 2. A Test of Three Methods to Reduce Cheating. Bar length represents the percentage of subjects self-reporting cheating (top), knowledge scores, measured as number of questions correct out of eight items (middle), and the median response time per question (bottom). Data comes from Student 2 sample ($N = 845$).

with 41 percent reporting cheating. Beginning with the Timer condition, subjects were significantly less likely to report cheating (16.4 percent, $p < .001$) and scored significantly lower on knowledge (4.3; $p < .001$, $d = .54$). One would expect this treatment (by design) to reduce time spent on the knowledge questions, and the results bear out this expectation ($p < .001$, $d = .71$).¹⁴

Turning to the Commitment condition, self-reported cheating was dramatically lower than in the control condition (9.5 percent, $p < .001$). In addition, knowledge scores (4.4; $p < .001$, $d = .49$) and time spent on the knowledge questions were significantly lower as well ($p = .015$, $d = .24$). Finally, we examine the Forgiving treatment, which was designed to be a stronger version

14. We use the log of the average time spent per question for all tests of reaction times.

of the Direct Request. Relative to the rate of cheating in the control condition (41.4 percent), the Forgiving treatment significantly decreased rates of self-reported cheating (22.4 percent, $p < .001$), knowledge scores (4.5; $p < .001$, $d = .40$), and time spent on the knowledge questions ($p = .004$, $d = .29$). However, this manipulation was less effective than both the Commitment and Timer conditions.

The three treatments from the Student 2 study significantly reduced cheating, knowledge scores, and time spent on the knowledge questions. The Timer and Commitment treatments led to the lowest levels of political knowledge, while the Commitment condition produced significantly lower levels of self-reported cheating ($p < .05$). This combination of results was unexpected, but it could have occurred if the Timer condition was less effective at reducing cheating *and* it interfered with non-cheaters' ability to answer the questions. Recall that even among those who reported cheating, the median subject spent less than 30 seconds answering each question in all of our samples and conditions. Thus, timers face the challenge of being short enough to prevent cheating, but not so short that they interfere with the regular response process. Overall, the Commitment language appears to be the most effective method for minimizing cheating on political knowledge questions.

In our next study, Student 3, we sought to replicate our findings regarding the Commitment mechanism and examine whether the treatment has the strongest effects among respondents who are the most likely to cheat (i.e., those high in SDE). In this study, 271 students were randomly assigned to the Commitment or control condition. The study also included the Balanced Inventory of Desirable Responding (BIDR-40; Paulhus 1991), which measures both self-deceptive enhancement (SDE) and impression management (IM). In the BIDR-40, subjects are asked to rate how true 40 statements are about themselves (e.g., "I am a completely rational person"; see [appendix D](#) for full scale). Each subscale (IM and SDE) is scored as the average agreement with each of the 20 corresponding items (Stöber, Dette, and Musch 2002).

To test whether the treatment had different effects across levels of SDE, we estimated a model in which knowledge was regressed on treatment status, SDE, IM, and interactions between each disposition and the treatment condition (full results shown in the online [appendix B](#); [table A2](#)). As expected, we find a negative interaction between SDE and the Commitment condition ($p = .07$), suggesting that the Commitment treatment was more effective among those high in SDE. This finding is consistent with our argument that these respondents are the most motivated to cheat.

DETERRING CHEATING IN ADULT SAMPLES

While our manipulations proved effective at reducing cheating in student samples, adult respondents may respond differently to the treatments. We investigate this possibility by testing the two most effective treatments from the

student studies, Timer and Commitment, in the adult sample collected through YouGov.¹⁵ Recall from [figure 1](#) that adult subjects were less likely than students to report cheating. This pattern is consistent with the claim that online panelists have weaker motives to cheat than student subjects. Yet, the weaker motivation to cheat may make it more difficult to find a significant effect for language designed to reduce cheating among adult subjects. In the control condition of the YouGov study, the rate of self-reported cheating was 14 percent. The Timer condition reduced self-reported cheating to 8 percent ($p = .03$) while the Commitment condition reduced cheating to 6 percent ($p = .009$). When it comes to levels of knowledge, the differences across conditions were more muted. On average, subjects in the control condition answered five out of eight questions correctly. Knowledge scores did not significantly differ in either the Timer (4.8; $p = .31$) or Commitment conditions (5.0; $p = .67$), though both the Timer and Commitment conditions decreased time spent on the questions ($p < .001$, $p = .08$, respectively). Overall, both manipulations appeared less effective on the adult sample than the student samples, though both reduced self-reported cheating and time spent on the questions. Moreover, as we show below, the Commitment condition also improved the validity of the political knowledge scale in the YouGov sample.

Although neither treatment had a main effect on political knowledge, the interventions might have an influence on the subgroup most inclined to cheat: those high in SDE. For this purpose, we included an abbreviated 16-item measure of SDE and IM ([Bobbio and Manganelli 2011](#)) in the YouGov study. Similar to our analysis in the previous section, we predicted political knowledge as a function of the Timer and Commitment conditions, SDE, IM, and interactions between each condition and both SDE and IM (see [table A3](#) in online [appendix B](#) for full model results). We find suggestive evidence of the expected negative interaction between SDE and the Commitment condition ($p = .12$) but weaker evidence for an interaction between SDE and the Timer condition ($p = .36$) and between IM and either condition (all $ps > .76$). Overall, there are similarities across student and adult respondents although the results are weaker among the YouGov sample. This may stem from YouGov respondents being less motivated to cheat.

DETERRING CHEATING IN MTURK SAMPLES

We now turn to the question of whether our treatments are effective in MTurk samples. Earlier, we argued that the low rate of self-reported cheating among MTurk subjects (see [figure 1](#)) is a function of opportunity costs that countervail the motivation to cheat (cf. [Goodman, Cryder, Cheema \[2013\]](#)). An alternative

15. We included an experiment involving the Commitment language in Student 4 and Campus Staff ($N = 66$ and $N = 59$, respectively). Despite our recruiting efforts, these studies were smaller than expected, which in turn reduced statistical power. In both studies, the magnitude of treatment effects was consistent with our other studies, but the differences were not statistically significant.

explanation is that MTurk subjects are simply more likely to *underreport* cheating for fear of having their pay rejected. If cheating is in fact low on MTurk (as we argue), the Commitment manipulation should have little effect on knowledge scores of MTurk respondents. To test this prediction, we randomly assigned subjects in our MTurk 3 sample ($N = 300$) to the Commitment condition or a control condition, followed by four knowledge questions. In line with our expectations, rates of self-reported cheating among MTurk subjects in the control condition were low, at 6.6 percent. This rate dropped to 1.5 percent in the treatment condition ($p < .05$). However, there was no significant difference in political knowledge scores across conditions ($p = .35$). This pattern bolsters the claim that cheating actually is rare among MTurk subjects. It also suggests that previous reports of low levels of cheating among MTurk respondents who had been instructed not to cheat (Berinsky, Huber, and Lenz 2012) may be driven by the sample more than the instructions.

The Effect of Cheating on the Validity of Political Knowledge Measures

Previous research provides some evidence that cheating harms the criterion validity of knowledge measures in student and adult samples (Clifford and Jerit 2014; Jensen and Thomsen 2014).¹⁶ Here, we investigate the related question of whether cheating interventions *improve* the predictive validity of these measures. In the survey conducted by YouGov, we included several open-ended thought-listing questions. Based upon previous research, political knowledge should be positively related to the number of considerations a person is able to list (e.g., Zaller [1992]; Zaller and Feldman [1992]). Thus, if the Commitment language improves the validity of knowledge measures, there should be a positive and significant interaction between the treatment and political knowledge in a model predicting the number of thoughts listed. Analysis of predictive validity in the YouGov sample provides a difficult test for the Commitment mechanism, as the effects on self-reported cheating and knowledge scores were smaller than in student samples.

At the beginning of the YouGov study, respondents reported their opinion to a closed-ended question on gun control and then were asked to list the thoughts that came to mind as they answered the question. Following the thought-listing, respondents were asked what factors they thought were responsible for the mass shootings that have occurred in the United States in recent years.

16. We also have evidence from Student 1, which consisted of a two-wave panel in which students answered political knowledge questions in a lab and then answered different knowledge items in a follow-up survey online. The correlation between attention to politics and political knowledge is significantly lower in the online wave, relative to the lab wave (see online appendix C).

There was a parallel set of questions on the topic of health care reform.¹⁷ Once again, we expected that political knowledge would be associated with listing a larger number of considerations for both the thought-listing and attribution items (Zaller 1992; Gomez and Wilson 2001). A coder who was blind to our expectations and treatment assignment coded the number of distinct considerations listed by respondents in each of the four questions (Krippendorff's $\alpha = .85$).¹⁸ The four items were summed to form an index of thought-listing (Cronbach's $\alpha = .85$).

We use a negative binomial regression model to predict the number of thoughts as a function of political knowledge scores, the treatment conditions, and interactions between knowledge and each condition.¹⁹ Full model results are shown in table A4 in online appendix B, but political knowledge is a strong predictor of holding more considerations ($p < .001$). As expected, however, there is a positive interaction between political knowledge and the Commitment condition, indicating that political knowledge is a stronger predictor of thought-listing in the Commitment condition than in the control condition ($p = .058$). In the Timer condition, the interaction between political knowledge is negative and statistically insignificant ($p = .86$). Moreover, a direct comparison between the Commitment and Timer conditions shows that the political knowledge scale has significantly higher predictive validity in the Commitment condition than in the Timer condition ($p < .05$). Overall, the results suggest that the Commitment condition, but not the Timer condition, improves the predictive validity of the political knowledge scale.

Are Cheating Interventions Off-Putting to Respondents?

Although interventions designed to reduce cheating are effective in most cases, researchers may worry that the treatments are off-putting. Timers may increase the stress on respondents while the commitment language may be interpreted as an accusation of dishonesty. We explored these concerns with two measures from the YouGov study. At the end of that survey, respondents were asked to rate how interested they would be “in taking another survey like this one” on a five-point scale. We reasoned that respondents would be less interested if they found the treatments distasteful. In the control

17. The fixed choice questions read as follows: “Do you favor or oppose stricter gun control laws?” and “Do you support or oppose the health care law passed by the President and Congress in 2010?” See appendix D for question wording of the closed- and open-ended items.

18. The reported Krippendorff's alpha comes from an intercoder reliability analysis based on a random subset of the data ($n = 60$).

19. We exclude inattentive respondents because they are less likely to give reliable responses to the SDE battery and are less responsive to experimental manipulations (Berinsky, Margolis, and Sances 2014; see online appendix A for details).

condition, the average interest in taking a similar survey in the future was 4.08, roughly corresponding with the “very interested” response option. Survey interest did not significantly differ in the Timer or Commitment conditions ($ps > .50$), which is at odds with the claim that these interventions upset respondents.

Additionally, we included an open-ended item asking if there was anything respondents disliked about the survey. Overall, 19 percent of the sample reported disliking something about the survey. Among that group, however, only 34 respondents (3 percent of the total sample) made a statement related to political knowledge. Several ($n = 19$) noted embarrassment over not being able to answer the questions; however, only three respondents referenced some aspect of the cheating manipulations. In the Timer condition, one person objected to not knowing when the 30 seconds was up and another complained that he could not take a break between questions (due to the automatic advance on the knowledge questions in that condition). The only dislike noted among respondents in the Commitment condition was “The quiz part without looking anything up.” Overall, respondents were more likely to dislike some feature of the knowledge questions in the control group (4.2 percent) compared to either the Timer (2.7 percent; $p = .28$) or the Commitment conditions (3.2 percent; $p = .47$).²⁰ Moreover, several people remarked that they were tempted to look up answers, or that they followed instructions in spite of their usual habit of searching for the answers. Comments such as these underscore the prevalence of cheating in online questionnaires and the importance of identifying this behavior in survey data.

Conclusion

Collectively, these studies demonstrate that respondents frequently look up the answers to political knowledge questions in online surveys and provide suggestive evidence that this behavior varies by sample. On MTurk, where respondents have a financial incentive to finish quickly, cheating rates seem relatively low. Among students, who face lower opportunity costs and may have a greater desire to self-enhance on political knowledge, cheating behavior is common. In this latter population, cheating rates are high enough to distort levels of political knowledge. Finally, in national samples, self-reported cheating rates were moderate, though we still found that a cheating intervention (i.e., the Commitment language) improved the validity of a knowledge scale.

Researchers and survey organizations have adopted practices such as time limits and instructions to discourage cheating, but have provided little evidence for their efficacy. According to our analyses, a direct request asking

20. Overall, 56 percent of respondents explicitly stated that there was nothing they disliked about the survey, and this figure did not vary across conditions ($p = .59$).

people not to cheat was the least effective of the techniques we examined, even when paired with forgiving language intended to reduce social desirability pressures. Timers were more effective, but they may interfere with honest subjects' ability to answer the questions. As a result, we recommend the commitment item—that is, asking respondents whether they are willing to answer knowledge questions without help, and requiring a yes or no answer. This technique yielded the lowest levels of cheating and the highest predictive validity—all without the potential disadvantages of timers or other unexpected disadvantages regarding respondent goodwill.

We also provide evidence for the utility of simple self-reports for assessing the prevalence of cheating behavior. Though the measure is no doubt a low estimate, it generally corresponded with two other indicators of cheating behavior: knowledge scores and reaction time. Thus, when measuring political knowledge in online surveys, we encourage researchers to employ a self-report item as an *aggregate* measure of cheating behavior. Such information will help diagnose potential problems and enhance our understanding of cheating behavior.²¹

While we have collected and analyzed a large amount of data, questions remain about the motives for cheating. We find some evidence that the motivation for this behavior varies by sample and by levels of self-deceptive enhancement, but it is unclear whether cheating varies by the type of knowledge question. Comparing the CCES and YouGov data, respondents are equally likely to cheat on political knowledge and vocabulary tests. Furthermore, among all of our political knowledge questions, item difficulty alone does not seem to drive cheating behavior (see note 8). Further research is needed to uncover how context, personality, and question-level factors interact to motivate cheating behavior.

The rise of online surveys has made data collection faster and more convenient than ever, particularly in combination with crowd-sourcing platforms and student samples. Although researchers have explored the generalizability of findings across different types of samples (Krupnikov and Levine 2014; Mullinix et al. 2015), there has been less attention to how differences in sample characteristics affect data quality, particularly when research is conducted online (Weinberg, Freese, and McElhattan [2014] is a notable exception). Our analyses suggest that financial incentives and personal motivations can have dramatic effects on data quality with regard to political knowledge. As researchers turn toward online research, and adopt new approaches to recruiting participants for research, it is increasingly important to understand how data quality is affected by the unique characteristics of different subject populations.

21. At present, researchers do not have a definitive way to identify cheating. The willingness to admit to this behavior may be correlated with characteristics that are related to knowledge (e.g., engagement, interest). Although we believe a self-reported cheating question can be useful in gauging the extent of this behavior in the aggregate, more caution is warranted when using this item as a measure of cheating at the individual level.

Appendix A. Details on Samples

Student Sample 1

A total of 127 subjects were recruited from political science courses at Stony Brook University for a panel study. The first wave of the study was conducted in the lab and included measures of political knowledge, attention to political news, and content for an unrelated study. Approximately one week later, subjects were recontacted to fill out an online survey, consisting of measures of political knowledge and attention to the news, along with unrelated content. Subjects were compensated with extra credit for completion of the first wave of the survey, and participants who completed the second wave were eligible for a cash prize (\$250) in a drawing. Eighty-four subjects completed the survey, for an attrition rate of 34 percent. Attrition is calculated as the percentage of respondents who drop out between the first and second time points. In Student 1, attrition is calculated as follows: $[(127 - 84) / 127] * 100$.

Student Sample 2

A total of 845 subjects were recruited for an online survey from required introductory political science courses at the University of Houston during the fall of 2014. Subjects were compensated with extra credit for completion of the survey. The survey was approximately 30 minutes long, with political knowledge questions coming near the end of the survey.

Student Sample 3

A total of 271 subjects were recruited for an online survey from required introductory political science courses at the University of Houston during the spring of 2015. Subjects were compensated with extra credit for completion of the survey. This survey was the second wave of a panel. The first wave ($N = 370$) measured demographics and material for an unrelated study approximately four weeks earlier. The attrition rate was 27 percent.

Student Sample 4

A total of 66 students were recruited from political science courses at Stony Brook University during the spring of 2015. This survey was the second wave of a panel study (first wave $N = 154$) focusing on an unrelated topic. The attrition rate was 57 percent.

Campus Staff Sample

A convenience sample of 59 campus staff was recruited from Stony Brook University during the spring of 2015. The sample was recruited through an IRB-approved broadcast e-mail to campus staff members in the College of Arts and Sciences, and respondents were paid \$25 for their participation. The data reported in this study were the second wave of a panel study (first wave $N = 81$) focusing on an unrelated topic. The attrition rate was 27 percent.

CCES 2014

A total of 1,000 respondents completed this survey as part of the Duke University pre-election wave of the 2014 Cooperative Congressional Election Study administered by YouGov/Polimetrix. Individuals are recruited into the YouGov/Polimetrix Internet panel using targeted online advertisements, and the CCES data are constructed from the thousands of people who have consented to be a part of this panel. More specifically, YouGov/Polimetrix selects a random sample of respondents from the US Census Bureau's American Community Survey (ACS). Each respondent from the YouGov-drawn ACS sample is matched with a YouGov panelist using a weighted absolute distance measure on four Census variables (age, race, gender, and education). More information on the survey methodology employed can be found at <http://projects.iq.harvard.edu/cces>. The participation rate for the full 2014 CCES was 22.7 percent, based upon the number of invitations that were sent and completions that were recorded.

YouGov

A total of 1,000 respondents were recruited by YouGov during December 2015 using the process described above. YouGov calculated a participation rate of 29.5 percent.

MTurk Sample 1

This convenience sample was collected in the spring of 2014 for an unrelated study. We recruited 500 respondents on an opt-in basis from Amazon's Mechanical Turk crowdsourcing platform. We restricted the posting to residents of the United States who were at least 18 years of age and had at least a 97 percent approval rate. The survey was advertised as a "short survey on political attitudes," and respondents were paid \$0.60 for completing the survey. All respondents who completed the survey were compensated and included in the analysis. Political knowledge was measured at the end of the 5–10-minute survey.

MTurk Sample 2

This convenience sample was collected in the summer of 2014 for an unrelated study. We recruited 505 respondents on an opt-in basis from Amazon's Mechanical Turk crowdsourcing platform. We restricted the posting to residents of the United States who were at least 18 years of age and had at least a 97 percent approval rate. The survey was advertised as a study about "political beliefs and attitudes," and respondents were paid \$2.00 for completing the survey. All respondents who completed the survey were compensated and included in the analysis. Political knowledge was measured at the end of the 30-minute survey.

MTurk Sample 3

This convenience sample was collected in the spring of 2015 for an unrelated study. We recruited 505 respondents on an opt-in basis from Amazon's Mechanical Turk crowdsourcing platform. We restricted the posting to

residents of the United States who were at least 18 years of age and had at least a 95 percent approval rate. The survey was advertised as a “short political survey,” and respondents were paid \$0.50 for completing the survey. All respondents who completed the survey were compensated and included in the analysis. Political knowledge was measured at the end of the 10-minute survey.

Appendix B. Knowledge Instructions and Treatment Language

All samples received the same control condition instructions for the knowledge section (shown below). Three samples that did not contain experimental manipulations used slightly different instructions, which are shown below.

Control

Now we have a set of questions concerning various public figures. We want to see how much information about them gets out to the public from television, newspapers, and the like.

Control – Student Sample 1, Wave 1 (Lab)

Now moving on to a different topic....next you will see some questions about politics and the economy.

Control – MTurk Sample 1

Now we'd like to ask you some factual questions about politics.

Control – MTurk Sample 2

Now we'd like to ask you some factual questions about politics. Many people have trouble answering questions like these, so if you can't think of the answer, select “don't know.”

Direct Request

Now we have a set of questions concerning various public figures. We want to see how much information about them gets out to the public from television, newspapers, and the like. Please give your best guess and do NOT use outside sources like the Internet to search for the correct answer.

Timer

Now we have a set of questions concerning various political issues. We want to see how much information about them gets out to the public from television, newspapers, and the like. Please do NOT use outside sources like the Internet to search for the correct answer. You will have 30 seconds to answer each question.

Commitment

Now we have a set of questions concerning various political issues. We want to see how much information about them gets out to the public from television,

newspapers, and the like. It is important to us that you do NOT use outside sources like the Internet to search for the correct answer. Will you answer the following questions without help from outside sources?

- Yes
- No

Forgiving

Now we have a set of questions concerning various political issues. We want to see how much information about them gets out to the public from television, newspapers, and the like. Many people will know the answers but won't be able to remember them at the moment. However, it is essential that we only measure what people are able to recall on their own. Thus, it is important to us that you do NOT use outside sources like the Internet to search for the correct answer.

Self-Reported Cheating Measure

Many people struggle to remember facts, even when they know them, and so they get help remembering. When you were answering the factual knowledge questions, did you get help from any other source, such as the internet or another person? (Please be honest, this is for statistical purposes only.)

Self-Reported Cheating Measure – CCES

Many people may know the meaning of these words but struggle to remember them, so they get help remembering. When you were answering these questions, did you get help remembering from any sources, such as the internet or another person?

Appendix C. Political Knowledge Scales

As noted in the text, each battery of political-knowledge questions appeared in multiple choice format. Answer categories below are presented horizontally for the sake of space.

Student 1, Wave 1 (Lab)

1. Do you happen to know the name of the current Speaker of the US House of Representatives? Is it...John Boehner; Nancy Pelosi; Newt Gingrich; Mitch McConnell
2. What is the name of the Chair of the Federal Reserve Board? Janet Yellen; Ben Bernanke; Kirsten Gillibrand; Hillary Clinton
3. What is the current unemployment rate in this country? Is it...3.7; 6.7; 10.7; 15.7
4. On which of the following does the US federal government currently spend the least? Foreign aid; Medicare; National defense; Social Security

Student 1, Wave 2 (Online)

1. For how many years is a United States Senator elected – that is, how many years are there in one full term of office for a US Senator? 2; 4; 6; 8
2. What is Medicare? A program run by the US federal government to pay for old people’s health care; A program run by state governments to provide health care to poor people; A private health insurance plan sold to individuals in all 50 states; A private, non-profit organization that runs free health clinics
3. Do you happen to know if children born to illegal immigrants in the US are automatically US citizens, or are they NOT automatically US citizens? Yes, automatically become citizens; No, do NOT automatically become citizens
4. Who is this? <image of John Boehner>John Boehner; Mitch McConnell; Harry Reid; Eric Cantor
5. What currency does this symbol represent? <image of Euro>Euro; Yen; Pound; Dollar

Student 2

1. Do you happen to know the name of the current Speaker of the U.S. House of Representatives? Is it...John Boehner; Nancy Pelosi; Harry Reid; Mitch McConnell
2. What is the name of the Chair of the Federal Reserve Board? Janet Yellen; Ben Bernanke; Kirsten Gillibrand; Eric Cantor
3. What is the current unemployment rate in this country? Is it...3.9; 5.9; 10.9; 15.9
4. On which of the following does the US federal government currently spend the least? Foreign aid; Medicare; National defense; Social Security
5. As far as you know, which foreign country holds the most US government debt? China; Japan; Saudi Arabia; Canada
6. How much of a majority is required for the US Senate and US House to override a presidential veto? One-half; Two-thirds; Three-fourths; Three-fifths
7. Do you happen to know what job or political office is now held by Ban Ki-Moon? Secretary of the US Department of Transportation; Prime Minister of Japan; Secretary of the United Nations; United States Ambassador to South Korea
8. In a recent address before the nation, President Obama outlined a four-point plan for addressing the Islamic State. Which of the following is *not* part of his plan? Humanitarian assistance to displaced civilians; Airstrikes against Islamic State targets; Training Iraq national guard units; Deployment of US ground combat units

Student 3

1. Do you happen to know if the share of Americans living at or below the federal poverty line is currently closer to: 5%; 15%; 25%; 35%
2. Who is the prime minister of Great Britain? Tony Abbott; David Cameron; Stephen Harper; Angela Merkel
3. Do you happen to know whether any of the following people have formally announced that they are running for president in 2016? Check all that apply. Hillary Clinton; Chris Christie; Rand Paul; Jeb Bush; None of the above
4. The United States recently announced that it would re-establish diplomatic relations with which of the following countries? Cuba; North Korea; Russia; Yemen
5. There are nine justices on the US Supreme Court. How many of them are women? 1; 2; 3; 4
6. What action did President Obama take earlier this year with regard to the Keystone Pipeline, which would carry oil from Canada to the Gulf of Mexico? He vetoed a bill that would have authorized its construction; He signed an executive order authorizing construction of the pipeline; He met with leaders of Canada and Mexico to discuss the project; He gave an address before the nation outlining the reasons why the United States should oppose the pipeline
7. What is the current unemployment rate in this country? Is it...3.5; 5.5; 10.5; 15.5
8. How much of a majority is required for the US Senate and US House to override a presidential veto? One-half; Two-thirds; Three-fourths; Three-fifths

Student 4, Campus Staff, MTurk 3 (Subjects randomly assigned to one of two knowledge batteries.)

Knowledge Battery Version 1

1. Do you happen to know if the share of Americans living at or below the federal poverty line is currently closer to: 5%; 15%; 25%; 35%
2. Who is the prime minister of Great Britain? Tony Abbott; David Cameron; Stephen Harper; Angela Merkel
3. How long is the term of office for a representative in the US House of Representatives? 2 years; 4 years; 6 years; 8 years
4. Do you happen to know if the national unemployment rate as reported by the government is currently closer to...1%; 6%; 10%; 15%

Knowledge Battery Version 2

1. Do you happen to know what the federal minimum wage is today? Is it: \$5.25; \$7.25; \$10.50; \$12.50
2. Who is the prime minister of Israel? Benjamin Netanyahu; Shimon Peres; Ariel Sharon; Hosni Mubarak

3. How long is the term of office for a senator in the US Senate? 2 years; 4 years; 6 years; 8 years
4. Do you happen to know if the national inflation rate as reported by the government is currently closer to...1%; 6%; 10%; 15%

CCES

We would like to know something about how people go about guessing words they do not know. Please choose the word that comes closest to the meaning of the word in capital letters. If you don't know, just use your best guess. [*Last two sentences of the instructions were repeated with each question.*]

ANIMOSITY: Hatred; Animation; Disobedience; Diversity; Friendship;
Don't know

CLOISTERED: Miniature; Bunched; Arched; Malady; Secluded;
Don't know

CAPRICE: Value; A star; Grimace; Whim; Inducement; Don't know

ACCUSTOM: Disappoint; Customary; Encounter; Get used to; Business;
Don't know

ALLUSION: Reference; Dream; Eulogy; Illusion; Aria; Don't know

MTurk 1

1. Do you happen to know the name of the current Speaker of the US House of Representatives? Is it...John Boehner; Nancy Pelosi; Newt Gingrich; Mitch McConnell
2. What is the name of the Chair of the Federal Reserve Board? Janet Yellen; Ben Bernanke; Kirsten Gillibrand; Hillary Clinton
3. What is the current unemployment rate in this country? Is it...3.3; 6.3; 10.3; 15.3
4. On which of the following does the US federal government currently spend the least? Foreign aid; Medicare; National defense; Social Security
5. Which political party has a majority in the US Senate? Democrats; Republicans
6. Which political party has a majority in the US House of Representatives? Democrats; Republicans

MTurk 2

1. Do you happen to know the name of the current Speaker of the US House of Representatives? Is it...John Boehner; Nancy Pelosi; Newt Gingrich; Mitch McConnell; Don't know
2. What is the name of the Chair of the Federal Reserve Board? Janet Yellen; Ben Bernanke; Kirsten Gillibrand; Hillary Clinton; Don't know
3. What is the current unemployment rate in this country? Is it currently closer to...3.3; 6.3; 10.3; 15.3; Don't know

4. What is ISIS? A Sunni extremist group seizing Syria and Iraq; A government program to monitor citizens; An energy company; A plan for stabilizing Iraq; Don't know
5. On which of the following does the US federal government currently spend the least? Foreign aid; Medicare; National defense; Social Security; Don't know
6. Which political party has a majority in the US Senate? Democrats; Republicans; Don't know
7. Which political party has a majority in the US House of Representatives? Democrats; Republicans; Don't know
8. Approximately how many people have died in the Syrian civil war? 150; 1,500; 15,000; 150,000; Don't know
9. What is the federal minimum wage for non-tipped employees? \$4.25; \$5.25; \$6.25; \$7.25; Don't know
10. What is the federal minimum wage for tipped employees? \$2.13; \$3.13; \$4.13; \$5.13; Don't know
11. Which of the following states has NOT legalized physician-assisted suicide? Oregon; Washington; Vermont; Nevada; Don't know
12. How many states currently have legal same-sex marriage? 5; 12; 19; 31; Don't know

YouGov

1. Do you happen to know the name of the current Speaker of the US House of Representatives? Is it...John Boehner; Kevin McCarthy; Paul Ryan; Mitch McConnell
2. The US and 11 Pacific Rim nations recently reached an agreement that would...Open up free trade between these nations; Increase immigration between these nations; Improve the quality of water in the Pacific Ocean; Reduce the emission of greenhouse gases
3. What is the name of the Chair of the Federal Reserve Board? Janet Yellen; Ben Bernanke; Kirsten Gillibrand; Elizabeth Warren
4. What is the current unemployment rate in this country? Is it closer to... 2%; 5%; 10%; 15%
5. How much of a majority is required for the US Senate and US House to override a presidential veto? One-half; Two-thirds; Three-fourths; Three-fifths
6. What does the term "Common Core" refer to? School curriculum standards for language and math; The military's code of conduct; A set of nutrition standards for school lunches; A plan to mandate English as the official language
7. What is the leading source of electricity in the United States? Is it... Renewable energy sources (wind, solar); Coal; Natural Gas; Nuclear power

8. Which party has the most seats in the US Senate? Is it the...Republican Party; Democratic Party

Appendix D. Criterion Validity Measures

BIDR-40; Paulhus 1991 (Student 4)

Items 1–20 measure self-deceptive enhancement, items 21–40 measure impression management. Items marked with “(R)” are reverse-scored (all even items).

How true are each of the following statements about yourself? Not true (1) – Somewhat true (4) – Very true (7)

1. My first impressions of people usually turn out to be right.
2. It would be hard for me to break any of my bad habits. (R)
3. I don't care to know what other people really think of me.
4. I have not always been honest with myself. (R)
5. I always know why I like things.
6. When my emotions are aroused, it biases my thinking. (R)
7. Once I've made up my mind, other people can seldom change my opinion.
8. I am not a safe driver when I exceed the speed limit. (R)
9. I am fully in control of my own fate.
10. It's hard for me to shut off a disturbing thought. (R)
11. I never regret my decisions.
12. I sometimes lose out on things because I can't make up my mind soon enough. (R)
13. The reason I vote is because my vote can make a difference.
14. My parents were not always fair when they punished me. (R)
15. I am a completely rational person.
16. I rarely appreciate criticism. (R)
17. I am very confident of my judgments.
18. I have sometimes doubted my ability as a lover. (R)
19. It's all right with me if some people happen to dislike me.
20. I don't always know the reasons why I do the things I do. (R)
21. I sometimes tell lies if I have to.
22. I never cover up my mistakes. (R)
23. There have been occasions when I have taken advantage of someone.
24. I never swear. (R)
25. I sometimes try to get even rather than forgive and forget.
26. I always obey laws, even if I'm unlikely to get caught. (R)
27. I have said something bad about a friend behind his or her back.
28. When I hear people talking privately, I avoid listening. (R)
29. I have received too much change from a salesperson without telling him or her.

30. I always declare everything at customs. (R)
31. When I was young I sometimes stole things.
32. I have never dropped litter on the street. (R)
33. I sometimes drive faster than the speed limit.
34. I never read sexy books or magazines. (R)
35. I have done things that I don't tell other people about.
36. I never take things that don't belong to me. (R)
37. I have taken sick-leave from work or school even though I wasn't really sick.
38. I have never damaged a library book or store merchandise without reporting it. (R)
39. I have some pretty awful habits.
40. I don't gossip about other people's business. (R)

BIDR 6 short form; Bobbio and Manganelli 2011 (YouGov)

How true are each of the following statements about yourself? Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree [Randomize order.]

1. My first impressions of people usually turn out to be right.
2. I always know why I like things.
3. Once I've made up my mind, other people can seldom change my opinion.
4. I am fully in control of my own fate.
5. I never regret my decisions.
6. I am a completely rational person.
7. I am very confident of my judgments.
8. It's all right with me if some people happen to dislike me.
9. I sometimes tell lies if I have to.
10. There have been occasions when I have taken advantage of someone.
11. I always obey laws, even if I'm unlikely to get caught
12. I have said something bad about a friend behind his or her back.
13. I have never dropped litter on the street.
14. I have done things that I don't tell other people about.
15. I have taken sick-leave from work or school even though I wasn't really sick.
16. I have some pretty awful habits.

Thought-listing and attribution questions

1. Do you favor or oppose stricter gun control laws? Favor; Oppose; Unsure
2. Still thinking about the question you just answered, what thoughts came to mind while you were answering that question? Please try to list everything that came to mind. (open-ended)
3. Thinking about the mass shootings that have occurred in the US in the last few years, what factors do you think are responsible for the shootings? (open-ended)

4. Do you support or oppose the health care law passed by the President and Congress in 2010? Support; Oppose; Unsure
5. Still thinking about the question you just answered, what thoughts came to mind while you were answering that question? Please try to list everything that came to mind. (open-ended)
6. For decades, experts have observed that the United States spends far more per person on health care than any other country. However, the US falls behind on most measures of health care outcomes, such as life expectancy. What factors do you think are responsible for the state of our health care system? (open-ended)

Supplementary Data

Supplementary data are freely available online at <http://poq.oxfordjournals.org/>.

References

- Ansolabehere, Stephen, and Brian Schaffner. 2014. "Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison." *Political Analysis* 22:285–303.
- . 2015. "Distractions: The Incidence and Consequences of Interruptions for Survey Respondents." *Journal of Survey Statistics and Methodology* 3:1–24.
- Barabas, Jason, Jennifer Jerit, William Pollock, and Carlisle Rainey. 2015. "The Question(s) of Political Knowledge." *American Political Science Review* 108:840–55.
- Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Using Mechanical Turk as a Subject Recruitment Tool for Experimental Research." *Political Analysis* 20:351–68.
- Berinsky, Adam J., Michele Margolis, and Michael Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Internet Surveys." *American Journal of Political Science* 58:739–53.
- Bobbio, Andrea, and Anna Maria Manganelli. 2011. "Measuring Social Desirability Responding: A Short Version of Paulhus' BIDR 6." *Testing, Psychometrics, Methodology* 18:117–35.
- Booth-Kewley, Stephanie, Gerald E. Larson, and Dina K. Miyoshi. 2007. "Social Desirability Effects on Computerized and Paper-and-Pencil Questionnaires." *Computers in Human Behavior* 23:463–77.
- Brown, Jonathan D. 2012. "Understanding the Better than Average Effect: Motives (Still) Matter." *Personality and Social Psychology Bulletin* 38:209–19.
- Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory A. Huber. 2015. "Partisan Bias in Factual Beliefs about the Economy." *Quarterly Journal of Political Science* 10:519–78.
- Burnett, Craig, M. 2015. "Exploring the Difference in Performance on Knowledge by Participants between Online and In-Person Modes." Working Paper, University of North Carolina–Wilmington.
- Cannell, Charles F., Peter V. Miller, and Lois Oskenbeurg. 1981. "Research on Interviewing Techniques." *Sociological Methodology* 12:389–437.
- Chandler, Jesse, Pam Mueller, and Gabriele Paolacci. 2014. "Nonnaiveté among Amazon Mechanical Turk Workers: Consequences and Solutions for Behavioral Researchers." *Behavioral Research* 46:112–30.
- Cialdini, Robert B., John T. Cacioppo, Rodney Bassett, and John A. Miller. 1978. "Low Ball Procedure for Producing Compliance: Commitment Then Cost." *Journal of Personality and Social Psychology* 36:463–76.

- Clifford, Scott, and Jennifer Jerit. 2014. "Is There a Cost to Convenience? An Experimental Comparison of Data Quality in Laboratory and Online Studies." *Journal of Experimental Political Science* 1:120–31.
- . 2015. "Do Attempts to Improve Respondent Attention Increase Social Desirability Bias?" *Public Opinion Quarterly* 79:790–802.
- Delli Carpini, Michael X., and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters*. New Haven, CT: Yale University Press.
- Duff, Brian, Michael J. Hanmer, Won-Ho Park, and Ismail K. White. 2007. "Good Excuses: Understanding Who Votes with an Improved Turnout Question." *Public Opinion Quarterly* 71:67–90.
- Fisher, Matthew, Mariel K. Goddu, and Frank C. Keil. 2015. "Searching for Explanations: How the Internet Inflates Estimates of Internal Knowledge." *Journal of Experimental Psychology: General* 144:674–87.
- Fricker, Scott, Mirta Galesic, Roger Tourangeau, and Ting Yan. 2005. "An Experimental Comparison of Web and Telephone Surveys." *Public Opinion Quarterly* 69:370–92.
- Gomez, Brad T., and J. Matthew Wilson. 2001. "Political Sophistication and Economic Voting in the American Electorate: A Theory of Heterogeneous Attribution." *American Journal of Political Science* 45:899–914.
- Gooch, Andrew, and Lynn Vavreck. 2015. "How Face-to-Face Interviews and Cognitive Skill Affect Item Non-Response: A Randomized Experiment Assigning Mode of Interview." Working Paper, UCLA.
- Goodman, Joseph K., Cynthia E. Cryder, and Amar Cheema. 2013. "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples." *Journal of Behavioral Decision Making* 26:213–24.
- Hauser, David J., and Norbert Schwarz. Forthcoming. "Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks Than Do Subject Pool Participants." *Behavior Research Methods*.
- Holbrook, Allyson L., and Jon A. Krosnick. 2010. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." *Public Opinion Quarterly* 74:37–67.
- Jacoby, William G. 2006. "Value Choices and American Public Opinion." *American Journal of Political Science* 50:706–23.
- Jensen, Carsten, and Jens Peter Frølund Thomsen. 2014. "Self-Reported Cheating in Web Surveys on Political Knowledge." *Quality and Quantity* 48:3343–54.
- Kam, Cindy. 2005. "Who Toes the Party Line? Cues, Values, and Individual Differences." *Political Behavior* 27:163–82.
- Kam, Cindy D., Jennifer R. Wilking, and Elizabeth J. Zechmeister. 2007. "Beyond the 'Narrow Database': Another Convenience Sample for Experimental Research." *Political Behavior* 29:415–40.
- Kleinberg, Mona and Richard R. Lau. 2014. "Cognitive Misers 2.0: Does Internet Access and the 24/7 Availability of Political Knowledge Create a Disincentive to Committing Information to Memory?" Paper presented at the Annual Meeting of the American Political Science Association, Washington, DC, USA.
- Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213–36.
- Krupnikov, Yanna, and Adam Seth Levine. 2014. "Cross-Sample Comparisons and External Validity." *Journal of Experimental Political Science* 1:59–80.
- Krupnikov, Yanna, Spencer Piston, and Nichole Bauer. 2016. "Saving Face: Identifying Voter Responses to Black Candidates and Female Candidates." *Political Psychology* 37:253–73.
- Lau, Richard R., and David P. Redlawsk. 2001. "Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making." *American Journal of Political Science* 45:951–71.
- Luskin, Robert C., and John G. Bullock. 2011. "'Don't Know' Means 'Don't Know': DK Responses and the Public's Level of Political Knowledge." *Journal of Politics* 73:547–57.

- Miller, Melissa K., and Shannon K. Orr. 2008. "Experimenting with a 'Third Way' in Political Knowledge Estimation." *Public Opinion Quarterly* 72:768–80.
- Mondak, Jeffery J. 2001. "Developing Valid Knowledge Scales." *American Journal of Political Science* 45:224–38.
- Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman, and Jeremy Freese. 2015. "The Generalizability of Survey Experiments." *Journal of Experimental Political Science* 2:109–38.
- Munzert, Simon, and Peter Selb. Forthcoming. "Measuring Political Knowledge in Web-Based Surveys: An Experimental Validation of Visual Versus Verbal Instruments." *Social Science Computer Review*.
- Paulhus, Delroy L. 1991. "Measurement and Control of Response Bias." In *Measures of Personality and Social Psychological Attitudes*, eds. John P. Robinson, Phillip R. Shaver, and Lawrence S. Wrightsman, pp. 17–59. San Diego, CA: Academic Press.
- Paulhus, Delroy L., P. D. Harms, M. Nadine Bruce, and Daria C. Lysy. 2003. "The Over-Claiming Technique: Measuring Self-Enhancement Independent of Ability." *Journal of Personality and Social Psychology* 84:890–904.
- Paulhus, Delroy L., and Oliver P. John. 1998. "Egoistic and Moralistic Bias in Self-Perceptions: The Interplay of Self-Deceptive Styles with Basic Traits and Motives." *Journal of Personality* 66:1024–60.
- Peter, Jochen, and Patti M. Valkenburg. 2011. "The Impact of 'Forgiving' Introductions on the Reporting of Sensitive Behavior in Surveys: The Role of Social Desirability Response Style and Developmental Status." *Public Opinion Quarterly* 75:779–87.
- Prior, Markus, Gaurav Sood, and Kabir Khanna. 2015. "You Cannot Be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions." *Quarterly Journal of Political Science* 10:489–518.
- Shulman, Hillary C., and Franklin J. Boster. 2014. "Effect of Test-Taking Venue and Response Format on Political Knowledge Tests." *Communication Methods and Measures* 8:177–89.
- Shulman, Hillary C., Franklin J. Boster, Christopher Carpenter, and Allison Shaw. 2014. "Why Do Students Completing a Political Knowledge Test Score Higher Online Than in the Classroom? A Series of Studies." Working Paper, North Central College.
- Sparrow, Betsy, Jenny Liu, and Daniel M. Wegner. 2011. "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips." *Science* 333:776–78.
- Stöber, Joachim, Dorothea E. Dette, and Jochen Musch. 2002. "Comparing Continuous and Dichotomous Scoring of the Balanced Inventory of Desirable Responding." *Journal of Personality Assessment* 78:370–89.
- Strabac, Zan, and Toril Aalberg. 2011. "Measuring Political Knowledge in Telephone and Web Surveys: A Cross-National Comparison." *Social Science Computer Review* 29:175–92.
- Taber, Charles, and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50:755–69.
- Ward, Adrian F. 2013. "Supernormal: How the Internet Is Changing Our Memories and Our Minds." *Psychological Inquiry* 24:341–48.
- Weinberg, Jill D., Jeremy Freese, and David McElhattan. 2014. "Comparing Data Characteristics and Results of an Online Factorial Survey between a Population-Based and a Crowdsourced-Recruited Sample." *Sociological Science* 1:292–310.
- Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.
- Zaller, John, and Stanley Feldman. 1992. "A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences." *American Journal of Political Science* 36(3):579.