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

Lucas Shen* Gaurav Sood† Daniel Weitzel‡

March 19, 2024§

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 toplines of commercial polls suggest.

Keywords: Political Knowledge; Partisan Gap; Motivated Skepticism

---

*Research Fellow, National University of Singapore, lucas@lucasshen.com

†Independent researcher, gsood07@gmail.com

‡Assistant Professor, Colorado State University, and Senior Research Fellow, University of Vienna, daniel.weitzel@colostate.edu

§Working paper, most recent version available at: https://github.com/soodoku/partisan-gaps

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 suggests that partisan gaps in political knowledge are large and common in the US (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, not including a 'Don't know' option, providing a partisan cue in the question stem, etc., affect partisan gaps in knowledge of facts with political implications. Removing these guessing-encouraging features yields knowledge scales with greater reliability and higher criterion validity. Substantively, we find that common design features of knowledge items on commercial polls are "inflationary"—they dramatically inflate the actual partisan gap in beliefs. On average, these features artificially widen the partisan gap in beliefs by 40% (14 percentage points; see 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 another 50% smaller (see Figure 4).

Our results contribute to a growing literature that suggests that a large fraction of partisan gaps are artifacts of survey design. Our results also 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.

Theory, Motivation, and Empirical Strategy

> "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 stem 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 federal deficit during the Obama administration, Republicans, thinking Democrats cause bad things, may infer that deficits rose under Obama. And we expect Democrats to come to the opposite conclusion. We thus expect guessing-encouraging designs or designs that prime partisanship to increase partisan gaps.

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). And about half of the items offered only two choices, a design choice that dramatically inflates estimates of knowledge (Bullock and Rader 2022). 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) (see also Bailey (2021) who shows that survey context (political items before knowledge questions or even just indications that a survey asks political questions) cause respondents to give more partisan answers).

To study the effect of "inflationary" features of survey and question design 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 respondents are confident about as evidence that the respondent knows the fact. Correct responses that respondents are confident about have higher test-retest reliability (Graham 2021), suggesting that these measures are also more valid. Finally, we analyze which item formats yield more reliable measures of knowledge and have greater criterion validity and find that items without the "inflationary" features have 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 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 be useful to study partisan gaps in misinformation.

Study 1: The Effect of guessing-encouraging features

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

1. the absence of a "Don't Know" option,
2. including additional neutral or partisan information in the question stem, and
3. the absence of a guessing discouraging preamble.

The omission of a Don't Know option, a lack of a guessing discouraging preamble, and the inclusion of neutral information of social proof inflate the partisan gap because they cause people to guess. On knowledge questions with partisan implications, the guesses tend to be biased, with partisans picking the partisan congenial option—the option that makes the party they identify with look good. For example, if we asked a Republican whether the unemployment rate increased, decreased, or remained the same over the Republican presidency, the partisan congenial response is that the unemployment rate decreased. And a Republican who doesn't know how the unemployment rate fared may guess it decreased. Some research suggests that the inclusion of a Don't Know option will induce bias (Mondak 1999; Mondak and Davis 2001; Mondak and Anderson 2004) as some people, e.g., women, may be less likely to guess and more likely to pick 'Don't Know' than men. However, more recent research shows that 'Don't Know' responses overwhelmingly represent ignorance (e.g. Luskin and Bullock 2011; Bullock and Rader 2022) and yield measures that have higher reliability and validity. Later in the paper we also show results consistent with this strand of research. In all, we conjecture that guessing-encouraging features artificially inflate partisan gaps.

Study 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 for a summary.)[1] In each condition, respondents answered nine misinformation items, ranging from President Obama's citizenship to whether global warming is happening or not. (For exact question wording, see Appendix SI 3.)

The four conditions are:

1. **Inflationary Design Approach (IDA)** The Inflationary Design Approach 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, so respondents cannot indicate that they don't know the answer. The questions also include social proof about the incorrect answer.[2] For instance, on a 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 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. The preamble simply

[1]For generalizability of effects in studies conducted on MTurk, see (Mullinix et al. 2015; Coppock, Leeper and Mullinix 2018).

[2]Social proof here refers to 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).

reads: "Now here are some questions about what you may know about politics and public affairs..."

2. **Commonly Used Design (CUD)** The Commonly Used Design makes one change to the Inflationary Design Approach. Like the IDA, the questions do not feature a 'Don't Know' option and include neutral information in the question stem that encourages guessing. However, the questions do not include social proof.

3. **Fewer Substantive Responses (FSR)** The Fewer Substantive Responses makes two changes to Commonly Used Design. First, the preamble discourages blind guessing and cheating. 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 'Don't Know' option (see, e.g., Luskin and Bullock 2011; Bullock et al. 2015). 'Don't Know' responses are then coded with incorrect answers as responses that were not correct.

4. **Improved Multiple Choice (IMC)** Improved Multiple Choice is the best version of these multiple choice questions. It offers respondents a 'Don't Know' option and does not include neutral information or social proof. Here, we also code 'Don't Know' responses together with incorrect answers as responses that were not correct.

Table 1: Experimental Treatments

		Treatments			
Condition	Label	Don't Know	Social Proof	Guessing Encouraged	Neutral Information
1	Inflationary Design Approach (IDA)	No	Yes	Yes	Yes
2	Commonly Used Design (CUD)	No	No	Yes	Yes
3	Fewer Substantive Response (FSR)	Yes	No	No	Yes
4	Improved Multiple Choice (IMC)	Yes	No	No	No

Study 1: Measures

Since the context of our study is the United States with a stable two-party system, 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. The question then branches and asks the partisans about the strength of their attachment to their 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.

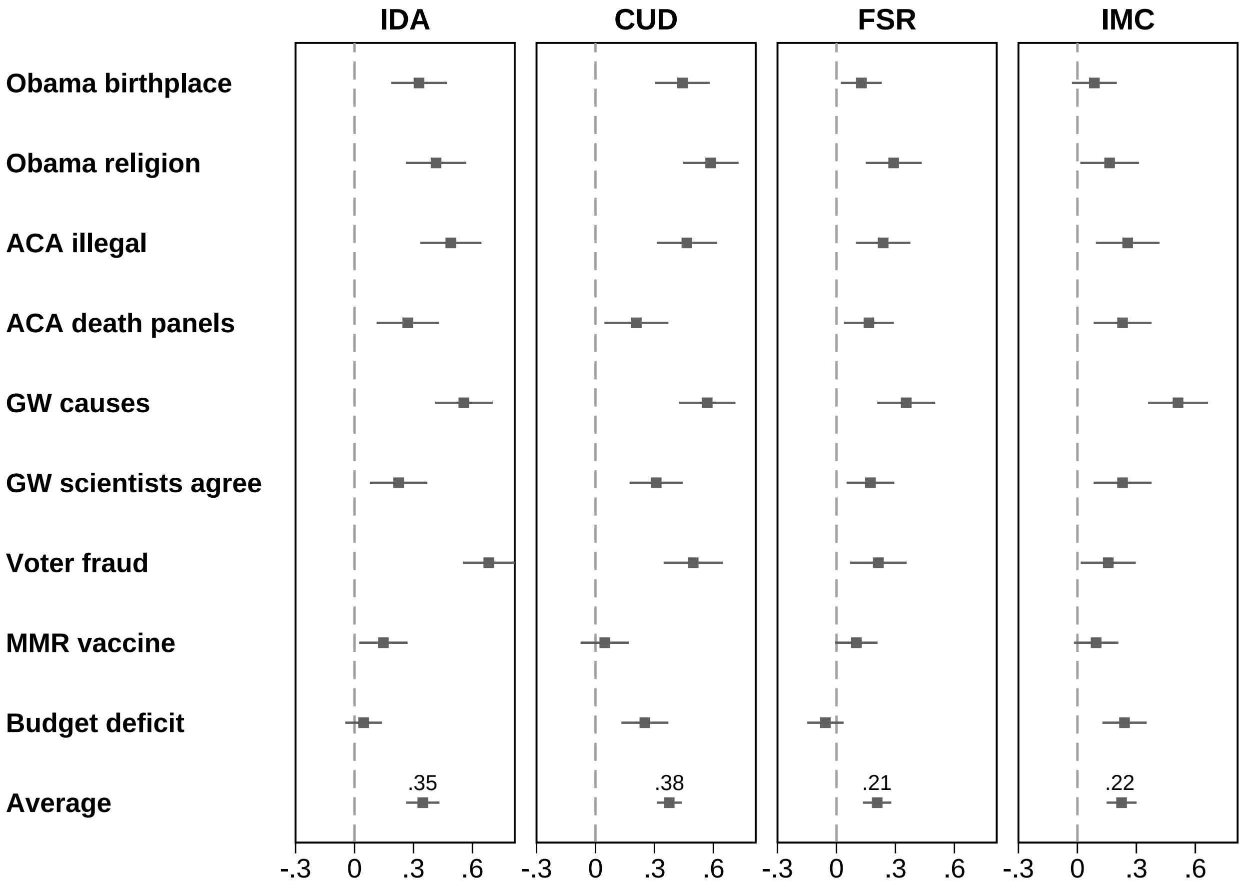
Study 1: Results

We start by summarizing the average partisan gap on each survey item in each treatment arm (see Figure 1).[3] In Condition 1, the baseline Inflationary Design Approach condition (first column), when the correct response is congenial to the respondents' party, respondents are 35 percentage points more likely to choose the correct response. The partisan gap is unresponsive to the removal of social proof in the question stem, as implemented in Condition 2, the Commonly Used Design. However, the estimates from Condition 3 and 4, the Fewer Substantive Responses and Improved Multiple Choice conditions, are approximately 14 percentage points lower than in Condition 1. The 14 percentage points reduction, due to the absence of guessing encouraging text and neutral information, translates to a 40% relative drop ($100 \times \frac{.35-.21}{.35}$).

To formally test our hypothesis, we regress whether the answer is correct on the interaction of the survey conditions and the congenial dummy. For respondent i, survey

[3]Balance tests suggest that the randomization was successful (see Figures SI 1.1 to SI 1.4).

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 SI 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.

item j, and condition k, we estimate the following equation:

$$\text{Correct}_{ijk} = \alpha + \beta\text{Congenial}_i + \sum_{k=1}^{3}\gamma\text{Condition}_k + \sum_{k=1}^{3}\delta_k(\text{Congenial}_i \times \text{Condition}_k) + \text{question}_j + \varepsilon_{ijk} \quad (1)$$

β captures the difference in the proportion of correct responses when the answer is congenial to the respondent's party under the baseline condition. The δ_ks capture how Con-

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]
Commonly Used Design (CUD)		0.010	0.000		0.011	0.002
		(0.028)	(0.022)		(0.028)	(0.021)
		[0.722]	[0.985]		[0.687]	[0.934]
Fewer Substantive Responses (FSR)		−0.064**	0.000		−0.063**	−0.001
		(0.024)	(0.019)		(0.024)	(0.019)
		[0.009]	[0.993]		[0.010]	[0.964]
Improved Multiple Choice (IMC)		−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 × CUD			0.024			0.024
			(0.046)			(0.045)
			[0.605]			[0.601]
Congenial × FSR			−0.173***			−0.163***
			(0.046)			(0.045)
			[0.000]			[0.000]
Congenial × IMC			−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 IDA, CUD, FSR, and IMC conditions. 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. Figure 1 visualizes partisan gaps by condition. Further alternative visualization of results in Figure SI 7.1 and Figure SI 7.2 in Appendix SI 7.

ditions 2-4 (CUD, FSR, IMC) affect the partisan knowledge gaps vs. the baseline condition (IDA). We include item fixed-effects and cluster standard errors by respondent.

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 conditions. The sharp negative coefficients on

Condition 3 (FSR) and 4 (IMC) 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 the interaction between congenial and the three conditions (baseline is the Inflationary Design Approach). Now, the congenial variable captures the knowledge gap in the IDA 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.

Columns (4)–(6) of Table 2 show that including self-reported characteristics of respondents 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, social proof, neutral information, and the explicit encouragement to guess, which are common in commercial polls. The gap drops on average from .35 and .38 in the more inflationary designs to .21 and .22 in designs that significantly clean questions from problematic features.

Study 2: The Effect of Partisan Cues on Partisan Gaps

In Study 2, we investigate the impact of partisan priming. We test it by manipulating whether the question stem has a partisan cue or not. We expect the presence of a partisan cue to exacerbate partisan gaps (Prior, Sood and Khanna 2015).

Study 2: Research Design and Data

To answer the question, we leverage data from two surveys: a national survey conducted by YouGov (Study 2) and a telephone survey in Texas (Study 3). 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 our third option:

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, meaning if an individual received the Republican cue in the unemployment question, they also received it in the deficit question. Congenial responses here are responses where the correct answer paired with the partisan prompt aligns with the partisanship of the respondent. Hence, the deficit question is congenial for Republican respondents if they see "when Republicans regained control of the U.S. Congress" but uncongenial if they see "when Democrats retained control of the Senate".

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 'no partisan cue' 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, …"

Study 2: YouGov Results

We estimate the impact of partisan cues by regressing whether the response is correct or not on the partisan congeniality of the cue. We code the cue as congenial if it increases the probability that the respondent would get the right correct by using partisan reasoning. For instance, if the right answer is that the objective conditions over some period worsened, then highlighting that the opposing party controlled Congress during that time would be a congenial cue.

$$\text{Correct}_i = \alpha + \beta(\textit{Congenial Cue})_i + \varepsilon_i, \tag{2}$$

Figure 2 plots the results. As Panel (a) of Figure 2 illustrates, showing a congenial cue instead of an uncongenial one causes the probability of the correct response on the unemployment question to increase by 14 percentage points ($p < 0.001$, reported in Table 3). Panel (b) of Figure 2 shows that this effect is not unique to the unemployment question. On the budget deficit question, the difference is 18 percentage points ($p < 0.001$). Partisans are, therefore, more likely to respond correctly to a survey question when there is a partisan cue in the question stem that frames the right answer as congenial to the party.[4]

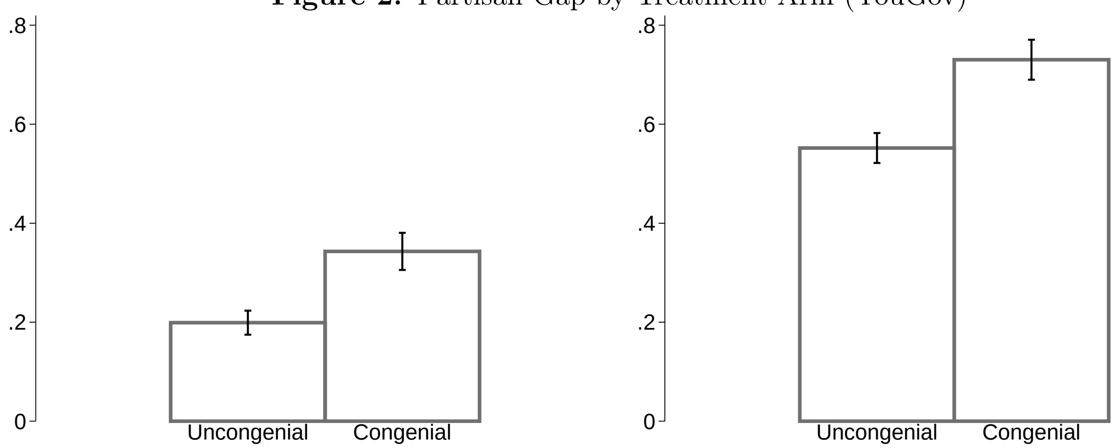
[4] Table SI 9.10 in Appendix SI 9 reports marginal effects from logistic regression models

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

	Unemployment has gone up		Deficit has gone up	
	(1)	(2)	(3)	(4)
Congenial	0.144***	0.147***	0.178***	0.188***
	(0.019)	(0.020)	(0.021)	(0.020)
	[0.000]	[0.000]	[0.000]	[0.000]
Constant	0.199***	3.569+	0.552***	7.636***
	(0.012)	(1.895)	(0.015)	(1.868)
	[0.000]	[0.060]	[0.000]	[0.000]
R^2	0.0262	0.0550	0.0346	0.167
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. Exact p-values in square brackets. All models are linear probability models. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. Alternative visualization of results in Figure 2 and Figure SI 7.3.

Figure 2: Partisan Gap by Treatment Arm (YouGov)



(a) Unemployment

(b) Budget deficit

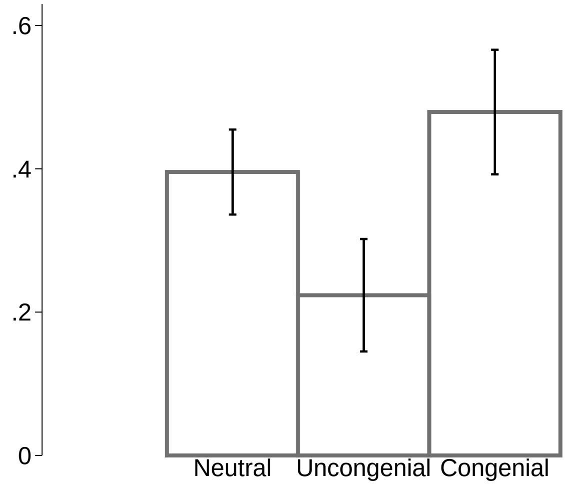
Bars indicate the predicted proportion of correct answers as reported in Table 3 (columns (1) and (4)). Capped vertical bars indicate 95% confidence intervals. Figure SI 7.3 in Appendix SI 7 plots the differences in effects.

Study 2: Texas Lyceum Results

We supplement our results with the Texas Lyceum survey. As Figure 3 shows, on the unemployment question, the pattern we saw on YouGov still holds when we include a neutral

with near identical coefficients.

Figure 3: Partisan Gap on Unemployment by Treatment Arm (Texas Lyceum)



Bars indicate the predicted proportion of responses saying that unemployment has gone up (correct response) as reported in column (1) of Table 4. Capped vertical bars indicate 95% confidence intervals. Figure SI 7.4 plots the differences in effects.

cue. Compared to respondents who received a neutral cue, respondents who received an uncongenial cue were 17 percentage points less likely to get the correct answer ($p < 0.001$). While respondents who received a congenial cue were eight percentage points more likely to get the correct answer ($p < 0.1$). These results are tabulated in Table 4.

Finally, we examine the federal tax rate question in the Texas Lyceum survey. As Table 5 shows, randomly receiving a congenial cue leads to a 21.5 percentage points increase in the chance of getting the answer right compared to the neutral cue condition ($p < 0.001$). On the other hand, an uncongenial cue leads to a 29.8 percent lower chance ($p < 0.001$). We also estimate how the cue that encourages guessing affects the "Don't Know" response rate. While the congenial (without the guessing) cue reduces the probability of responding "Don't Know", the effect of congenial with guessing cue is effectively the same (a post-estimation test for the null that the effect for Congenial and Congenial w/ guessing have the same effect

Table 4: Partisan Gap on Unemployment by Treatment Arm (Texas Lyceum)

	Unemployment has gone up	
	(1)	(2)
Congenial	0.084^{+}	0.085^{+}
	(0.044)	(0.044)
	[0.059]	[0.055]
Uncongenial	-0.172^{***}	-0.195^{***}
	(0.040)	(0.042)
	[0.000]	[0.000]
Constant	0.395^{***}	0.057
	(0.030)	(0.175)
	[0.000]	[0.746]
R^2	0.0483	0.153
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. Exact p-values in square brackets. Alternative visualization of results in Figure 3 and Figure SI 7.4.

on Responding "Don't Know" cannot be rejected, with a p-value = .705).[5] [6]

Overall, results from survey experiments with YouGov and Texas Lyceum 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 data show that the partisan gap in the presence of partisan cues is upwardly biased.

[5] A post-estimation test for the null that Uncongenial and Uncongenial w/ guessing has the same effect on Responding "Don't Know" cannot be rejected, with a p-value = .616.

[6] Appendix SI 9 reports marginal effects from logistic regression models with similar conclusions.

Table 5: Impact of Various Treatments on Partisan Gap on Federal Taxes (Texas Lyceum)

	Responded "Gone up"		Responded "Don't Know"	
	(1)	(2)	(3)	(4)
Congenial	0.215***	0.171**	−0.077*	−0.081*
	(0.051)	(0.056)	(0.036)	(0.038)
	[0.000]	[0.002]	[0.032]	[0.034]
Uncongenial	−0.298***	−0.228***	−0.063	−0.077
	(0.042)	(0.048)	(0.042)	(0.050)
	[0.000]	[0.000]	[0.134]	[0.127]
Congenial w/ guessing	0.091+	0.042	−0.074*	−0.066+
	(0.052)	(0.057)	(0.036)	(0.038)
	[0.082]	[0.462]	[0.043]	[0.078]
Uncongenial w/ guessing	−0.290***	−0.234***	−0.038	−0.051
	(0.040)	(0.047)	(0.041)	(0.043)
	[0.000]	[0.000]	[0.356]	[0.240]
Constant	0.381***	−0.223	0.187***	0.884***
	(0.031)	(0.177)	(0.025)	(0.180)
	[0.000]	[0.208]	[0.000]	[0.000]
R^2	0.151	0.219	0.00868	0.126
Demographic controls	.	Yes	.	Yes
Respondents	758	752	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. Exact p-values in square brackets. Alternative visualization of results in Figure SI 7.5 and Figure SI 7.6.

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

Lastly, we examine the consequences of scoring decisions on partisan gaps. We introduce an assessment that considers respondents' confidence in their answers. We aim to score only actual political knowledge, the confidently held correct beliefs about political facts.

Study 3: Research Design and Data

Knowledge questions are commonly offered as multiple-choice items, and conventionally, if a respondent marks the right answer, it is taken as evidence that the respondent truly 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 takes respondents' confidence in their answers into account, 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 Confidence Coding Design condition builds on the first four conditions and does not encourage guessing, and features no social proof. (The question wording for the items is presented in Appendix SI 3.) Since the items in the multiple choice questions are dichotomous choice, we only offer a true and an incorrect response; the CCD scoring 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 U.S." 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.[7]

[7]The questions and their design are included in Appendix SI 3. In Appendix SI 6, we try less stringent criteria, and the main picture remains broadly unchanged.

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 had a "Don't Know" option available to them. (For question text, see Appendix SI 5.)

The scoring for this survey is more nuanced, as the multiple-choice questions had four potential response options. In the CCD treatment, survey participants see the same question as in the multiple choice treatment but have to rank 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. But 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 cannot be as confident about the correct answer as any other option. For instance, it cannot be that the four options are all rated 10.

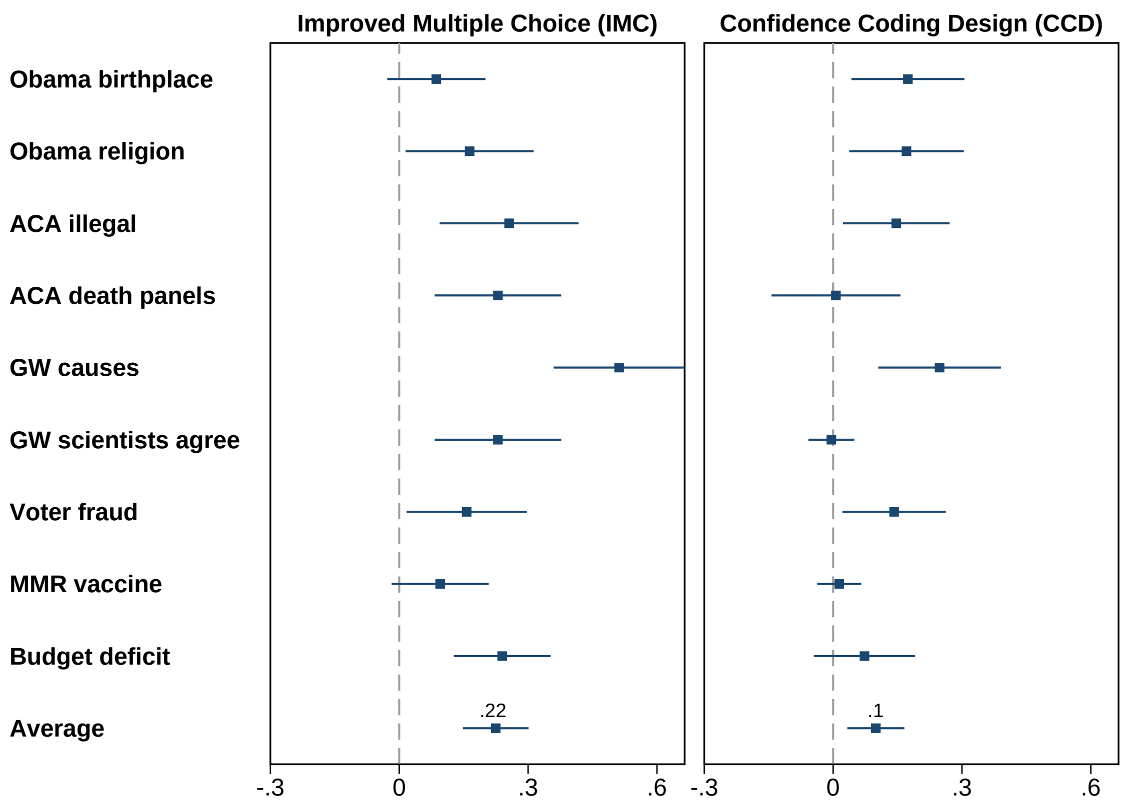
3. The respondent must be at least c confident in the correct answer. In the main text, we use a c of 10, but in Appendix SI 6, we try less stringent criteria.

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

Study 3: MTurk 1 Results

The best version of the dichotomous multiple-choice items (Condition 4, Improved Multiple Choice) showed a partisan gap of .22 (see Figure 1). As Figure 4 shows, nearly half of the gap vanishes under the confidence scoring of the Confidence Coding Design. 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.

Figure 4: 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 SI 5). See Tables SI 2.1 to SI 2.5 in Appendix SI 2 for the regression estimates of the multiple-choice conditions to the confidence coding condition. See Figure SI 2.6 for the same analysis with all four multiple-choice conditions pooled together. Figure SI 6.1 implements a robustness check, setting the relative scoring threshold t to 8.

Study 3: MTurk 2 Results

We use data from our last survey experiment to once again shed light on the question of how treating answers a respondent is confident about as evidence that the respondent knows the fact changes our understanding of the magnitude of partisan gaps. To analyze the data, we regress the dependent variable, an indicator of whether the response is correct, on the interaction between Relative Scoring (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. As 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 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 that relative scoring is associated with more correct responses and a negative one with fewer. δ captures the difference in how the two scoring treatments, multiple choice and confidence coding, 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 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

Table 6: Confidence Scoring and Knowledge Gaps: MTurk 2

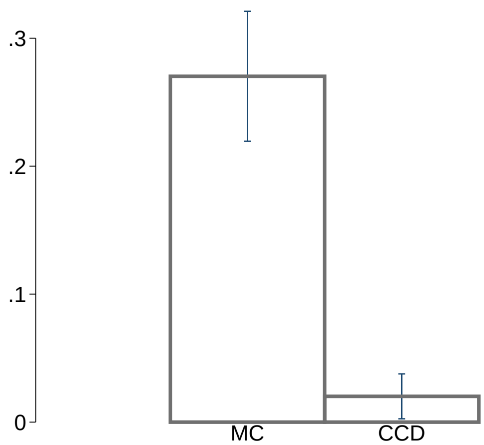
	Individual survey question				
	Affordable Care Act (1)	Affordable Care Act 2 (2)	Greenhouse gases (3)	Donald Trump (4)	All (5)
Congenial	0.091*	0.084*	0.087*	0.005	0.025
	(0.038)	(0.040)	(0.041)	(0.038)	(0.023)
	[0.018]	[0.036]	[0.033]	[0.895]	[0.270]
Confidence Coding Design (CCD)	−0.179***	−0.201***	−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 SI 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 full confidence of $c = 10$ (see Study 3: Research Design and Data in the Study 3: The Effect of the Scoring Method on Partisan Gaps section). The baseline are the multiple choice designs. Table SI 6.7 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. Exact p-values in square brackets.

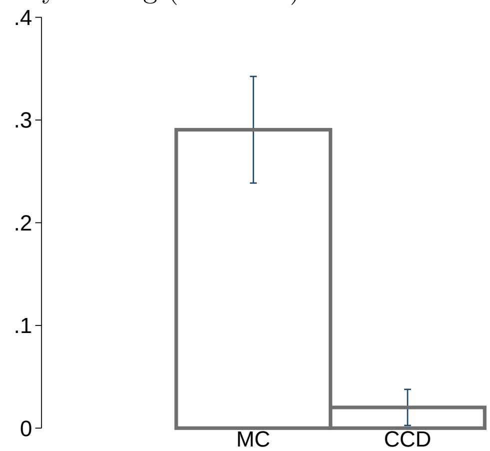
are associated with a higher proportion of correct responses. In the multiple choice scoring treatment, partisans are more likely to get questions correct when answers are congenial to their partisanship. For the first three models focusing on the Affordable Care Act and Greenhouse Gas questions, the effects are statistically significant. This is not the case for model 4 and the pooled model. γ shows us that this is not the case for congenial questions that are scored with the relative scoring rule of the CCD approach. In this treatment, all but the Greenhouse Gas question see the partisan gap in knowledge disappear.

In all, if we pool evidence across the two MTurk studies, the data suggest that treating only confident correct answers as evidence that the respondent knows the answer shrinks the partisan gap substantially. The U.S. partisan gap in political knowledge decreases substantially if we only code answers that are completely (or at least strongly) confident in.

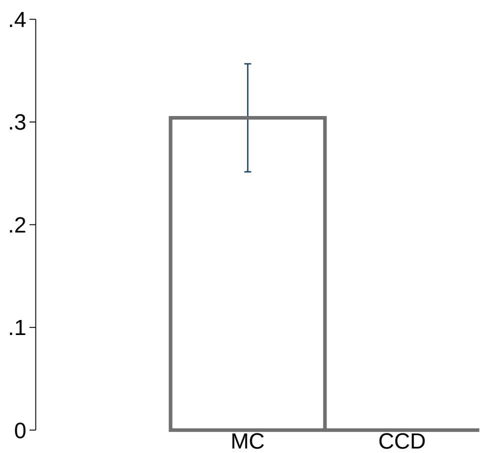
Figure 5: Partisan Gaps by Coding (MTurk 2)



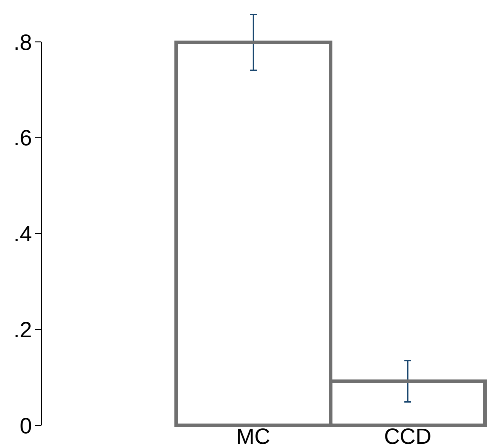
(a) Affordable Care Act



(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. CCD bar indicates the predicted proportion of correct responses for the Confidence Coding Design with congenial responses. Capped vertical bars indicate 95% confidence intervals.

Validity and Reliability of Question Designs

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 answers as evidence of knowledge. But which item and survey design choices lead to 'better' measures? The presumption is that better political knowledge measures are also better instruments for measuring partisan gaps.

To answer which design choices lead to better measures, 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 SI SI 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 (Inflationary Design Approach) 3 (Fewer Substantive Responses) having greater reliability than Conditions 2 (Commonly Used Design) and 4 (Improved Multiple Choice). One of the reasons for this mixed picture may be that partisan guessing increases reliability without increasing validity because it introduces correlated errors. 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 Improved Multiple Choice condition and the Confidence Coding Design and criterion variables is markedly higher (.34) than in conditions 1-3. The Inflationary Design Approach (.11), Commonly Used Design (.20), and Fewer Substantive Responses (.26) all score lower.[8]

Table 7: Validity and Reliability

	Conditions				
	No DK		With DK		
	IDA (1)	CUD (2)	FSR (3)	IMC (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
	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 coefficient of scale reliability 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 obtained above are consistent with those obtained by Graham (2021), which finds that the test-retest reliability of confident correct answers is much higher. In all, the data suggest that the substantially smaller partisan gap that we see in CCD is also the best estimate of the partisan gap.

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

Discussion and Conclusion

Since at least the publication of Bartels (2002), the conventional wisdom has been that partisan gaps in 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 United States is not founded in differences in beliefs. We find that standard features of commercial polls, like not asking don't know, inserting a partisan cue, and treating unconfident answers as knowledge, inflate the partisan gaps. Offering respondents a 'Don't Know' option, not offering social proof or neutral information, and not encouraging them to take guesses decreases the partisan gap by 14 percentage points. Offering a question design that allows respondents to express their confidence in their answers further reduces the gap by 12 percentage points. Once we remove the opportunities and incentives for our respondents to substitute knowledge responses with partisan expressions, the knowledge differences between Democrats and Republicans decrease substantially.

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 in the United States are partly a consequence of 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, and the downside is that people know even less than we thought.

Research focusing on multi-party systems around the world has found that they are increasingly presidentializing[9] 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. We hence expect that the conclusions that we are drawing based on studies conducted in the United States would travel to other contexts.

[9]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.

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

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.

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

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 Belinda Creel Davis. 2001. "Asked and answered: Knowledge levels when we will not take "don't know" for an answer." *Political Behavior* 23:199–224.

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*

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)* .

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.

Supporting Information

SI 1 Balance Tests

Figure SI 1.1: MTurk 1—IDA and CUD

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

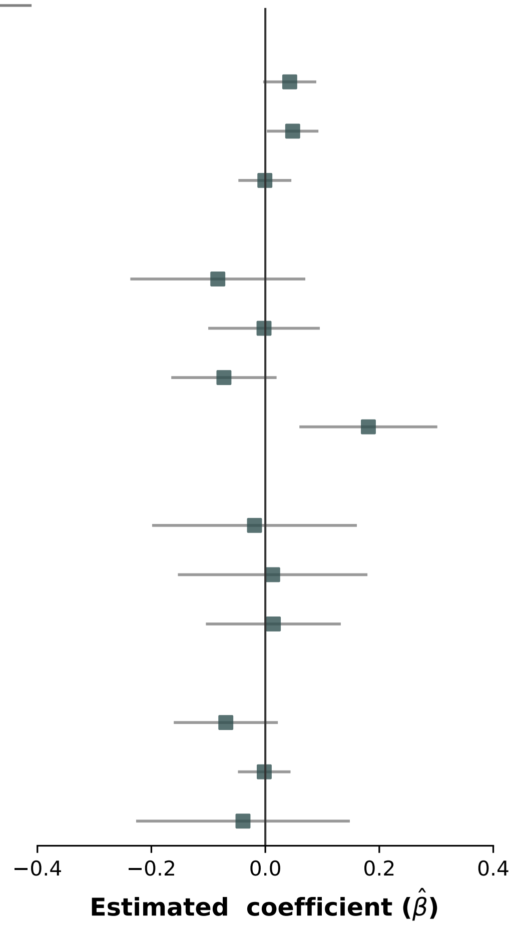


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compare respondents assigned to the IDA condition vs. respondents assigned to the CUD condition. See Table 1 in Study 1: The Effect of guessing-encouraging features. Rows are self-reported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

Figure SI 1.2: MTurk 1—IDA and FSR

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

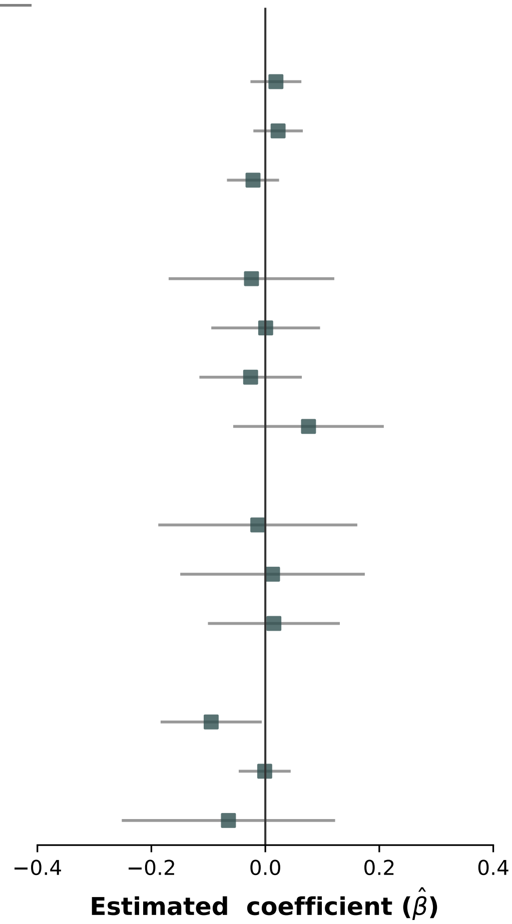


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compare respondents assigned to the IDA condition vs. respondents assigned to the FSR condition. See Table 1 in Study 1: The Effect of guessing-encouraging features. Rows are self-reported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

Figure SI 1.3: MTurk 1—IDA and IMC

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

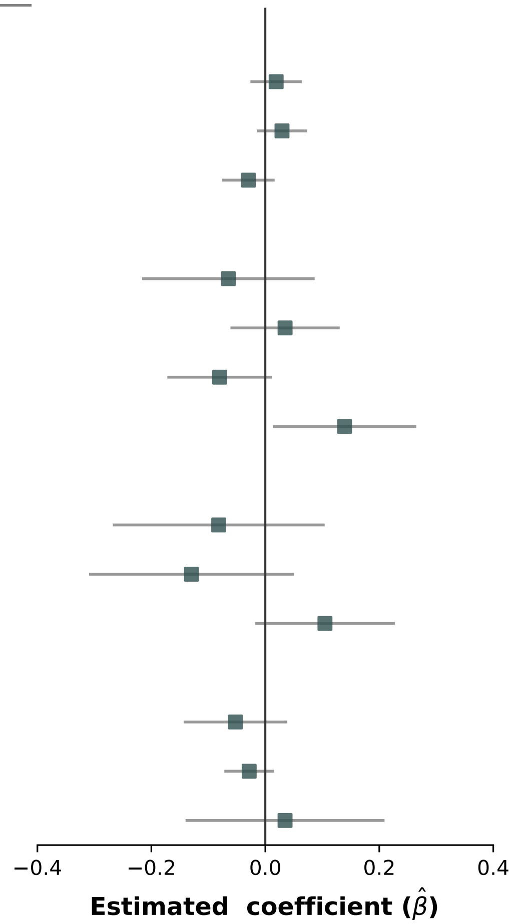


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compare respondents assigned to the IDA condition vs. respondents assigned to the IMC condition. See Table 1 in Study 1: The Effect of guessing-encouraging features. Rows are self-reported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

Figure SI 1.4: MTurk 1—IDA and 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

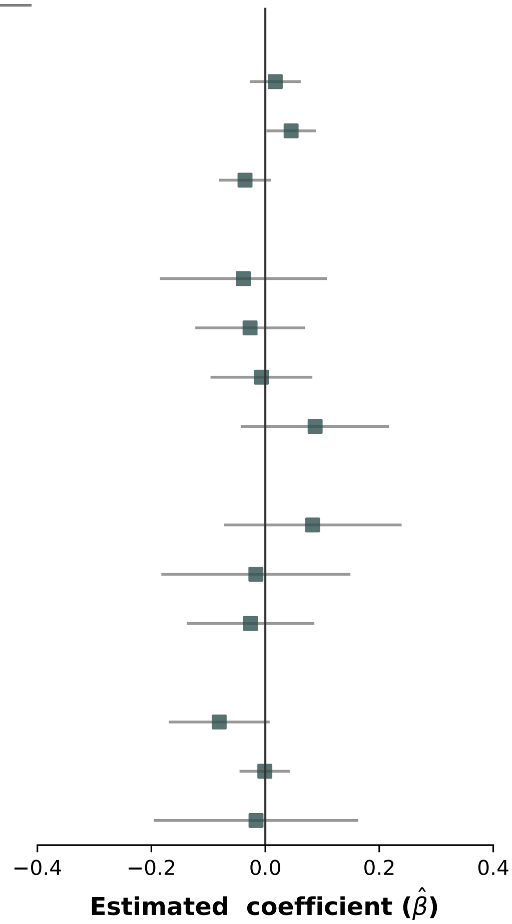


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compare respondents assigned to the IDA condition vs. respondents assigned to the CCD condition. See Table 1 in Study 1: The Effect of guessing-encouraging features. Rows are self-reported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

SI 2 Additional Results for Confidence Coding (Mturk 1)

Table SI 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 the IDA, CUD, FSR, and IMC conditions 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 SI 2.2 to SI 2.5 for the results comparing the Confidence Coding condition to each of the four other individual survey conditions. See Figure SI 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 SI 2.2: Confidence Coding vs. IDA (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 the IDA condition (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 SI 2.1 for the results comparing the Confidence Coding condition with all the four other conditions (IDA, CUD, FSR, IMC) pooled together. See Figure SI 2.2 for the visualization of how Confidence Coding mediates the effect that congenial responses have. See Figure SI 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 SI 2.3: Confidence Coding vs. CUD (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 the CUD condition (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 SI 2.1 for the results comparing the Confidence Coding condition with all the four other conditions (IDA, CUD, FSR, IMC) pooled together. See Figure SI 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 SI 2.4: Confidence Coding vs. FSR (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

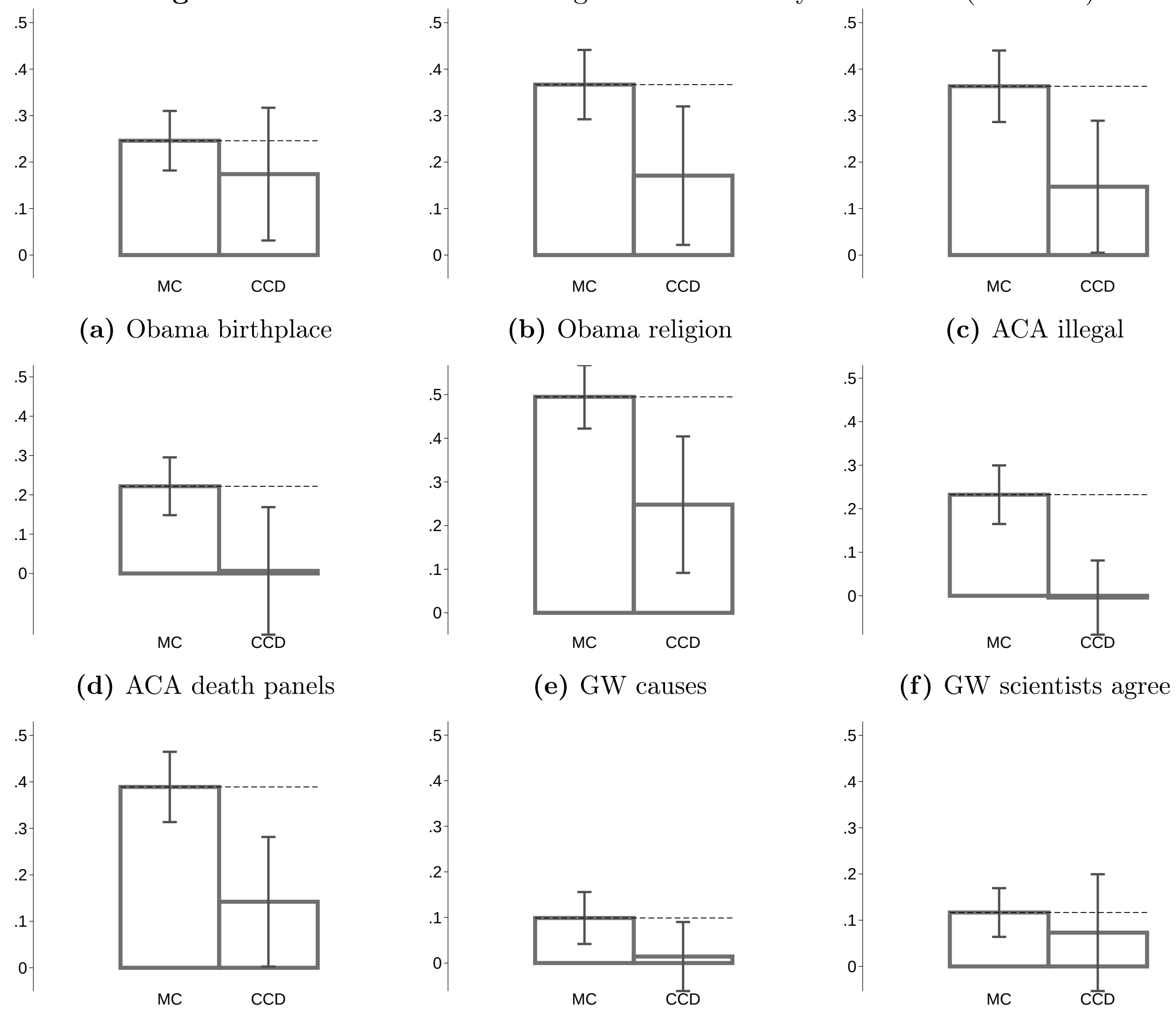
All models are linear probability models where the dependent variable indicates whether the response to a survey item is correct. Under the Confidence Coding condition, we only consider responses as correct when they are chosen with complete confidence (10 on a 0–10 scale). The baseline condition is the FSR condition (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 SI 2.1 for the results comparing the Confidence Coding condition with all the four other conditions (IDA, CUD, FSR, IMC) pooled together. See Figure SI 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 SI 2.5: Confidence Coding vs. IMC (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 the IMC condition (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 SI 2.1 for the results comparing the Confidence Coding condition with all the four other conditions (IDA, CUD, FSR, IMC) pooled together. See Figure SI 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.

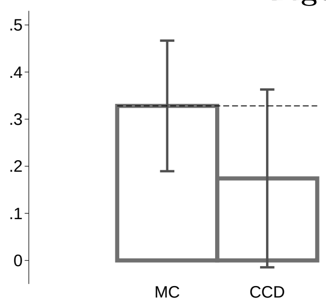
Figure SI 2.1: Confidence Coding vs. Other Survey Conditions (MTurk 1)



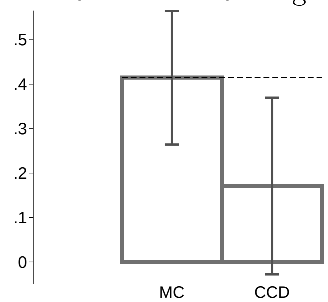
(a) Obama birthplace **(b)** Obama religion **(c)** ACA illegal **(d)** ACA death panels **(e)** GW causes **(f)** GW scientists agree **(g)** Voter fraud **(h)** MMR vaccine **(i)** Budget deficit

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 Multiple Choice conditions (IDA, CUD, FSR, IMC; see Table 1 for the descriptions). Reconstructed from the estimates from Table SI 2.1. Capped vertical bars indicate 95% confidence intervals.

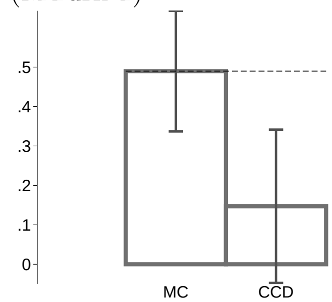
Figure SI 2.2: Confidence Coding vs. IDA (MTurk 1)



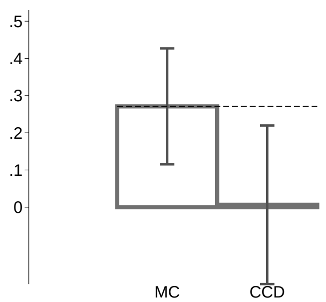
(a) Obama birthplace



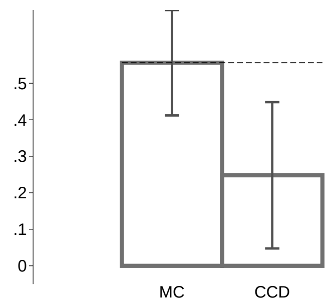
(b) Obama religion



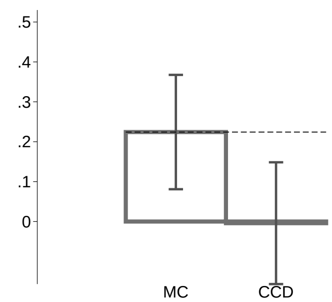
(c) ACA illegal



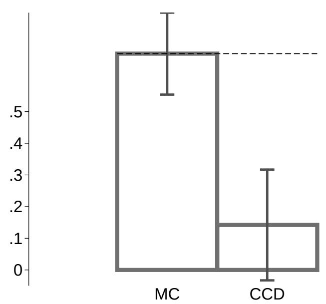
(d) ACA death panels



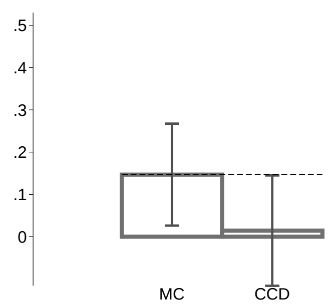
(e) GW causes



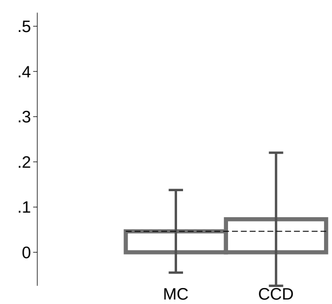
(f) GW scientists agree



(g) Voter fraud



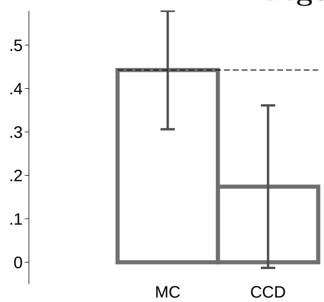
(h) MMR vaccine



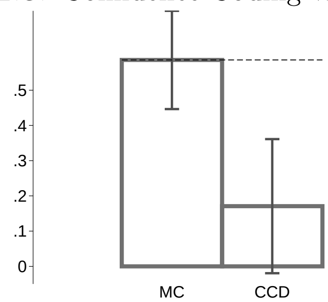
(i) Budget deficit

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 multiple choice IDA condition (see Table 1 for the descriptions). Reconstructed from the estimates from Table SI 2.2. Capped vertical bars indicate 95% confidence intervals.

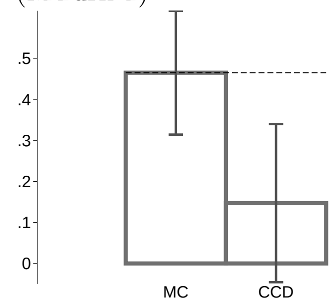
Figure SI 2.3: Confidence Coding vs. CUD (MTurk 1)



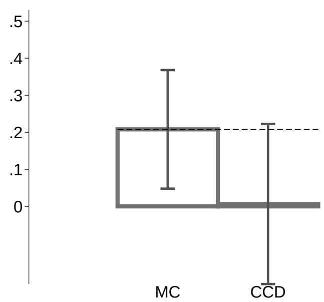
(a) Obama birthplace



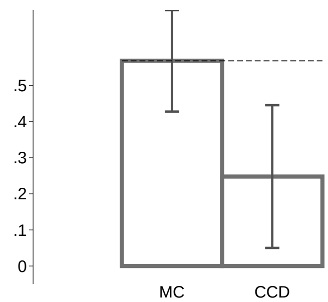
(b) Obama religion



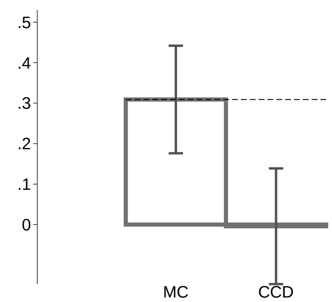
(c) ACA illegal



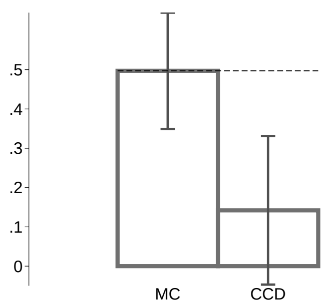
(d) ACA death panels



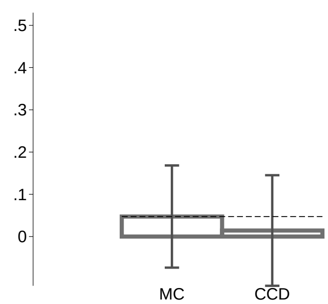
(e) GW causes



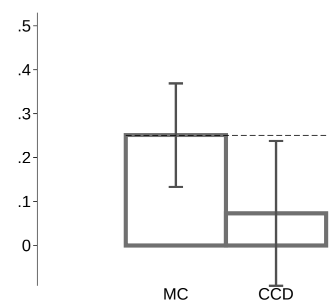
(f) GW scientists agree



(g) Voter fraud



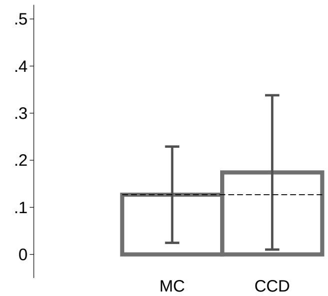
(h) MMR vaccine



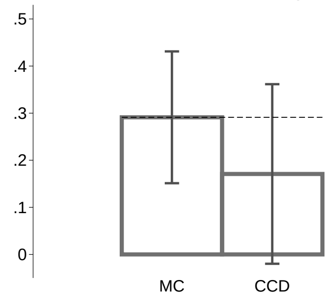
(i) Budget deficit

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 multiple choice CUD condition (see Table 1 for the descriptions). Reconstructed from the estimates from Table SI 2.3. Capped vertical bars indicate 95% confidence intervals.

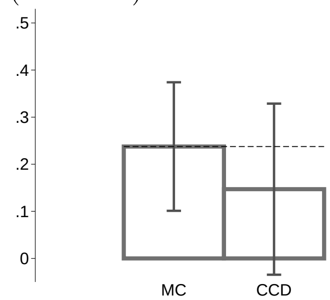
Figure SI 2.4: Confidence Coding vs. FSR (MTurk 1)



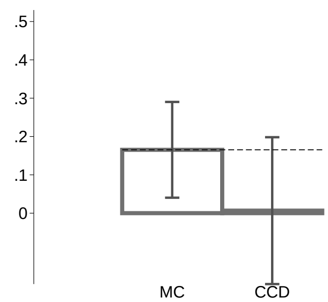
(a) Obama birthplace



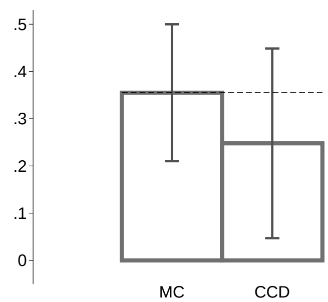
(b) Obama religion



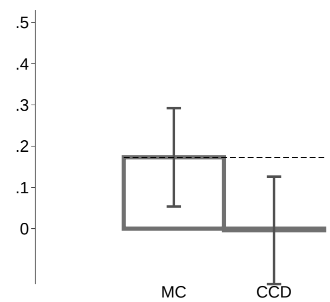
(c) ACA illegal



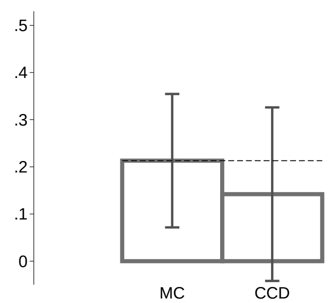
(d) ACA death panels



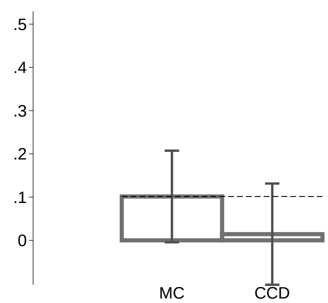
(e) GW causes



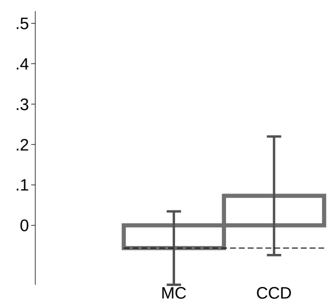
(f) GW scientists agree



(g) Voter fraud



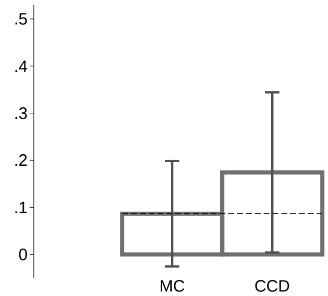
(h) MMR vaccine



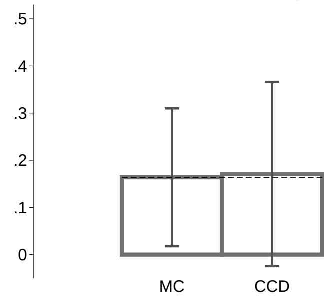
(i) Budget deficit

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 multiple choice CUD condition (see Table 1 for the descriptions). Reconstructed from the estimates from Table SI 2.4. Capped vertical bars indicate 95% confidence intervals.

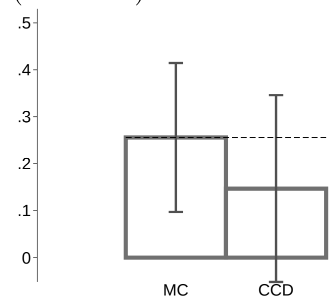
Figure SI 2.5: Confidence Coding vs. IMC (MTurk 1)



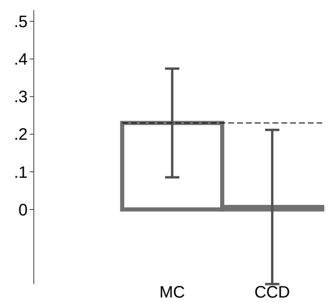
(a) Obama birthplace



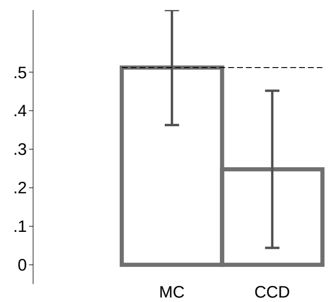
(b) Obama religion



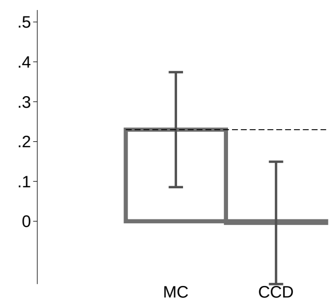
(c) ACA illegal



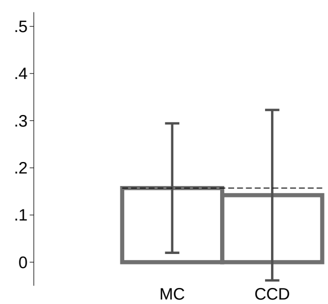
(d) ACA death panels



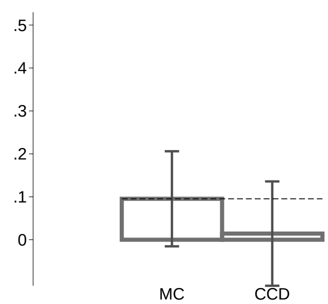
(e) GW causes



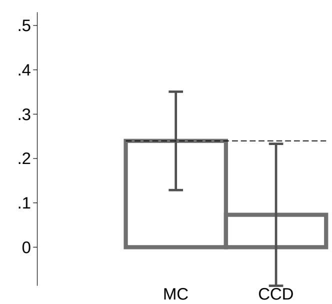
(f) GW scientists agree



(g) Voter fraud



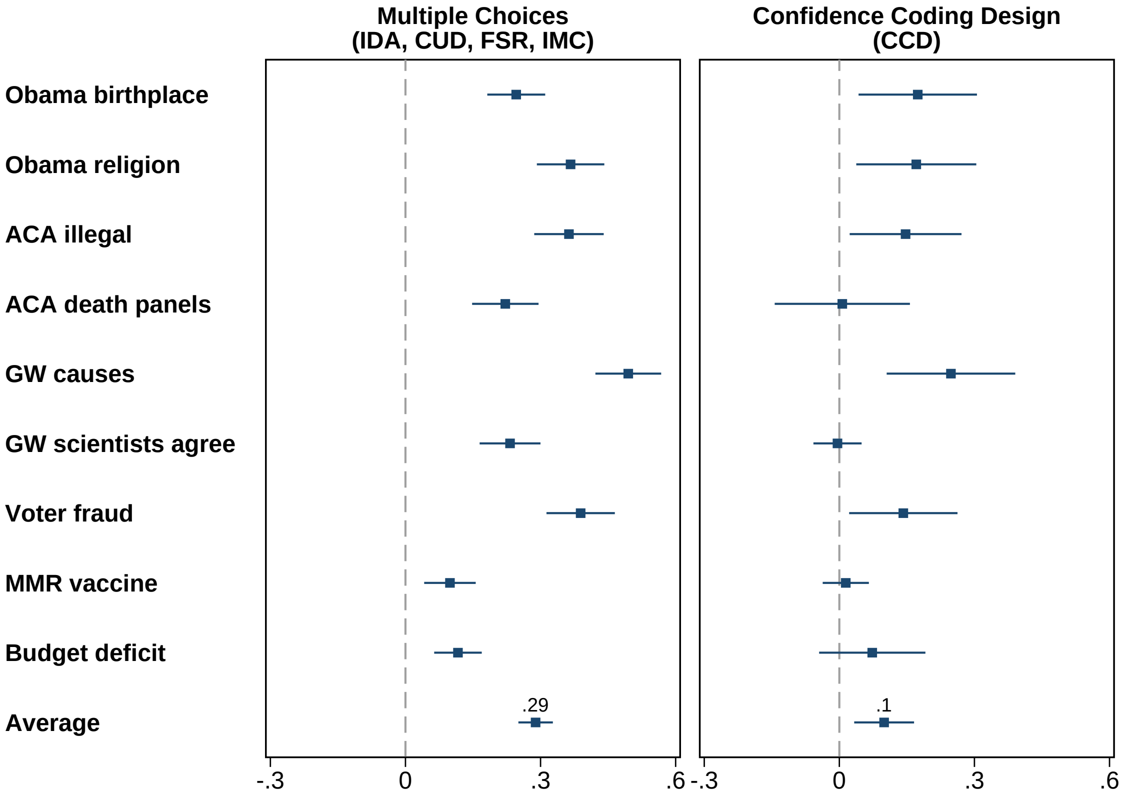
(h) MMR vaccine



(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 multiple choice CUD condition (see Table 1 for the descriptions). Reconstructed from the estimates from Table SI 2.4. Capped vertical bars indicate 95% confidence intervals.

Figure SI 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 IDA, CUD, FSR, and IMC) and the confidence coding design (CCD). Corresponds to Figure 4. The CCD condition only considers selecting the right answer with complete confidence as evidence that the respondent knows the answer (see Appendix SI 5). See Tables SI 2.1 to SI 2.5 in Appendix SI 2 for the regression estimates of the multiple-choice conditions to the confidence coding condition.

SI 3 Item Text for MTurk 1

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

- **IP = IDA**
- **RW = CUD**

Preface for Different Conditions

RW, IP

Now here are some questions about what you may know about politics and public affairs.

FSR, IMC, CCD

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 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

  RW and IP

  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

  FSR
  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)

  IMC

  Was Barack Obama born in . . . ?

  - the US
  - Another country
  - DK (plus DK pref)

- Obama Religion
  RW

  Do you personally believe that Barack Obama is a . . . ?

  - Muslim
  - Christian

  IP

  Most people have a religion. Some people believe Barack Obama is a Muslim. Do you personally believe that Barack Obama is a . . . ?

  - Muslim
  - Christian

FSR

Some people believe Barack Obama is a Muslim. Is he a . . . ?

- Muslim
- Christian
- DK (+ DK pref)

IMC

Is Barack Obama a . . . ?

- Muslim
- Christian
- DK (plus DK pref)

- ACA Illegal
  RW

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

IP

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

FSR

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)

IMC
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
RW
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

IP

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

FSR

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)

IMC
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

  RW

  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

  IP

  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

  FSR

  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)

  IMC

  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
  RW
  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

  IP
  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

  FSR
  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)

  IMC
  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
  RW
  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

  IP
  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

  FSR
  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)

  IMC
  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
  RW
  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

IP
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

FSR
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)

IMC
Does the vaccine for Measles, Mumps, and Rubella (MMR) . . . ?

- Cause autism in children
- Not cause autism in children.
- DK (+ DK pref)

• Obama—Budget Deficit
RW
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

IP
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

FSR
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)

IMC

Since 2012, has the annual federal budget deficit . . .

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

SI 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, Postgraduate Degree

SI 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 SI 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 SI 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 SI 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 SI 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 SI 5.1: Affordable Care Act 1 Scale Question

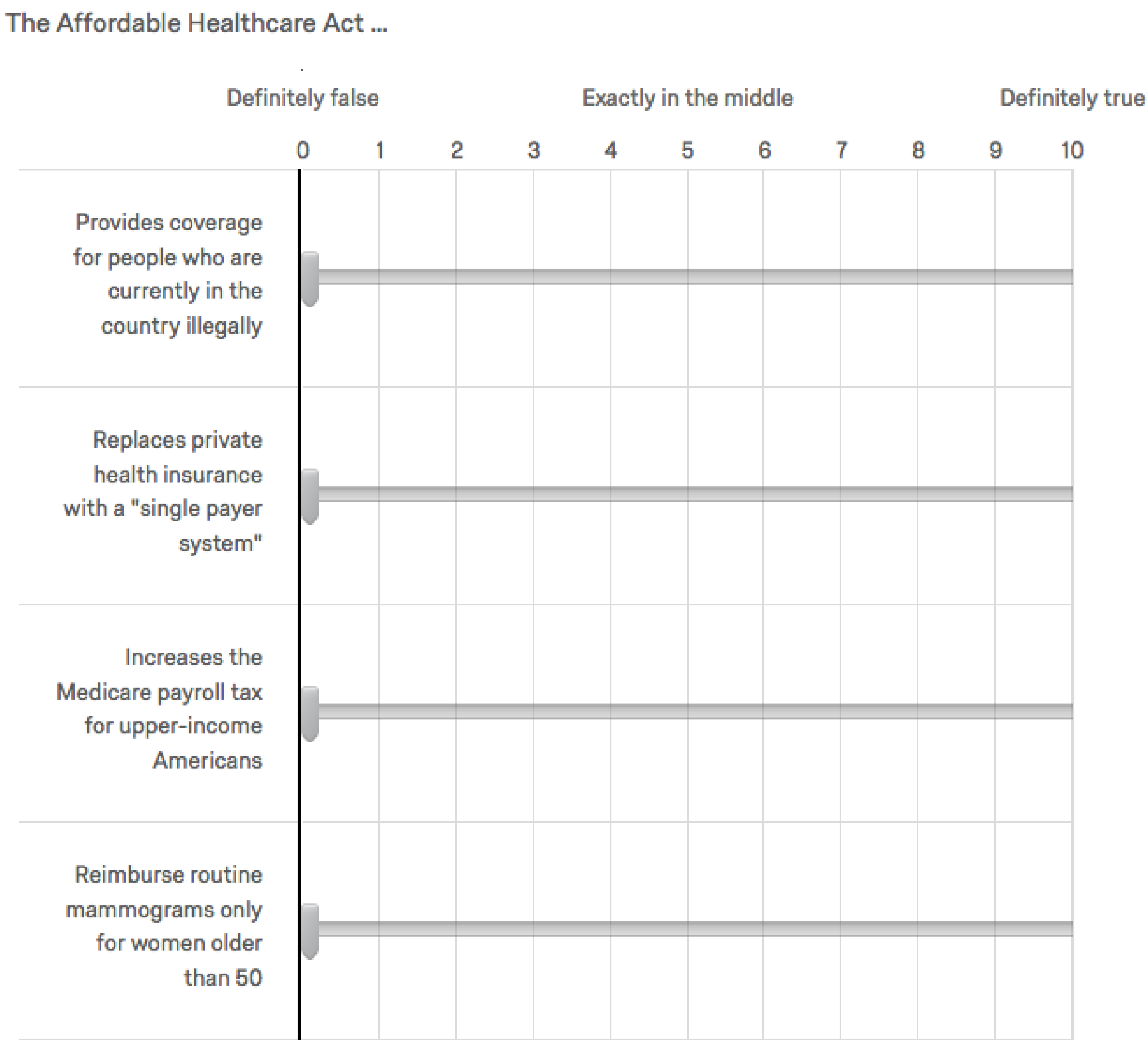


Figure SI 5.2: Greenhouse Gases Scale Question

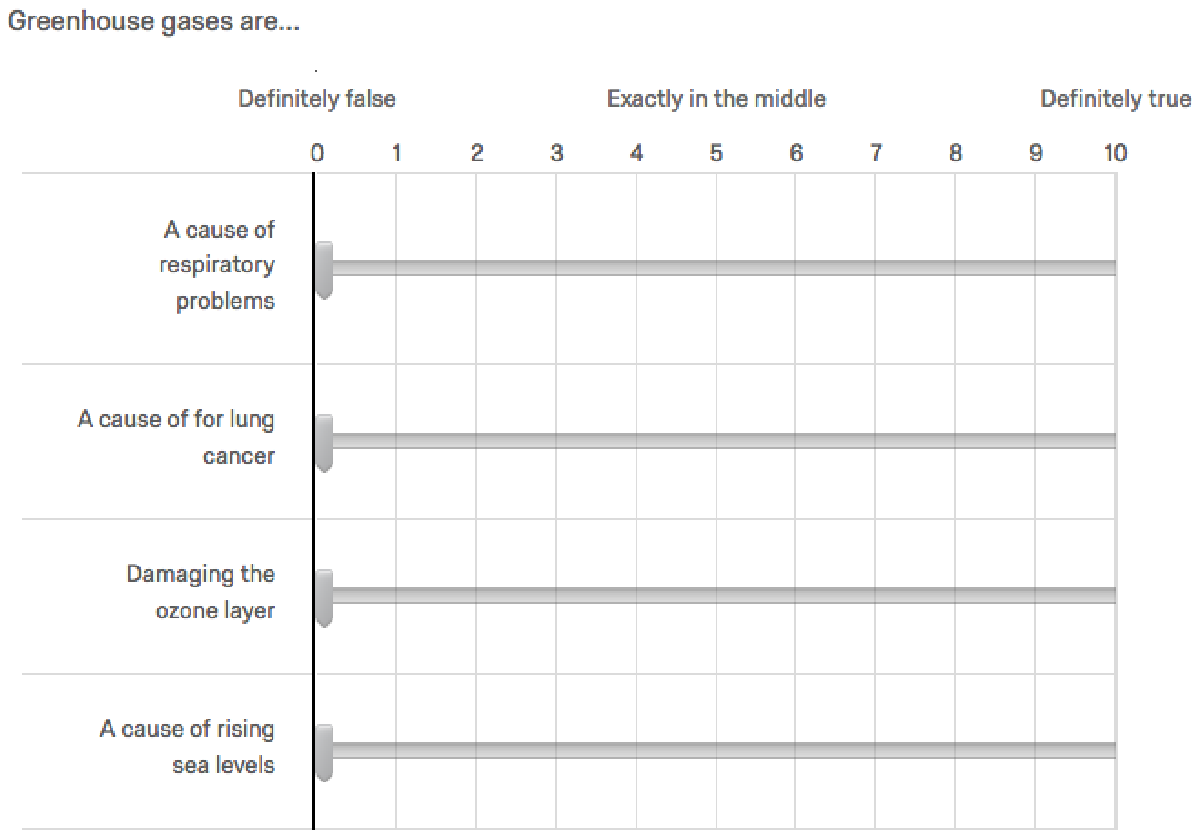


Figure SI 5.3: Affordable Care Act 2 Scale Question

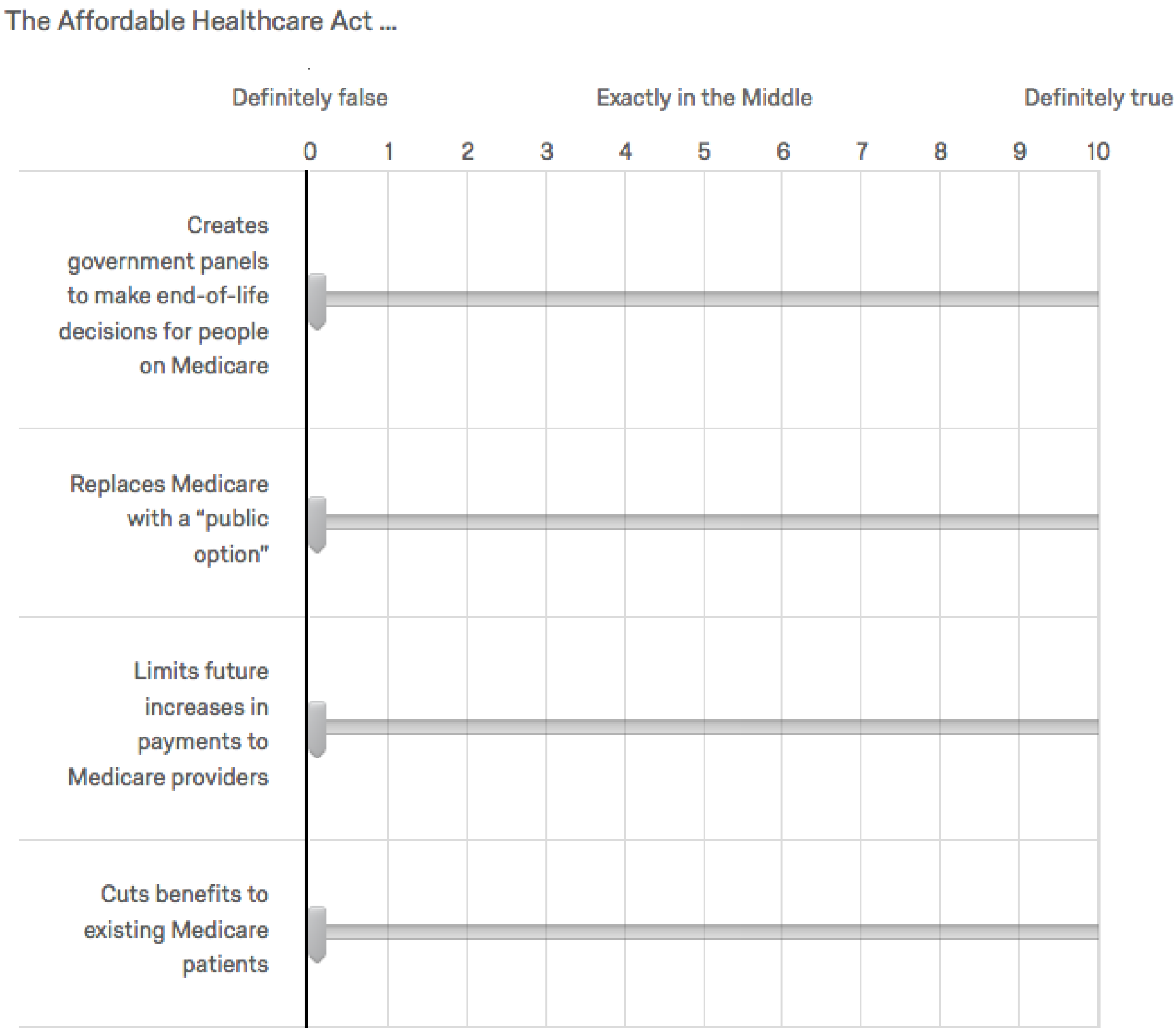
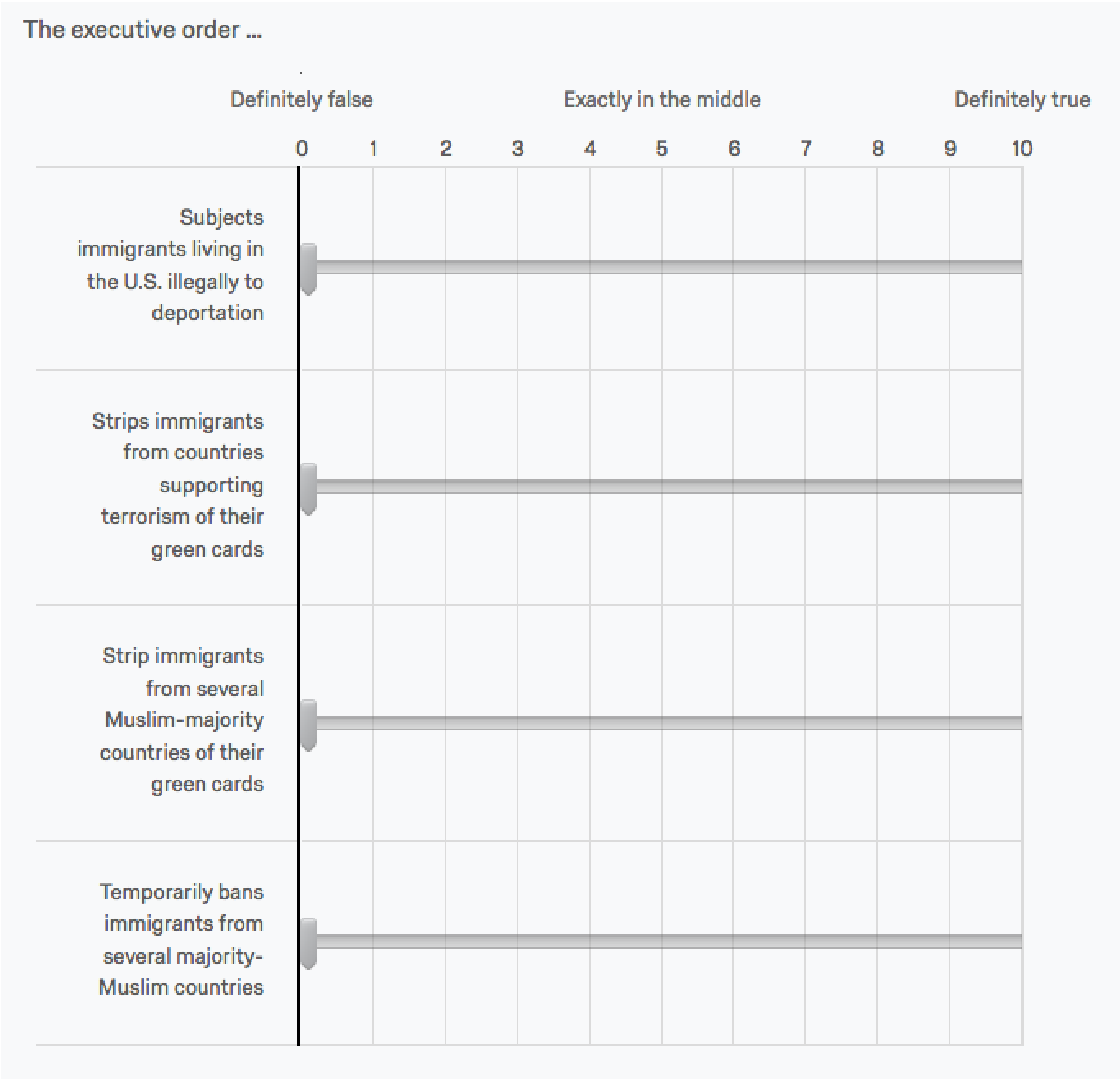


Figure SI 5.4: Executive Order Scale Question



SI 6 Alternate Scoring Criteria for CCD

Table SI 6.6 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 SI 6.6: Proportion correct across questions and scoring

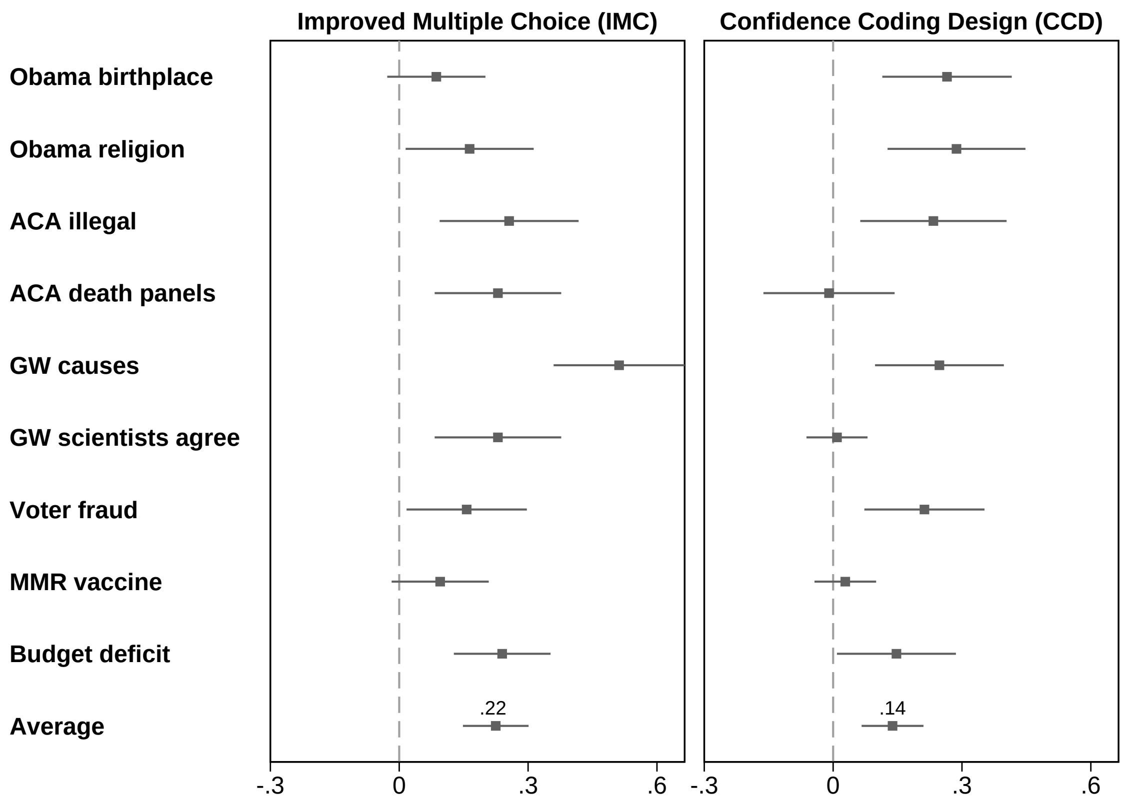
Question	**Closed**	**Relative Scoring**	
		8	**10**
ACA	0.24	0.01	0.01
ACA2	0.26	0.04	0.01
GG	0.25	0.02	0.01
DT	0.78	0.10	0.07

Table SI 6.7: 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.

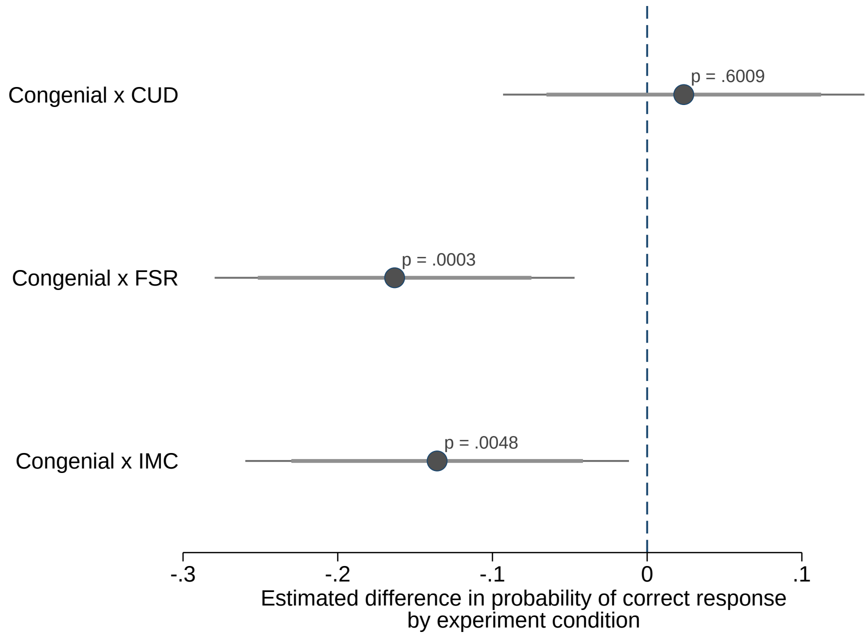
Figure SI 6.1: Robustness check for Confidence Coding and Knowledge Gaps: MTurk 1



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 SI 5). Corresponds to Figure 4, the difference here is that the analysis implements a Confidence Coding threshold of 8. See Table SI 6.7 for the analogous table for Study 3: MTurk 2 Results.

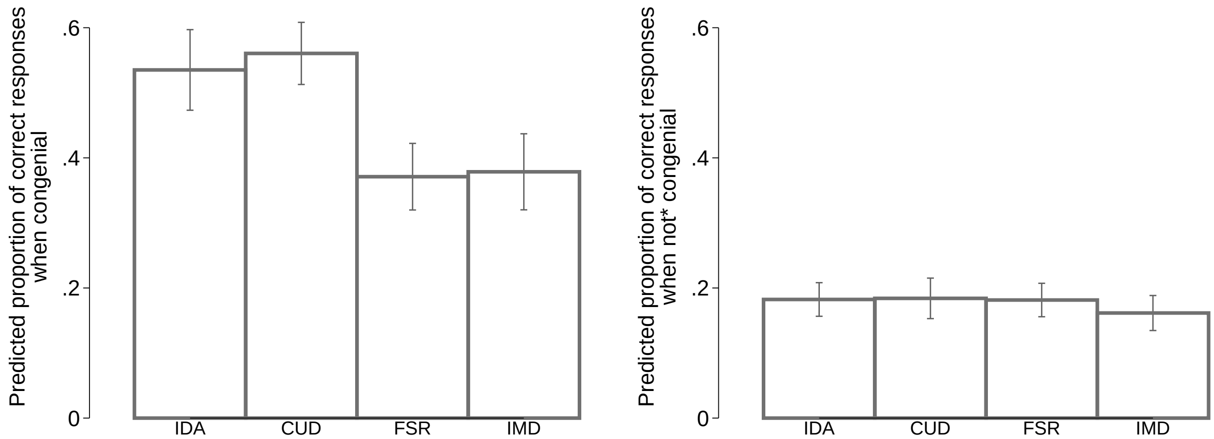
SI 7 Alternative Visualizations of Results

Figure SI 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 the IDA (Inflationary Design Approach) condition. 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 SI 7.2 visualizes absolute effects by condition.

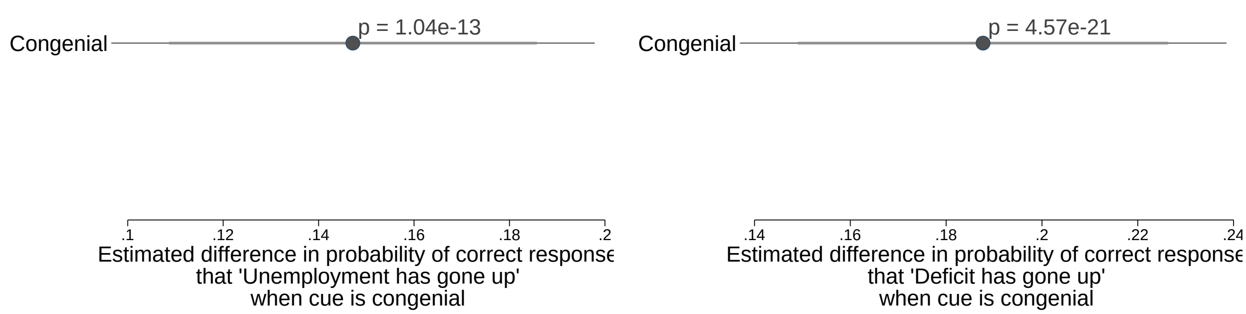
Figure SI 7.2: Partisan Gap by Treatment Arm (YouGov)



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

Bars indicate the estimated proportion of correct answers by whether the correct response is congenial (1st column) vs 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 SI 7.1 visualizes differences in effects of the CUD, FSR, and IMC conditions relative to the IDA condition when the response is congenial.

Figure SI 7.3: The Impact of Partisan Cues on Partisan Gaps (YouGov)



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

The figure shows the estimated difference in the probability of getting the correct response for the two questions in the YouGov survey when the correct response is congenial. Coefficients are as estimated in Table 3. Demographic controls include age cohort, gender, education level, marital status, employment status, news interest, family income, and race. Horizontal lines indicate 99% and 95% confidence intervals (by gradation). Figure 2 visualizes absolute effects by condition.

Figure SI 7.4: Partisan Gap on Unemployment by Treatment Arm (Texas Lyceum)

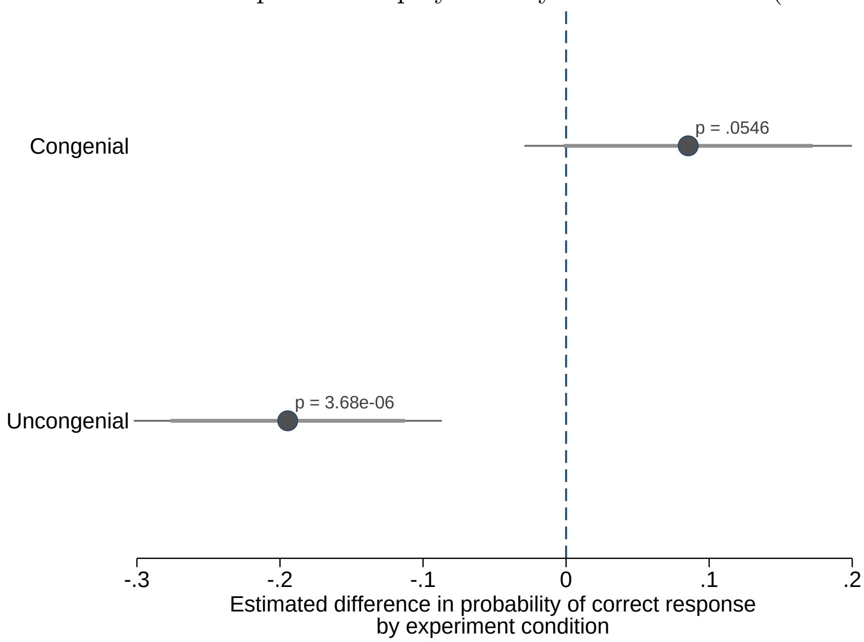
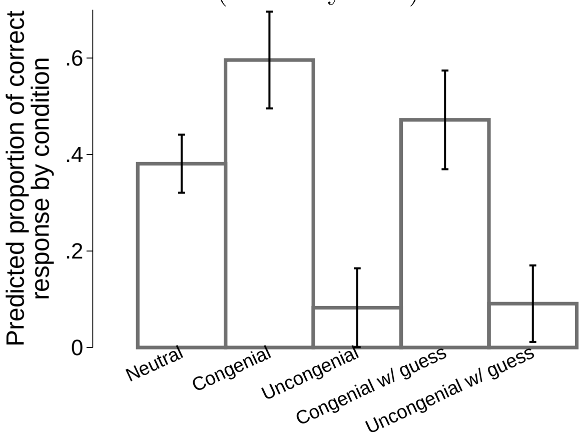
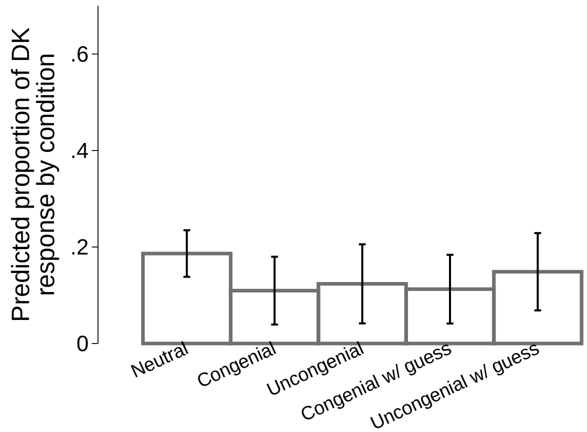


Figure shows the estimated difference in the probability of getting the correct response by congenial vs uncognenial (base categorical is neutral cue). The baseline category is the neutral cue. Coefficients are as estimated in Table 4 (column (2))—demographic covariates include age cohort, gender, education level, marital status, number of children, children's school enrollment, family income, religion, liberalism/conservatism, and race. Horizontal lines indicate 99% and 95% confidence intervals (by gradation). Figure 3 visualizes absolute effects by condition.

Figure SI 7.5: Impact of congenial/uncongenial cues, with/without guessing on Federal Taxes (Texas Lyceum)



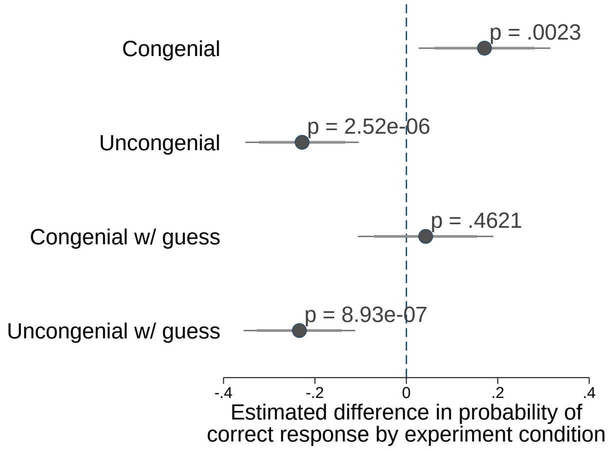
(a) Responded "Gone up"



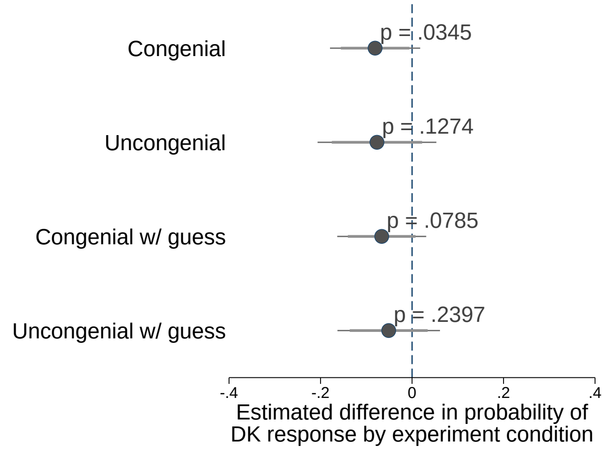
(b) Responded "Don't Know"

The figure reports the estimated proportion of correct answers in the five conditions, with the Neutral condition (no congenial cues) as the baseline. This figure is an alternative visualization of Table 5 (columns (2) and (4)). Demographic controls include age cohort, gender, education level, marital status, number of children, children's school enrollment, family income, religion, liberalism/conservatism, and race. Capped vertical bars indicate 95% confidence intervals. Figure SI 7.6 visualizes differences in the effects of the four conditions relative to the baseline neutral condition.

Figure SI 7.6: Impact of Various Treatments on Partisan Gap on Federal Taxes (Texas Lyceum)



(a) Responded "Gone up"



(b) Responded "Don't Know"

Figure shows the estimated difference in the probability of getting responding "Gone up" in subfigure (a) and "Don't Know" in subfigure (b). Coefficients are as estimated in Table 5 (columns (2) and (4)). Demographic controls include age cohort, gender, education level, marital status, number of children, children's school enrollment, family income, religion, liberalism/conservatism, and race. Horizontal lines indicate 99% and 95% confidence intervals (by gradation). Figure SI 7.5 visualizes absolute effects by condition.

SI 8 Differences by Subgroups (Gender/Race)

In this appendix, we test if there is sub-group-specific variation in how the different question designs affect the proportion of correct responses we receive. We interact the variables in Table 2 first with gender and then with race in order to determine if our design varies in its effect. As the two tables below show, this is not the case. We are now examining whether statistically significant differences occur across subgroups in Table 2. As a reminder, Table 2 examines whether congenial questions and different question designs affect the proportion of correct responses. We update the interaction in Table 2 to a triple interaction with gender (Table SI 8.8) and race (Table SI 8.9) and can report that empirically, we do not find systematic hidden knowledge across sub-groups. The only statistically significant coefficient exists for the interaction between the fewer substantive responses and the female dummy. Central to our argument is that the improved multiple-choice design is not affected by the interaction with gender or race. The improved designs we are proposing do not exhibit sub-group-specific variation.

Table SI 8.8: 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]
CUD	0.033	0.035
	(0.029)	(0.026)
	[0.249]	[0.189]
FSR	0.053+	0.048+
	(0.031)	(0.029)
	[0.089]	[0.093]
IMC	−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 × CUD	−0.033	−0.036
	(0.069)	(0.066)
	[0.626]	[0.586]
Congenial × FSR	−0.223**	−0.212**
	(0.068)	(0.066)
	[0.001]	[0.001]
Congenial × IMC	−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]
CUD × Female	−0.056	−0.057
	(0.042)	(0.040)
	[0.185]	[0.152]
FSR × Female	−0.093*	−0.087*
	(0.039)	(0.037)
	[0.018]	[0.019]
IMC × Female	0.007	0.001
	(0.037)	(0.036)
	[0.852]	[0.975]
(Congenial × CUD) × Female	0.123	0.123
	(0.092)	(0.090)
	[0.183]	[0.174]
(Congenial × FSR) × Female	0.093	0.088
	(0.095)	(0.093)
	[0.324]	[0.343]
(Congenial × IMC) × 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 IDA, CUD, FSR, and IMC conditions. 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 SI 8.9: 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]
CUD	0.003	−0.000
	(0.025)	(0.024)
	[0.907]	[0.992]
FSR	0.009	0.008
	(0.021)	(0.021)
	[0.678]	[0.701]
IMC	−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 × CUD	0.039	0.046
	(0.048)	(0.047)
	[0.418]	[0.336]
Congenial × FSR	−0.179***	−0.169***
	(0.047)	(0.047)
	[0.000]	[0.000]
Congenial × IMC	−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]
CUD × Non-White	−0.013	0.005
	(0.045)	(0.045)
	[0.765]	[0.910]
FSR × Non-White	−0.036	−0.037
	(0.049)	(0.045)
	[0.458]	[0.410]
IMC × Non-White	0.018	0.028
	(0.047)	(0.048)
	[0.704]	[0.557]
(Congenial × CUD) × Non-White	−0.269+	−0.308+
	(0.161)	(0.164)
	[0.094]	[0.061]
(Congenial × FSR) × Non-White	−0.035	−0.044
	(0.141)	(0.142)
	[0.807]	[0.757]
(Congenial × IMC) × 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 IDA, CUD, FSR, and IMC conditions. 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 on vertical space.

SI 9 Linear Probability Models

Table SI 9.10: The Impact of Partisan Cues on Partisan Gaps (YouGov)

	Unemployment has gone up		Deficit has gone up	
	(1)	(2)	(3)	(4)
Congenial	0.144***	0.148***	0.178***	0.189***
	(0.019)	(0.019)	(0.021)	(0.019)
	[0.000]	[0.000]	[0.000]	[0.000]
Demographic controls	.	Yes	.	Yes
Respondents	2,104	2,063	2,104	2,063

Table is the same as Table 3, except that all models are logit models. Reported coefficients are the marginal effects (directly comparable to Table 3). 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. Exact p-values in square brackets. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001.

Table SI 9.11: Partisan Gap on Unemployment by Treatment Arm (Texas Lyceum)

	Unemployment has gone up	
	(1)	(2)
Congenial	0.077+	0.079*
	(0.042)	(0.040)
	[0.064]	[0.049]
Uncongenial	−0.178***	−0.199***
	(0.040)	(0.039)
	[0.000]	[0.000]
Demographic controls	.	Yes
Respondents	758	747

The table is the same as Table 4, except that all models are logit models. Reported coefficients are the marginal effects (directly comparable to Table 4). 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. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. Exact p-values in square brackets.

Table SI 9.12: Impact of Various Treatments on Partisan Gap on Federal Taxes (Texas Lyceum)

	Responded "Gone up"		Responded "Don't Know"	
	(1)	(2)	(3)	(4)
Congenial	0.170***	0.129**	−0.076*	−0.080*
	(0.039)	(0.043)	(0.038)	(0.038)
	[0.000]	[0.003]	[0.045]	[0.036]
Uncongenial	−0.374***	−0.298***	−0.059	−0.076
	(0.073)	(0.072)	(0.043)	(0.048)
	[0.000]	[0.000]	[0.164]	[0.109]
Congenial w/ guessing	0.072+	0.028	−0.072+	−0.074*
	(0.041)	(0.043)	(0.038)	(0.037)
	[0.077]	[0.525]	[0.057]	[0.045]
Uncongenial w/ guessing	−0.353***	−0.297***	−0.033	−0.049
	(0.063)	(0.066)	(0.037)	(0.038)
	[0.000]	[0.000]	[0.369]	[0.190]
Demographic controls	.	Yes	.	Yes
Respondents	758	752	758	722

The table is the same as Table 5, except that all models are logit models. Reported coefficients are the marginal effects (directly comparable to Table 5). 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. All models are linear probability models. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001. Exact p-values in square brackets.

SI 10 Hierarchical Models

Model 1 has fixed effects by item and random effects by the respondent, while Model 2 takes guidance from Gelman et al. and estimates that model.

Table SI 10.13: 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)
Commonly Used Design (CUD)	−0.023	−0.023
	(0.020)	(0.025)
Fewer Substantive Responses (FSR)	0.0002	0.0002
	(0.020)	(0.026)
Improved Multiple Choice (IMC)	0.0004	0.0004
	(0.022)	(0.029)
Congenial × CUD	−0.132***	−0.132***
	(0.048)	(0.048)
Congenial × FSR	−0.173***	−0.173***
	(0.046)	(0.046)
Congenial × IMC	0.024	0.024
	(0.046)	(0.046)
Constant	0.030	0.184**
	(0.019)	(0.081)
Observations	5,652	5,652