A Measurement Gap?
The Effect of Survey Instrument and Scoring on the Partisan Knowledge Gap

Lucas Shen* Gaurav Sood† Daniel Weitzel‡

September 25, 2025

Abstract

Research suggests that partisan gaps in political knowledge with partisan implications are wide and widespread in the US. Using a series of experiments, we estimate the extent to which the partisan gaps in commercial surveys reflect differences in confidently held beliefs rather than motivated guessing. Knowledge items on commercial surveys often have guessing-encouraging features. Removing such features yields scales with greater reliability and higher criterion validity. More substantively, partisan gaps on scales without these "inflationary" features are roughly 40% smaller. Thus, contrary to Prior, Sood and Khanna (2015), who find that the upward bias is explained by the knowledgeable deliberately marking the wrong answer (partisan cheerleading), our data suggest, in line with Bullock et al. (2015), that partisan gaps on commercial surveys in the United States are strongly upwardly biased by motivated guessing by the ignorant. Relatedly, we also find that partisans know less than what the toplines of commercial polls suggest.

Keywords: Political Knowledge; Partisan Gap; Motivated Skepticism

*Senior Scientist, Institute for Human Development and Potential, Agency for Science, Technology and Research, Singapore, lucas@lucasshen.com

†Independent researcher, Seattle, WA, USA gsood07@gmail.com

‡Assistant Professor, Colorado State University, Fort Collins, CO, USA, daniel.weitzel@colostate.edu

Wide and widespread partisan gaps in political knowledge challenge the idea that citizens can hold representatives accountable (Hochschild and Einstein 2015; Bailey 2021). Hence, the alarm over research that finds as much (Bartels 2002; Campbell et al. 1980; Jerit and Barabas 2012) (though see Roush and Sood (2023)). However, an emerging line of research argues that a large fraction of the partisan knowledge gap is an artifact of the survey response process (Bullock et al. 2015; Huber and Yair 2018; Prior, Sood and Khanna 2015; Graham and Yair 2023) (though see Berinsky (2017), Peterson and Iyengar (2021), and Malka and Adelman (2022)). In this paper, we extend this investigation.

Our starting point is commercial polls in the US. In particular, we examine how common features of knowledge items on commercial polls, e.g., presenting social proof about the less socially desirable option, including information about the topic, not including a 'Don't know' option, providing a partisan cue in the question stem, etc., affect estimates of partisan gaps in knowledge of facts with political implications. Removing these guessing-encouraging features yields knowledge scales with greater reliability and higher criterion validity. When we measure partisan gaps using these scales, they are 14 percentage points smaller (Figure 1). To further ablate response biases, we use an instrument and scoring scheme inspired by Pasek, Sood and Krosnick (2015) and Graham (2021) that considers respondents' confidence in their answers. Using the scoring scheme that credits only confidently held beliefs as knowledge, we find that partisan gaps are 50% smaller still (see Figure 2).

Our results contribute to a growing literature that suggests that a large fraction of partisan gaps are artifacts of survey design. The results further clarify the source of bias in estimates of partisan gaps. While some previous research shows that the partisan gap is due to partisan cheerleading—deliberate selection of congenial incorrect answers by the knowledgeable (Prior, Sood and Khanna 2015), our data suggests that the bias in the estimate of the partisan gap is primarily a result of partisan guessing by the ignorant (see also Bullock et al. (2015) who reach similar conclusions).

Our results suggest that some concerns about democratic health are overstated and some are underappreciated. Reducing guessing-related error reveals that partisan gaps on partisan knowledge items are not as wide, but also that partisans know less about politics than what the toplines of commercial polls suggest.

1 Theory and Motivation

> "Has unemployment increased, decreased, or stayed the same since President Joe Biden took office in 2021?"

How knowledge about this fact and other such politically consequential facts is distributed across the population is relevant to the health of a democracy. If there are wide gaps in partisans' knowledge of politically relevant facts, citizens' ability to hold politicians accountable might be limited.

Concerningly, a large body of research finds that partisan gaps in political knowledge with partisan implications are both wide and widespread (Bartels 2002; Jerit and Barabas 2012; Laloggia 2018; Lodge and Taber 2013) (though see Roush and Sood (2023)). Some recent research, however, shows that a large part of the partisan gaps stems from partisan responding rather than differences in what partisans know to be true about the world (Bullock et al. 2015; Prior, Sood and Khanna 2015; Huber and Yair 2018; Graham and Yair 2023) (though see Peterson and Iyengar (2021), Berinsky (2017), and Malka and Adelman (2022)).

More generally, researchers argue that partisan gaps in political knowledge with partisan implications are inflated by:

1. **Partisan Cheerleading:** Partisans who know the right uncongenial answer deliberately pick the wrong partisan congenial answer to register their support for their party or to influence the survey results (Prior, Sood and Khanna 2015).

2. **Partisan Guessing:** Partisans who don't know the answer offer substantive responses congenial to their party (Bullock et al. 2015; Graham and Yair 2023). For instance, when asked about what happened to the unemployment rate during the Biden administration, Republicans ignorant about the unemployment rate may still respond that unemployment rose during the Biden administration because they viscerally dislike Democrats or because they believe that Democrats mishandle the economy.

In this paper, we interrogate the latter explanation in the context of commercial polls. An analysis of 180 media polls by Luskin et al. (2018) found that guessing-encouraging features were exceedingly common. For instance, less than 9% of the surveys offered an explicit 'Don't Know' or 'Not Sure' option, which causes a positive bias in the estimates of political knowledge (Luskin and Bullock 2011; Cor and Sood 2016). And about half of the items offered only two choices, a design choice that dramatically inflates estimates of knowledge (Bullock and Rader 2022; Fortin-Rittberger 2016). An overwhelming majority of the items (168) also included wording encouraging guessing by framing the factual question as a 'matter of opinion.' They also found that the scoring rules used by analysts treated all correct responses–even when the respondent is unconfident about their answer—as evidence of knowledge. Doing so conflates guesses and on-the-spot inferences with knowledge (Pasek, Sood and Krosnick 2015). Other research finds that the partisan context of the survey cues directional motivations and increases the partisan gap (Prior, Sood and Khanna 2015; Bailey 2021).

1.1 Guessing Vs. Diffidence

Survey measurement of political knowledge using closed-ended items faces two competing challenges: respondents marking a substantive option when they don't know the correct answer, and respondents not marking the correct answer when they do know. If you take

each incorrect answer in a multiple choice item as evidence of a random guess by a respondent who doesn't know the answer, the percentage of correct answers that stem from guessing on political knowledge questions is 22% (Luskin and Bullock 2011).[1] Guessing-related error doesn't just distort the description of how much people know but also abrades correlations with the latent construct. The rationale is as follows: the less you know, the more items you must guess on, and the more the positive error in your score absent guessing adjustment in scoring (Cor and Sood 2016). In effect, the guessing-related error is negatively correlated with the latent construct (knowledge). These concerns motivate many researchers to look for a solution. Some researchers argue that offering a 'Don't Know' option is a reasonable solution. They point out that 'Don't Know' responses hide very little knowledge (Sturgis, Allum and Smith 2008; Luskin and Bullock 2011; Jessee 2017; Sanchez and Morchio 1992), the tendency to mark don't know doesn't vary by gender (Ferrín, Fraile and García-Albacete 2017) or personality (Jessee 2017), and hence produces more descriptively and correlationally valid estimates (Luskin and Bullock 2011; Jessee 2017). Others, however, contend that the cure is worse than the problem. They implicitly argue that Don't Know responses hide a lot of knowledge and that the amount of knowledge that Don't Know responses hide varies by the type of person (Mondak 1999; Mondak and Anderson 2004; Dolan and Hansen 2020; Kraft and Dolan 2023).

Don't Know is but one way to reduce guessing-related errors. Other ways include removing social proof and neutral information from the question stem, using self-assessed confidence, and increasing the number and difficulty of options (Fortin-Rittberger 2016; Bullock and Rader 2022).

We hypothesize that guessing-encouraging features inflate partisan gaps and yield

[1]In the DK Discouraging condition, the percentage correct is 60.7. Correcting for guessing reduces the number to 47.1.

measures with lower correlational and descriptive validity.[2]

2 Empirical Strategy

To study the effect of "inflationary" features of survey and knowledge items on the partisan knowledge gap, we conduct a series of survey experiments that modify various guessing-encouraging features. To study the effect of taking respondents' confidence into account, we draft an instrument and scoring rule inspired by Pasek, Sood and Krosnick (2015), which uses self-assessed confidence to rescore the answers, taking only correct answers that respondents are confident about as evidence that the respondent knows the fact. We also analyze which item formats produce measures with greater reliability and higher criterion validity.

In all, we use data from four surveys. The results of these four surveys are presented as part of three studies:

- In Study 1, we use data from a survey experiment conducted on Amazon Mechanical Turk (MTurk) (*MTurk 1*) to examine how guessing-encouraging features affect the partisan gap.
- In Study 2, we use survey experiments conducted on a *YouGov* and a telephone survey (*Texas Lyceum*) to examine the effect of partisan cues on the partisan gap.
- Lastly, in Study 3, we use data from *MTurk 1* and another survey fielded on MTurk (*MTurk 2*) to study the impact of taking respondents' confidence in their answers into

[2]In Appendix SM 8, we analyze the impact of removing social proof, neutral information about the topic, etc., on the gender gap. In Appendix SM 10, using data from Bullock and Rader (2022), we assess the impact of increasing the difficulty and number of choices on the gender gap. The results are inconsistent, suggesting that the effect is unlikely to be large and consistent.

account on the partisan gap.

Before we proceed further, we would like to note that many of our questions are on topics on which people can be misinformed—know the wrong thing confidently. This includes partisan retrospection items like those used by Bartels (2002). However, on all of these 'misinformation' items, we can also ask how many people know the correct answer. Like Bartels (2002) and Prior, Sood and Khanna (2015)—and for much the same reasons—we are interested in measuring the partisan gap in knowledge, though we believe that it would also be useful to study partisan gaps in misinformation.

3 Study 1: The Effect of Guessing-Encouraging Features

The first study focuses on four survey design features that we suspect inflate the partisan gap. These features are:

1. the absence of a "Don't Know" (DK) option
2. including additional neutral information in the question stem
3. providing social proof for an answer
4. the absence of a guessing discouraging preamble

3.1 Research Design and Data

We conducted a survey experiment on MTurk in mid-2017 in which we randomly assigned 1,253 respondents to one of four conditions (see Table 1).[3] In each condition, respondents

[3]For generalizability of effects in studies conducted on MTurk, see (Mullinix et al. 2015; Coppock, Leeper and Mullinix 2018). Balance tests suggest that the randomization was successful (see Figures SM 1.1 to SM 1.4).

answered nine misinformation items, ranging from President Obama's citizenship to whether global warming is happening or not. (For exact question wording, see Appendix SM 3.) The conditions reduce the number of inflationary features from four to zero and are labeled with the abbreviated features.

The four conditions are:

1. **Condition 1 (NDK+SP+GE+NI):** The design includes four features that encourage guessing. It serves as our baseline condition. The items in this condition include all the common features of commercial polls. In this design, the 'Don't Know' option is never presented (NDK; prefix N indicates 'Not presented'), so the respondents must guess even if they don't know. The questions also include social proof about the incorrect answer (SP). By social proof, we mean information about what other people believe. Seeing that some people believe in an option can cause more people to select that option (see Cialdini 2009; Sherif 1935). For instance, on the question about where Mr. Obama was born, we add, "Some people believe Barack Obama was not born in the United States but was born in another country." In other cases, we provide some neutral information (NI) about the topic, like "According to the Constitution, American presidents must be natural-born citizens." Lastly, the preamble to the knowledge questions is neutral and doesn't discourage guessing or cheating (GE). (Please see Luskin and Bullock (2011) for data on how the DK neutral preamble has the same effect as a DK discouraging one.) The preamble simply reads: "Now here are some questions about what you may know about politics and public affairs..."

2. **Condition 2 (NDK+NSP+GE+NI):** By removing social proof (SP), we arrive at a very commonly used design in commercial polling. Like the baseline condition, the questions in this design do not feature a 'Don't Know' option (NDK) but include neutral information (NI) in the question stem.

3. **Condition 3 (DK+NSP+GD+NI):** The next design removes two inflationary features. First, the preamble now discourages blind guessing and cheating (GD). The preamble reassures respondents that it is okay not to know the answers to these questions, asks respondents to commit not to look up answers or ask anyone, and asks respondents to mark don't know when they don't know the answer. Second, the items now include a DK option (see, e.g., Luskin and Bullock 2011; Bullock et al. 2015). (We code DK the same as an incorrect answer.)

4. **Condition 4 (DK+NSP+GD+NNI):** Our final version offers respondents a DK option, discourages guessing and cheating (DG), and does not include neutral information (NI) or social proof (NSP) in the question stem. (We code DK the same as an incorrect answer.)

Table 1: Experimental Treatments

		Treatments				
Condition	Label	Don't Know	Social Proof	Guessing Encouraged	Neutral Information	Inflationary Features
1	NDK+SP+GE+NI	No	Yes	Yes	Yes	4
2	NDK+NSP+GE+NI	No	No	Yes	Yes	3
3	DK+NSP+GD+NI	Yes	No	No	Yes	1
4	DK+NSP+GD+NNI	Yes	No	No	No	0

3.2 Measures

Since our study was fielded in the United States, we measure partisanship using the conventional branched seven-point partisan self-identification scale. Respondents are first asked if they identify as Republicans, Democrats, or Independents. If respondents pick a party, they are asked about the strength of their attachment to the party. Independents are asked if they lean toward one party or the other. In our study, independents who lean toward one of the two major parties are coded as supporters of that party. A knowledge item is coded as congenial if the correct answer is congenial to the partisanship of the respondent.

3.3 Results

We start by summarizing the average partisan gap on each survey item in each treatment arm (see Figure 1). In the baseline condition (Condition 1) with all inflationary attributes, when the correct response is congenial to the respondents' party, respondents are 35 percentage points more likely to choose the correct response. As Figure 1 shows, the partisan gap is unresponsive to the removal of social proof in the question stem (Condition 2). However, the estimates in Conditions 3 and 4 are approximately 14 percentage points lower than in the baseline condition. The 14 percentage points reduction, stemming from the inclusion of guessing discouraging text and the exclusion of neutral information in the question stem, translates to a 40% relative drop ($100 \times \frac{.35-.21}{.35}$).

To formally test our hypothesis, we estimate the following equation:

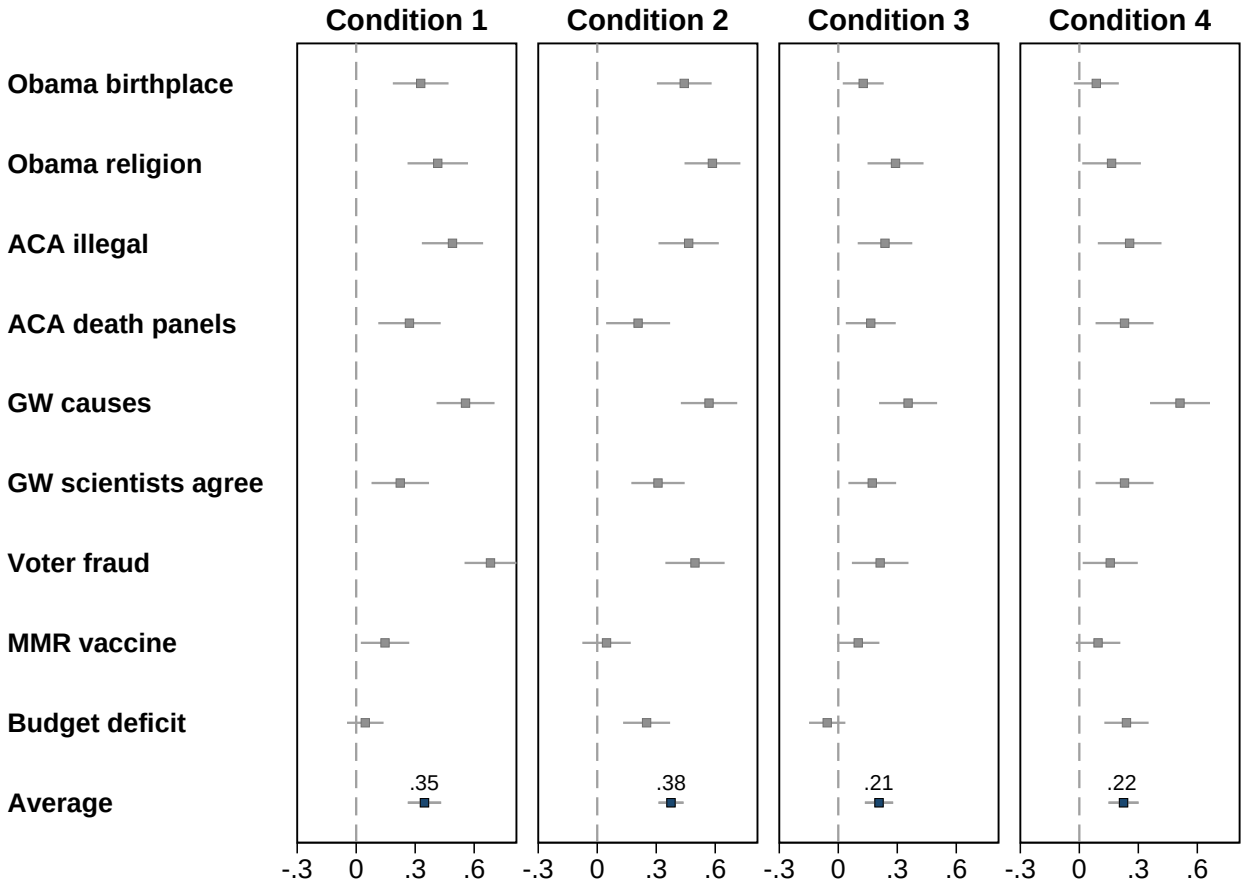
$$\text{Correct}_{ijk} = \alpha + \beta \text{Congenial}_i + \sum_{k=1}^{4} \gamma \text{Condition}_k + \sum_{k=1}^{4} \delta_k (\text{Congenial}_i \times \text{Condition}_k) + \text{question}_j + \varepsilon_{ijk} \tag{1}$$

In the above equation, i iterates over respondents, j over survey items, and k over treatment arms. β captures the difference in the proportion of correct responses when the answer is congenial to the respondent's party in the baseline condition. δ_ks capture how Conditions 2-4 affect the partisan knowledge gaps versus the baseline condition. s

Table 2 reports the results. Column (1) includes just the congenial variable, which is significant and consistent with conventional wisdom about gaps in partisan knowledge (e.g. Bullock et al. 2015; Laloggia 2018).

Column (2) only includes the survey treatments. The negative coefficients on Condition 3 (DK+NSP+GD+NI) and 4 (DK+NSP+GD+NNI) show that respondents' estimated knowledge is sharply lower in the two conditions, which remove first any encouragement to guess and then any neutral information, compared to the baseline. In column (3), we include

Figure 1: Partisan Gap by Treatment Arm (MTurk 1)



The figure shows the estimated partisan gap in each of the nine knowledge items (see Appendix SM 3 for the details of the items) and the average partisan gap across the four conditions (Table 1). The partisan gap is estimated using the linear model Correct response$_i = \alpha + \beta$congenial$_i + \varepsilon_i$ where 'congenial' is a dummy variable that takes the value 1 when the correct response is congenial to the party. All four columns have the same horizontal axis scale. Horizontal bars are 95% confidence intervals constructed from robust standard errors.

the interaction between the congenial dummy and the three conditions (baseline is Condition 1 with all inflationary features). Now, the congenial variable captures the knowledge gap in the baseline condition (corresponding to column (1) of Figure 1). The congenial and survey condition interactions reveal the extent to which partisan knowledge gaps change across the different survey conditions. The gap drops from .35 and .38 in the more inflationary designs to .21 and .22 in designs with fewer problematic features.

Columns (4)–(6) of Table 2 show that including self-reported characteristics of re-

Table 2: The Effect of Various Treatments on the Partisan Gap (MTurk 1)

	(1)	(2)	(3)	(4)	(5)	(6)
Congenial	0.281***		0.351***	0.284***		0.353***
	(0.017)		(0.035)	(0.017)		(0.034)
	[0.000]		[0.000]	[0.000]		[0.000]
Condition 2		0.010	0.000		0.011	0.002
		(0.028)	(0.022)		(0.028)	(0.021)
		[0.722]	[0.985]		[0.687]	[0.934]
Condition 3		−0.064**	0.000		−0.063**	−0.001
		(0.024)	(0.019)		(0.024)	(0.019)
		[0.009]	[0.993]		[0.010]	[0.964]
Condition 4		−0.080**	−0.023		−0.079**	−0.021
		(0.025)	(0.019)		(0.025)	(0.019)
		[0.002]	[0.245]		[0.002]	[0.281]
Congenial × (Cond. 2)			0.024			0.024
			(0.046)			(0.045)
			[0.605]			[0.601]
Congenial × (Cond. 3)			−0.173***			−0.163***
			(0.046)			(0.045)
			[0.000]			[0.000]
Congenial × (Cond. 4)			−0.132**			−0.136**
			(0.048)			(0.048)
			[0.006]			[0.005]
Constant	0.179***	0.306***	0.184***	0.050	1.331	0.227
	(0.007)	(0.020)	(0.014)	(1.056)	(1.255)	(1.010)
	[0.000]	[0.000]	[0.000]	[0.962]	[0.289]	[0.823]
R^2	0.315	0.234	0.328	0.324	0.243	0.337
Survey item FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	.	.	.	Yes	Yes	Yes
Items	9	9	9	9	9	9
Respondents	628	628	628	627	627	627
Respondent-items	5,652	5,652	5,652	5,643	5,643	5,643

All models are linear probability models where the dependent variable is whether the response is correct. See Table 1 for the description of the four conditions. Condition 1, with four inflationary features, is the baseline. Demographic controls include age, gender, education, and race. Standard errors are clustered at the respondent level. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. P-values in square brackets. Figure 1 visualizes partisan gaps by condition. Alternate visualization of results in Figure SM 7.1 and Figure SM 7.2 in Appendix SM 7.

spondents does not change the conclusion. Overall, Study 1 suggests that partisan gaps are much larger in surveys with guessing-encouraging features, like the absence of a 'Don't Know' option, the inclusion of social proof and neutral information in the question stem, and no discouragement to guess, all features that are common in commercial polls.

4 Study 2: The Effect of Partisan Cues on Partisan Gaps

In Study 2, we investigate the impact of partisan cues. We test this by manipulating the partisan cue in the question stem.

4.1 Research Design and Data

To answer the question, we leverage data from a national survey conducted by YouGov and a telephone survey in Texas. The YouGov survey includes data from 2,000 respondents interviewed between July 10th and 12th, 2012. The Texas survey has data from 1,003 respondents who were interviewed between September 10th and 21st, 2012.

In the YouGov survey, we asked respondents two retrospective economic evaluation questions: unemployment and the budget deficit. To manipulate congeniality, we randomly inserted a Republican or a Democratic cue into the question stem. In particular, we asked the following two questions:

1. Since the 2010 midterm elections, ("when Republicans regained control of the U.S. Congress" or "when Democrats retained control of the Senate"), the unemployment rate [had] gone up, down, or remained the same, or couldn't you say?

2. Since the 2010 midterm elections, ("when Republicans regained control of the U.S. Congress" or "when Democrats retained control of the Senate"), has the budget deficit gone up, gone down, remained the same, or couldn't you say?

In the Texas survey, we added a 'no partisan cue' condition to the unemployment rate question. A third of the respondents saw:

3. Since the 2010 midterm elections, has the unemployment rate gone up, gone down, or remained the same? Or couldn't you say?

The partisan cue was randomized between participants. If someone received the Republican cue in the unemployment question, they also received it in the deficit question.

We made two more changes to the second and final question on the Texas survey. First, we switched the question from one about budget deficits to one about federal tax rates. Second, we changed the treatment conditions to 1. no partisan cue, 2. Democratic cue, and 3. Democratic cue with a substantive response encouraging phrase. Respondents assigned to the 'no partisan cue' condition saw, "Since January 2009, have federal taxes increased, decreased, or remained the same, or couldn't you say?." The Democratic cue condition prepended "Since Barack Obama took office..." to the question. The last version prepended a substantive response encouraging phrase. The question now read: "Based on what you have heard, since Barack Obama took office, ..."

For national macroeconomic retrospective evaluation questions like the ones we ask in these two surveys, we assume that the answer is congenial when the correct answer has positive implications for the president's party. On such a question, we expect to obtain the largest estimate of the partisan gap when the partisan cue nudges the president's co-partisans toward the right answer and the president's main opposing partisans toward the wrong answer. For instance, if, say, under President Obama, the unemployment rate went down over some years, we expect a cue that highlights the Democratic responsibility to lead Democrats to mark more correct answers and Republicans to mark fewer correct answers. Conversely, a partisan cue that highlights Republican responsibility will lead Democrats toward the wrong answer and Republicans toward the right answer, attenuating the partisan gap. Conditional on there being a partisan cue in the question stem, theoretically, the best estimate of the partisan gap is the difference between the proportion correct between groups when the partisan cue directs partisans to the wrong answer. The rationale is that adversarial cues yield estimates of knowledge that are least contaminated with partisan guessing. Partisan guessing is but one force affecting partisan gaps. Another is partisan

cheerleading. Having a partisan cue increases partisan cheerleading—deliberately marking the wrong answer even when you know the correct answer. Hence, questions with a neutral stem likely yield a better estimate of partisan gaps than ones with a partisan cue though we expect question stems with adversarial cues to do the best.

4.2 YouGov Results

We estimate the impact of randomly inserting partisan cues on the estimated partisan gap by regressing whether the response is correct on the interaction of partisan cue and whether the correct response is congenial (to the party):

$$\text{Correct}_i = \alpha + \beta\text{Congenial}_i + \gamma(\text{Dem. cue})_i + \delta\text{Congenial}_i \times (\text{Dem. cue})_i + \varepsilon_i, \quad (2)$$

where the constant is the proportion of correct responses in the baseline condition. β captures the partisan gap in the baseline condition (Republican cue). We are interested in δ, which captures how randomly receiving the Democratic cue (which leads Democrats to mark the right answer and Republicans to mark the wrong one) widens the estimated partisan gap.

Table 3 reports the estimated coefficients for the two questions in YouGov. Randomly receiving a Democratic cue increases the probability of getting the correct response in the unemployment question by 16 percentage points ($p < .001$) and a seven percentage points increase in the gap on the federal deficit question ($p = .002$). Adjusting for demographics (see columns (2) and (4)) doesn't appreciably change the coefficients.

And as we write above (in Section 4.1), the best estimate of partisans' stores of knowledge is under an adversarial partisan cue. Compared to the theoretical maximal partisan gap (Democrats getting a cue that makes them more likely to get the correct answer and Republicans getting a cue that makes them more likely to get the wrong answer), the par-

Table 3: The Impact of Partisan Cues on Partisan Gaps (YouGov)

	"Unemployment has gone down"		"Deficit has gone down"	
	(1)	(2)	(3)	(4)
Congenial	0.044	0.102***	0.027+	0.029+
	(0.029)	(0.030)	(0.014)	(0.016)
	[0.129]	[0.001]	[0.051]	[0.072]
Democratic cue	−0.016	−0.015	−0.009	−0.006
	(0.029)	(0.029)	(0.012)	(0.012)
	[0.588]	[0.612]	[0.449]	[0.615]
Congenial × Democratic cue	0.155***	0.147***	0.067**	0.061**
	(0.041)	(0.041)	(0.021)	(0.022)
	[0.000]	[0.000]	[0.002]	[0.005]
Constant	0.297***	0.017	0.041***	−1.292
	(0.021)	(2.000)	(0.009)	(1.117)
	[0.000]	[0.993]	[0.000]	[0.247]
R^2	0.0273	0.0887	0.0214	0.0383
Demographic controls	.	Yes	.	Yes
Respondents	2,104	2,066	2,104	2,066

Dependent variables indicate whether or not the respondent chose the correct answer. Demographic controls include age cohort, gender, education level, marital status, employment status, news interest, family income, and race. Standard errors are heteroskedasticity-robust. P-values in square brackets. All models are linear probability models. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001.

tisan gap in the unemployment question obtained under the adversarial cue is, on average, 14 percentage points smaller ($p < .001$).

4.3 Texas Lyceum Results

Table 4 shows that the pattern we saw in YouGov (Section 4.2) still holds when we include a neutral cue. Here, we let the Republican cue be the baseline condition and examine the extent to which a neutral cue and a Democratic cue change the estimated partisan gap (Section 4.1). The specification is otherwise similar to Equation (2) (with an additional Neutral cue arm). Consistent with the theory, randomly receiving the Democratic cue increases the gap by 20 percentage points ($p = .013$) (column(1) in Table 4). Again, little changes when we add the baseline demographics (column (2) of Table 4). A meta-analysis of the Texas Lyceum

Table 4: Impact of Partisan Cues on Proportion Correct on Unemployment (Texas Lyceum)

	"Unemployment has gone down"	
	(1)	(2)
Congenial	0.089	0.048
	(0.054)	(0.073)
	[0.103]	[0.511]
Neutral cue	0.013	0.017
	(0.046)	(0.048)
	[0.771]	[0.725]
Democratic cue	−0.053	−0.047
	(0.044)	(0.045)
	[0.233]	[0.299]
Congenial x Neutral cue	0.072	0.074
	(0.078)	(0.084)
	[0.356]	[0.376]
Congenial x Democratic cue	0.196*	0.216**
	(0.079)	(0.082)
	[0.013]	[0.009]
Constant	0.189***	−0.159
	(0.032)	(0.181)
	[0.000]	[0.380]
R^2	0.0502	0.123
Demographic controls	.	Yes
Respondents	758	752

The dependent variable is whether or not the respondent got the answer correct. Demographic controls include age cohort, gender, education level, marital status, number of children, children's school enrollment, family income, religion, liberalism/conservatism, and race. Standard errors are heteroskedasticity-robust. All models are linear probability models. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. P-values in square brackets.

estimates (Table 4) and the YouGov estimate (Table 3) suggests a 16 percentage point increase in the partisan gap when individuals randomly receive a Democratic cue (Figure SM 7.3).

Finally, we examine the federal tax rate question in the Texas Lyceum survey in Table 5. In this question, the baseline condition is without any partisan cue, and we compare how the estimated partisan gap changes with a Democratic cue and a Democratic cue with wording that encourages guessing (Section 4.1). The estimates in column (1) of Table 5 have

Table 5: Impact of Partisan Cues and Guessing-Encouraging Wording on Proportion Correct on Federal Taxes (Texas Lyceum)

	"Federal taxes remained the same"	
	(1)	(2)
Congenial	0.098	0.105
	(0.061)	(0.080)
	[0.109]	[0.188]
Democratic cue	−0.050	−0.049
	(0.053)	(0.054)
	[0.344]	[0.364]
Democratic cue w/ guess cue	0.070	0.071
	(0.056)	(0.058)
	[0.214]	[0.226]
Congenial x Democratic cue	0.112	0.133
	(0.088)	(0.093)
	[0.202]	[0.154]
Congenial x Democratic cue w/ guess cue	−0.058	−0.037
	(0.086)	(0.091)
	[0.499]	[0.687]
Constant	0.303***	0.131
	(0.038)	(0.211)
	[0.000]	[0.535]
R^2	0.0214	0.0847
Demographic controls	.	Yes
Respondents	758	752

The dependent variable is whether or not the respondent got the answer correct. Demographic controls include age cohort, gender, education level, marital status, number of children, children's school enrollment, family income, religion, liberalism/conservatism, and race. Standard errors are heteroskedasticity-robust. All models are linear probability models. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. P-values in square brackets.

large standard errors that mean we cannot confidently conclude that much about the effect of Democratic Cue on the partisan gap.

Overall, survey experiments show that partisan cues dramatically affect the size of partisan gaps. If partisan gaps only reflected partisans' existing stores of knowledge, the gaps would be unresponsive to these cues. Thus, the experiments highlight the role of partisan inference and cheerleading in the estimates of partisan gaps that we see.

5 Study 3: The Effect of the Scoring Method on Partisan Gaps

Lastly, we examine the consequences of scoring decisions on partisan gaps. We introduce a scoring scheme that considers respondents' confidence in their answers. We propose scoring only confidently held correct beliefs about political facts as knowledge.

5.1 Research Design and Data

Knowledge questions are commonly offered as multiple-choice items, and conventionally, if a respondent marks the correct answer, it is taken as evidence that the respondent knows the answer. Such scoring does not differentiate between confidently held beliefs, hunches, inferences, blind guesses, and expressive responses. To distinguish between hunches, guesses, and confidently held beliefs, we use the design from studies like Pasek, Sood and Krosnick (2015). In our Confidence Coding Design (CCD), respondents rate claims on a Likert scale, going from 'definitely false' (0) to 'definitely true' (10).

To estimate the impact of the question and scoring design that considers respondents' confidence in their answers, we use data from two separate survey experiments. Our first survey experiment is the one underlying Study 1 (*MTurk 1*). The survey had a fifth condition in addition to the four above-mentioned closed-ended multiple-choice conditions. The fifth condition offered the same questions, except respondents were asked to respond on a Likert scale ranging from 0 (definitely not true) to 10 (definitely true). The CCD condition builds on the first four conditions. It discourages guessing and features no social proof or neutral information in the question stem. (See Appendix SM 3 for the question text.) Since the items in the multiple choice questions are dichotomous choice, we only offer a true and an incorrect response; the confidence coding is straightforward. One of the response options from the multiple-choice question becomes its own Likert scale item. Respondents can indicate

whether they think the statement (e.g., "Barack Obama was born in the US" or "Barack Obama is a Muslim") is "definitely false (0)" or "definitely true." We scored respondents who marked 'definitely true' for the right answer or "definitely false" for the wrong answers as knowledgeable. (See Appendix SM 3 for the question text.)

For the second survey experiment, we turn to another MTurk survey (*MTurk 2*). In this survey, we randomly assigned 1,059 respondents to two conditions. The preamble, topics, and answer options of these questions were identical to the first survey and included questions about the Affordable Care Act (2 questions), the effect of greenhouse gases (1 question), and the consequences of Mr. Trump's executive order on immigration (1 question). In the multiple-choice version of the item, participants received three options. In two of the four conditions, respondents also saw a "Don't Know" option. (See Appendix SM 5 for the question text.)

The scoring for this survey is more nuanced, as the multiple-choice questions had four potential response options. In the confidence coding treatment, survey participants see the same question as in the multiple choice treatment, but have to score the correctness of all the n answer options from the multiple choice treatment. Broadly, we code an answer as correct if the respondent indicates that they are confident that the correct answer is correct and when they do not indicate that any of the incorrect options might also be correct. More precisely, we code a response as correct if four conditions are met:

1. The respondent is most confident about the correct answer. For instance, it shouldn't be the case that the respondent is more confident about an incorrect answer.

2. The respondent is not as confident about the correct answer as another option. For instance, it cannot be that the four options are all rated 10.

3. The respondent must have at least c level of confidence in the correct answer. We use a c of 10 in the main text, but in Appendix SM 6, we try less stringent criteria.

4. The confidence in the incorrect answers cannot be above the threshold t. We use a t of 0 in the main text, but in the Appendix SM 6, we try less stringent criteria.

5.2 MTurk 1 Results

The partisan gap on the best version of the dichotomous multiple-choice items (Condition 4; DK+NSP+GD+NNI) was .22 (see Figure 1). As Figure 2 shows, nearly half of the gap vanishes under confidence scoring. Furthermore, the number of items with no statistically significant gap between partisans doubles from two to four. In all, there is a nearly 11 percentage point drop in the size of the partisan gap when we treat only confident correct answers as evidence that the respondent knows the answer.

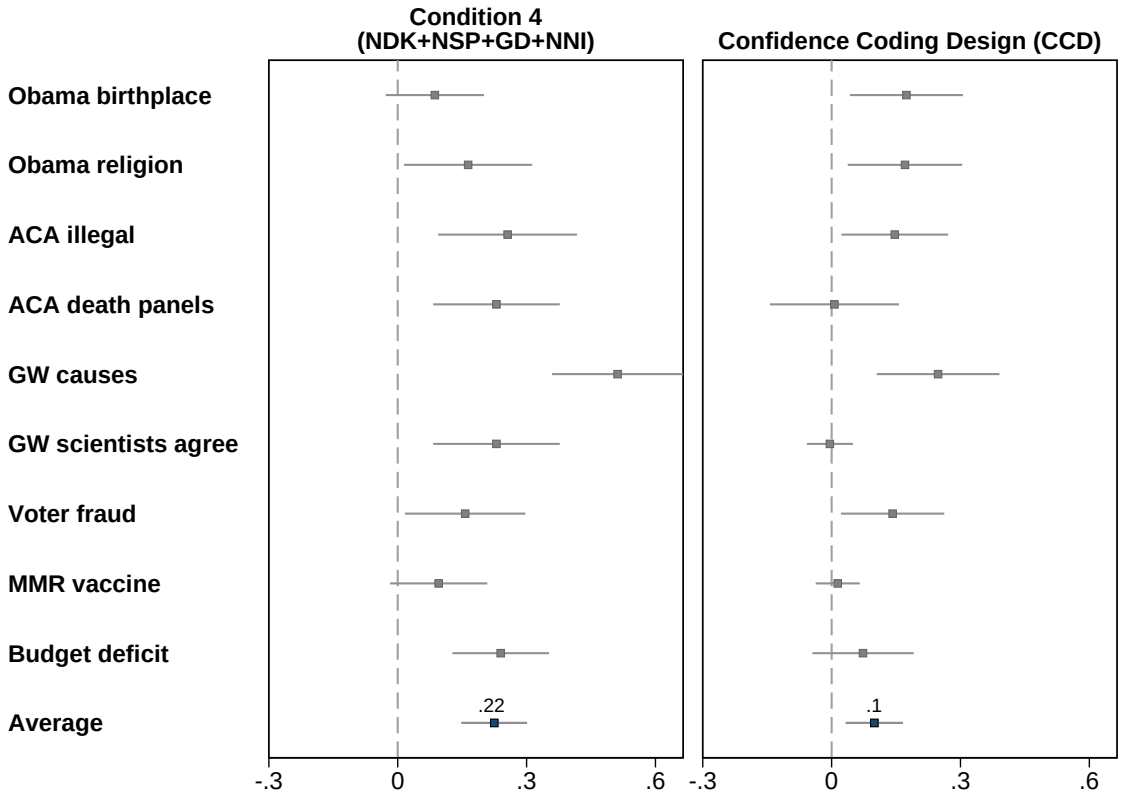
5.3 MTurk 2 Results

To further illuminate how treating answers a respondent is confident about as evidence that the respondent knows the fact affects partisan gaps, we regress the dependent variable, an indicator of whether the response is correct, on the interaction between CCD (with conventional scoring serving as the baseline) and the congenial dummy:

$$\text{Correct}_{ij} = \alpha + \beta\text{Congenial}_i + \gamma\text{Scoring} + \delta(\text{Congenial}_i \times \text{Scoring}) + \varepsilon_{ij} \tag{3}$$

for respondents i and survey item j. In Equation (1), β captures the difference in the proportion of correct responses when the answer to the question is congenial to the respondent's party affiliation under the baseline conventional scoring condition. A positive estimate indicates that respondents are likelier to choose the correct response when it is

Figure 2: Partisan Gaps in Knowledge in Different Question Designs



The figure shows the estimated partisan gaps in knowledge from MTurk 1 for two different survey conditions. The CCD condition only considers selecting the right answer with complete confidence as evidence that the respondent knows the answer (see Appendix SM 5). See Tables SM 2.1 to SM 2.5 in Appendix SM 2 for the regression estimates of the multiple-choice conditions to the confidence coding condition. See Figure SM 2.6 for the same analysis with all four multiple-choice conditions pooled together. Figure SM 6.1 implements a robustness check, setting the relative scoring threshold t to 8.

congenial to their party affiliation in the multiple-choice treatment. γ captures the effect of relative scoring in the CCD scheme. A positive coefficient indicates confidence scoring is associated with more correct responses, and a negative one with fewer. δ captures how the two scoring treatments, multiple choice, and CCD, affect the knowledge gaps across partisans for congenial questions. In the pooled equation, which includes all questions, we also include question fixed effects, question_j.

Table 6 reports the results from Equation (3). Columns 1 through 4 report the question-specific estimates. Column 5 pools all questions and adds question fixed-effects to the model. In this specification, the intercept term reports the proportion correct for

Table 6: Confidence Scoring and Knowledge Gaps: MTurk 2

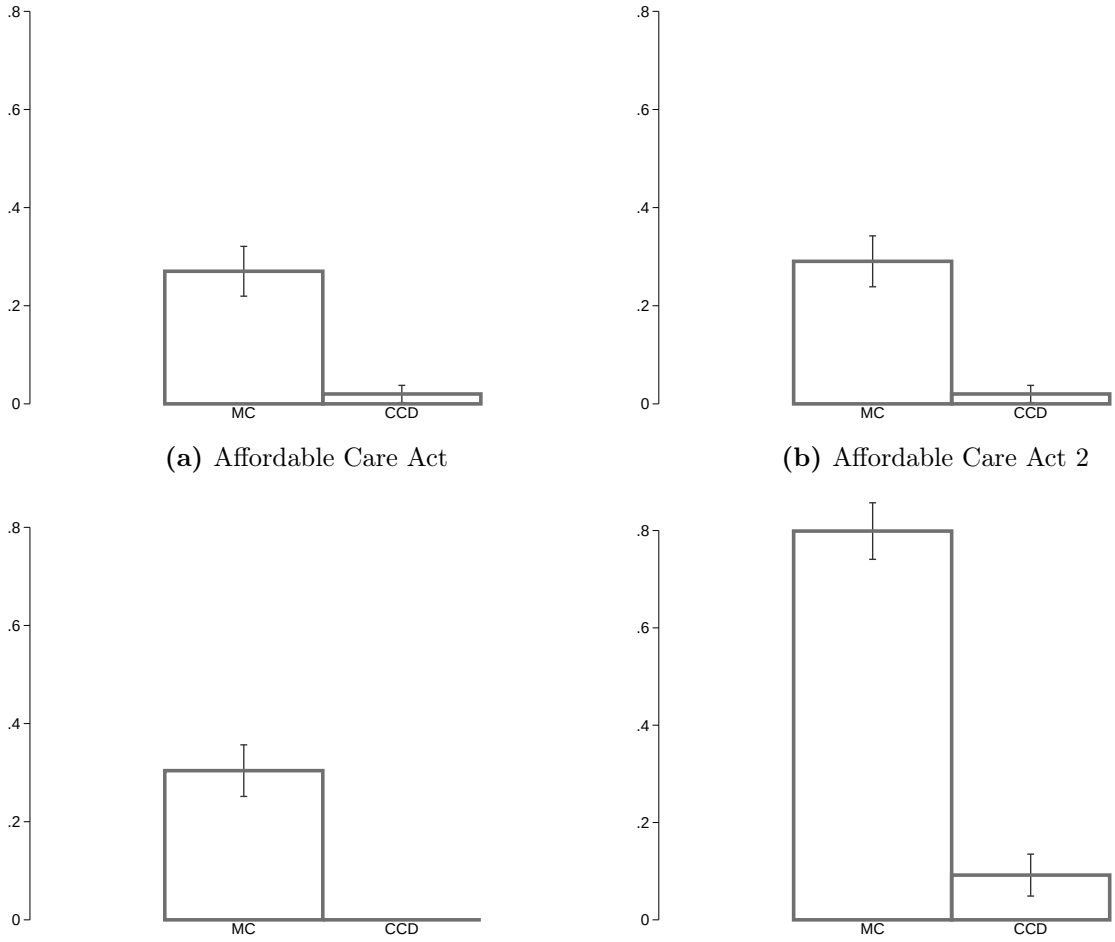
	Individual survey question				
	Affordable Care Act	Affordable Care Act 2	Greenhouse gases	Donald Trump	All
	(1)	(2)	(3)	(4)	(5)
Congenial	0.091*	0.084*	0.087*	0.005	0.025
	(0.038)	(0.040)	(0.041)	(0.038)	(0.023)
	[0.018]	[0.036]	[0.033]	[0.895]	[0.270]
Confidence Coding Design (CCD)	−0.179***	−0.201***	−0.206***	−0.737***	−0.377***
	(0.028)	(0.030)	(0.032)	(0.028)	(0.018)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Congenial × CCD	−0.071+	−0.070+	−0.098*	0.031	0.024
	(0.039)	(0.041)	(0.041)	(0.046)	(0.026)
	[0.073]	[0.092]	[0.018]	[0.509]	[0.351]
Constant	0.179***	0.207***	0.217***	0.794***	0.376***
	(0.028)	(0.030)	(0.030)	(0.024)	(0.017)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
R^2	0.119	0.128	0.149	0.528	0.305
Survey item FE	No	No	No	No	Yes
Items	1	1	1	1	4
Respondents	902	902	902	902	902
Respondent-items	902	902	902	902	3,608

Dependent variables indicate whether the respondent answered the question(s) correctly. See Appendix SM 5 for the exact wording of the four questions. Columns (1)–(4) estimates by the individual survey questions. Column (5) includes all questions and adds the survey question fixed effects. All models are linear probability models. In the confidence coding scheme, a response is correct only if the correct answer is selected with complete confidence of $c = 10$ (see Research Design and Data in the Study 3: The Effect of the Scoring Method on Partisan Gaps section). The baseline is the multiple choice designs. Table SM 6.2 implements a robustness check, setting the relative scoring threshold to $c = 8$. Standard errors are clustered at the respondent level. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. P-values in square brackets.

uncongenial questions that were scored with multiple choice rules. For β, we can see across all but one column (column 4, Donald Trump) that congenial questions in multiple choice scoring are associated with a higher proportion of correct responses. In the multiple-choice treatment, partisans are more likely to get questions correct when answers are congenial to their partisanship. γ shows that the partisan gap in knowledge nearly disappears in the CCD.

Pooling evidence across the two studies, it appears that treating only confident correct answers as evidence that the respondent knows the answer yields dramatically lower estimates of the partisan gap.

Figure 3: Partisan Gaps by Coding (MTurk 2)



(a) Affordable Care Act

(b) Affordable Care Act 2

(c) Greenhouse Gases

(d) Donald Trump

Bars indicate the predicted proportion of correct responses as reported in Table 6. MC bar indicates the predicted proportion of correct responses for multiple-choice with congenial responses. The CCD bar indicates the predicted proportion of correct responses for the Confidence Coding Design with congenial responses. Note that the scales of the vertical axes vary. Capped vertical bars indicate 95% confidence intervals.

6 Validity and Reliability of Different Ways of Measuring Political Knowledge

We have shown that survey and item design choices that encourage guessing have larger partisan gaps than ones that discourage guessing and where the scoring scheme codes only confident correct answers as evidence of knowledge. But which item and survey design choices lead to 'better' measures? To answer that, we use data from the first MTurk survey to assess the reliability and criterion validity of different designs. Specifically, we use average inter-item correlation and Cronbach's α to measure the scale's reliability. To measure criterion validity, we use the correlation of the scale with three criteria thought to correlate heavily with political knowledge: education, political interest, and political participation (see Appendix SM 4 for the question text). We expect items that discourage guessing to have higher reliability and greater criterion validity.

Table 7 reports results for each of the four conditions (see Table 1) and the confidence coding condition (CCD) that scores a response as correct when the respondent is completely confident about the correct answer. CCD has better reliability than other versions. However, the picture is more mixed for the other conditions, with Conditions 1 (NDK+SP+GE+NI), 3 (DK+NSP+GD+NI) having greater reliability than Conditions 2 (NDK+NSP+GE+NI), and 4 (DK+NSP+GD+NNI). One of the reasons for this mixed picture may be that partisan guessing increases reliability without increasing validity because it introduces correlated errors (a point we discussed in Section 1.1). A more diagnostic test for the quality of the instrument, hence, is criterion validity. As Panel A of Table 7 shows, the average correlation between the Condition 4 (multiple choice with no inflationary features) condition and the Confidence Coding Design and criterion variables is markedly higher (.35) than in conditions 1-3. The baseline Condition 1 (.11), Condition 2 (.20), and Condition 3 (.26) all score lower.[4]

[4]We did one more test to get at the validity. We hypothesized that partisan guessing

Table 7: Validity and Reliability

	Conditions				
	No DK		With DK		
	Cond. 1 (1)	Cond. 2 (2)	Cond. 3 (3)	Cond. 4 (4)	CCD (5)
	Panel A. Criterion correlational validity				
Political interest	.115	.278	.271	.412	.379
Political participation	.138	.168	.276	.298	.356
Education	.077	.167	.23	.18	.302
Average of 3 items	.11	.204	.259	.297	.346
	Panel B. Inter-item correlation				
Average inter-item correlation	.237	.163	.248	.172	.325
	Panel C. Scale reliability				
Cronbach's alpha	.737	.637	.748	.652	.812

Panel A reports the correlation coefficient between each condition and the three criterion variables. Political interest and political participation (voting) are coded on an 11-point scale. Education is coded from 1–5 by education qualification. Panel B reports the inter-item correlation for the nine items (see Figure 1). Panel C reports the Cronbach's alpha for the nine items. See Table 1 for a brief description of the first four conditions and Study 3: The Effect of the Scoring Method on Partisan Gaps for the confidence coding design.

The results above are consistent with those obtained by Graham (2021), which finds that the test-retest reliability of confident correct answers is much higher, and Vidigal (2024), which finds "incorporating belief certainty results in a knowledge scale that displays theoretically expected relationships with a range of outcome variables while also having superior psychometric properties."

would lead to a greater negative correlation between congenial and uncongenial items on items that encouraged guessing. And indeed, the item-rest correlations between uncongenial and congenial items are the smallest for CCD.)

7 Discussion and Conclusion

Since at least the publication of Bartels (2002), the conventional wisdom has been that partisan gaps in confidently held beliefs about politically consequential facts are wide and widespread. The conventional wisdom in academia has also become the received wisdom for the mass public—nearly 80% of Americans believe that Democrats and Republicans disagree on facts (Laloggia 2018).

In line with some other research on this topic (Bullock et al. 2015; Prior, Sood and Khanna 2015; Schaffner and Luks 2018, though see Berinsky 2017 and Peterson and Iyengar 2020), our results suggest that a big chunk of the partisan gap in the knowledge of politically consequential facts in the US is not founded in differences in confidently held beliefs. We find that standard features of commercial polls, like not offering don't know, inserting a partisan cue, and treating unconfident answers as knowledge, inflate the partisan gaps. Removing such 'inflationary' features reduces the partisan gap dramatically.

The fact that partisan gaps are smaller may seem at odds with some political behavior research. For instance, selective exposure theory posits vast imbalances in the consumption of partisan news. However, recent studies show that most people consume scant political news (Prior 2007; Flaxman, Goel and Rao 2016), and the news that they do consume is relatively balanced (Flaxman, Goel and Rao 2016; Garz et al. 2018; Gentzkow and Shapiro 2011; Guess 2020). Other evidence points to the fact that Democrats and Republicans update similarly in light of events (Gerber and Green 1999; Kernell and Kernell 2019; Coppock 2021).

In the end, the results of our studies paint a mixed picture of democratic competence. Smaller partisan gaps partly result from the fact that the average respondent doesn't know the facts. It is primarily partisan guessing masquerading as partisan gaps. The upside is that partisan gaps are small; the downside is that people know even less than we thought.

Lastly, while the data are from the US, we think there is reason to believe that similar

concerns vitiate partisan gaps in knowledge measured in other countries. Research focusing on multi-party systems around the world has found that they are increasingly presidentializing[5] and affectively polarizing (Hobolt, Leeper and Tilley 2021; Wagner 2021; 2024). See also Bailey (2021) that shows that self-reported economic perceptions by British voters are shaped by political cues in surveys, and Bisgaard and Slothuus (2018) that shows how central partisan cues have become for economic perceptions in Denmark. Hence, we expect the conclusions from the paper to travel to other contexts.

[5]See Poguntke and Webb (2005) for the general argument, Krauss and Nyblade (2005) for Japan, and Poguntke and Webb (2015) for Italy and Germany.

References

Bailey, Jack. 2021. "Political Surveys Bias Self-Reported Economic Perceptions." *Public Opinion Quarterly* 85(4):987–1008.

Bartels, Larry M. 2002. "Beyond the Running Tally: Partisan Bias in Political Perceptions." *Political Behavior* 24(2):1061–1078.

Berinsky, Adam J. 2017. "Telling the Truth About Believing the Lies? Evidence for the Limited Prevalence of Expressive Survey Responding." *Journal of Politics* 80(1):211–224.

Bisgaard, Martin and Rune Slothuus. 2018. "Partisan elites as culprits? How party cues shape partisan perceptual gaps." *American Journal of Political Science* 62(2):456–469.

Bullock, John G., Alan S. Gerber, Seth J. Hill and Gregory A. Huber. 2015. "Partisan Bias in Factual Beliefs About Politics." *Quarterly Journal of Political Science* 10:519–578.

Bullock, John G and Kelly Rader. 2022. "Response options and the measurement of political knowledge." *British Journal of Political Science* 52(3):1418–1427.

Campbell, Angus, Philip E Converse, Warren E Miller and Donald E Stokes. 1980. *The american voter*. University of Chicago Press.

Cialdini, Robert B. 2009. *Influence: Science and practice*. Vol. 4 Pearson education Boston.

Coppock, Alexander. 2021. "Persuasion in Parallel." *Chicago studies in American politics. University of Chicago Press. Forthcoming* .

Coppock, Alexander, Thomas J Leeper and Kevin J Mullinix. 2018. "Generalizability of heterogeneous treatment effect estimates across samples." *Proceedings of the National Academy of Sciences* 115(49):12441–12446.

Cor, M Ken and Gaurav Sood. 2016. "Guessing and forgetting: a latent class model for measuring learning." *Political Analysis* 24(2):226–242.

Dolan, Kathleen and Michael A Hansen. 2020. "The variable nature of the gender gap in political knowledge." *Journal of Women, Politics & Policy* 41(2):127–143.

Ferrín, Monica, Marta Fraile and Gema García-Albacete. 2017. "The gender gap in political knowledge: is it all about guessing? An experimental approach." *International Journal of Public Opinion Research* 29(1):111–132.

Flaxman, Seth, Sharad Goel and Justin M. Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80(S1):298–320.

Fortin-Rittberger, Jessica. 2016. "Cross-national gender gaps in political knowledge: How much is due to context?" *Political research quarterly* 69(3):391–402.

Garz, Marcel, Gaurav Sood, Daniel F. Stone and Justin Wallace. 2018. "What Drives Demand for Media Slant?" Working paper.

Gentzkow, Matthew and Jesse M. Shapiro. 2011. "Ideological Segregation Online and Offline." *Quarterly Journal of Economics* 126(4):1799–1839.

Gerber, Alan and Donald Green. 1999. "Misperceptions About Perceptual Bias." *Annual Review of Political Science* 2:189–210.

Graham, Matthew H. 2021. "Measuring Misperceptions?" *American Political Science Review* .

Graham, Matthew H and Omer Yair. 2023. Less Partisan but No More Competent: Expressive Responding and Fact-Opinion Discernment. Technical report Working Paper.

Guess, Andrew M. 2020. "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets." *American Journal of Political Science* forthcoming.

Hobolt, Sara B, Thomas J Leeper and James Tilley. 2021. "Divided by the vote: Affective polarization in the wake of the Brexit referendum." *British Journal of Political Science* 51(4):1476–1493.

Hochschild, Jennifer and Katherine Levine Einstein. 2015. "'It isn't what we don't know that gives us trouble, it's what we know that ain't so': Misinformation and democratic politics." *British Journal of Political Science* 45(3):467–475.

Huber, Gregory A. and Omer Yair. 2018. How Robust is Evidence of Partisan Perceptual Bias in Survey Responses? A New Approach for Studying Expressive Responding. In *Annual Meeting of the Midwest Political Science Association*. Chicago: .

Jerit, Jennifer and Jason Barabas. 2012. "Partisan Perceptual Bias and the Information Environment." *The Journal of Politics* 74(3):672–684.

Jessee, Stephen A. 2017. ""Don't know" responses, personality, and the measurement of political knowledge." *Political Science Research and Methods* 5(4):711–731.

Kernell, Georgia and Samuel Kernell. 2019. "Monitoring the Economy." *Journal of Elections, Public Opinion, and Parties* forthcoming.

Kraft, Patrick W and Kathleen Dolan. 2023. "Asking the right questions: A framework for developing gender-balanced political knowledge batteries." *Political Research Quarterly* 76(1):393–406.

Krauss, Ellis S and Benjamin Nyblade. 2005. "'Presidentialization'in Japan? The prime minister, media and elections in Japan." *British journal of political science* 35(2):357–368.

Laloggia, John. 2018. "Republicans and Democrats Agree: They Can't Agree on Basic Facts." *Pew Research FactTank* August 23.

Lodge, Milton and Charles S. Taber. 2013. *The Rationalizing Voter*. New York: Cambridge University Press.

Luskin, Robert C., Gaurav Sood, Yul Min Park and Joshua Blank. 2018. "Misinformation about Misinformation? Of Headlines and Survey Design." Working paper.

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." *The Journal of Politics* 73(2):547–557.

Malka, Ariel and Mark Adelman. 2022. "Expressive survey responding: A closer look at the evidence and its implications for American democracy." *Perspectives on Politics* pp. 1–12.

Mondak, Jeffery J. 1999. "Reconsidering the measurement of political knowledge." *Political analysis* 8(1):57–82.

Mondak, Jeffery J and Mary R Anderson. 2004. "The knowledge gap: A reexamination of gender-based differences in political knowledge." *The Journal of Politics* 66(2):492–512.

Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman and Jeremy Freese. 2015. "The generalizability of survey experiments." *Journal of Experimental Political Science* 2(2):109–138.

Pasek, Josh, Gaurav Sood and Jon A Krosnick. 2015. "Misinformed about the affordable care act? Leveraging certainty to assess the prevalence of misperceptions." *Journal of Communication* 65(4):660–673.

Peterson, Erik and Shanto Iyengar. 2020. "Partisan Gaps in Political Information and Information-Seeking Behavior: Motivated Reasoning or Cheerleading?" *American Journal of Political Science* forthcoming.

Peterson, Erik and Shanto Iyengar. 2021. "Partisan Gaps in Political Information and Information-Seeking Behavior: Motivated Reasoning or Cheerleading?" *American Journal of Political Science* 65(1):133–147.

Poguntke, Thomas and Paul Webb. 2005. "The presidentialization of politics in democratic societies: A framework for analysis." *The presidentialization of politics: a comparative study of modern democracies* 1:1–25.

Poguntke, Thomas and Paul Webb. 2015. "Presidentialization and the politics of coalition: lessons from Germany and Britain." *Italian Political Science Review/Rivista Italiana di Scienza Politica* 45(3):249–275.

Prior, Markus. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. New York: Cambridge University Press.

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(4):489–518.

Roush, Carolyn E. and Gaurav Sood. 2023. "A Gap in Our Understanding? Reconsidering the Evidence for Partisan Knowledge Gaps." *Quarterly Journal of Political Science* 18(1):131–151.
URL: *http://dx.doi.org/10.1561/100.00020178*

Sanchez, Maria Elena and Giovanna Morchio. 1992. "Probing "dont know" answers: Effects on survey estimates and variable relationships." *Public Opinion Quarterly* 56(4):454–474.

Schaffner, Brian F. and Samantha Luks. 2018. "Misinformation or Expressive Responding? What an Inauguration Crowd Can Tell Us About the Source of Political Misinformation in Surveys." *Public Opinion Quarterly* 82(1):135–147.

Sherif, Muzafer. 1935. "A study of some social factors in perception." *Archives of Psychology (Columbia University)* .

Sturgis, Patrick, Nick Allum and Patten Smith. 2008. "An experiment on the measurement of political knowledge in surveys." *Public Opinion Quarterly* 72(1):90–102.

Vidigal, Robert. 2024. "Measuring Belief Certainty in Political Knowledge." *Political Behavior* pp. 1–23.

Wagner, Markus. 2021. "Affective polarization in multiparty systems." *Electoral Studies* 69:102199.

Wagner, Markus. 2024. "Affective polarization in Europe." *European Political Science Review* pp. 1–15.

Supplementary Materials

SM 1 Balance Tests

Figure SM 1.1: MTurk 1—Condition 1 (Baseline) vs. Condition 2

Variable	Est. (95% Conf. Int.)	P-value
Politics		
Political interest	0.04(-0.00 to 0.09)	0.07
Intention to vote	0.05(0.00 to 0.09)	0.04
Political leaning	-0.00(-0.05 to 0.05)	0.97
Education		
HS diploma	-0.08(-0.24 to 0.07)	0.29
Some college	-0.00(-0.10 to 0.10)	0.96
Four-year college	-0.07(-0.17 to 0.02)	0.12
Post-graduate	0.18(0.06 to 0.30)	0.0
Race		
Asian	-0.02(-0.20 to 0.16)	0.84
Black	0.01(-0.15 to 0.18)	0.88
White	0.01(-0.10 to 0.13)	0.82
Other		
Female	-0.07(-0.16 to 0.02)	0.14
Age cohort	-0.00(-0.05 to 0.04)	0.94
Hispanic/Latin	-0.04(-0.23 to 0.15)	0.68

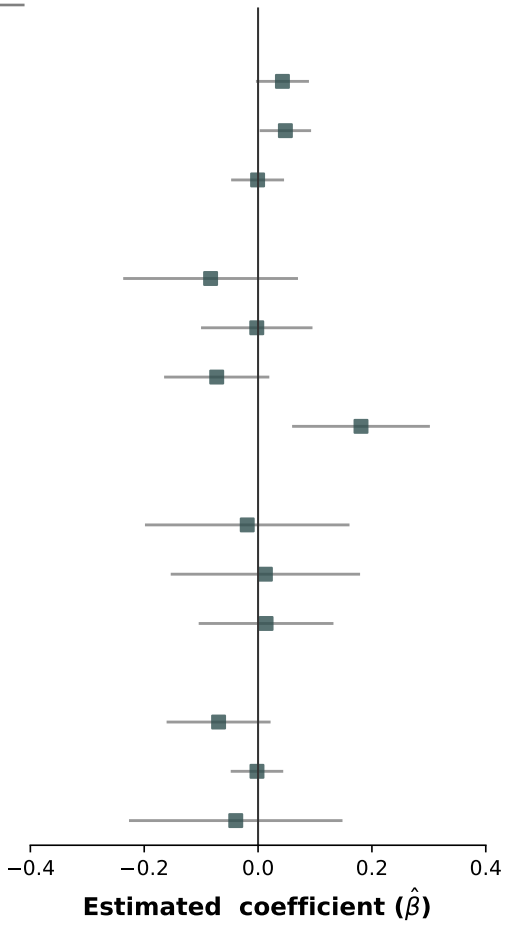


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compare respondents assigned to Condition 1 vs. respondents assigned to Condition 2. See Table 1 in Study 1: The Effect of Guessing-Encouraging Features. Rows are self-reported characteristics. The second column reports the estimates from regressing the characteristics on the Condition 2 dummy, with Condition 1 as the baseline. The third column reports the p-values. Horizontal bars are 95% confidence intervals constructed from robust standard errors.

Figure SM 1.2: MTurk 1—Condition 1 (baseline) vs. Condition 3

Variable	Est. (95% Conf. Int.)	P-value
Politics		
Political interest	0.02(-0.03 to 0.06)	0.41
Intention to vote	0.02(-0.02 to 0.07)	0.31
Political leaning	-0.02(-0.07 to 0.02)	0.35
Education		
HS diploma	-0.02(-0.17 to 0.12)	0.74
Some college	0.00(-0.09 to 0.10)	0.99
Four-year college	-0.03(-0.12 to 0.06)	0.57
Post-graduate	0.08(-0.06 to 0.21)	0.26
Race		
Asian	-0.01(-0.19 to 0.16)	0.88
Black	0.01(-0.15 to 0.17)	0.88
White	0.01(-0.10 to 0.13)	0.8
Other		
Female	-0.10(-0.18 to -0.01)	0.04
Age cohort	-0.00(-0.05 to 0.04)	0.96
Hispanic/Latin	-0.06(-0.25 to 0.12)	0.5

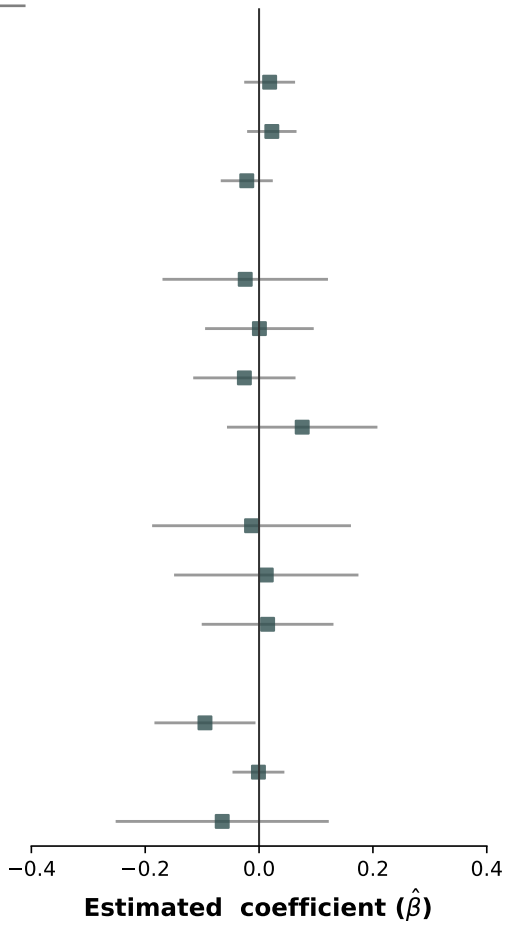


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compare respondents assigned to Condition 1 vs. respondents assigned to Condition 3. See Table 1 in Study 1: The Effect of Guessing-Encouraging Features. Rows are self-reported characteristics. The second column reports the estimates from regressing the characteristics on the Condition 3 dummy, with Condition 1 as the baseline. The third column reports the p-values. Horizontal bars are 95% confidence intervals constructed from robust standard errors.

Figure SM 1.3: MTurk 1—Condition 1 (Baseline) vs. Condition 4

Variable	Est. (95% Conf. Int.)	P-value
Politics		
Political interest	0.02(-0.03 to 0.06)	0.41
Intention to vote	0.03(-0.01 to 0.07)	0.19
Political leaning	-0.03(-0.08 to 0.02)	0.21
Education		
HS diploma	-0.06(-0.22 to 0.09)	0.4
Some college	0.03(-0.06 to 0.13)	0.48
Four-year college	-0.08(-0.17 to 0.01)	0.09
Post-graduate	0.14(0.01 to 0.26)	0.03
Race		
Asian	-0.08(-0.27 to 0.10)	0.39
Black	-0.13(-0.31 to 0.05)	0.16
White	0.10(-0.02 to 0.23)	0.09
Other		
Female	-0.05(-0.14 to 0.04)	0.26
Age cohort	-0.03(-0.07 to 0.02)	0.2
Hispanic/Latin	0.03(-0.14 to 0.21)	0.7

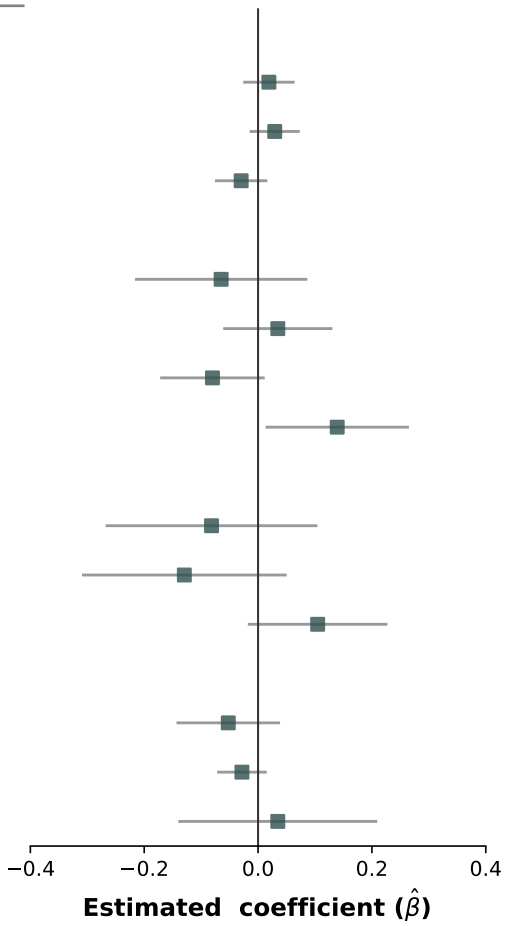


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compare respondents assigned to the Condition 1 vs. respondents assigned to Condition 4. See Table 1 in Study 1: The Effect of Guessing-Encouraging Features. Rows are self-reported characteristics. The second column reports the estimates from regressing the characteristics on the Condition 4 dummy, with Condition 1 as the baseline. The third column reports the p-values. Horizontal bars are 95% confidence intervals constructed from robust standard errors.

Figure SM 1.4: MTurk 1—Condition 1 (baseline) vs. CCD

Variable	Est. (95% Conf. Int.)	P-value
Politics		
Political interest	0.02(-0.03 to 0.06)	0.44
Intention to vote	0.05(0.00 to 0.09)	0.04
Political leaning	-0.04(-0.08 to 0.01)	0.12
Education		
HS diploma	-0.04(-0.19 to 0.11)	0.61
Some college	-0.03(-0.12 to 0.07)	0.58
Four-year college	-0.01(-0.10 to 0.08)	0.88
Post-graduate	0.09(-0.04 to 0.22)	0.19
Race		
Asian	0.08(-0.07 to 0.24)	0.3
Black	-0.02(-0.18 to 0.15)	0.84
White	-0.03(-0.14 to 0.09)	0.65
Other		
Female	-0.08(-0.17 to 0.01)	0.07
Age cohort	-0.00(-0.05 to 0.04)	0.97
Hispanic/Latin	-0.02(-0.20 to 0.16)	0.86

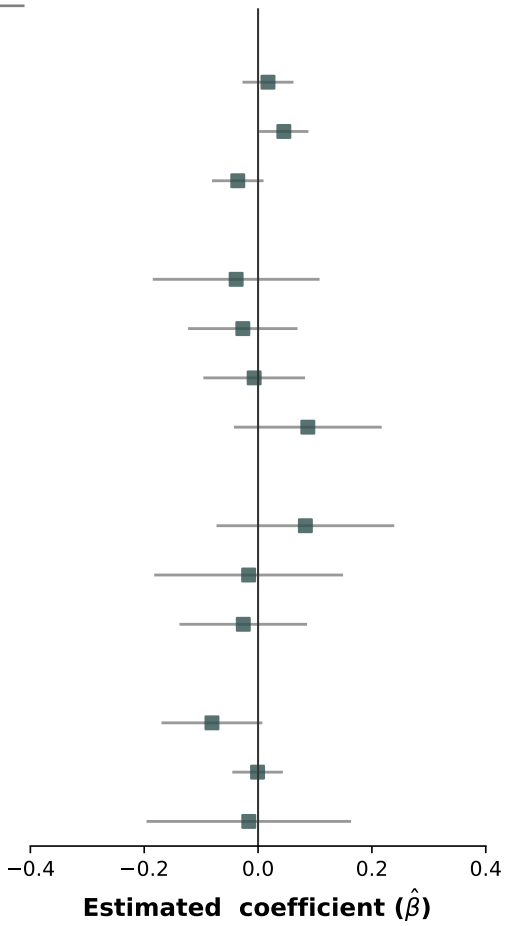


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compare respondents assigned to Condition 1 vs. respondents assigned to the CCD condition. See Table 1 in Study 1: The Effect of Guessing-Encouraging Features. Rows are self-reported characteristics. The second column reports the estimates from regressing the characteristics on the CCD dummy, with Condition 1 as the baseline. The third column reports the p-values. Horizontal bars are 95% confidence intervals constructed from robust standard errors.

SM 2 Additional Results for Confidence Coding (Mturk 1)

Table SM 2.1: Confidence Coding vs. Other Survey Conditions (MTurk 1)

	Obama birthplace	Obama religion	ACA illegal	ACA death panels	GW causes GW causes	GW scientists agree	Voter fraud	MMR vaccine	Budget deficit	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Congenial	0.246***	0.367***	0.363***	0.222***	0.495***	0.232***	0.389***	0.099***	0.117***	0.281***
	(0.033)	(0.038)	(0.039)	(0.037)	(0.037)	(0.034)	(0.039)	(0.029)	(0.027)	(0.017)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]
Confidence Coding (CCD)	−0.010	−0.091***	−0.161***	−0.011	−0.079***	−0.042*	−0.095***	−0.062***	0.044	−0.058***
	(0.017)	(0.020)	(0.018)	(0.043)	(0.016)	(0.019)	(0.024)	(0.016)	(0.044)	(0.010)
	[0.576]	[0.000]	[0.000]	[0.804]	[0.000]	[0.030]	[0.000]	[0.000]	[0.319]	[0.000]
Congenial × CCD	−0.072	−0.196*	−0.216**	−0.215**	−0.247**	−0.236***	−0.247***	−0.085*	−0.044	−0.171***
	(0.073)	(0.076)	(0.072)	(0.083)	(0.080)	(0.043)	(0.071)	(0.039)	(0.064)	(0.034)
	[0.324]	[0.010]	[0.003]	[0.009]	[0.002]	[0.000]	[0.001]	[0.029]	[0.499]	[0.000]
Constant	0.036***	0.109***	0.161***	0.137***	0.088***	0.069***	0.130***	0.071***	0.806***	0.176***
	(0.009)	(0.015)	(0.018)	(0.017)	(0.014)	(0.012)	(0.016)	(0.013)	(0.019)	(0.007)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
R^2	0.127	0.185	0.171	0.0636	0.301	0.111	0.190	0.0379	0.0220	0.343
Survey item FE	No	No	No	No	No	No	No	No	No	Yes
Items	1	1	1	1	1	1	1	1	1	9
Respondents	784	774	728	729	784	787	785	775	747	794
Respondent-items	784	774	728	729	784	787	785	775	747	6,893

All models are linear probability models where the dependent variable indicates whether the response to a survey item is correct. Under the Confidence Coding condition, we only consider responses as correct when they are chosen with complete confidence (10 on a 0–10 scale). The baseline conditions are Conditions 1–4 pooled together (see Table 1 for the descriptions). Columns (1)–(9) are for each of the survey questions. The model in column (10) pools all nine survey questions. See Table 6 for a similar result using MTurk 2. See Tables SM 2.2 to SM 2.5 for the results comparing the Confidence Coding condition to each of the four other individual survey conditions. See Figure SM 2.1 for the visualization of how Confidence Coding mediates the effect that congenial responses have. Standard errors are clustered at the respondent level. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. Exact p-values in square brackets.

Table SM 2.2: Confidence Coding vs. Condition 1 (MTurk 1)

	Obama birthplace	Obama religion	ACA illegal	ACA death panels	GW causes GW causes	GW scientists agree	Voter fraud	MMR vaccine	Budget deficit	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Congenial	0.328***	0.415***	0.490***	0.271***	0.556***	0.224**	0.683***	0.147*	0.046	0.351***
	(0.071)	(0.077)	(0.078)	(0.080)	(0.074)	(0.073)	(0.066)	(0.062)	(0.047)	(0.035)
	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.002]	[0.000]	[0.018]	[0.321]	[0.000]
Confidence Coding (CCD)	−0.006	−0.067*	−0.170***	−0.022	−0.055*	−0.069*	−0.081*	−0.044+	−0.044	−0.063***
	(0.024)	(0.032)	(0.039)	(0.054)	(0.027)	(0.034)	(0.038)	(0.025)	(0.051)	(0.015)
	[0.797]	[0.036]	[0.000]	[0.684]	[0.040]	[0.042]	[0.032]	[0.084]	[0.397]	[0.000]
Congenial × CCD	−0.154	−0.244*	−0.343***	−0.264*	−0.308**	−0.228**	−0.541***	−0.133*	0.027	−0.243***
	(0.096)	(0.101)	(0.099)	(0.109)	(0.102)	(0.078)	(0.089)	(0.067)	(0.075)	(0.046)
	[0.111]	[0.017]	[0.001]	[0.016]	[0.003]	[0.004]	[0.000]	[0.048]	[0.723]	[0.000]
Constant	0.032+	0.085**	0.170***	0.149***	0.064*	0.096**	0.117***	0.053*	0.894***	0.177***
	(0.018)	(0.029)	(0.039)	(0.037)	(0.025)	(0.031)	(0.033)	(0.023)	(0.032)	(0.014)
	[0.081]	[0.004]	[0.000]	[0.000]	[0.012]	[0.002]	[0.001]	[0.023]	[0.000]	[0.000]
R^2	0.169	0.236	0.316	0.0823	0.360	0.126	0.435	0.0816	0.0117	0.436
Survey item FE	No	No	No	No	No	No	No	No	No	Yes
Items	1	1	1	1	1	1	1	1	1	9
Respondents	300	290	244	245	300	303	301	291	263	310
Respondent-items	300	290	244	245	300	303	301	291	263	2,537

All models are linear probability models where the dependent variable indicates whether the response to a survey item is correct. Under the Confidence Coding condition, we only consider responses as correct when they are chosen with complete confidence (10 on a 0–10 scale). The baseline condition is Condition 1 (see Table 1 for the descriptions). Columns (1)–(9) are for each of the survey questions. The model in column (10) pools all nine survey questions. See Table 6 for a similar result using MTurk 2. See Table SM 2.1 for the results comparing the Confidence Coding condition with all the four other conditions pooled together. See Figure SM 2.2 for the visualization of how Confidence Coding mediates the effect that congenial responses have. See Figure SM 2.2 for the visualization of how Confidence Coding mediates the effect that congenial responses have. Standard errors are clustered at the respondent level. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. Exact p-values in square brackets.

Table SM 2.3: Confidence Coding vs. Condition 2 (MTurk 1)

	Obama birthplace	Obama religion	ACA illegal	ACA death panels	GW causes GW causes	GW scientists agree	Voter fraud	MMR vaccine	Budget deficit	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Congenial	0.443***	0.586***	0.465***	0.208*	0.569***	0.309***	0.497***	0.047	0.251***	0.375***
	(0.070)	(0.071)	(0.077)	(0.082)	(0.072)	(0.068)	(0.075)	(0.062)	(0.060)	(0.030)
	[0.000]	[0.000]	[0.000]	[0.011]	[0.000]	[0.000]	[0.000]	[0.443]	[0.000]	[0.000]
Confidence Coding (CCD)	0.016	−0.094**	−0.214***	−0.118*	−0.083**	−0.004	−0.128**	−0.113**	0.177**	−0.063***
	(0.018)	(0.035)	(0.042)	(0.059)	(0.031)	(0.023)	(0.042)	(0.035)	(0.062)	(0.018)
	[0.385]	[0.007]	[0.000]	[0.047]	[0.007]	[0.853]	[0.002]	[0.001]	[0.005]	[0.000]
Congenial × CCD	−0.268**	−0.415***	−0.318**	−0.201+	−0.321**	−0.313***	−0.355***	−0.033	−0.178*	−0.264***
	(0.095)	(0.097)	(0.098)	(0.110)	(0.101)	(0.073)	(0.096)	(0.067)	(0.084)	(0.042)
	[0.005]	[0.000]	[0.001]	[0.069]	[0.002]	[0.000]	[0.000]	[0.621]	[0.035]	[0.000]
Constant	0.010	0.112***	0.214***	0.245***	0.092**	0.031+	0.163***	0.122***	0.673***	0.178***
	(0.010)	(0.032)	(0.042)	(0.044)	(0.029)	(0.018)	(0.038)	(0.033)	(0.048)	(0.017)
	[0.319]	[0.001]	[0.000]	[0.000]	[0.002]	[0.081]	[0.000]	[0.000]	[0.000]	[0.000]
R^2	0.262	0.380	0.308	0.0790	0.369	0.187	0.287	0.0592	0.0761	0.377
Survey item FE	No	No	No	No	No	No	No	No	No	Yes
Items	1	1	1	1	1	1	1	1	1	9
Respondents	307	297	251	252	307	310	308	298	270	317
Respondent-items	307	297	251	252	307	310	308	298	270	2,600

All models are linear probability models where the dependent variable indicates whether the response to a survey item is correct. Under the Confidence Coding condition, we only consider responses as correct when they are chosen with complete confidence (10 on a 0–10 scale). The baseline condition is Condition 2 (see Table 1 for the descriptions). Columns (1)–(9) are for each of the survey questions. The model in column (10) pools all nine survey questions. See Table 6 for a similar result using MTurk 2. See Table SM 2.1 for the results comparing the Confidence Coding condition with all the four other conditions pooled together. See Figure SM 2.3 for the visualization of how Confidence Coding mediates the effect that congenial responses have. Standard errors are clustered at the respondent level. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. Exact p-values in square brackets.

Table SM 2.4: Confidence Coding vs. Condition 3 (MTurk 1)

	Obama birthplace	Obama religion	ACA illegal	ACA death panels	GW causes GW causes	GW scientists agree	Voter fraud	MMR vaccine	Budget deficit	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Congenial	0.127*	0.291***	0.238***	0.165*	0.355***	0.173**	0.213**	0.101+	−0.056	0.179***
	(0.052)	(0.071)	(0.070)	(0.064)	(0.074)	(0.061)	(0.072)	(0.054)	(0.046)	(0.029)
	[0.016]	[0.000]	[0.001]	[0.010]	[0.000]	[0.005]	[0.003]	[0.061]	[0.225]	[0.000]
Confidence Coding (CCD)	−0.008	−0.083**	−0.102***	0.042	−0.119***	−0.033	−0.108**	−0.050*	−0.099*	−0.065***
	(0.022)	(0.031)	(0.028)	(0.047)	(0.032)	(0.027)	(0.037)	(0.024)	(0.045)	(0.015)
	[0.720]	[0.007]	[0.000]	[0.376]	[0.000]	[0.215]	[0.004]	[0.038]	[0.029]	[0.000]
Congenial × CCD	0.047	−0.120	−0.091	−0.159	−0.107	−0.177**	−0.071	−0.087	0.129+	−0.069+
	(0.084)	(0.097)	(0.093)	(0.098)	(0.102)	(0.066)	(0.094)	(0.060)	(0.075)	(0.041)
	[0.572]	[0.217]	[0.330]	[0.106]	[0.296]	[0.008]	[0.450]	[0.146]	[0.085]	[0.096]
Constant	0.034*	0.102***	0.102***	0.085**	0.127***	0.059**	0.144***	0.059**	0.949***	0.179***
	(0.017)	(0.028)	(0.028)	(0.026)	(0.031)	(0.022)	(0.033)	(0.022)	(0.020)	(0.014)
	[0.044]	[0.000]	[0.000]	[0.001]	[0.000]	[0.007]	[0.000]	[0.007]	[0.000]	[0.000]
R^2	0.0680	0.146	0.117	0.0326	0.202	0.0810	0.0935	0.0521	0.0200	0.428
Survey item FE	No	No	No	No	No	No	No	No	No	Yes
Items	1	1	1	1	1	1	1	1	1	9
Respondents	330	320	274	275	330	333	331	321	293	340
Respondent-items	330	320	274	275	330	333	331	321	293	2,807

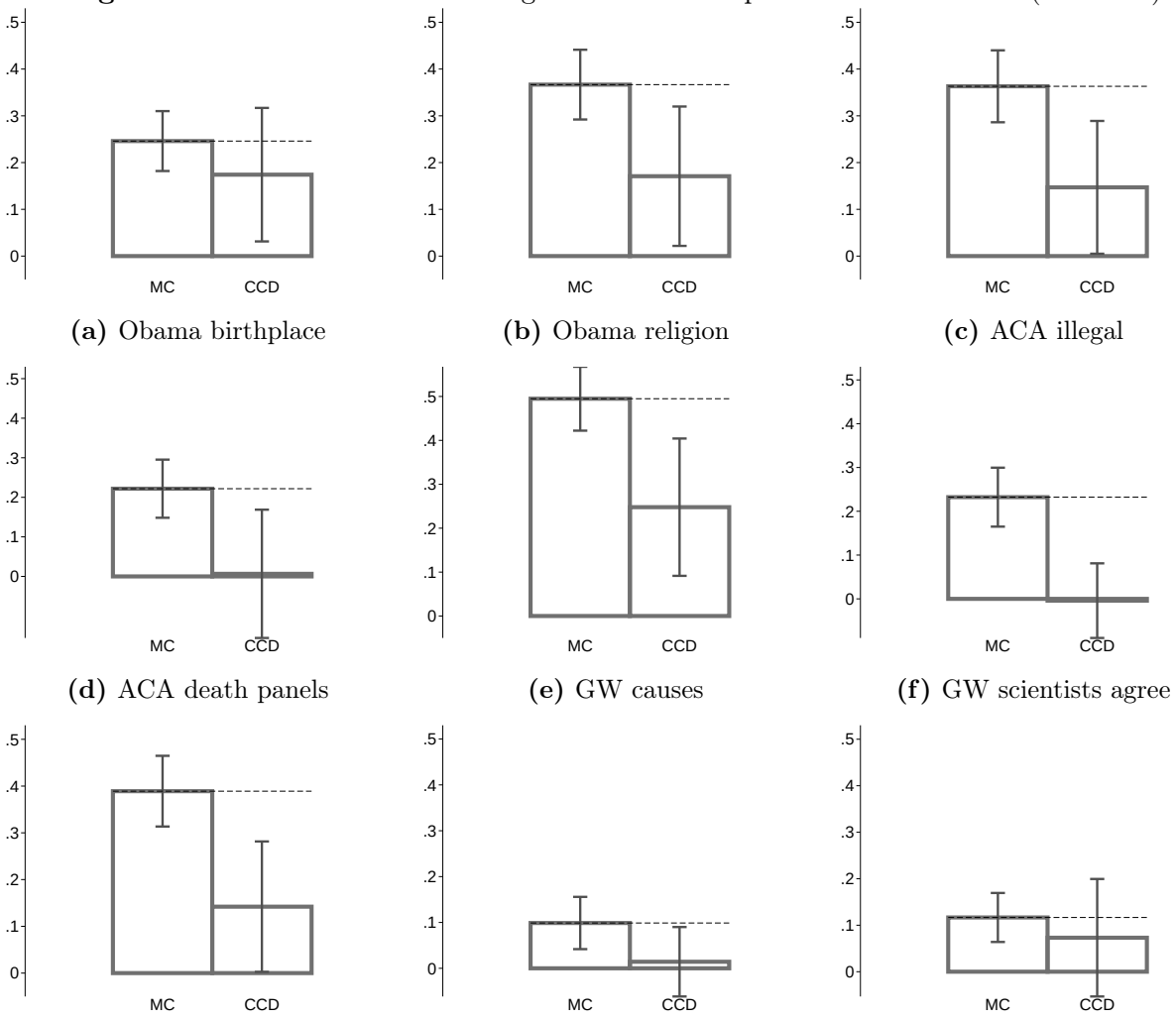
All models are linear probability models where the dependent variable indicates whether the response to a survey item is correct. Under the Confidence Coding condition, we only consider responses as correct when they are chosen with complete confidence (10 on a 0–10 scale). The baseline condition is Condition 3 (see Table 1 for the descriptions). Columns (1)–(9) are for the survey questions. The model in column (10) pools all nine survey questions. See Table 6 for a similar result using MTurk 2. See Table SM 2.1 for the results comparing the Confidence Coding condition with all the four other conditions pooled together. See Figure SM 2.4 for the visualization of how Confidence Coding mediates the effect that congenial responses have. Standard errors are clustered at the respondent level. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. Exact p-values in square brackets.

Table SM 2.5: Confidence Coding vs. Condition 4 (MTurk 1)

	Obama birthplace	Obama religion	ACA illegal	ACA death panels	GW causes GW causes	GW scientists agree	Voter fraud	MMR vaccine	Budget deficit	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Congenial	0.086	0.164*	0.256**	0.230**	0.512***	0.230**	0.157*	0.095+	0.240***	0.219***
	(0.057)	(0.075)	(0.081)	(0.074)	(0.076)	(0.074)	(0.070)	(0.056)	(0.057)	(0.033)
	[0.131]	[0.029]	[0.002]	[0.002]	[0.000]	[0.002]	[0.025]	[0.092]	[0.000]	[0.000]
Confidence Coding (CCD)	−0.037	−0.116**	−0.170***	0.037	−0.054*	−0.063*	−0.063+	−0.044+	0.154*	−0.042**
	(0.027)	(0.035)	(0.036)	(0.048)	(0.025)	(0.031)	(0.033)	(0.023)	(0.059)	(0.015)
	[0.182]	[0.001]	[0.000]	[0.437]	[0.029]	[0.043]	[0.062]	[0.061]	[0.010]	[0.008]
Congenial × CCD	0.088	0.007	−0.109	−0.223*	−0.264*	−0.234**	−0.015	−0.081	−0.167*	−0.109*
	(0.087)	(0.100)	(0.102)	(0.105)	(0.104)	(0.078)	(0.092)	(0.062)	(0.082)	(0.044)
	[0.313]	[0.945]	[0.285]	[0.034]	[0.012]	[0.003]	[0.871]	[0.192]	[0.042]	[0.014]
Constant	0.063**	0.134***	0.170***	0.089**	0.062**	0.089**	0.098***	0.054*	0.696***	0.155***
	(0.023)	(0.032)	(0.036)	(0.027)	(0.023)	(0.027)	(0.028)	(0.021)	(0.044)	(0.014)
	[0.007]	[0.000]	[0.000]	[0.001]	[0.007]	[0.001]	[0.001]	[0.013]	[0.000]	[0.000]
R^2	0.0510	0.0843	0.137	0.0555	0.314	0.119	0.0589	0.0464	0.0666	0.363
Survey item FE	No	No	No	No	No	No	No	No	No	Yes
Items	1	1	1	1	1	1	1	1	1	9
Respondents	315	305	259	260	315	318	316	306	278	325
Respondent-items	315	305	259	260	315	318	316	306	278	2,672

All models are linear probability models where the dependent variable indicates whether the response to a survey item is correct. Under the Confidence Coding condition, we only consider responses as correct when they are chosen with complete confidence (10 on a 0–10 scale). The baseline condition is Condition 4 (see Table 1 for the descriptions). Columns (1)–(9) are for each of the survey questions. The model in column (10) pools all nine survey questions. See Table 6 for a similar result using MTurk 2. See Table SM 2.1 for the results comparing the Confidence Coding condition with all the four other conditions pooled together. See Figure SM 2.5 for the visualization of how Confidence Coding mediates the effect that congenial responses have. Standard errors are clustered at the respondent level. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. Exact p-values in square brackets.

Figure SM 2.1: Confidence Coding vs. Other Multiple Choice Conditions (MTurk 1)



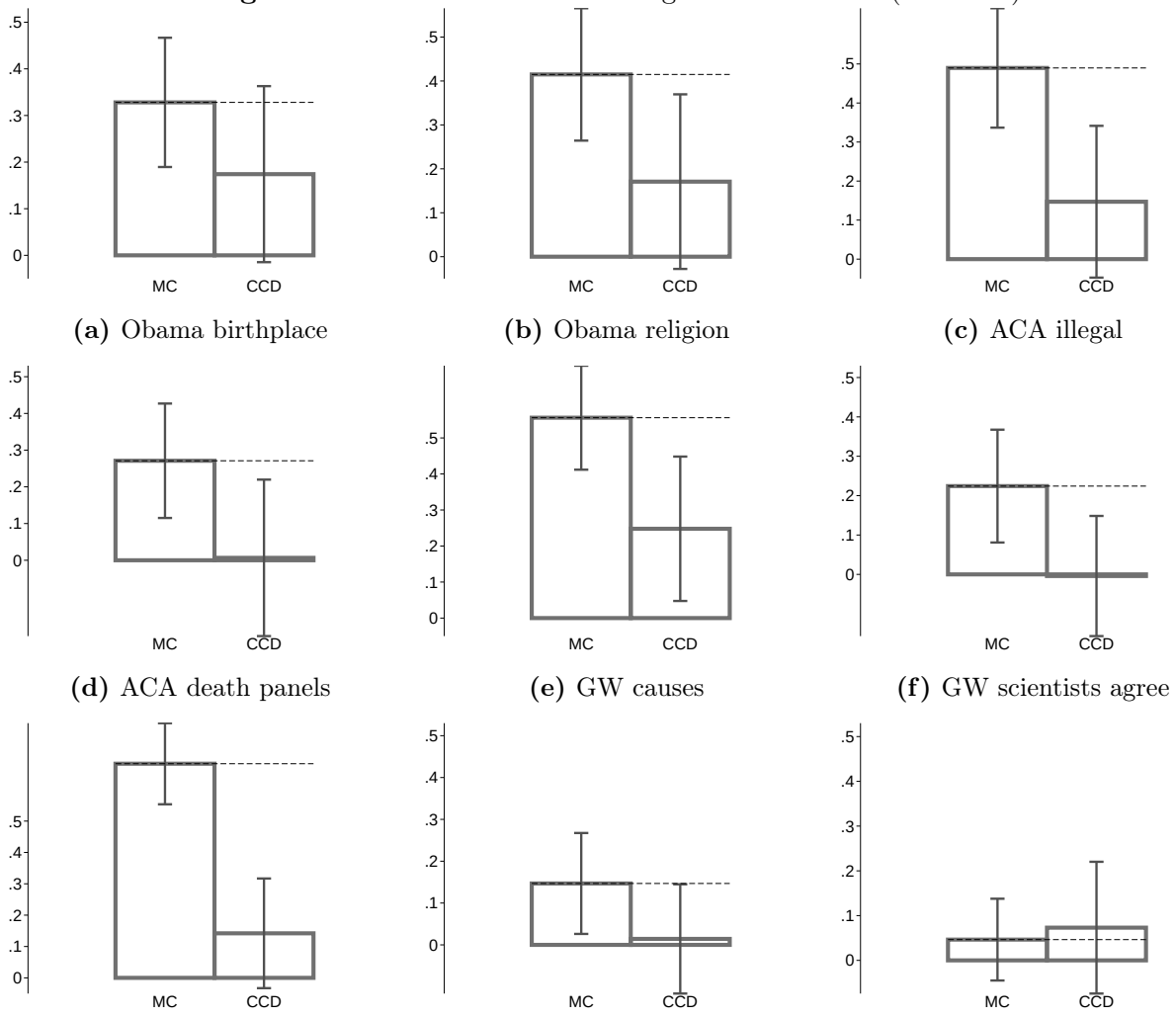
(a) Obama birthplace **(b)** Obama religion **(c)** ACA illegal

(d) ACA death panels **(e)** GW causes **(f)** GW scientists agree

(g) Voter fraud **(h)** MMR vaccine **(i)** Budget deficit

Bars indicate the predicted percent of correct responses when the correct response is congenial to the party, depending on whether the survey condition is based on Confidence Coding (CCD) or from the four Multiple Choice conditions (see Table 1 for the descriptions). Reconstructed from the estimates from Table SM 2.1. Capped vertical bars indicate 95% confidence intervals.

Figure SM 2.2: Confidence Coding vs. Condition 1 (MTurk 1)



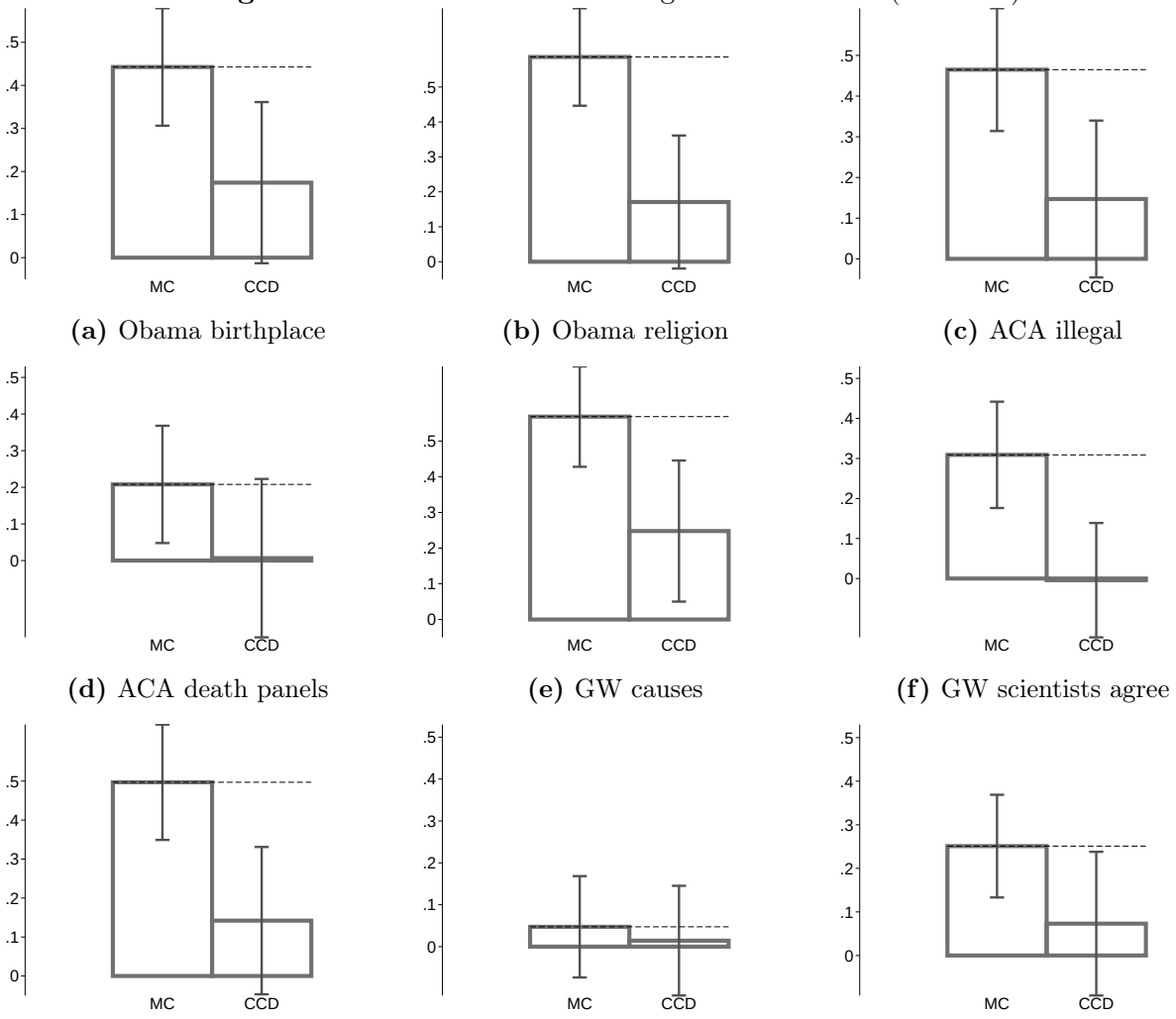
(a) Obama birthplace **(b)** Obama religion **(c)** ACA illegal

(d) ACA death panels **(e)** GW causes **(f)** GW scientists agree

(g) Voter fraud **(h)** MMR vaccine **(i)** Budget deficit

Bars indicate the predicted percent of correct responses when the correct response is congenial to the party, depending on whether the survey condition is based on Confidence Coding (CCD) or from the multiple choice Condition 1 (see Table 1 for the descriptions). Reconstructed from the estimates from Table SM 2.2. Capped vertical bars indicate 95% confidence intervals.

Figure SM 2.3: Confidence Coding vs. Condition 2 (MTurk 1)



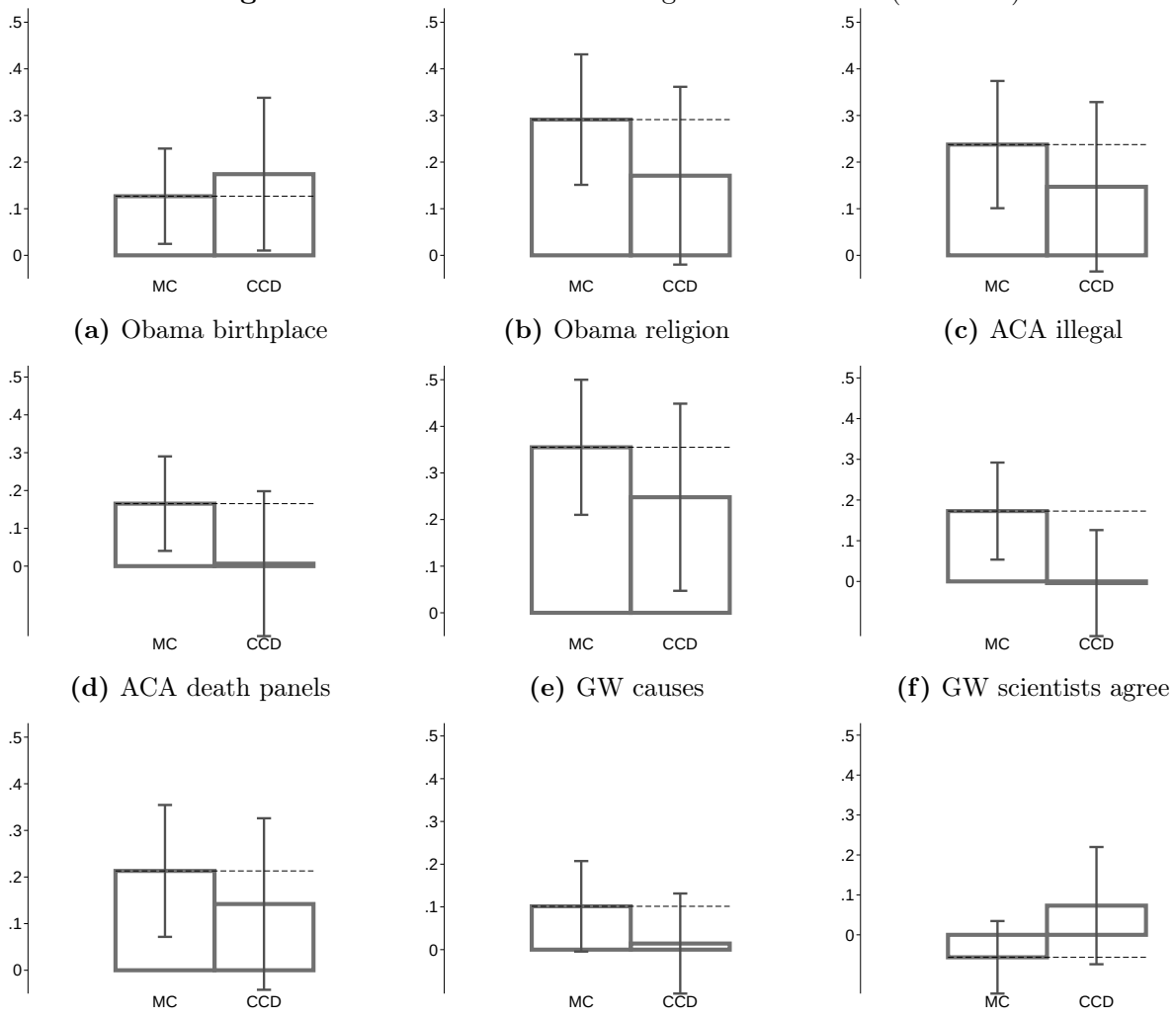
(a) Obama birthplace **(b)** Obama religion **(c)** ACA illegal

(d) ACA death panels **(e)** GW causes **(f)** GW scientists agree

(g) Voter fraud **(h)** MMR vaccine **(i)** Budget deficit

Bars indicate the predicted percent of correct responses when the correct response is congenial to the party, depending on whether the survey condition is based on Confidence Coding (CCD) or from the multiple choice Condition 2 (see Table 1 for the descriptions). Reconstructed from the estimates from Table SM 2.3. Capped vertical bars indicate 95% confidence intervals.

Figure SM 2.4: Confidence Coding vs. Condition 3 (MTurk 1)



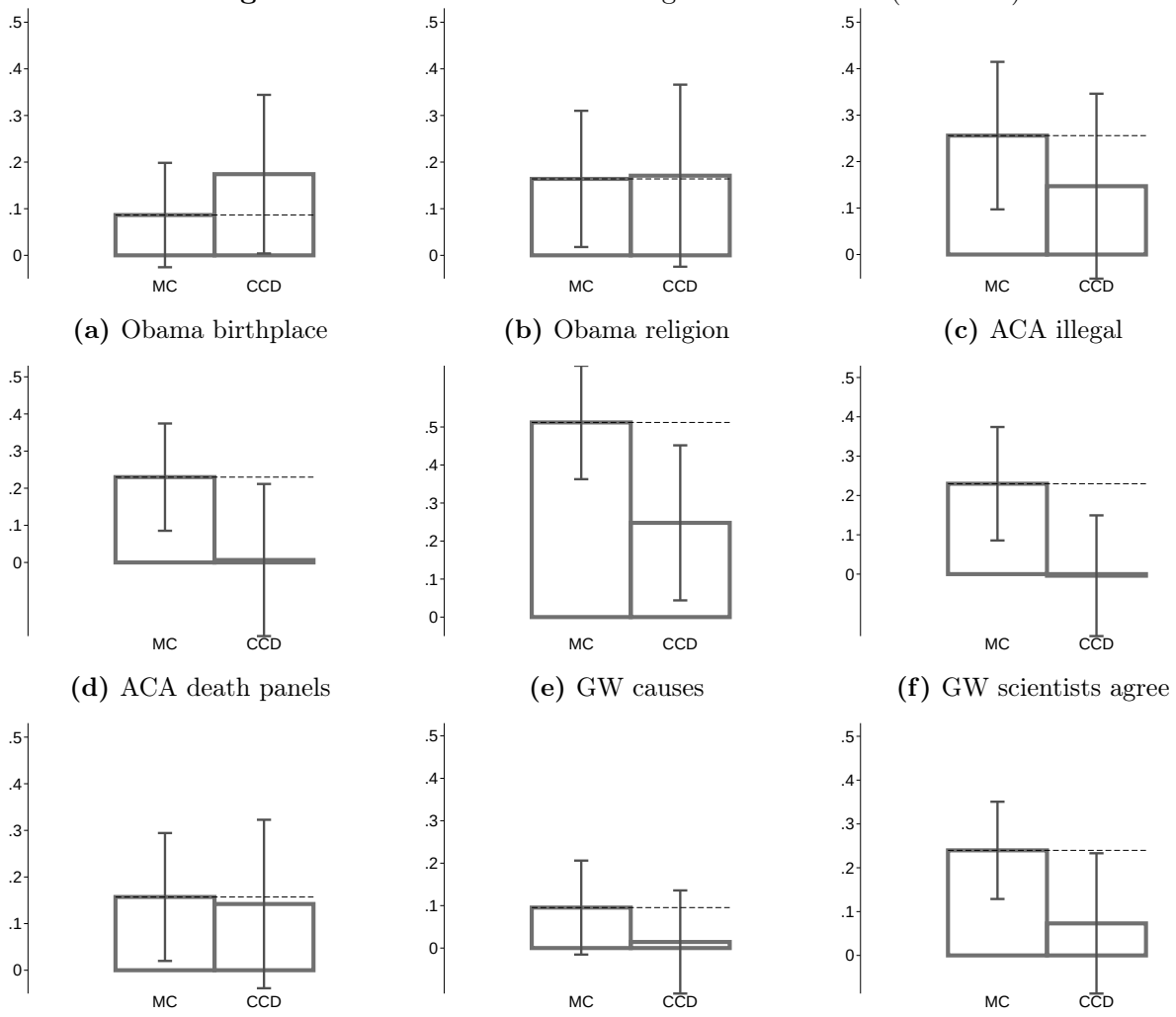
(a) Obama birthplace **(b)** Obama religion **(c)** ACA illegal

(d) ACA death panels **(e)** GW causes **(f)** GW scientists agree

(g) Voter fraud **(h)** MMR vaccine **(i)** Budget deficit

Bars indicate the predicted percent of correct responses when the correct response is congenial to the party, depending on whether the survey condition is based on Confidence Coding (CCD) or from the multiple choice Condition 3 (see Table 1 for the descriptions). Reconstructed from the estimates from Table SM 2.4. Capped vertical bars indicate 95% confidence intervals.

Figure SM 2.5: Confidence Coding vs. Condition 4 (MTurk 1)



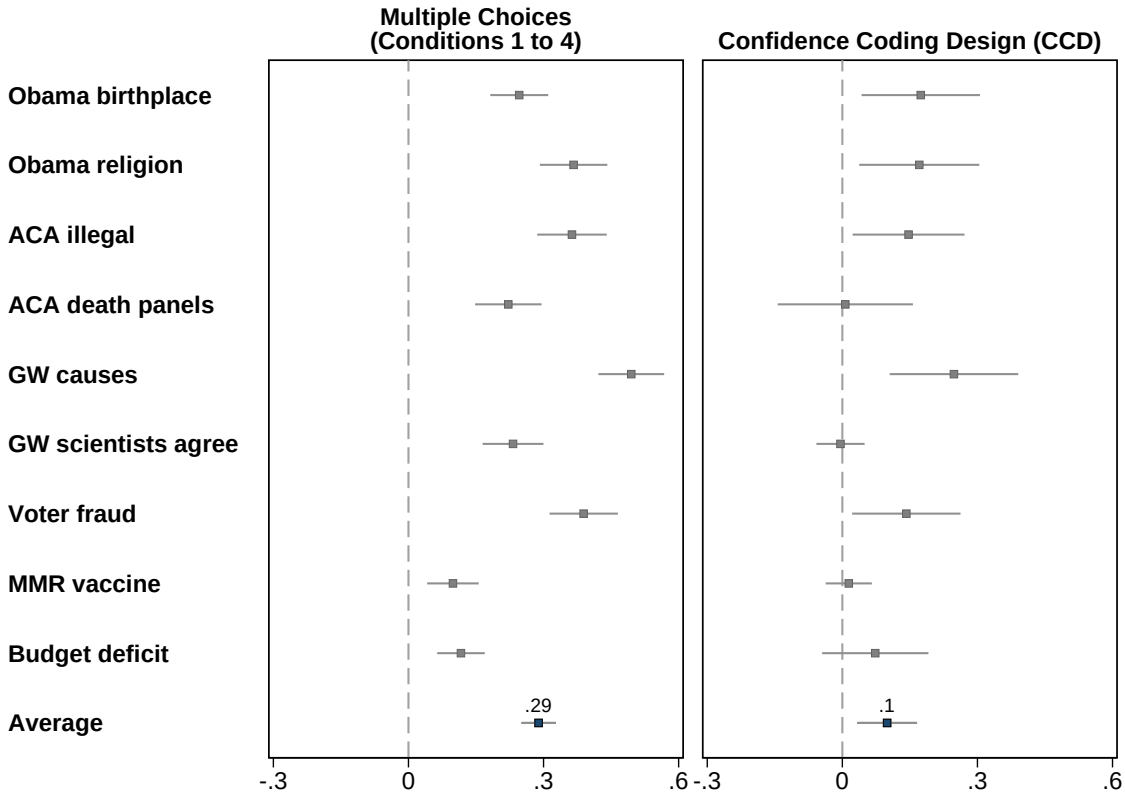
(a) Obama birthplace **(b)** Obama religion **(c)** ACA illegal

(d) ACA death panels **(e)** GW causes **(f)** GW scientists agree

(g) Voter fraud **(h)** MMR vaccine **(i)** Budget deficit

Bars indicate the predicted percent of correct responses when the correct response is congenial to the party, depending on whether the survey condition is based on Confidence Coding (CCD) or from the multiple choice Condition 4 (see Table 1 for the descriptions). Reconstructed from the estimates from Table SM 2.4. Capped vertical bars indicate 95% confidence intervals.

Figure SM 2.6: Partisan Gaps in Knowledge Across Question Designs (Pooled multiple choices)



The figure shows the estimated partisan gaps in knowledge from MTurk 1 for the four multiple-choice survey conditions (pooling Conditions 1–4, see Table 1) and the confidence coding design (CCD). Corresponds to Figure 2. The CCD condition only considers selecting the right answer with complete confidence as evidence that the respondent knows the answer (see Appendix SM 5). See Tables SM 2.1 to SM 2.5 in Appendix SM 2 for the regression estimates of the multiple-choice conditions to the confidence coding condition.

SM 3 Item Text for MTurk 1

For exposition, we present the conditions here using different terms than in the main body (see Table 1). The following shows how our terminologies for conditions map to the MTurk questionnaires.

- **Condition 1 = NDK+SP+GE+NI**
- **Condition 2 = NDK+NSP+GE+NI**
- **Condition 3 = DK+NSP+GD+NI**
- **Condition 4 = DK+NSP+GD+NNI**
- **Condition 5 = CCD**

Preface for Different Conditions

- **Conditions 1 and 2**
  Now here are some questions about what you may know about politics and public affairs.
- **Conditions 3, 4, and 5**
  Now here are some questions about what you may know about politics and public affairs.

  We are interested in measuring what people currently know and can recall on their own and are just as interested in what people don't know as in what they do know. So we'd like your agreement to just say "don't know" if you don't know the answer—without looking anything up or talking with anyone about it.

Item Text for Condition 5 (CCD)

Now here are a series of statements. On a scale of 0 to 10, where 0 means definitely false, 10 means definitely true, and 5 is exactly in the middle, how definitely true or false is each statement?

- Barack Obama was born in the US (T)
- Barack Obama is a Muslim (F)
- The Affordable Care Act gives illegal immigrants financial help to buy health insurance (F)
- The Affordable Care Act does not create government panels to make decisions about end-of-life care (T)
- Temperatures around the world are increasing because of human activity, like burning coal and gasoline (T)

- Most climate scientists believe that global warming is not occurring (F)
- In the 2016 presidential election, President Trump won the majority of the legally cast votes (F)
- The vaccine for measles, mumps, and rubella (MMR) causes autism in children. (F)
- Since 2012, the annual federal budget deficit has increased. (T)

Rest of the Conditions, By Item

- Obama's Birthplace
  - **Conditions 1 and 2**
    According to the Constitution, American presidents must be "natural-born citizens." Some people believe Barack Obama was not born in the United States but was born in another country. Do you think Barack Obama was born in . . . ?
    * The US
    * Another country
  - **Condition 3**
    Some people believe Barack Obama was not born in the United States but was born in another country. Was he born in . . . ?
    * The US
    * Another country
    * DK (plus DK pref)
  - **Condition 4**
    Was Barack Obama born in . . . ?
    * the US
    * Another country
    * DK (plus DK pref)
- Obama Religion
  - **Condition 1**
    Most people have a religion. Some people believe Barack Obama is a Muslim. Do you personally believe that Barack Obama is a . . . ?
    * Muslim
    * Christian
  - **Condition 2**
    Do you personally believe that Barack Obama is a . . . ?
    * Muslim

    - Christian
  - **Condition 3**
    Some people believe Barack Obama is a Muslim. Is he a . . . ?
    - Muslim
    - Christian
    - DK (+ DK pref)
  - **Condition 4**
    Is Barack Obama a . . . ?
    - Muslim
    - Christian
    - DK (plus DK pref)

- ACA Illegal
  - **Condition 1**
    As you may know, there is currently talk of changing the Affordable Care Act (ACA), enacted in 2010. Some people believe that the ACA gives illegal immigrants financial help to buy health insurance. To the best of your knowledge, would you say the ACA. . . ?
    - Gives illegal immigrants financial help to buy health insurance
    - Does not give illegal immigrants financial help to buy health insurance
  - **Condition 2**
    To the best of your knowledge, would you say the Affordable Care Act . . . ?
    - Gives illegal immigrants financial help to buy health insurance
    - Does not give illegal immigrants financial help to buy health insurance
  - **Condition 3**
    Some people believe that the Affordable Care Act gives illegal immigrants financial help to buy health insurance. Does the Affordable Care Act . . . ?
    - Give illegal immigrants financial help to buy health insurance
    - Not give illegal immigrants financial help to buy health insurance
    - DK (+ DK pref)
  - **Condition 4**
    Does the Affordable Care Act . . . ?
    - Give illegal immigrants financial help to buy health insurance
    - Not Give illegal immigrants financial help to buy health insurance
    - Don't know (+ DK pref)

- ACA—Death Panels

- **Condition 1**
  Some people believe that the Affordable Care Act establishes a government panel to make decisions about end-of-life care. To the best of your knowledge, would you say that the Affordable Care Act . . . ?
  * Creates government panels to make decisions about end-of-life care
  * Does not create government panels to make decisions about end-of-life care
- **Condition 2**
  To the best of your knowledge, would you say that the Affordable Care Act . . . ?
  * Creates government panels to make decisions about end-of-life care
  * Does not create government panels to make decisions about end-of-life care
- **Condition 3**
  Some people believe that the Affordable Care Act establishes a government panel to make decisions about end-of-life care. Does the Affordable Care Act . . . ?
  * Creates government panels to make decisions about end-of-life care
  * Does not create government panels to make decisions about end-of-life care
  * DK (+ DK pref)
- **Condition 4**
  Does the Affordable Care Act . . . ?
  * Creates government panels to make decisions about end-of-life care
  * Does not create government panels to make decisions about end-of-life care
  * DK (+ DK pref)

- Global Warming—Happening + Causes
  - **Condition 1**
    Recently, you may have noticed that global warming has been getting some attention in the news. Some people believe that temperatures are increasing around the world because of natural variation over time, such as that produced the ice age. Which of the following best fits your view about this? Would you say that temperatures around the world are. . . ?
    * Increasing because of the natural variation over time, such as produced by the ice age
    * Increasing because of human activity, like burning coal and gasoline
    * Staying about the same as they have been
  - **Condition 2**
    Which of the following best fits your view about this? Are temperatures around the world . . . ?
    * Increasing because of the natural variation over time, such as produced by the ice age

    - Increasing because of human activity, like burning coal and gasoline
    - Staying about the same as they have been
  - **Condition 3**
    Some people believe that temperatures are increasing around the world because of natural variation over time, such as produced the ice age. Are temperatures around the world . . . ?
    - Increasing because of the natural variation over time, such as produced by the ice age
    - Increasing because of human activity, like burning coal and gasoline
    - Staying about the same as they have been
    - DK (+ DK pref)
  - **Condition 4**
    Are temperatures around the world . . . ?
    - Increasing because of natural variation over time, such as produced by the ice age
    - Increasing because human activity, like burning coal and gasoline
    - Staying about the same as they have been
    - DK (+ DK pref)

- GW—Scientist Agreement
  - **Condition 1**
    As you may know, the term "global warming" refers to the claim that temperatures have been increasing around the world. Some people believe that most climate scientists believe that global warming is not occurring. Just your impression, which one of the following statements do you think is most accurate?
    - Most climate scientists believe that global warming is occurring.
    - Most climate scientists believe that global warming is not occurring.
    - Climate scientists are about equally divided about whether global warming is occurring or not
  - **Condition 2**
    Just your impression, which one of the following statements do you think is most accurate?
    - Most climate scientists believe that global warming is occurring.
    - Most climate scientists believe that global warming is not occurring.
    - Climate scientists are about equally divided about whether global warming is occurring or not
  - **Condition 3**
    Some people believe that most climate scientists believe that global warming is not occurring. Which one of the following statements is most accurate?

    - Most climate scientists believe that global warming is occurring.
    - Most climate scientists believe that global warming is not occurring.
    - Climate scientists are about equally divided about whether global warming is occurring or not
    - DK (+ DK pref)
  - **Condition 4**
    Which one of the following statements is most accurate?
    - Most climate scientists believe that global warming is occurring.
    - Most climate scientists believe that global warming is NOT occurring.
    - Climate scientists are about equally divided about whether global warming is occurring or not
    - DK (+ DK pref)

- Voter Fraud
  - **Condition 1**
    As you may know, not everyone living in the US has the legal right to vote. President Trump has said that several million people voted illegally in the 2016 presidential election and that he won the majority of the legally cast votes. Do think that President Trump . . . ?
    - Won the majority of the legally cast votes
    - Did not win the majority of the legally cast votes
  - **Condition 2**
    As you may know, President Trump has said that several million people voted illegally in the 2016 presidential election and that he won the majority of the legally cast votes. Do you believe that President Trump . . . ?
    - Won the majority of the legally cast votes
    - Did not win the majority of the legally cast votes
  - **Condition 3**
    As you may know, President Trump has said that several million people voted illegally in the 2016 presidential election and that he won the majority of the legally cast votes. Did President Trump . . . ?
    - Won the majority of the legally cast votes
    - Did not win the majority of the legally cast votes
    - DK (+ DK pref)
  - **Condition 4**
    In the 2016 presidential election, did President Trump . . . ?
    - Won the majority of the legally cast votes
    - Did not win the majority of the legally cast votes

    * DK (+ DK pref)

- Vaccines
  - **Condition 1**
    As you may know, most children receive the vaccine for Measles, Mumps, and Rubella (MMR). Some people believe that the MMR vaccine causes autism in children. From what you have read or heard, do you personally think that the MMR vaccine:
    * Causes autism in children
    * Does not cause autism in children
  - **Condition 2**
    From what you have read or heard, do you personally think that the vaccine for Measles, Mumps, and Rubella (MMR):
    * Causes autism in children
    * Does not cause autism in children
  - **Condition 3**
    Some people believe that the vaccine for Measles, Mumps, and Rubella (MMR) causes autism in children. Does the MMR vaccine . . . ?
    * Cause autism in children
    * Not cause autism in children.
    * DK (+ DK pref)
  - **Condition 4**
    Does the vaccine for Measles, Mumps, and Rubella (MMR) . . . ?
    * Cause autism in children
    * Not cause autism in children.
    * DK (+ DK pref)

- Obama—Budget Deficit
  - **Condition 1**
    As you may know, the federal government runs a deficit when it spends more than it takes in. Since 2012, with the Republicans having the majority in the U.S. House of Representatives, would you say that the annual federal budget deficit has . . .
    * Increased
    * Stayed about the same
    * Decreased

– **Condition 2**
As you may know, the federal government runs a deficit when it spends more than it takes in. Since 2012, would you say that the annual federal budget deficit has ...

  * Increased
  * Stayed about the same
  * Decreased

Condition 3
Since 2012, with the Republicans having the majority in the U.S. House of Representatives,

  * has the annual federal budget deficit ....
  * Increased
  * Stayed about the same
  * Decreased
  * DK (+ DK pref)

– **Condition 4**
Since 2012, has the annual federal budget deficit ...

  * Increased
  * Stayed about the same
  * Decreased
  * DK (+ DK pref)

SM 4 Criterion Variables (MTurk 1)

- Political Interest: On a scale from 0 to 10, where 0 is not at all, 10 is passionately, and 5 is exactly in the middle, how interested would you say you generally are in politics and public affairs?
- Vote: Again on a scale from 0 to 10, where now 0 means certain not to vote, 10 means certain to vote, and 5 is exactly in the middle, how likely would you say you are to vote in the next Congressional elections?
- What's the highest level of education you have obtained? No High School Diploma, High School Diploma or Equivalent, Some College, Four-year College Graduate, Post-graduate Degree

SM 5 Item Text for MTurk 2

The second Amazon MTurk survey was fielded in April 2017. We interviewed 1,059 participants. In this survey, we used new questions and probes to examine the effect of question design on (partisan) knowledge. We asked the participants four questions about the Affordable Care Act (2), the effect of greenhouse gases (1), and Donald Trump's recent executive order on immigration (1).

One-half of the survey respondents got a conventional closed-ended item with five options, including the opportunity to mark Don't know. The other half of the respondents had to assess the truth of statements on a scale from definitely false (0) to definitely true (10).

1. Does the Affordable Care Act ...?

- CE: Provide coverage for people who are currently in the country illegally, Replace private health insurance with a "single-payer system", **Increase the Medicare payroll tax for upper-income Americans**, Reimburse routine mammograms only for women older than 50, Don't know (5)
- Scale: Rating each response option above from definitely false (0) to definitely true (10). Don't know was not included. See Figure SM 5.1.

2. Are greenhouse gases ...?

- CE: A cause of respiratory problems, A cause of lung cancer, Damaging the ozone layer, **A cause of rising sea levels**, or Don't know
- Scale: Rating each response option above from definitely false (0) to definitely true (10). Don't know was not included. See Figure SM 5.2.

3. And does the Affordable Care Act ...?

- CE: Create government panels to make end-of-life decisions for people on Medicare, Replace Medicare with a "public option", **Limit future increases in payments to Medicare providers**, Cut benefits to existing Medicare patients, Don't know
- Scale: Rating each response option above from definitely false (0) to definitely true (10). Don't know was not included. See Figure SM 5.3.

4. Does President Trump's most recent executive order on immigration ...?

- CE: Subject immigrants living in the U.S. illegally to deportation, Strip immigrants from countries supporting terrorism of their green cards, Strip immigrants from several Muslim-majority countries of their green cards, **Temporarily ban immigrants from several majority-Muslim countries**, Don't know

- Scale: Rating each response option above from definitely false (0) to definitely true (10). Don't know was not included. See Figure SM 5.4.

If the close-ended questions 3 and 4 were not answered with Don't know the respondents received one of two follow-up questions:

- OE: What made you choose that response?
- CE: What made you choose that response? I asked someone I know, I looked it up, I've read, seen, or heard that, It makes me feel good to think that, It makes sense, in view of other things I know, I just thought I'd take a shot

Figure SM 5.1: Affordable Care Act 1 Scale Question

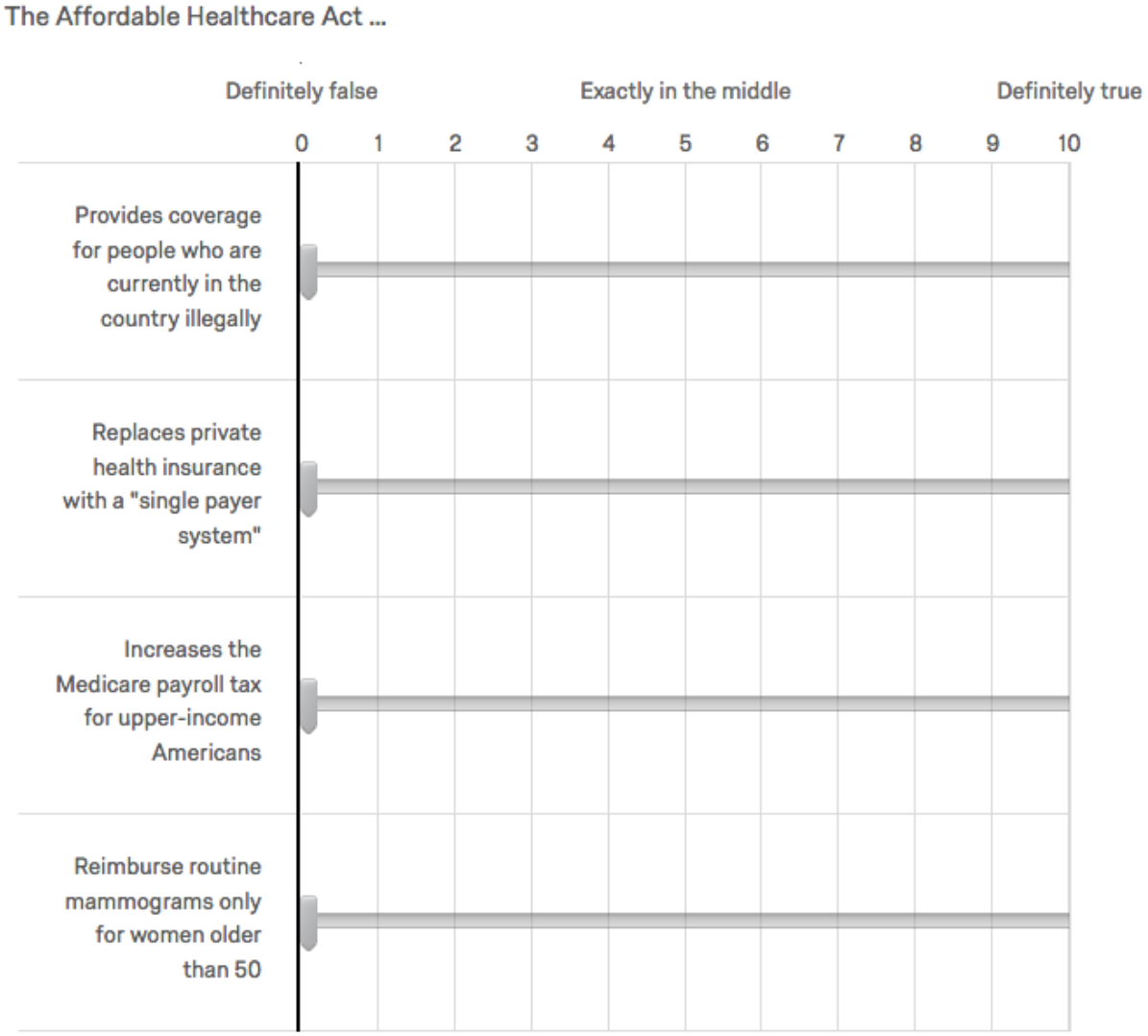


Figure SM 5.2: Greenhouse Gases Scale Question

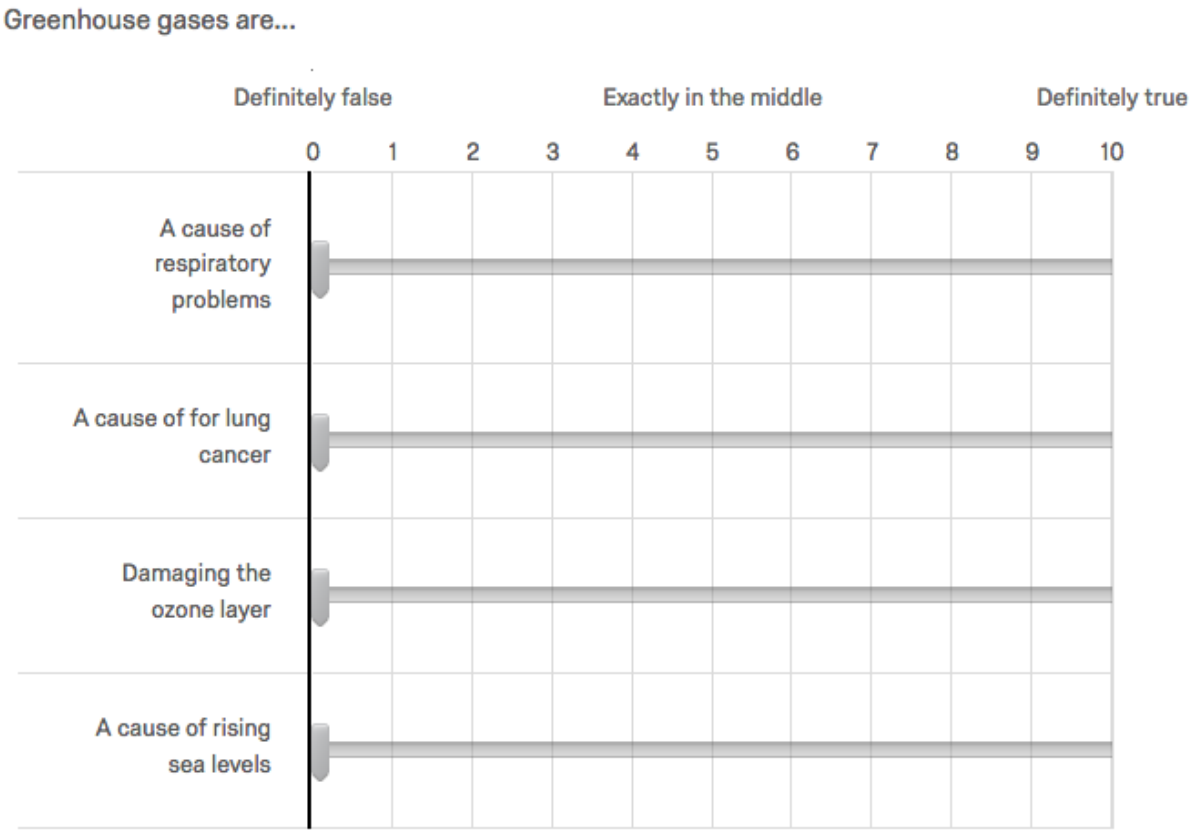


Figure SM 5.3: Affordable Care Act 2 Scale Question

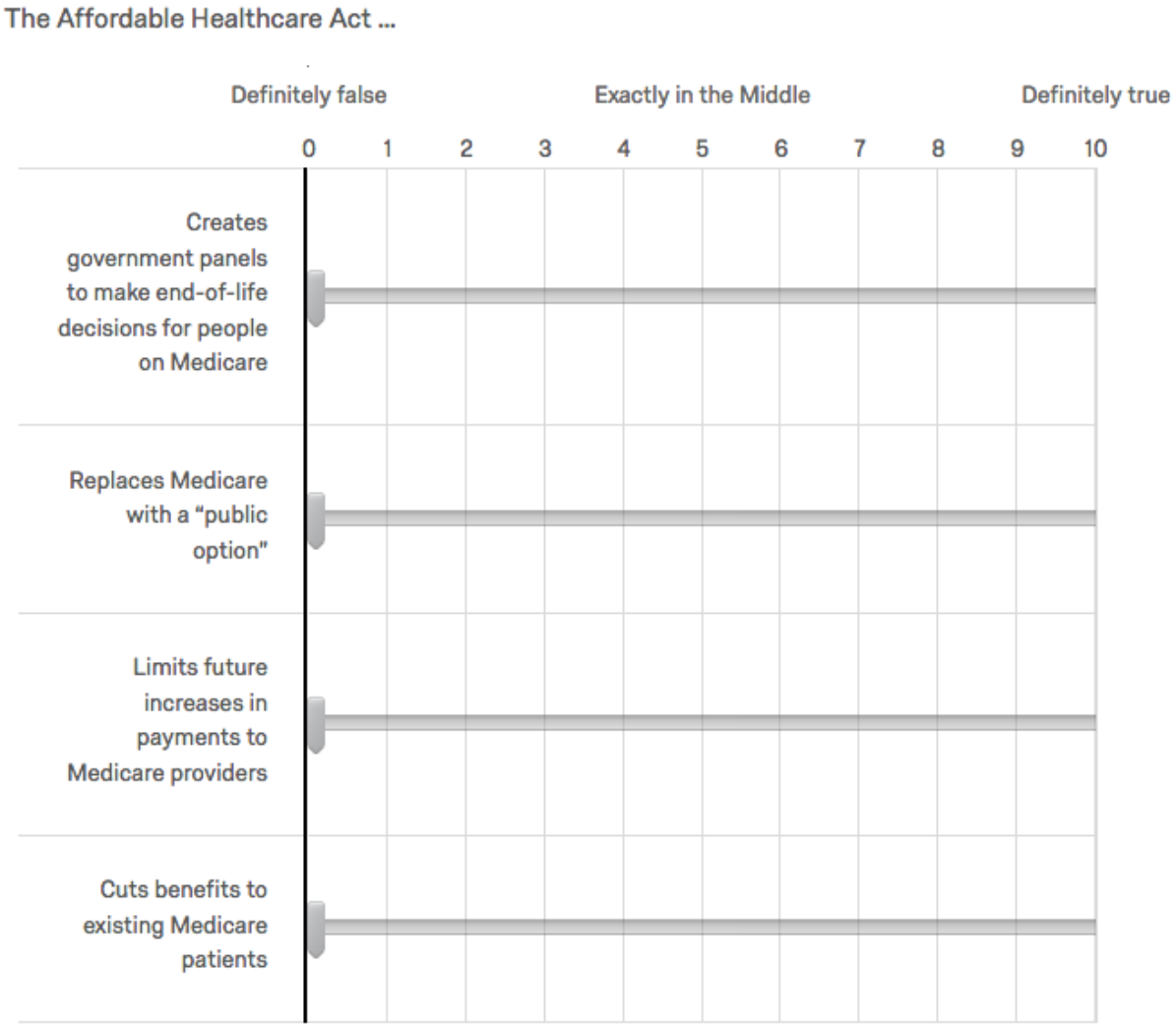
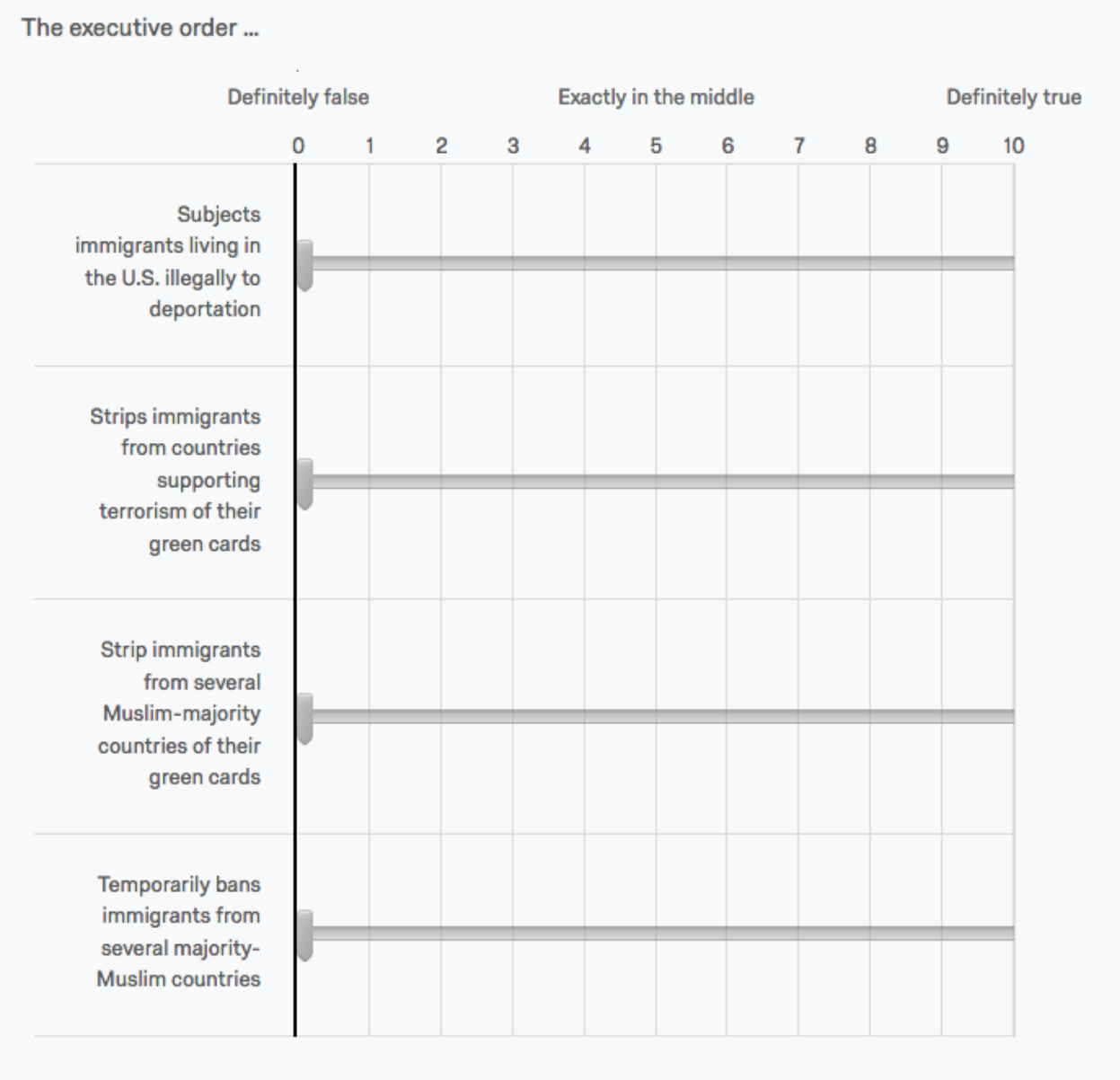


Figure SM 5.4: Executive Order Scale Question



SM 6 Alternate Scoring Criteria for CCD

Table SM 6.1 shows the proportion of correct answers across the Affordable Care Act questions (ACA and ACA2), the Greenhouse Gas question, and the question about Donald Trump's executive order. We report the proportion correct for closed questions in the multiple-choice format and the Confidence Coding at the thresholds of 8 and 10. For the Confidence Coding (CCD) to code an answer as correct the confidence for the correct answer had to be 8 (or 10), the scoring had to be the maximum number given, it had to be unique, and incorrect answers were not allowed to be scored higher than 2 (or 0).

Table SM 6.1: Proportion correct across questions and scoring

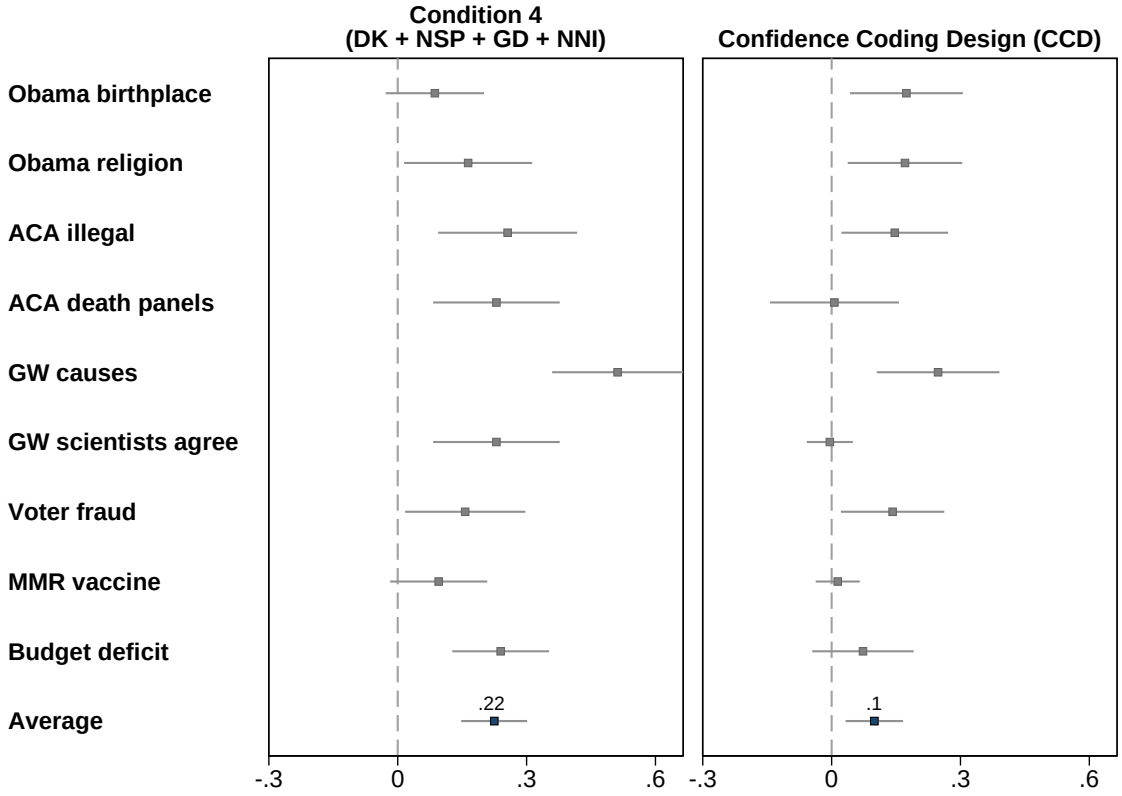
Question	**Closed**	**Relative Scoring**	
		8	**10**
ACA	0.24	0.01	0.01
ACA2	0.26	0.04	0.01
GG	0.25	0.02	0.01
DT	0.78	0.10	0.07

Table SM 6.2: Robustness check for Confidence Scoring and Knowledge Gaps: MTurk Study 2

	ACA	ACA2	GG	DT	All
Congenial	0.09*	0.08*	0.09*	0.00	0.03
	[0.02; 0.17]	[0.01; 0.16]	[0.01; 0.17]	[−0.07; 0.08]	[−0.02; 0.07]
Rel. Scoring (RS)	−0.18*	−0.20*	−0.20*	−0.71*	−0.37*
	[−0.23; −0.12]	[−0.26; −0.14]	[−0.26; −0.14]	[−0.76; −0.65]	[−0.40; −0.33]
Congenial x RS	−0.07	−0.03	−0.09*	0.03	0.03
	[−0.14; 0.01]	[−0.11; 0.06]	[−0.17; −0.01]	[−0.06; 0.13]	[−0.02; 0.09]
Intercept	0.18*	0.21*	0.22*	0.79*	0.28*
	[0.12; 0.23]	[0.15; 0.27]	[0.16; 0.28]	[0.75; 0.84]	[0.24; 0.31]
R^2	0.12	0.10	0.14	0.48	0.29
Survey item FE	No	No	No	No	Yes
Items	1	1	1	1	4
Respondents	902	902	902	902	902
Respondent-items	902	902	902	902	3608

* Null hypothesis value outside the confidence interval.

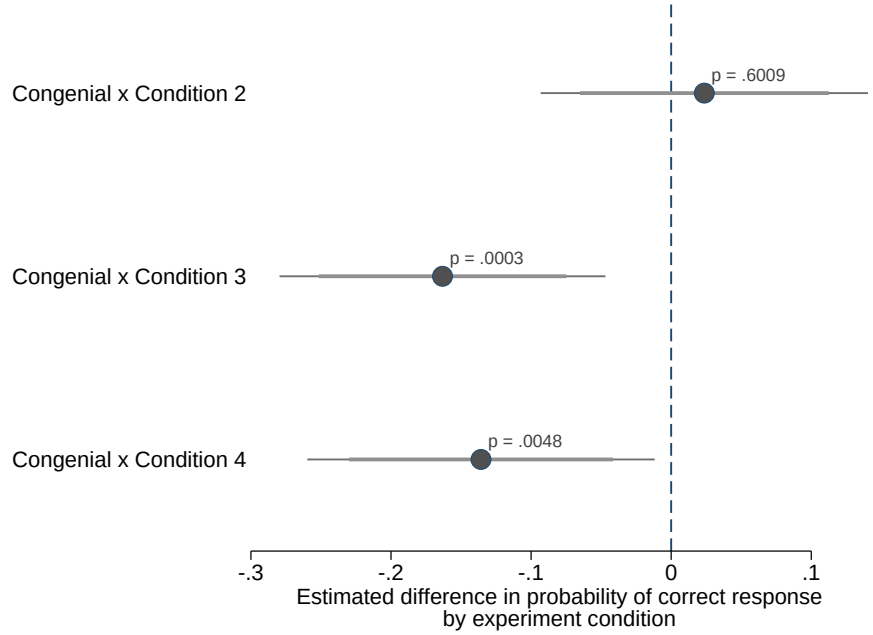
Figure SM 6.1: Robustness Check for Confidence Coding and Knowledge Gaps: MTurk 1



The figure shows the estimated partisan gaps in knowledge from MTurk 1 for two different survey conditions. The CCD condition only considers selecting the right answer with confidence larger than seven as evidence that the respondent knows the answer (see Appendix SM 5). Corresponds to Figure 2, the difference here is that the analysis implements a Confidence Coding threshold of 8. See Table SM 6.2 for the analogous table for Study 3: MTurk 2 Results.

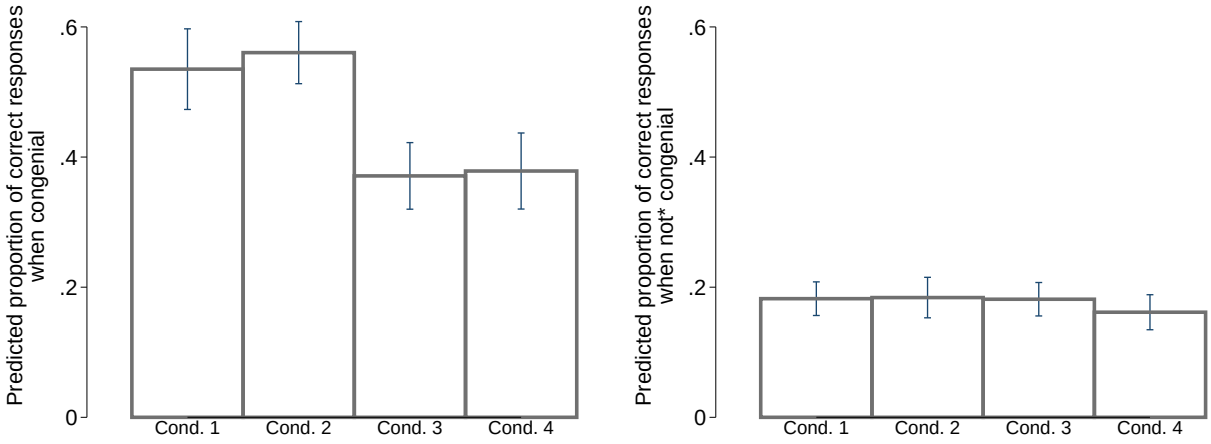
SM 7 Alternate Visualizations of Results

Figure SM 7.1: The Effect of Various Treatments on the Partisan Gap (MTurk 1)



The figure shows the estimated difference in the probability of getting the correct response by the different experimental conditions (see Table 1 for the four conditions). The baseline experiment condition is Condition 1. Coefficients are as estimated in Table 2 (column (6))—survey item fixed effects and demographic covariates (age, gender, education, and race) are included. Horizontal lines indicate 99% and 95% confidence intervals (by gradation). Figure SM 7.2 visualizes absolute effects by condition.

Figure SM 7.2: Partisan Gap by Treatment Arm (MTurk 1)



(a) Congenial **(b)** Uncongenial

Bars indicate the estimated proportion of correct answers by whether the correct response is congenial (1st column) or uncongenial (2nd column). See Table 1 for descriptions of the four conditions. This figure is an alternative visualization of Table 2, with all eight estimates coming from the same model (Table 2, column (6)). All axes have the same scale. Capped vertical bars indicate 95% confidence intervals. Figure SM 7.1 visualizes differences in effects of Conditions 2–4 relative to the baseline Condition 1 (Table 1) when the response is congenial.

Figure SM 7.3: Weighted Estimate of Inflation of Partisan Gap (Study 2)

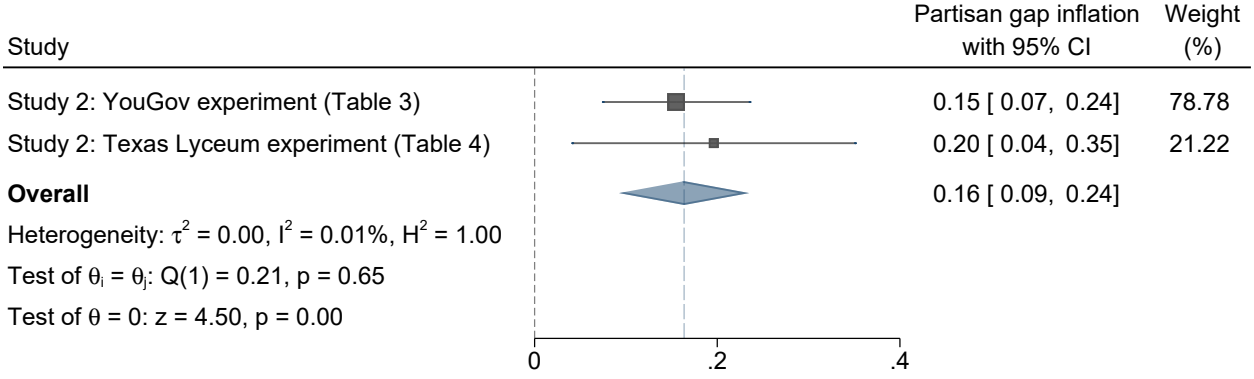


Figure summarizes a precision-weighted estimate of the estimates from the YouGov (column (1) of Table 3) and Texas Lyceum (column (1) of Table 4) experiments examining how random assignment to a Democratic cue in the question artificially inflates the estimated partisan gap in the question on unemployment (the shared question between the two experiments). Specifically, the coefficients are the estimated interaction of the congenial and Democratic cue; the baseline is always the Republican cue. Using the estimates with demographic baselines (column (2) of Table 3 and column (2) of Table 4) leads to similar conclusions.

SM 8 Differences by Subgroups (Gender/Race)

How does the item format affect the gender or race gap in knowledge? To study the question, we expand the equation behind Table 2 and add interactions (separately) for gender and race. As before, we would like to note that our treatments compound in that we cannot uniquely identify the effect of including a Don't Know option. However, the concern is whether discouraging guessing expands the gender (race) gap. If true, we would see a smaller gender (racial) gap in Condition 2, 3, and 4 vis-a-vis Condition 1, which serves as our baseline condition. As Table SM 8.1 and Table SM 8.2 show, there is scant evidence for the hypothesis. The only statistically significant coefficient for the interaction between condition and gender is for the interaction between Condition 3 and the female dummy. (Note also that there is no statistically significant gender gap in the baseline condition and that whatever evidence we have here is that women know more than men.)

Table SM 8.1: The Effect of Various Treatments on the Partisan Gap (MTurk 1), by Gender

	(1)	(2)
Congenial	0.363***	0.367***
	(0.056)	(0.055)
	[0.000]	[0.000]
Condition 2	0.033	0.035
	(0.029)	(0.026)
	[0.249]	[0.189]
Condition 3	0.053+	0.048+
	(0.031)	(0.029)
	[0.089]	[0.093]
Condition 4	−0.028	−0.021
	(0.023)	(0.023)
	[0.225]	[0.368]
Female	0.040	0.038
	(0.026)	(0.025)
	[0.127]	[0.126]
Congenial × Cond. 2	−0.033	−0.036
	(0.069)	(0.066)
	[0.626]	[0.586]
Congenial × Cond. 3	−0.223**	−0.212**
	(0.068)	(0.066)
	[0.001]	[0.001]
Congenial × Cond. 4	−0.140+	−0.151*
	(0.072)	(0.072)
	[0.053]	[0.036]
Congenial × Female	−0.019	−0.023
	(0.072)	(0.070)
	[0.787]	[0.738]
Cond. 2 × Female	−0.056	−0.057
	(0.042)	(0.040)
	[0.185]	[0.152]
Cond. 3 × Female	−0.093*	−0.087*
	(0.039)	(0.037)
	[0.018]	[0.019]
Cond. 4 × Female	0.007	0.001
	(0.037)	(0.036)
	[0.852]	[0.975]
(Congenial × Cond. 2) × Female	0.123	0.123
	(0.092)	(0.090)
	[0.183]	[0.174]
(Congenial × Cond. 3) × Female	0.093	0.088
	(0.095)	(0.093)
	[0.324]	[0.343]
(Congenial × Cond. 4) × Female	0.026	0.033
	(0.096)	(0.095)
	[0.784]	[0.732]
Constant	0.161***	0.141
	(0.017)	(0.999)
	[0.000]	[0.888]
R^2	0.332	0.339
Survey item FE	Yes	Yes
Demographic controls	.	Yes
Items	9	9
Respondents	627	627
Respondent-items	5,643	5,643

Same as Table 2, except with addition interaction by gender (base category is Male). All models are linear probability models where the dependent variable is whether or not the response is correct. See Table 1 for the description of the conditions. Condition 1 is the baseline. Demographic controls include age, gender, education, and race. Standard errors are clustered at the respondent level. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. Exact p-values in square brackets.

Table SM 8.2: The Effect of Various Treatments on the Partisan Gap (MTurk 1), by White vs Non-White

	(1)	(2)
Congenial	0.354***	0.353***
	(0.036)	(0.036)
	[0.000]	[0.000]
Condition 2	0.003	−0.000
	(0.025)	(0.024)
	[0.907]	[0.992]
Condition 3	0.009	0.008
	(0.021)	(0.021)
	[0.678]	[0.701]
Condition 4	−0.022	−0.023
	(0.022)	(0.021)
	[0.309]	[0.293]
Non-White	0.069**	0.058*
	(0.026)	(0.027)
	[0.008]	[0.034]
Congenial × Cond. 2	0.039	0.046
	(0.048)	(0.047)
	[0.418]	[0.336]
Congenial × Cond. 3	−0.179***	−0.169***
	(0.047)	(0.047)
	[0.000]	[0.000]
Congenial × Cond. 4	−0.120*	−0.120*
	(0.049)	(0.049)
	[0.015]	[0.015]
Congenial × Non-White	0.149	0.167
	(0.117)	(0.120)
	[0.203]	[0.165]
Cond. 2 × Non-White	−0.013	0.005
	(0.045)	(0.045)
	[0.765]	[0.910]
Cond. 3 × Non-White	−0.036	−0.037
	(0.049)	(0.045)
	[0.458]	[0.410]
Cond. 4 × Non-White	0.018	0.028
	(0.047)	(0.048)
	[0.704]	[0.557]
(Congenial × Cond. 2) × Non-White	−0.269+	−0.308+
	(0.161)	(0.164)
	[0.094]	[0.061]
(Congenial × Cond. 3) × Non-White	−0.035	−0.044
	(0.141)	(0.142)
	[0.807]	[0.757]
(Congenial × Cond. 4) × Non-White	0.000	0.000
	(.)	(.)
	[.]	[.]
Constant	0.168***	0.445
	(0.016)	(1.016)
	[0.000]	[0.662]
R^2	0.332	0.341
Survey item FE	Yes	Yes
Demographic controls	.	Yes
Items	9	9
Respondents	628	627
Respondent-items	5,652	5,643

Same as Table 2, except with additional interaction by Non-White vs White (base category is White). All models are linear probability models where the dependent variable is whether or not the response is correct. See Table 1 for the description of the conditions. Condition 1 is the baseline. Demographic controls include age, gender, education, and race. Standard errors are clustered at the respondent level. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. Exact p-values not reported as in Table 2 to conserve vertical space.

SM 9 Hierarchical Models

In Table SM 9.1, Model 1 includes item fixed effects and random effects by the respondent, while Model 2 takes guidance from Gelman et al. (2012) and estimates the prescribed model.

Table SM 9.1: Comparison of Linear Mixed-Effects Models

	Dependent Variable: Correct	
	item1	
	Model 1	Model 2
	(1)	(2)
Congenial	0.351***	0.351***
	(0.035)	(0.035)
Cond. 2	0.0004	0.0004
	(0.022)	(0.029)
Cond. 3	0.0002	0.0002
	(0.020)	(0.026)
Cond. 4	−0.023	−0.023
	(0.020)	(0.025)
Congenial × Cond. 2	0.024	0.024
	(0.046)	(0.046)
Congenial × Cond. 3	−0.173***	−0.173***
	(0.046)	(0.046)
Congenial × Cond. 4	−0.132***	−0.132***
	(0.048)	(0.048)
Constant	0.030	0.184**
	(0.019)	(0.081)
Observations	5,652	5,652

SM 10 Impact of Difficulty and Number of Options on the Gender Gap

Bullock and Rader (2022) examine the effect of difficulty and number of response options on the estimates of political knowledge. They manipulate the difficulty and number of options to reduce the impact of guessing. In the article, the authors do not test for the impact of the gender gap on political knowledge. However, their replication files allow us to test for those. The hypotheses are that as the number of response options and response option difficulty increase, the gender gap will go down as men will find it harder to guess correctly. Appendix SM 10 presents results of interacting response option difficulty and number of options with gender. The table shows that the treatment effectively reduces the proportion correct. Five response options (vs. three) reduce success by 12 percentage points, while having more difficult response options causes a 22 percentage points drop. However, the coefficients of interest are those on the interaction terms. For neither model are they statistically significant. For Model 2, the coefficient on the interaction term is not too far away from the twice the standard error of the mean but note that the coefficient is in the opposite direction—in that reducing success from guessing increases the gender gap.

Table SM 10.1: Impact of Increasing the Difficulty and Number of Response Options on the Gender Gap

	Model 1	Model 2
Intercept	0.57^{***}	0.59^{***}
	(0.02)	(0.02)
Female	-0.06^{***}	-0.03
	(0.01)	(0.02)
Five Response Options	-0.12^{***}	
	(0.01)	
Five Response Options*Female	0.00	
	(0.01)	
Difficult Response Option		-0.22^{***}
		(0.02)
Difficult Response Option*Female		-0.04
		(0.02)
Deviance	4460.32	1999.82
Dispersion	0.22	0.23
Observations	20551	8706

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$