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

Lucas Shen[*] Gaurav Sood[†] Daniel Weitzel[‡]

March 13, 2023[§]

Abstract

Research shows that partisan gaps in political knowledge are wide and widespread. Using a series of experiments, we investigate the extent to which partisan gaps are a result of differences in beliefs than motivated guessing, on-the-spot inferences, cheerleading, and other such processes. We manipulate common features of knowledge items in commercial surveys and find that they inflate the partisan gap in beliefs by 40%. The artificially large partisan gaps in commercial polls are partly a result of item features that cause people who don't know to offer a substantive response. In all, we find that partisans know far less and the absolute magnitude of the partisan gap in beliefs is substantially smaller.

Keywords: 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 challenge the idea that citizens can hold representatives accountable (Hochschild and Einstein 2015). Hence the alarm over research that suggests as much (Bartels 2002; Campbell et al. 1980; Jerit and Barabas 2012). 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). In this paper, we extend this investigation.

Using data from four different surveys, we examine how common features of knowledge items on commercial polls, e.g., encouraging respondents to guess, and scoring rules, e.g., accounting for respondents' self-assessed confidence about the answer, affect partisan gaps. Results from survey experiments consistently show 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). To further ablate response biases, we use an instrument and scoring scheme inspired by Pasek, Sood and Krosnick (2015); Graham (2021) that takes into account respondents' confidence in their answers. Using the scoring scheme, we find that partisan gaps are up to 50% smaller.

Our results contribute to a growing literature that suggests that a large fraction of partisan gaps are artifacts of survey design. The results alleviate concerns about democratic health. However, our study also provides reasons to be pessimistic. Reducing guessing and other such response biases reveals a yet more sobering picture of how much partisans know.

Theory, Motivation, and Empirical Strategy

A large body of research suggests that partisan gaps in political knowledge are wide and widespread (Bartels 2002; Jerit and Barabas 2012; Lodge and Taber 2013). One explanation for partisan gaps is that they are a result of differences in what partisans know. More recently, scholars have offered a second explanation: partisan responding (Bullock et al. 2015; Prior, Sood and Khanna 2015). They find that nearly half of the partisan gap is a result of biases in the measurement process. In the following sections, we elaborate on both of these explanations.

Partisan Differences in Beliefs

Partisan gaps in survey measures of political knowledge and misinformation may reflect differences in what partisans believe to be true. Differences in beliefs may, in turn, stem from selective exposure to information or motivated reasoning and retention.

Selective exposure to information–being exposed to more congenial than uncongenial information–can affect what facts people know about the world (Redlawsk 2002; Stroud 2010). Conventionally, partisan gaps are thought to stem from cognitive dissonance—people find information that is dissonant to their worldview to be painful and work to avoid it (e.g., Abelson 1959; Festinger 1962).

Partisans, however, do not have to prefer congenial information to consume more of it. For example, African Americans, who overwhelmingly identify as Democrats, may be more exposed to the negative consequences of economic downturns than White Americans, a majority of whom identify as Republicans, and as a result may have different beliefs about economic conditions than White Americans. By the same token, selective exposure may stem from different 'tastes' in politics. For instance, partisans of different stripes may be interested in different policies. Viewed thus, the partisan gap is similar to other types of knowledge gaps across groups—see research on gaps in gender (Dolan 2011; Barabas et al. 2014) and race (Abrajano 2015).

Whatever the cause, the effect of selective exposure is undoubtedly made worse by 'motivated skepticism' (Taber and Lodge 2006; Stroud 2008). People are more skeptical of uncongenial than congenial information (Zaller 1992). As a result, people are more likely to question uncongenial information and work to disprove it. People may also be more inclined to distrust and ignore uncongenial information entirely (Peterson and Iyengar 2021). Separately, but relatedly, people may be less likely to remember uncongenial information (see, for example Bayes et al. 2020; Hill 2017; Flynn, Nyhan and Reifler 2017; Taber and Lodge 2006). In all, it is possible that selective exposure, motivated skepticism, motivated retention, and related processes are the sole explanations for the observed partisan gaps in political knowledge.

Artifact of the Survey Design And Scoring

Measuring quantities in mind-in our case factual beliefs-using surveys is imperfect. On factual questions where the answer has implications about a party, responses are likely to be biased because of:

- **Partisan Cheerleading.** Surveys can encourage respondents to respond 'expressively' by highlighting partisan motivations over accuracy motivations (Bullock et al. 2015; Huber and Yair 2018; Prior, Sood and Khanna 2015). This can cause partisans to pick the partisan congenial answer even when they know it is wrong.
- **Affect Based Inference.** For example, 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.
- **Stereotype Based Inference.** For instance, Republicans may think of Democrats as generally indifferent to deficits, and may hence infer, without actually knowing, that deficits increased under Mr. Obama (e.g. Rahn 1993; Goggin, Henderson and Theodoridis 2020). In a highly polarized political environment, minimal information can be enough to switch individuals from answering a knowledge question to using

affect or expressive motivations to answer a question (Klar 2014; Merkley and Stecula 2018).

These biases are likely affected by survey features. For instance, adding a partisan cue to the question causes partisans to pick the partisan congenial answer plausibly by priming partisan considerations (see, e.g., Prior, Sood and Khanna 2015). More generally, survey responses to partisan consequential factual items on political surveys, which likely increase the salience of politics in respondents' minds, may be contaminated by affect-based inference. Secondly, survey features that encourage people to guess when they don't know likely inflate partisan gaps by increasing the share of biased guesses (Bullock et al. 2015).

Lastly, conventionally, all correct answers are taken as evidence that the respondent knows the fact (Luskin and Bullock 2011). This conflates guesses and on-the-spot inferences with knowledge and inflates the partisan gap. Pasek, Sood and Krosnick (2015) use self-assessed confidence to rescore the answers, taking only correct answers respondents are confident about as evidence that the respondent knows the item. More recently Graham (2021) finds that self-assessed confidence in answers is correlated with the reliability of the answers. We postulate that taking only confident correct answers as knowledge would reduce the partisan gap (see also Graham 2021).

Features of Political Knowledge Items in Media Polls

Based on an analysis of 180 media polls, Luskin and Bullock (2011) find that many media polls include guessing encouraging features such as providing background information and social proof in the question stem. For instance, less than 9% of the surveys offered an explicit 'Don't Know' or 'Not Sure' option. And about half of the items offered only two choices. As the number of response options increases, the probability of correct answers decreases (Bullock and Rader 2022). An overwhelming majority of the items (168) also included wording that encouraged guessing, by framing the factual question as one of a 'matter of opinion.' They also find that the scoring rules used by analysts treat all correct responses–even when the respondent is inconfident about their answer—as evidence of knowledge. In all, as we note above, these guessing encouraging features are likely to inflate partisan gaps. We expect that removing these inflationary features will diminish the partisan gaps in political knowledge.

Empirical Strategy

To test the effect of various "inflationary" features of the survey and question design on the partisan knowledge gap, we fielded four surveys. 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 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 scoring rules on the partisan gap.

Before we proceed further, a note. Many of the questions we use in our analysis 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 right answer. Like Bartels (2002); Prior, Sood and Khanna (2015), etc., 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. We believe that the results obtained here are likely generalizable to other 'less polarized' knowledge questions as the number of politically consequential questions with partisan implications that are not polarized is rapidly declining. Having said that, we believe more research is needed to test the generalizability of these results.

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.

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.) (For generalizability of effects in studies conducted on MTurk, see (Mullinix et al. 2015; Coppock, Leeper and Mullinix 2018).) In each condition, respondents answered nine misinformation items, ranging from Mr. Obama's citizenship to whether global warming is happening or not. (For the question wording for each of the items, see Appendix SI 2.)

The four conditions are:

Inflationary Design Approach (IDA) The IDA 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. 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 reads: "Now here are some questions about what you may know about politics and public affairs..."

Commonly Used Design (CUD) CUD makes one change to the IDA. 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.

Fewer Substantive Responses (FSR) FSR makes two changes to CUD. 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 to not 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).

Improved Multiple Choice (IMC) IMC is the best version of these multiple choice questions. It offers respondents a 'Don't Know' option and does not include guessing encouraging neutral information or social proof.

Table 1: Experimental Treatments

		Treatments			
Condition	Label	Don't Know	Social Proof	Guessing Encouraged	Neutral Information
1	IDA	No	Yes	Yes	Yes
2	CUD	No	No	Yes	Yes
3	FSR	Yes	No	No	Yes
4	IMC	Yes	No	No	No

Measures

We measure partisanship using the conventional branched seven-point partisan self-identification scale. 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.

Results

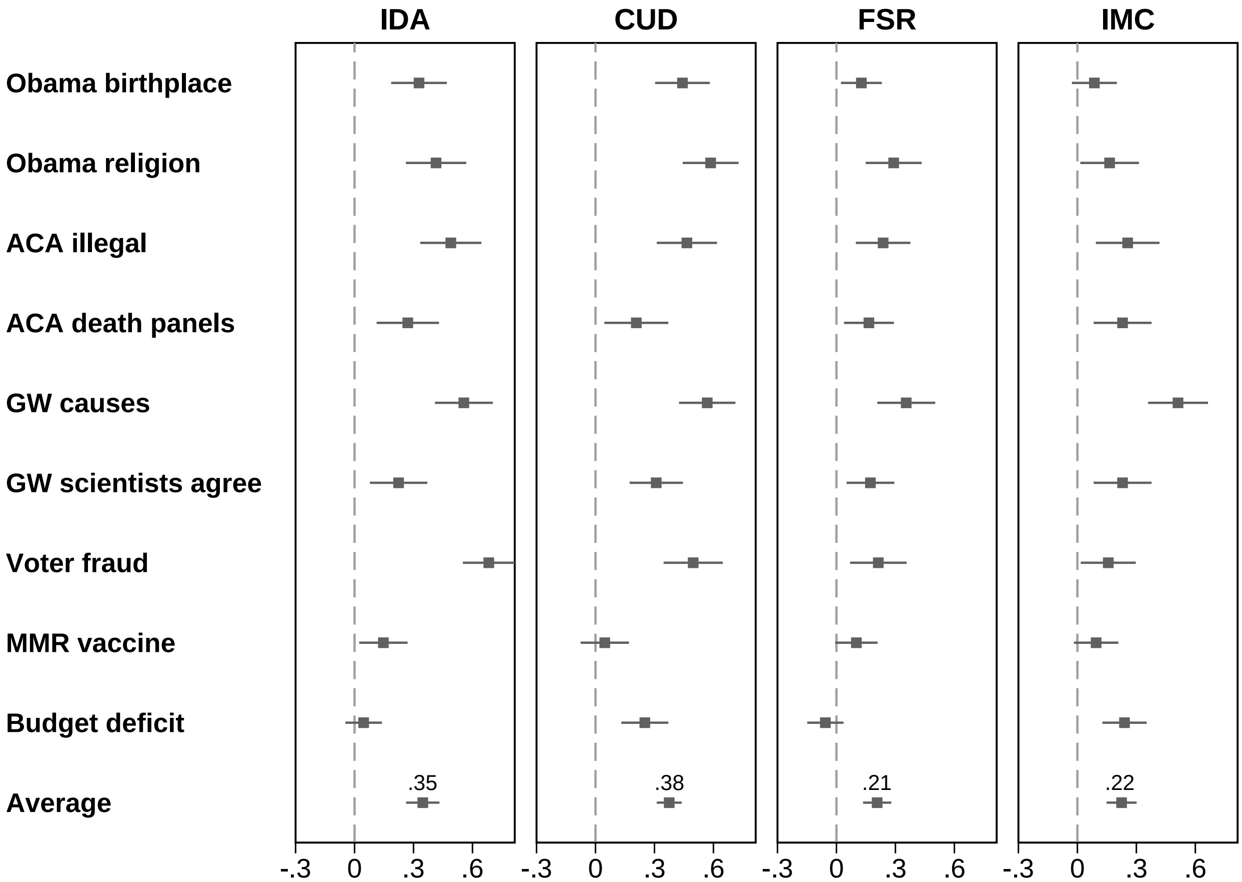
We start by summarizing the average partisan gap on each survey item in each treatment arm (see Figure 1). (See Figures SI 1.1 to SI 1.4 for balance tests.) In the baseline IDA condition (first column), when the correct response is congenial to the respondents' party, respondents are 35 percentage points more likely to choose the correct response. The partisan gap is irresponsive to the changes made in CUD. However, the estimates from the FSR and IMC conditions are approximately 14 percentage points lower than in the IDA. This means that removing inflationary features from the questions decreases the partisan gap in political knowledge. The 14 percentage points reduction translates to a 40% relative drop ($100 \times \frac{.35-.21}{.35}$).

To formally test our hypothesis, we regress whether or not the answer is correct, on the interaction of the survey conditions and the congenial dummy. For respondent i, survey item j, and condition k, we estimate the following equation

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

β captures the difference in the proportion of correct responses when the answer is congenial to the respondent's party, see Figure 1. A positive estimate suggests that respondents are more likely to choose the correct answer when it is congenial to their party. We focus on the δ_k's, which capture how the different conditions affect observed knowledge gaps (difference between columns in Figure 1). The baseline treatment arm is always IDA, so the δ_k's capture how the three conditions (CUD, FSR, IMC) affect partisan knowledge gaps. We include item fixed-effects and cluster standard errors by respondents.

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



The figure shows the estimated partisan gap in each of the four conditions Table 1. Rows indicate the nine knowledge items (see Appendix SI 2) and their average. The partisan gap is estimated using the linear model $1\{\text{Correct response}\}_i = \alpha + \beta \text{congenial}_i + \varepsilon_i$. Congenial is a dummy variable that takes the value 1 when the correct response is congenial to the party. Horizontal bars are 95% confidence intervals constructed from robust standard errors.

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; Roush and Sood 2023).

Column (2) includes only the survey conditions. Two of them (FSR, IMC) produce partisan gaps that are significantly smaller than the baseline. In column (3), we include the interaction between congenial and the four conditions (baseline is IDA). Now the congenial variable captures the knowledge gap in the IDA condition (corresponding to column (1)

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)
CUD		0.010	0.000		0.011	0.002
		(0.028)	(0.022)		(0.028)	(0.021)
FSR		-0.064^{**}	0.000		-0.063^{**}	-0.001
		(0.024)	(0.019)		(0.024)	(0.019)
IMC		-0.080^{**}	-0.023		-0.079^{**}	-0.021
		(0.025)	(0.019)		(0.025)	(0.019)
Congenial × CUD			0.024			0.024
			(0.046)			(0.045)
Congenial × FSR			-0.173^{***}			-0.163^{***}
			(0.046)			(0.045)
Congenial × IMC			-0.132^{**}			-0.136^{**}
			(0.048)			(0.048)
Constant	0.179^{***}	0.306^{***}	0.184^{***}	0.156^{***}	0.303^{***}	0.164^{***}
	(0.007)	(0.020)	(0.014)	(0.013)	(0.024)	(0.016)
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

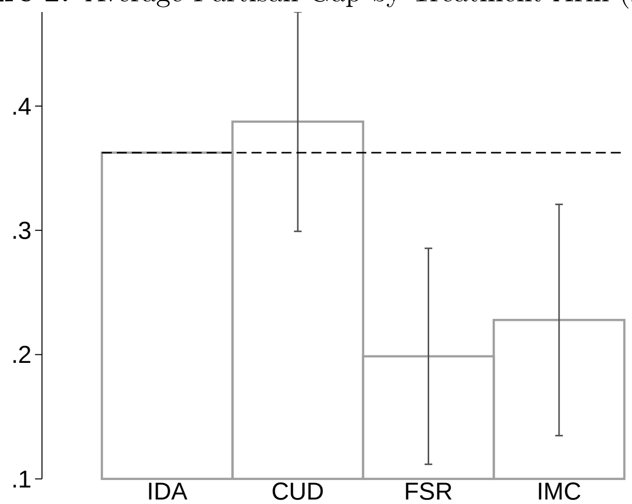
All models are linear probability models where the dependent variable is whether the response is correct or not. See Table 1 for the description of the IDA, CUD, FSR, IMC conditions. Demographic controls include age cohort, gender, education level (college degree, high school, no high school, post-graduate, and some college), and race (Hispanic, Asian, Black, White, Others). Standard errors are clustered at the respondent level. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001.

of Figure 1). The congenial and survey condition interactions reveal the extent to which partisan knowledge gaps change across the different survey conditions.

Figure 2 shows, in absolute terms, the estimates of how the different survey conditions attenuate the partisan gap. For the FSR interaction term, just adding a 'Don't Know' response option reduces the estimated knowledge gap by more than 49% ($p < 0.001$). For the IMC interaction term, adding a 'Don't Know' without social proof and without encouragement to guess reduces the estimated knowledge gap by more than 37% ($p < 0.01$).

Columns (4)–(6) of Table 2 show that including self-reported characteristics of respondents does not change the conclusion. Overall, Study 1 suggests that a large chunk of

Figure 2: Average Partisan Gap by Treatment Arm (MTurk 1)



The figure shows the predicted partisan gap in each of the conditions. The IDA is the baseline condition. (See Table 1 for a summary of the key features of the conditions.) The plot was constructed from the coefficient of the interaction term between the *congenial* indicator and the treatment arms as reported in column (3) of Table 2. Vertical bars show the 95% confidence intervals.

partisan gaps is a consequence of the inflationary questionnaire design features common in commercial polls.

Study 2: The Effect of Partisan Cues on Partisan Gaps

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

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 who were 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:

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?

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

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 or not the response is correct 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 time period became worse, 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 3 plots the results. As Panel (a) of Figure 3 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 3 shows that this effect is not unique to the unemployment question. On the budget deficit question, the difference is 18 percentage points ($p < 0.001$).[1]

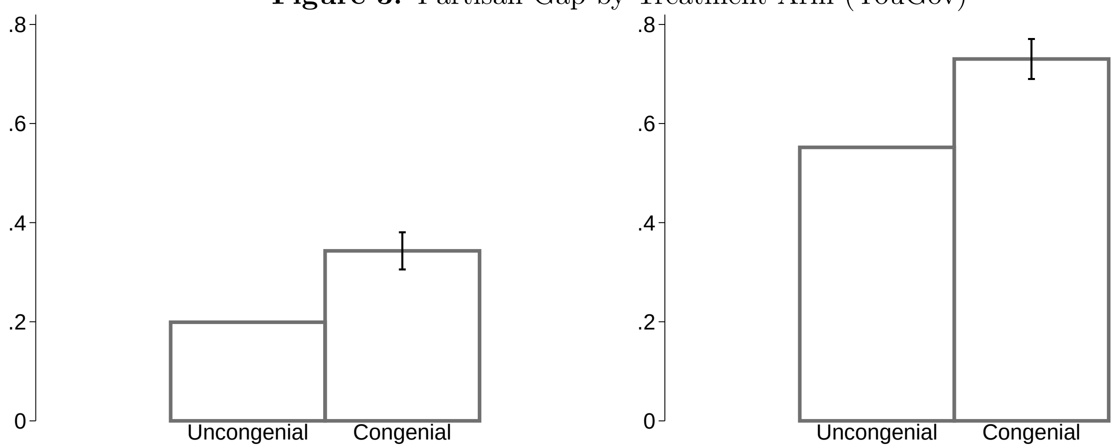
[1] Figure SI 1.5 show that there is some heterogeneity in how the congenial cue affects Republicans as opposed to Democrats. However, the effect is not unique to either party

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)
Constant	0.199***	3.569+	0.552***	7.636***
	(0.012)	(1.895)	(0.015)	(1.868)
R^2	0.026	0.055	0.035	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. All models are linear probability models. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001.

Figure 3: Partisan Gap by Treatment Arm (YouGov)



(a) Unemployment

(b) Budget deficit

Bars indicate the predicted percent of correct answers as reported in Table 3 (columns (1) and (4)). Capped vertical bars indicate 95% confidence intervals.

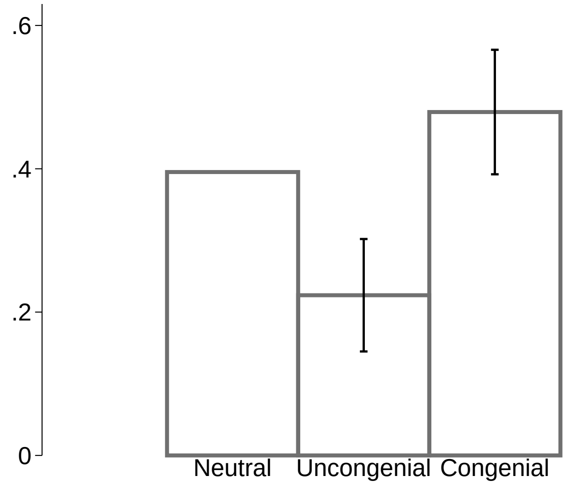
Study 2: Texas Lyceum Results

We supplement our results with the Texas Lyceum survey. As Figure 4 shows, on the unemployment question, the pattern that we saw on YouGov still holds when we include a

---

since partisans of both types are more likely to get the correct response when randomly assigned the congenial cue.

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



Bars indicate the predicted percent 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.

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)
Uncongenial	-0.172^{***}	-0.195^{***}
	(0.040)	(0.042)
Constant	0.395^{***}	0.057
	(0.030)	(0.175)
R^2	0.048	0.153
Demographic controls	.	Yes
Respondents	758	752

Dependent variable is whether 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.

neutral cue. Compared to respondents who received a neutral cue, respondents who received an uncongenial cue are 17 percentage points less likely to get the correct answer ($p < 0.001$). While respondents who receive a congenial cue are 8 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: 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)
Uncongenial	-0.298^{***}	-0.228^{***}	-0.063	-0.077
	(0.042)	(0.048)	(0.042)	(0.050)
Congenial w/ guessing	0.091^{+}	0.042	-0.074^{*}	-0.066^{+}
	(0.052)	(0.057)	(0.036)	(0.038)
Uncongenial w/ guessing	-0.290^{***}	-0.234^{***}	-0.038	-0.051
	(0.040)	(0.047)	(0.041)	(0.043)
Constant	0.381^{***}	-0.223	0.187^{***}	0.884^{***}
	(0.031)	(0.177)	(0.025)	(0.180)
R^2	0.151	0.219	0.009	0.126
Demographic controls	.	Yes	.	Yes
Respondents	758	752	758	752

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.

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. Including a substantive response encouraging cue does not have a stark effect. Overall, results from Studies 2 and 3 show that partisan cues dramatically affect the size of the partisan gaps.

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 takes into account respondents' confidence in their answers.

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 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 collected data in two separate surveys. Our first survey is the one underlying Study 1. The survey had a fifth condition in addition to the four conditions presented above. The fifth condition offered the same questions except this time respondents were asked to respond on a Likert scale. The CCD condition builds on the first four conditions and does not encourage guessing and features no social proof. (The exact question wording for each of the items is presented in Appendix SI 2.) Since the items are dichotomous choice, the CCD scoring is straightforward. We scored respondents who marked 'definitely true' about the right answer as knowledgeable. In Appendix SI 3, we try less stringent criteria.

For the second study, we turn to another MTurk survey. In the 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), the effect of greenhouse gases (1), and the consequences of Mr. Trump's executive order on immigration (1). 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. (The exact question wording for each of the items is presented in Appendix SI 3.)

The scoring for this study 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 of 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.

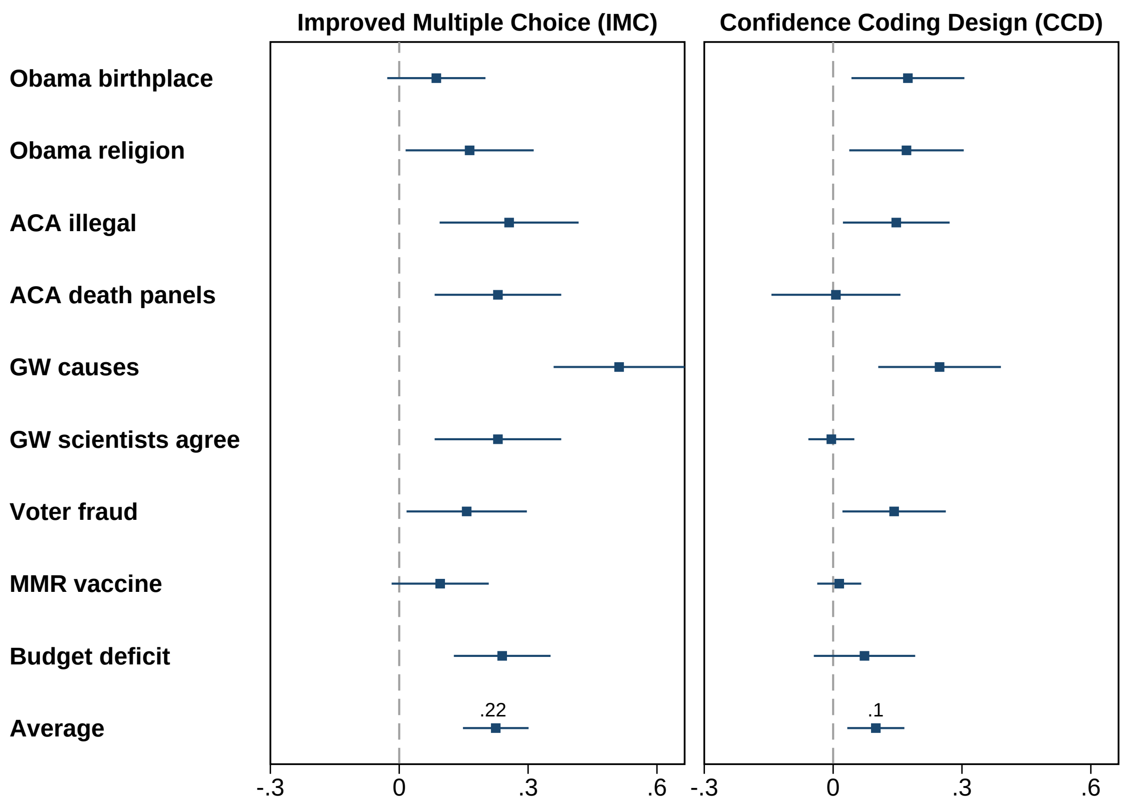
3. The respondent must be at least β confident in the correct answer. In the main text, we use a β of 10 but in Appendix SI 3, we try less stringent criteria.

4. The confidence in the incorrect answers cannot be above θ. In the main text, we use a θ of 0 but in the Appendix SI 3, we try less stringent criteria.

Study 3: MTurk 1 Results

The best version of the dichotomous multiple-choice items (IMC) showed a partisan gap of .22 (see Figure 1). As Figure 5 shows, nearly half of the gap vanishes under CCD. 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 5: Partisan Gaps in Knowledge in different question designs



The figure shows the estimated partisan gaps in knowledge from the MTurk sample for Study 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 3). See Tables SI 1.1 to SI 1.5 in Appendix SI 1.1 for the regression estimates of the multiple-choice conditions to the confidence coding condition.

Study 3: MTurk 2 Results

We use data from our last study 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:

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

for respondents i, survey item j, and scoring condition k. 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. A positive estimate indicates that respondents are more likely 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

Table 6: Confidence Scoring and Knowledge Gaps: MTurk Study 2

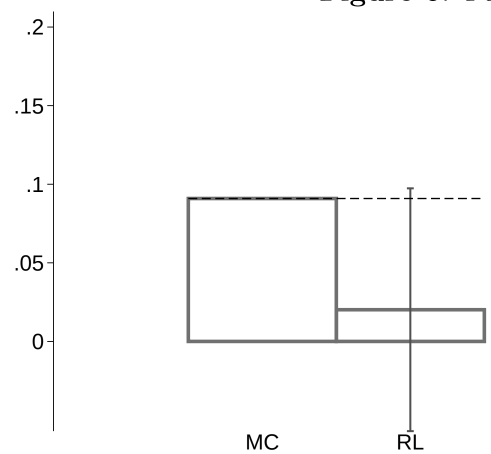
	Individual survey question				
	Affordable Care Act (1)	Affordable Care Act 2 (2)	Greenhouse gases (3)	Donald Trump (4)	All (5)
Congenial	0.091*	0.084*	0.087*	0.005	0.025
	(0.038)	(0.040)	(0.041)	(0.038)	(0.023)
Relative Scoring (RS)	−0.179***	−0.201***	−0.206***	−0.737***	−0.377***
	(0.028)	(0.030)	(0.032)	(0.028)	(0.018)
Congenial × RS	−0.071+	−0.070+	−0.098*	0.031	0.024
	(0.039)	(0.041)	(0.041)	(0.046)	(0.026)
Constant	0.179***	0.207***	0.217***	0.794***	0.376***
	(0.028)	(0.030)	(0.030)	(0.024)	(0.017)
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 3 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 relative scoring scheme, a response is correct only if the correct answer is selected with full confidence of 10 (see Research Design and Data in the Study 3: The Effect of the Scoring Method on Partisan Gaps section). The baseline is the multiple choice designs. Table SI 3.7 implements a robustness check setting the relative scoring threshold to 8. Standard errors are clustered at the respondent level. Significance levels: + 0.1 * 0.05 ** 0.01 *** 0.001.

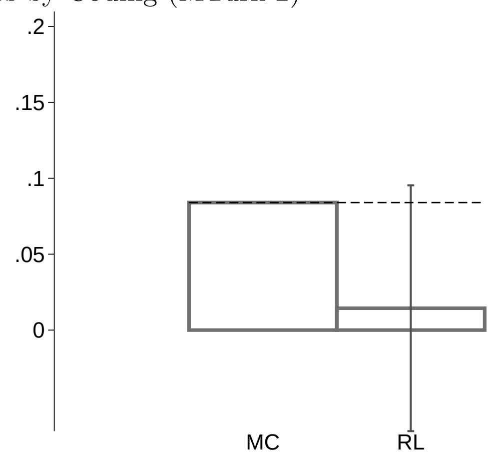
scoring are associated with a higher proportion of correct responses. In the MC scoring treatment partisans are more likely to get questions correct when answers are congenial to their partisanship. For the first three models focusing on the Affordable Care Act and Greenhouse Gas questions the effects 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.

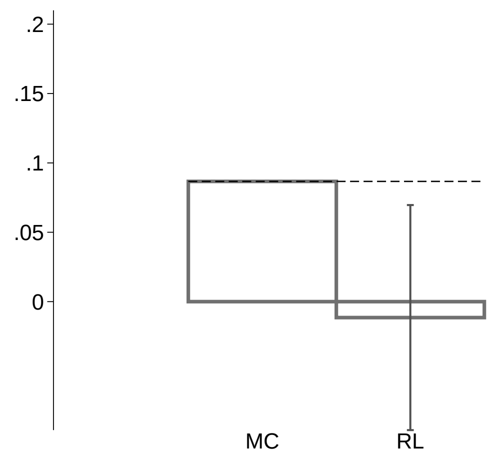
Figure 6: Partisan Gaps by Coding (MTurk 2)



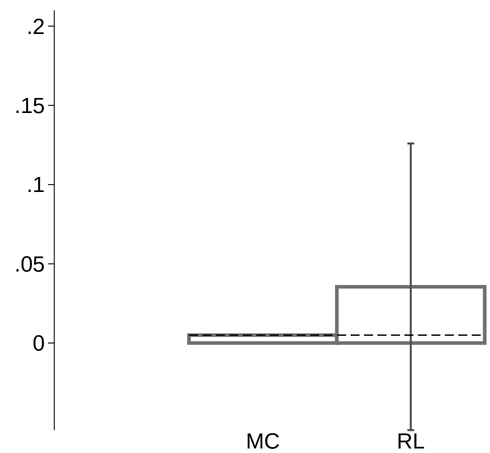
(a) Affordable Care Act



(b) Affordable Care Act 2



(c) Greenhouse Gases



(d) Donald Trump

Bars indicate the predicted percent of correct responses as reported in Table 6. MC bar indicates the predicted effect of multiple choice with congenial responses on getting the correct response. RL bar indicates the effect of relative scoring with congenial responses on getting the correct response relative to the multiple choice (MC) scheme. Capped vertical bars indicate 95% confidence intervals.

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 both 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 is not founded in differences in beliefs. We find that common features of commercial polls like not asking don't know, inserting a partisan cue, and treating inconfident answers as knowledge inflate the partisan gaps.

The fact that partisan gaps are smaller may seem at odds with some political behavior research. For instance, the theory of selective exposure 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 in light of events in a similar fashion (Gerber and Green 1999; Kernell and Kernell 2019; Coppock 2021).

In the end, the results paint a mixed picture of democratic competence. Smaller partisan gaps are partly a consequence of the fact that the average respondent doesn't know the facts. It is mostly partisan responding masquerading as partisan gaps. The upside is that partisan gaps are small and the downside is that people know even less than we thought.

References

Abelson, Robert P. 1959. "Modes of Resolution of Belief Dilemmas." *Journal of Conflict Resolution* 3(4):343–352.

Abrajano, Marisa. 2015. "Reexamining the "racial gap" in political knowledge." *The Journal of Politics* 77(1):44–54.

Barabas, Jason, Jennifer Jerit, William Pollock and Carlisle Rainey. 2014. "The question (s) of political knowledge." *American Political Science Review* 108(4):840–855.

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

Bayes, Robin, James N Druckman, Avery Goods and Daniel C Molden. 2020. "When and how different motives can drive motivated political reasoning." *Political Psychology* 41(5):1031–1052.

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.

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.

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.

Dolan, Kathleen. 2011. "Do women and men know different things? Measuring gender differences in political knowledge." *The Journal of Politics* 73(1):97–107.

Festinger, Leon. 1962. *A Theory of Cognitive Dissonance*. Palo Alto: Stanford University Press.

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

Flynn, DJ, Brendan Nyhan and Jason Reifler. 2017. "The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics." *Political Psychology* 38:127–150.

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.

Goggin, Stephen N, John A Henderson and Alexander G Theodoridis. 2020. "What goes with red and blue? Mapping partisan and ideological associations in the minds of voters." *Political Behavior* 42(4):985–1013.

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

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

Hill, Seth J. 2017. "Learning together slowly: Bayesian learning about political facts." *The Journal of Politics* 79(4):1403–1418.

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.

Klar, Samara. 2014. "Partisanship in a social setting." *American journal of political science* 58(3):687–704.

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. 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.

Merkley, Eric and Dominik A Stecula. 2018. "Party elites or manufactured doubt? The informational context of climate change polarization." *Science Communication* 40(2):258–274.

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.

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.

Rahn, Wendy M. 1993. "The role of partisan stereotypes in information processing about political candidates." *American Journal of Political Science* pp. 472–496.

Redlawsk, David P. 2002. "Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making." *Journal of Politics* 64(4):1021–1044.

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.

Stroud, Natalie Jomini. 2008. "Media use and political predispositions: Revisiting the concept of selective exposure." *Political Behavior* 30(3):341–366.

Stroud, Natalie Jomini. 2010. "Polarization and Partisan Selective Exposure." *Journal of Communication* 60:556–576.

Taber, Charles S and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50(3):755–769.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion.* Cambridge: Cambridge University Press.

SUPPORTING INFORMATION

SI 1 Supporting figures

Figure SI 1.1: MTurk Study 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

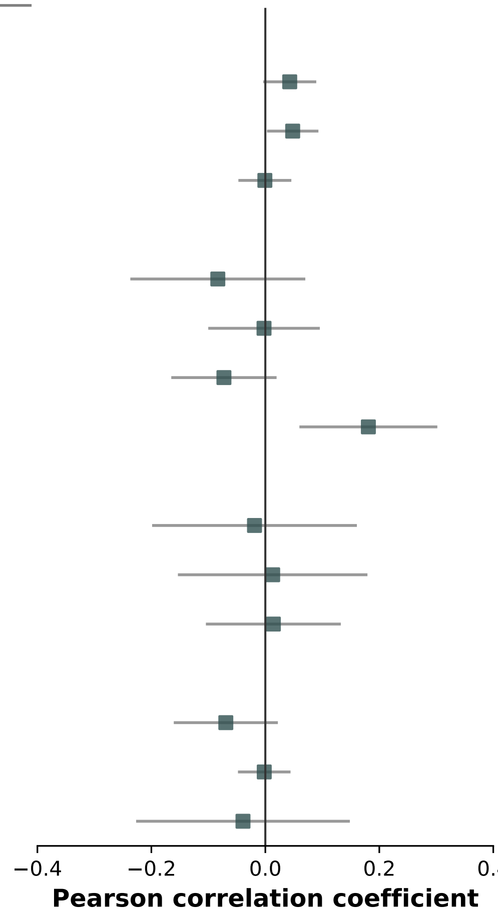


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compares respondents assigned to the IDA condition vs. respondents assigned to the CUD condition. See Table 1 in Study 1: The Effect of Guessing Encouraging Features. Rows are self-reported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

Figure SI 1.2: MTurk Study 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

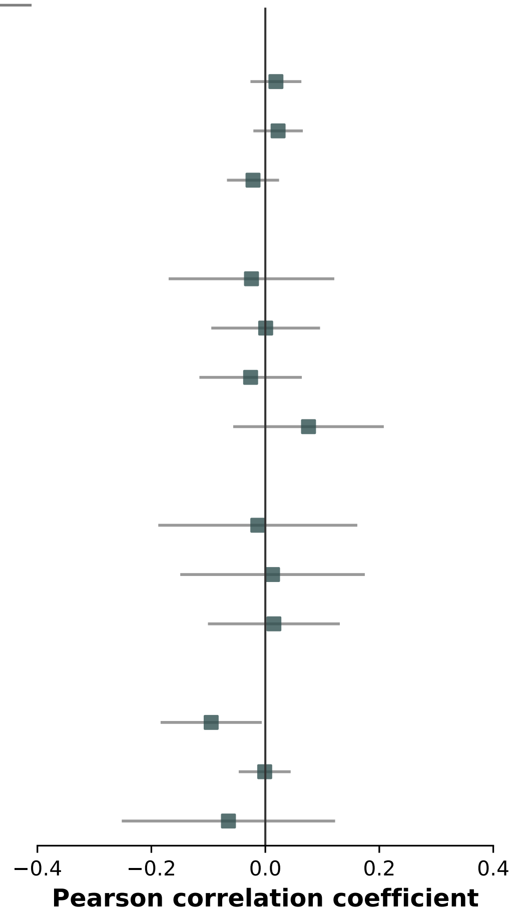


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compares respondents assigned to the IDA condition vs. respondents assigned to the FSR condition. See Table 1 in Study 1: The Effect of Guessing Encouraging Features. Rows are self-reported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

Figure SI 1.3: MTurk Study 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

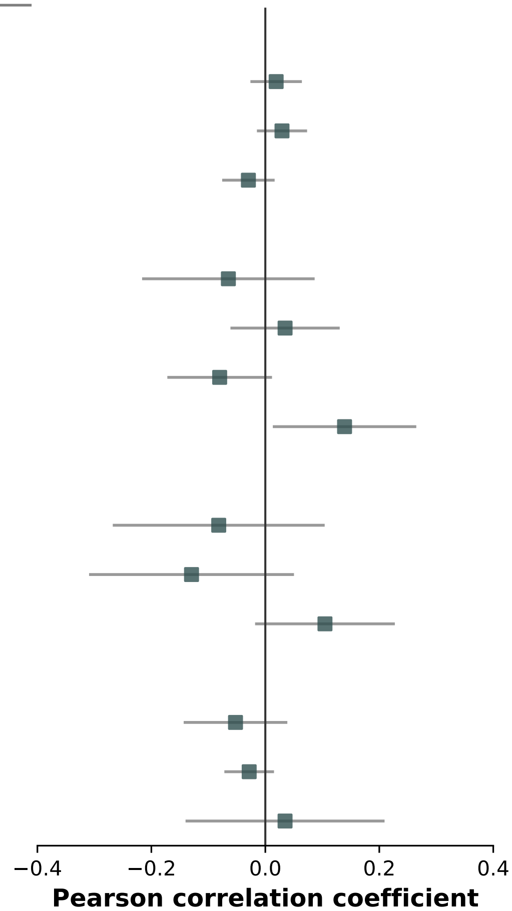


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compares respondents assigned to the IDA condition vs. respondents assigned to the IMC condition. See Table 1 in Study 1: The Effect of Guessing Encouraging Features. Rows are self-reported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

Figure SI 1.4: MTurk Study 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

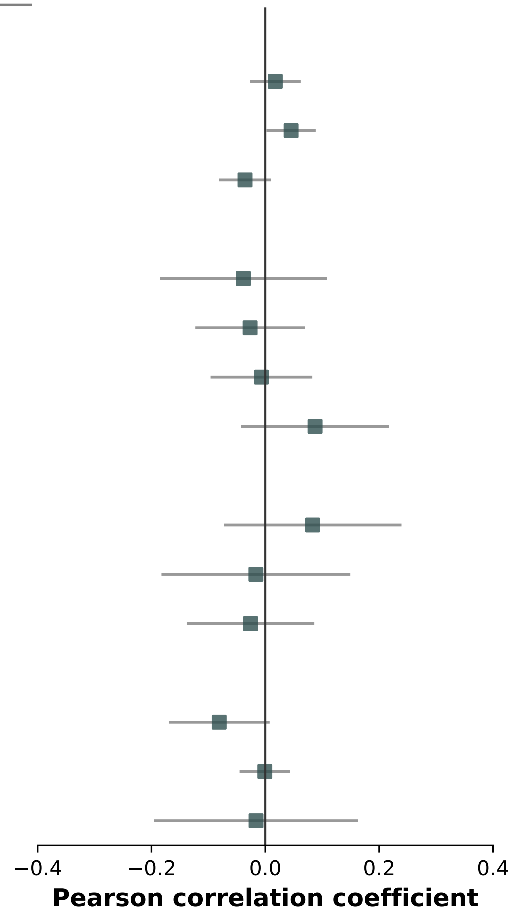
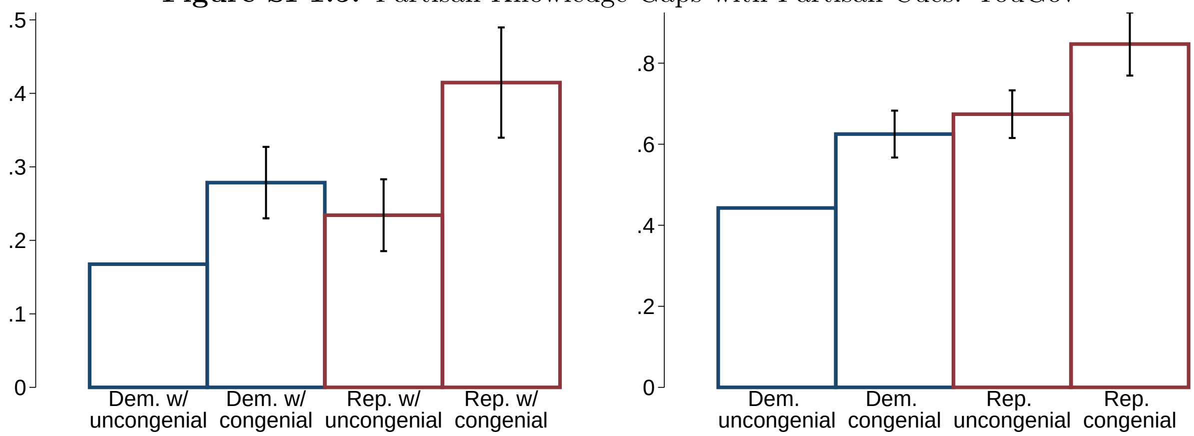


Figure shows the balance tests of respondent characteristics for the Amazon Mechanical Turk Study 1 sample. The tests compares respondents assigned to the IDA condition vs. respondents assigned to the CCD condition. See Table 1 in Study 1: The Effect of Guessing Encouraging Features. Rows are self-reported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

Figure SI 1.5: Partisan Knowledge Gaps with Partisan Cues: YouGov



(a) Unemployment   **(b)** Budget deficit

Figure shows the effect of congenial cues for the YouGov survey by partisanship. Bars indicate the predicted percent of responses saying that unemployment have gone up (correct response) as retrieved from the estimates in Table 3 (columns (2) and (5)). The estimates are obtained by estimating:

$$\text{correct response}_i = \alpha + \beta(\textit{congenial cue})_i + \gamma(\textit{Rep})_i + \delta(\textit{congenial cue} \times \textit{Rep})_i + \varepsilon_i.$$

Capped vertical bars indicate 95% confidence intervals.

SI 1.1 Confidence Scoring for Mturk Study 1

Table SI 1.1: Confidence Scoring vs. Other Survey Conditions (MTurk Study 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)
Confidence scoring (CS)	−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)
Congenial × CS	−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)
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)
R^2	0.127	0.185	0.171	0.064	0.301	0.111	0.190	0.038	0.022	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 Scoring condition, we only consider responses as correct when they are chosen with a full confidence of 10 (on a 0–10 point scale). The 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 Study 2. See Tables SI 1.2 to SI 1.5 for the results comparing the Confidence Scoring condition to each of the four other individual survey conditions. See Figure SI 1.6 for the visualization of how Confidence Scoring 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.

Table SI 1.2: Confidence Scoring vs. IDA (MTurk Study 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)
Confidence scoring (CS)	−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)
Congenial × CS	−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)
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)
R^2	0.169	0.236	0.316	0.082	0.360	0.126	0.435	0.082	0.012	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 Scoring condition, we only consider responses as correct when they are chosen with a full confidence of 10 (on a 0–10 point scale). The 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 Study 2. See Table SI 1.1 for the results comparing the Confidence Scoring condition with all the four other conditions (IDA, CUD, FSR, IMC) pooled together. See Figure SI 1.7 for the visualization of how Confidence Scoring mediates the effect that congenial responses have. See Figure SI 1.7 for the visualization of how Confidence Scoring 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.

Table SI 1.3: Confidence Scoring vs. CUD (MTurk Study 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)
Confidence scoring (CS)	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)
Congenial × CS	−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)
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)
R^2	0.262	0.380	0.308	0.079	0.369	0.187	0.287	0.059	0.076	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 Scoring condition, we only consider responses as correct when they are chosen with a full confidence of 10 (on a 0–10 point scale). The 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 Study 2. See Table SI 1.1 for the results comparing the Confidence Scoring condition with all the four other conditions (IDA, CUD, FSR, IMC) pooled together. See Figure SI 1.8 for the visualization of how Confidence Scoring 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.

Table SI 1.4: Confidence Scoring vs. FSR (MTurk Study 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)
Confidence scoring (CS)	−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)
Congenial × CS	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)
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)
R^2	0.068	0.146	0.117	0.033	0.202	0.081	0.094	0.052	0.020	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

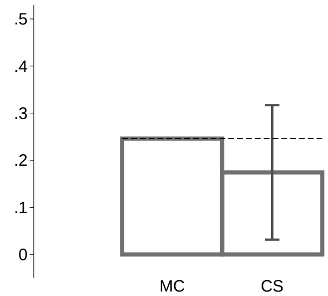
All models are linear probability models where the dependent variable indicates whether the response to a survey item is correct. Under the Confidence Scoring condition, we only consider responses as correct when they are chosen with a full confidence of 10 (on a 0–10 point scale). The the baseline condition is the FSR 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 Study 2. See Table SI 1.1 for the results comparing the Confidence Scoring condition with all the four other conditions (IDA, CUD, FSR, IMC) pooled together. See Figure SI 1.9 for the visualization of how Confidence Scoring 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.

Table SI 1.5: Confidence Scoring vs. IMC (MTurk Study 1)

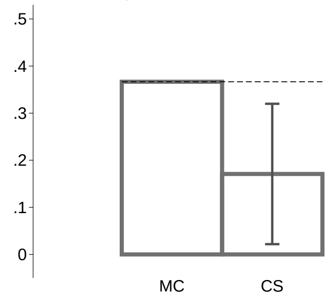
	Obama birthplace	Obama religion	ACA illegal	ACA death panels	GW causes GW causes	GW scientists agree	Voter fraud	MMR vaccine	Budget deficit	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Congenial	0.086	0.164*	0.256**	0.230**	0.512***	0.230**	0.157*	0.095+	0.240***	0.219***
	(0.057)	(0.075)	(0.081)	(0.074)	(0.076)	(0.074)	(0.070)	(0.056)	(0.057)	(0.033)
Confidence scoring (CS)	−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)
Congenial × CS	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)
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)
R^2	0.051	0.084	0.137	0.055	0.314	0.119	0.059	0.046	0.067	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 Scoring condition, we only consider responses as correct when they are chosen with a full confidence of 10 (on a 0–10 point scale). The 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 Study 2. See Table SI 1.1 for the results comparing the Confidence Scoring condition with all the four other conditions (IDA, CUD, FSR, IMC) pooled together. See Figure SI 1.10 for the visualization of how Confidence Scoring 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.

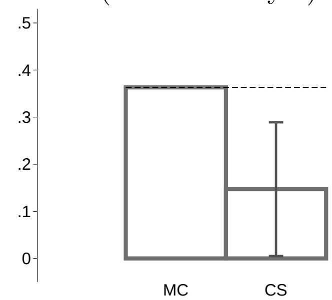
Figure SI 1.6: Confidence Scoring vs. Other Survey Conditions (MTurk Study 1)



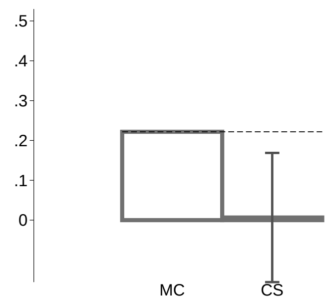
(a) Obama birthplace



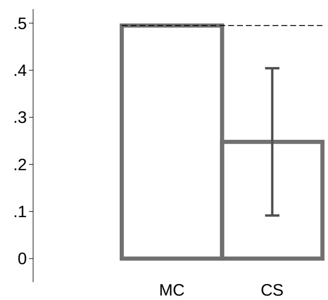
(b) Obama religion



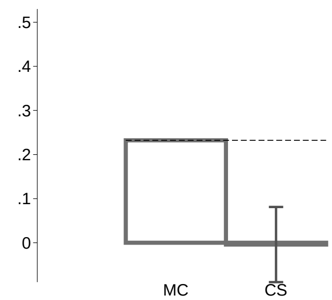
(c) ACA illegal



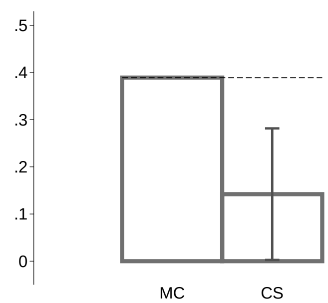
(d) ACA death panels



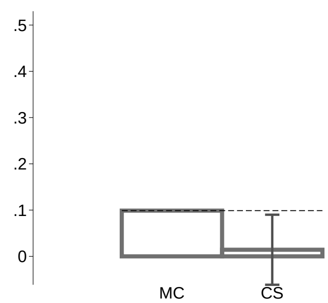
(e) GW causes



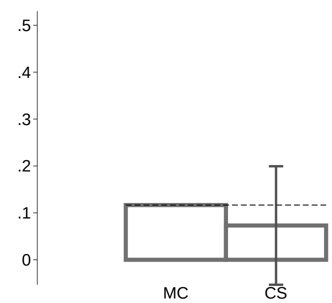
(f) GW scientists agree



(g) Voter fraud



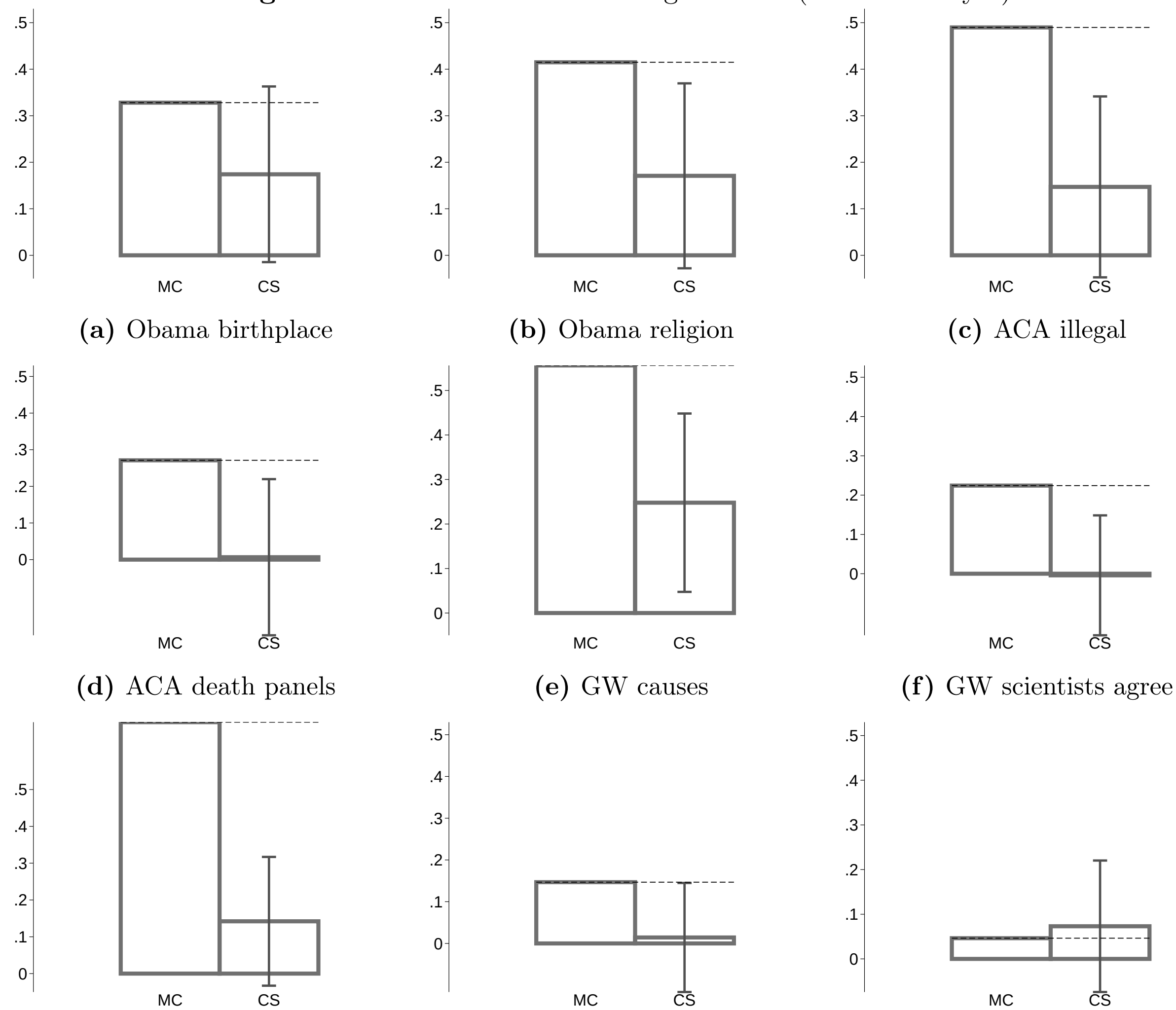
(h) MMR vaccine



(i) Budget deficit

Bars indicate the predicted percent of correct responses when the correct response is congenial to party, depending on whether the survey condition is based on Confidence Scoring (CS) or from Multiple Choice conditions (IDA, CUD, FSR, IMC; see Table 1 for the descriptions). Reconstructed from the estimates from Table SI 1.1. Capped vertical bars indicate 95% confidence intervals.

Figure SI 1.7: Confidence Scoring vs. IDA (MTurk Study 1)



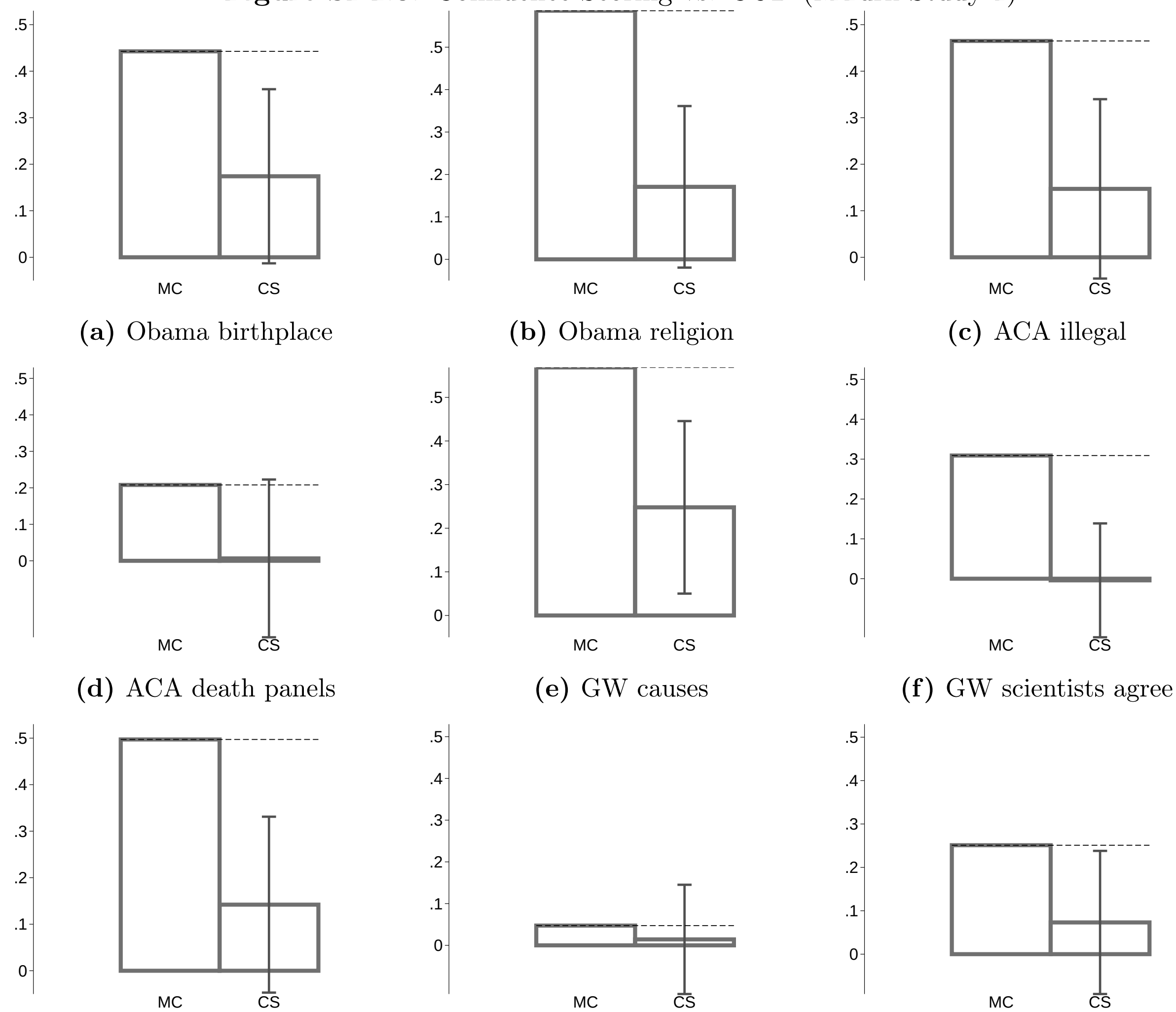
(a) Obama birthplace **(b)** Obama religion **(c)** ACA illegal

(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 party, depending on whether the survey condition is based on Confidence Scoring (CS) or from multiple choice IDA condition (see Table 1 for the descriptions). Reconstructed from the estimates from Table SI 1.2. Capped vertical bars indicate 95% confidence intervals.

Figure SI 1.8: Confidence Scoring vs. CUD (MTurk Study 1)



(a) Obama birthplace

(b) Obama religion

(c) ACA illegal

(d) ACA death panels

(e) GW causes

(f) GW scientists agree

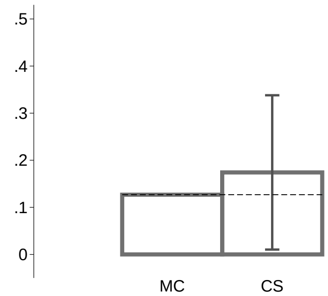
(g) Voter fraud

(h) MMR vaccine

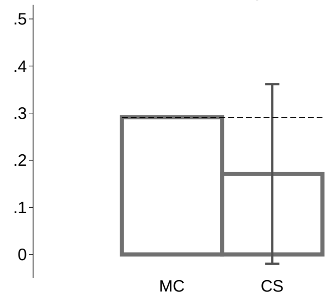
(i) Budget deficit

Bars indicate the predicted percent of correct responses when the correct response is congenial to party, depending on whether the survey condition is based on Confidence Scoring (CS) or from multiple choice CUD condition (see Table 1 for the descriptions). Reconstructed from the estimates from Table SI 1.3. Capped vertical bars indicate 95% confidence intervals.

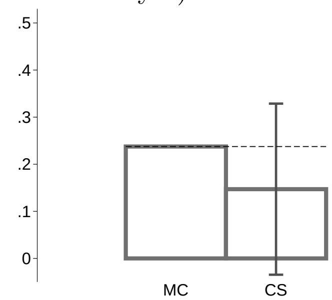
Figure SI 1.9: Confidence Scoring vs. FSR (MTurk Study 1)



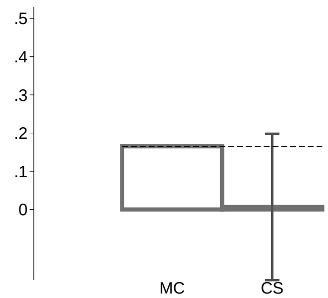
(a) Obama birthplace



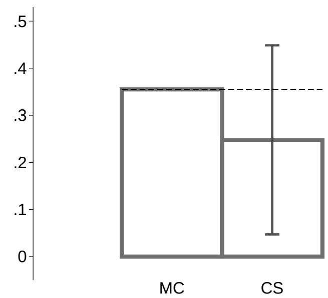
(b) Obama religion



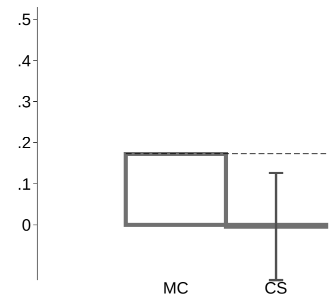
(c) ACA illegal



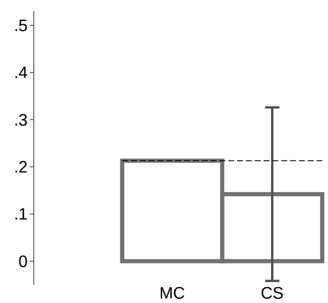
(d) ACA death panels



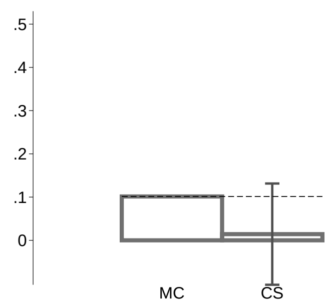
(e) GW causes



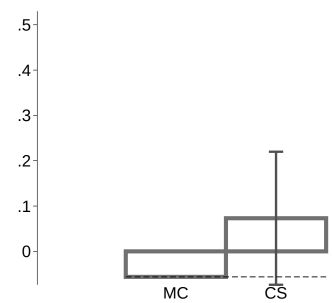
(f) GW scientists agree



(g) Voter fraud



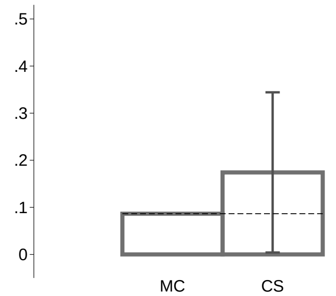
(h) MMR vaccine



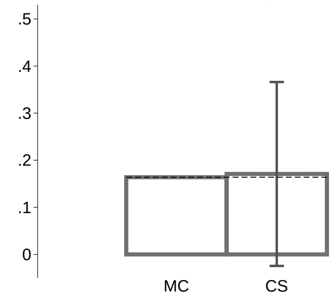
(i) Budget deficit

Bars indicate the predicted percent of correct responses when the correct response is congenial to party, depending on whether the survey condition is based on Confidence Scoring (CS) or from multiple choice CUD condition (see Table 1 for the descriptions). Reconstructed from the estimates from Table SI 1.4. Capped vertical bars indicate 95% confidence intervals.

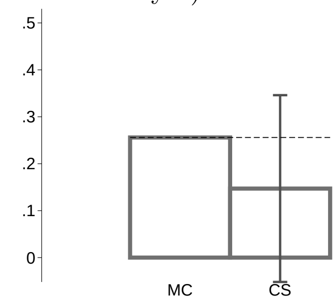
Figure SI 1.10: Confidence Scoring vs. IMC (MTurk Study 1)



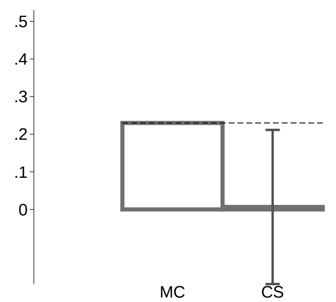
(a) Obama birthplace



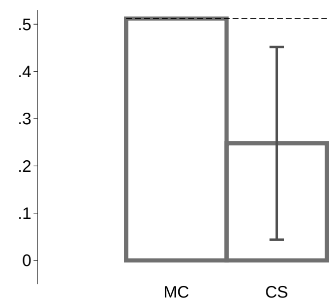
(b) Obama religion



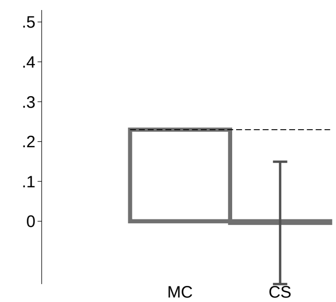
(c) ACA illegal



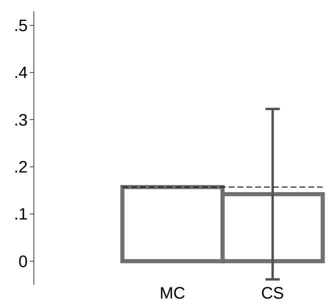
(d) ACA death panels



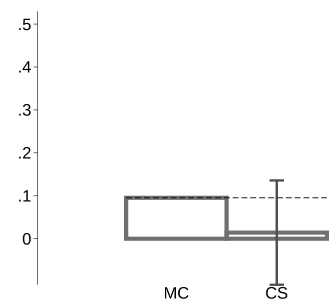
(e) GW causes



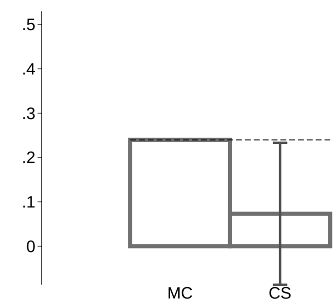
(f) GW scientists agree



(g) Voter fraud



(h) MMR vaccine



(i) Budget deficit

Bars indicate the predicted percent of correct responses when the correct response is congenial to party, depending on whether the survey condition is based on Confidence Scoring (CS) or from multiple choice CUD condition (see Table 1 for the descriptions). Reconstructed from the estimates from Table SI 1.4. Capped vertical bars indicate 95% confidence intervals.

SI 2 Item Text for the MTurk Study

Preface for Different Conditions

RW, IP

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

FSR, 14k, 24k

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 24k

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)

14k

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)

14k

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 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)

  14k
  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 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 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)

  14k
  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 the natural variation over time, such as 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)

14k
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)

14k
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)

14k
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 is 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 is 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)

14k
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)

  14k

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

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

SI 3 Item Text for the Second MTurk Study

The second Amazon MTurk survey was fielded in April 2017 and had 1,059 participants. In this survey, we made use of 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 3.1.

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 3.2.

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 3.3.

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 3.4.

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 3.1: Affordable Care Act 1 Scale Question

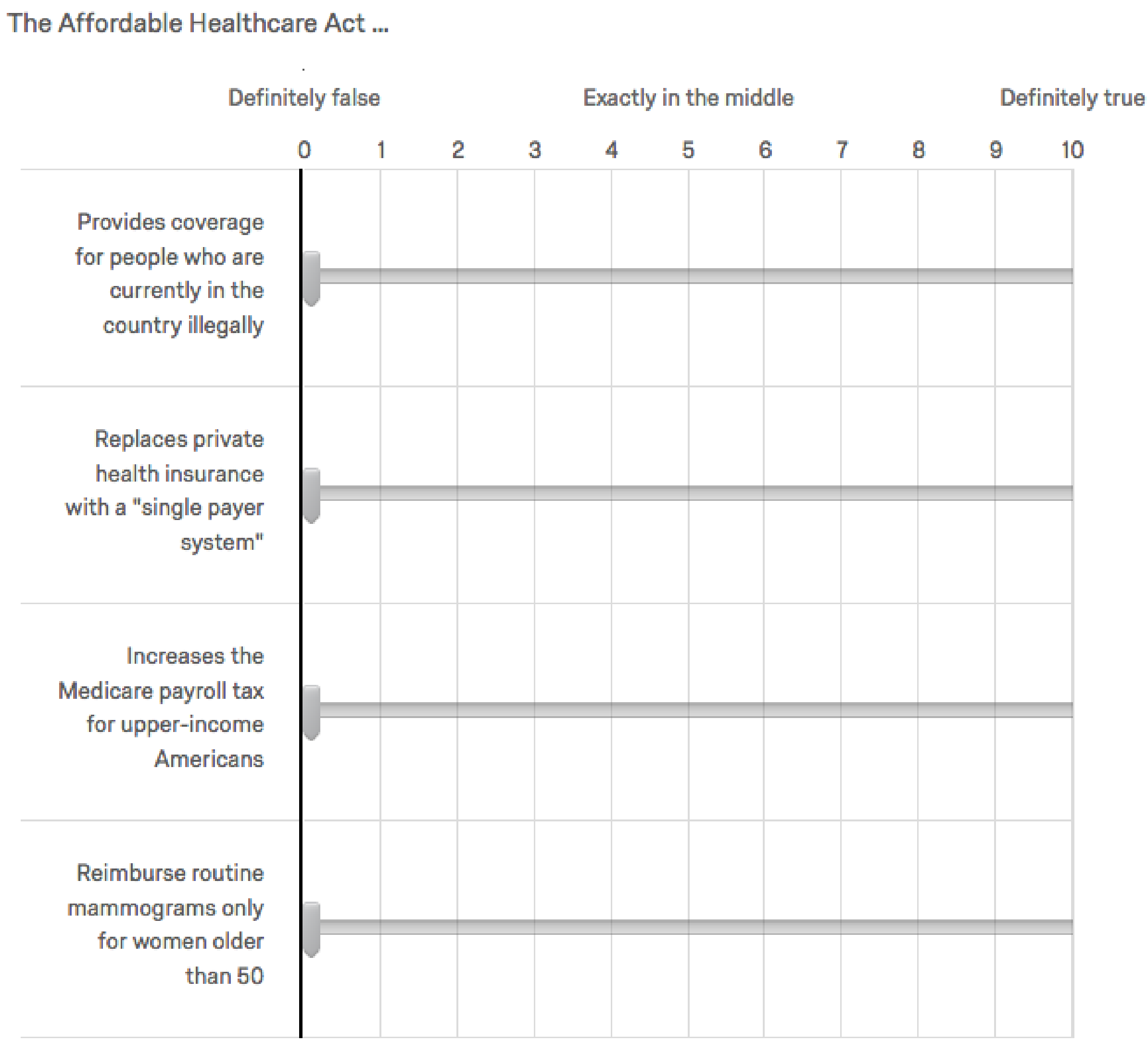


Figure SI 3.2: Greenhouse Gases Scale Question

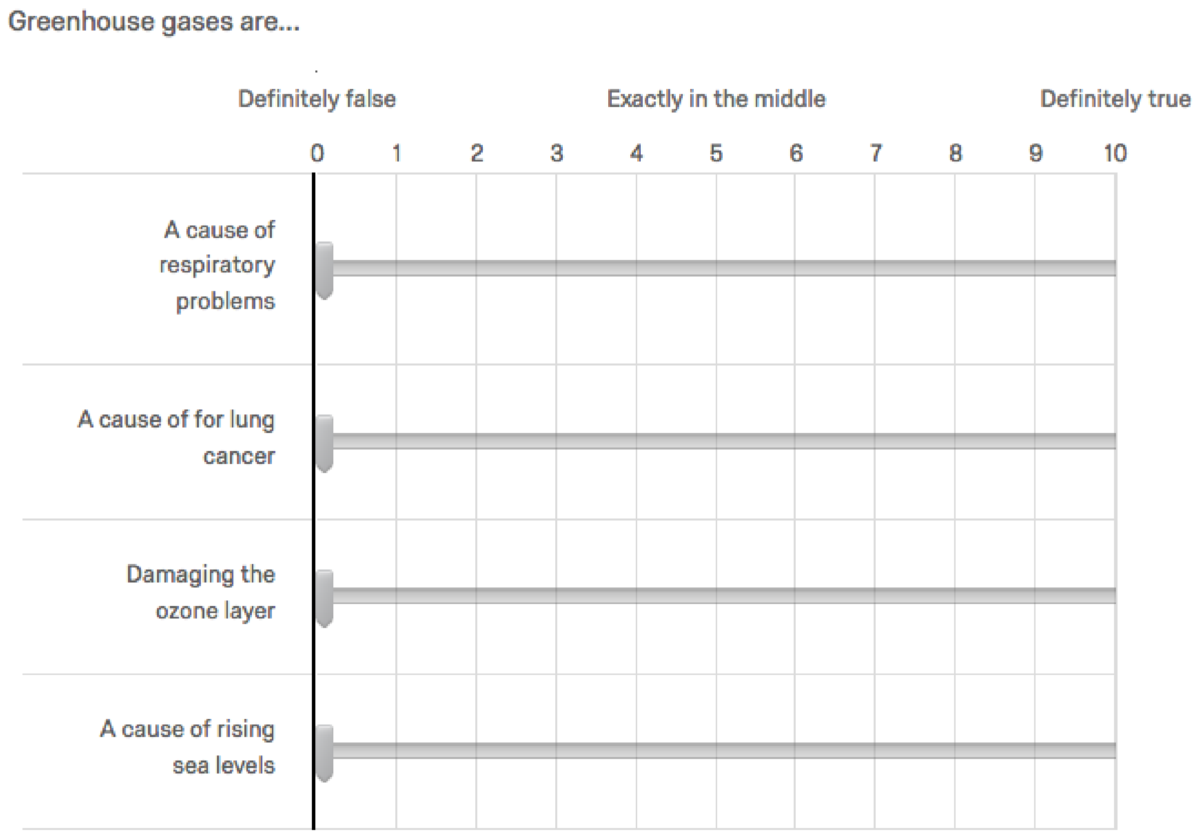


Figure SI 3.3: Affordable Care Act 2 Scale Question

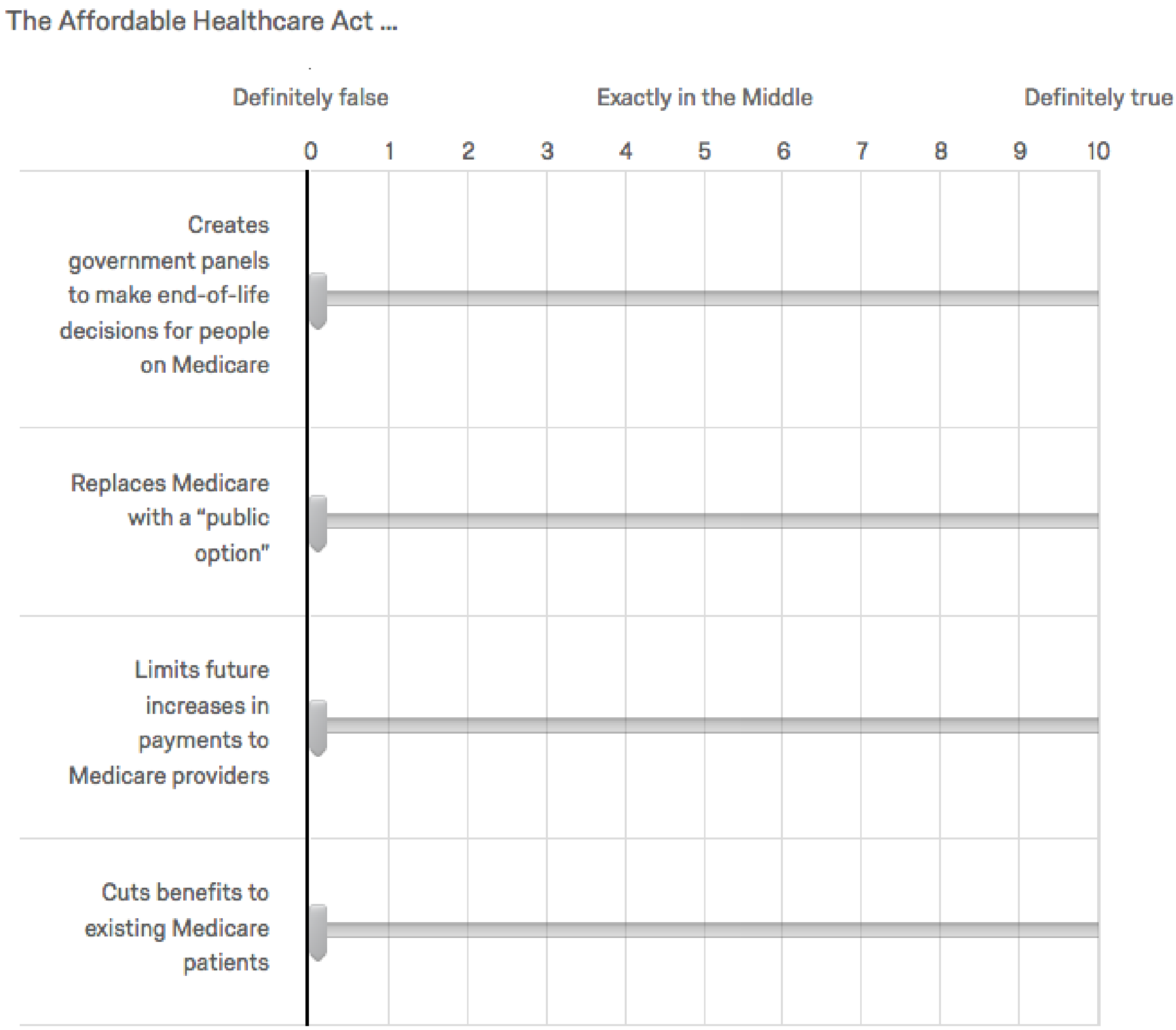
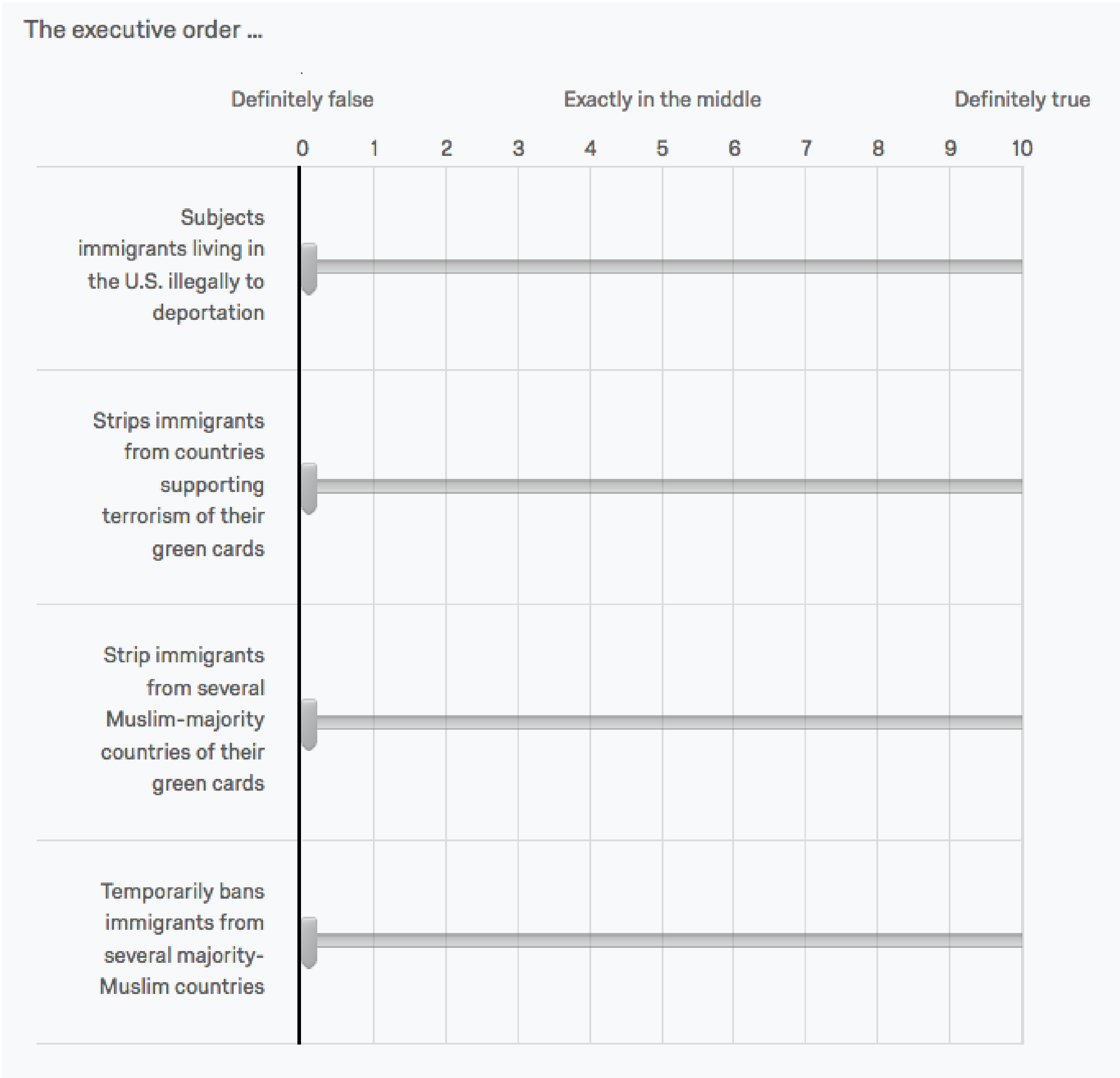


Figure SI 3.4: Executive Order Scale Question



Proportion Correct across Questions

Table SI 3.6 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 relative scoring at the thresholds of 8 and 10. For the relative scoring 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 3.6: 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

Less Stringent Coding Criteria for CCD

Table SI 3.7: 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	3608
Respondent-items	902	902	902	902	902

* Null hypothesis value outside the confidence interval.