Quarterly Journal of Political Science, 2015, 10: 519–578

Partisan Bias in Factual Beliefs about Politics

John G. Bullock[1], Alan S. Gerber[2], Seth J. Hill[3] and Gregory A. Huber[4*]

[1] *Department of Government, University of Texas at Austin, 158 West 21st Street, Stop A1800, Austin, TX 78712, USA; john.bullock@utexas.edu*
[2] *Department of Political Science, Institution for Social and Policy Studies, Yale University, 77 Prospect Street, PO Box 208209, New Haven, CT 06520-8209, USA; alan.gerber@yale.edu*
[3] *Department of Political Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0521, USA; sjhill@ucsd.edu*
[4] *Department of Political Science, Institution for Social and Policy Studies, Yale University, 77 Prospect Street, PO Box 208209, New Haven, CT 06520-8209, gregory.huber@yale.edu*

ABSTRACT

Partisanship seems to affect factual beliefs about politics. For example, Republicans are more likely than Democrats to say that the deficit rose during the Clinton administration; Democrats are more likely to say that inflation rose under Reagan. What remains unclear is whether such patterns reflect differing beliefs among partisans or instead reflect a desire to praise one party or criticize another. To shed light on this question, we present a model of survey response

[*] An earlier version of this paper was circulated as NBER Working Paper 19080 and was presented at Harvard, MIT, NYU, Princeton, Stanford, UCSD, UT-Austin, Yale, and the Annual Meetings of the American Political Science and Midwest

Online Appendix available from:
http://dx.doi.org/10.1561/100.00014074_app
Supplementary Material available from:
http://dx.doi.org/10.1561/100.00014074_supp
MS submitted 14 June 2014; final version received 11 May 2015
ISSN 1554-0626; DOI 10.1561/100.00014074

in the presence of partisan cheerleading and payments for correct and "don't know" responses. We design two experiments based on the model's implications. The experiments show that small payments for correct and "don't know" answers sharply diminish the gap between Democrats and Republicans in responses to "partisan" factual questions. Our conclusion is that the apparent gulf in factual beliefs between members of different parties may be more illusory than real. The experiments also bolster and extend a major finding about political knowledge in America: we show (as others have) that Americans know little about politics, but we also show that they often recognize their own lack of knowledge.

A persistent pattern in American public opinion is the presence of large differences between Democrats and Republicans in statements of factual beliefs. Partisan divisions are expected for questions about political tastes, but they extend even to evaluations of economic trends during a president's tenure (Bartels, 2002, pp. 133–138). What do these differences mean? One view is that Democrats and Republicans see "separate realities" (Kull *et al.*, 2004), with differences arising because of partisanship's effect as a "perceptual screen" in information acquisition and processing (e.g., Campbell *et al.*, 1960; Gerber *et al.*, 2010, esp. ch. 8). By this account, scholars and commentators are correct to take survey respondents' statements at face value (e.g., Bartels, 2002; Jerit and Barabas, 2012; Shapiro and Bloch-Elkon, 2008), because those statements reveal respondents' beliefs. Partisan differences in responses to questions about important facts therefore raise concerns about polarization in the mass electorate. Such differences also threaten defenses of democracy that are based on retrospective voting (Fiorina, 1981): voters may be unable to hold elected officials accountable for

Political Science Associations. We thank seminar participants at those institutions, as well as Kevin Arceneaux, Matias Bargsted, Michael Cobb, David Doherty, Conor Dowling, Mo Fiorina, Matt Levendusky, Neil Malhotra, Markus Prior, Bob Shapiro, Gaurav Sood, and the editors and reviewers, for useful comments and advice.

their performance in office if even their views of economic performance are colored by their partisanship (see also Healy and Malhotra, 2009).

An alternative view is that survey responses are not entirely sincere. Instead, they may reflect the expressive value of making statements that portray one's party in a favorable light (Brennan and Lomasky, 1997; Hamlin and Jennings, 2011; see also Schuessler, 2000). Partisan divergence in surveys may therefore measure the joy of partisan "cheerleading" rather than sincere differences in beliefs about the truth. Furthermore, divergence in expressed survey responses may occur under two different conditions: when partisans are aware that their responses are inaccurate, or when they understand that they simply don't know the truth. In either of these cases, partisan differences in factual assessments would be of less concern than is suggested by prior work, because survey responses would not reveal actual beliefs about factual matters. Despite this possibility, almost no research has attempted to determine the extent to which partisan divergence in responses to factual questions reflects sincere differences in beliefs.

This paper reports results from two novel experiments designed to distinguish sincere from expressive partisan differences in responses to factual survey questions. We motivate our experiments with a model in which respondents value both partisan responding and incentives for correct and "don't know" responses. The model shows that incentives can reduce partisan divergence when expressive responding would otherwise mask shared (i.e., bipartisan) beliefs about factual matters. In both experiments, all subjects were asked factual questions, but some were given financial incentives to answer correctly. We find that even small incentives reduce partisan divergence substantially — on average, by about 55% and 60% across the questions for which partisan gaps appear when subjects are not incentivized.

Our model also reveals that incentives for correct responses may not deter cheerleading among those who recognize that they don't know the correct response. Even when paid to answer correctly, those who are unsure expect to earn less for offering the response that they think is most likely to be correct (relative to those who are sure of the correct response), and so they are more likely to continue offering an expressive partisan response. In our second experiment, we therefore offer to pay some participants both for correct responses and a smaller amount for admitting that they do not know the correct response. We

find that large proportions of respondents choose "don't know" under these conditions. Furthermore, partisan gaps are even smaller in this condition — about 80% smaller than for unincentivized responses. This finding shows that partisan divergence in responses to these questions is driven by expressive behavior *and* by respondents understanding that they do not actually know the correct answers. To the best of our knowledge, this is the first analysis which demonstrates that people are aware of their own ignorance of political facts.[1]

These results speak to questions about the meaning of public opinion and the mechanisms through which partisanship affects important outcomes. Most importantly, they call into question the claim that partisan divergence in expressed beliefs about factual matters is cause for concern about voters' abilities to judge incumbent performance. To the extent that factual beliefs are determined by partisanship, paying partisans to answer correctly should not affect their responses to factual questions. But it does. We find that even modest payments substantially reduce the observed gaps between Democrats and Republicans, which suggests that Democrats and Republicans do not hold starkly different beliefs about many important facts. It also suggests that, when using survey data to understand *why* people make the political choices that they do, analysts should be cautious in interpreting correlations between factual assessments and those choices. Survey responses to factual questions may not accurately reflect beliefs, and the correlation between vote choice and factual assessments (of candidates or political conditions) observed in surveys may be in part artifactual.[2] Thus, even if partisanship is a crucial influence on votes and other political outcomes (Gerber *et al.*, 2010), it may operate more through its effects on tastes than through its effects on perceptions of reality.

These results also affect our interpretation of partisan polarization in the mass electorate. Republicans and Democrats do hold different factual beliefs, but their differences are likely not as large as naïve analysis of survey data suggests. Just as people enjoy rooting for their

[1]In this regard, the most closely related work is Bishop *et al.* (1984) and Luskin and Bullock (2011).

[2]Our results confirm concerns in the literature on economic voting (e.g., Ansolabehere *et al.*, 2013) that survey reports of economic conditions may be contaminated by expressive partisan responding.

favorite sports teams and arguing that their teams' players are superior, even when they are not, surveys give citizens an opportunity to cheer for their partisan teams (Green *et al.*, 2002). Deep down, however, many individuals understand the true merits of different teams and players — or, at minimum, they understand that they don't know enough to support their expressive responding as correct. And while our experimental approach cannot be used to discern whether partisan divergence in attitudes is sincere, an implication of our work is that if respondents misstate their factual beliefs in surveys because of their partisan leanings, they may misstate their attitudes in surveys for the same reason. We return to this point in the discussion section.

Our work is also of significance for survey methodology. In particular, how should one interpret experiments which show that partisan cues increase partisan divisions in survey response? Such results are commonly taken to show that partisanship affects attitudes (e.g., Cohen, 2003). Our results raise the possibility, however, that partisan cues merely remind participants about the expressive utility that they gain from offering partisan-friendly survey responses. One implication is that studies in which partisan cues bring about partisan variation in survey response may not be showing that partisanship alters actual attitudes or beliefs. A key task for researchers is thus to understand when survey responses reflect real attitudes and when they reflect more expressive tendencies.

On the whole, then, understanding whether polarization in survey responses is real or artificial speaks to core concerns of subfields throughout political science, for scholars who rely on survey methods to study attitudes and behavior and for those who are interested in polarization of mass attitudes. Finally, it speaks to political psychologists because it bears directly on the connection between partisan identity and perceptions of political reality.

1 Theory and Prior Evidence

Prior research documents partisan differences in expressed factual beliefs (e.g., Gaines *et al.*, 2007; Jacobson, 2006; Jerit and Barabas, 2012), and some of it focuses on differences in evaluations of retrospective economic

conditions (e.g., Bartels, 2002; Conover *et al.*, 1986, 1987, pp. 133–38).[3] Many of these differences arise because members of one party issue economic assessments that deviate starkly from objective conditions. For example, despite the large improvement in unemployment and inflation during Reagan's presidency, Bartels (2002) shows that, in 1988, Democrats were especially likely to report that unemployment and inflation had increased since 1980. This pattern was reversed in 2000, when Republicans were more likely to offer negative retrospective evaluations at the conclusion of the presidency of Democrat Clinton.[4]

How should we interpret these partisan gaps? Bartels presents one common view when he argues that partisans likely believe their divergent assessments: "Absent some complicated just-so story involving stark differences in the meaning of 'unemployment' and inflation. . . these large differences can only be interpreted as evidence of partisan biases in perceptions" (Bartels, 2002, pp. 136–137). An alternative view is that differences in survey responses are the result of a combination of motivations. Individuals may offer responses that are consistent with their partisanship not solely because they believe those responses, but also because doing so gives them the opportunity to support their "team" (e.g., Gerber *et al.*, 2010; Green *et al.*, 2002).

Many social scientists have wrestled with the problem of insincere survey responses (e.g., Berinsky, 2005). But they typically focus on responses to sensitive topics like race rather than on problems that may be caused by "expressive benefits" in survey response.[5] And the methods used to overcome problems associated with responses to sensitive topics — for example, "list experiments" (Kuklinski *et al.*, 1997) — may not apply to the problem of eliciting sincere responses when people derive expressive benefits from answering insincerely.

[3]A related but distinct literature concerns partisan differences in responses to nonfactual questions (see Berinsky, 2015).

[4]Additional work examines conditions that can exacerbate apparent partisan gaps. Asking political questions prior to economic ones increases the correlation between partisanship and subjective economic evaluations (Lau *et al.*, 1990; Palmer and Duch, 2001; Sears and Lau, 1983; Wilcox and Wlezien, 1993), and partisan gaps are larger when elections are more salient (Lavine *et al.*, 2012, Chapter 5; see also Stroud, 2008). As we note earlier, what is unclear is how to interpret these patterns. Do circumstances that make partisanship more salient call relevant information to mind, or do they simply increase the expressive value of partisan responses?

[5]An exception to this characterization is the literature on economic voting discussed earlier.

Instead, scholars have long used incentives to elicit honest or rational responses. In a review of relevant experiments, Morton and Williams (2010, pp. 358–361) argue that incentives often reduce the size and frequency of decision-making errors. But almost all of the studies that they review are apolitical and do not involve tests of factual knowledge. Prior and Lupia (2008) do study the effect of financial incentives on responses to factual questions about politics, and they find that the effects are real but weak.[6] However, they do not examine the effects of incentives on partisan patterns in responding.

To date, only Prior (2007) and Prior *et al.* (2015) have examined the effects of incentives on partisan response patterns to factual questions about politics. Prior (2007) asked subjects 14 questions about politics; some were assigned at random to receive $1 for each correct answer. The results were mixed, but they suggest that $1 incentives can reduce party differences in responses to such questions.[7] Prior *et al.* (2015) present two experiments in which they urge people to answer correctly or provide relatively large financial incentives ($1 or $2 for each correct response). Both treatments reduce errors in answers to questions about the performance of the U.S. economy during the George W. Bush administration. Across the two experiments, financial incentives appear to reduce the rate of error by about 40%; simply urging people to answer correctly may be still more effective. An important unanswered question from that work, however, is how respondents who do not know the correct answers behave in the presence and absence of incentives for correct responses. It may be, for example, that partisan responses are insincere, but that respondents continue to offer them when given incentives because they do not know which other answer might be correct. If respondents could express their lack of knowledge about the truth, would partisan gaps be even smaller?

To address these questions, we present a model of survey response which incorporates the possibility that individuals (a) receive utility

[6] All subjects in the Prior and Lupia (2008) study were asked 14 factual questions about politics. Subjects in a control condition averaged 4.5 correct answers, while those who were paid $1 for each correct answer averaged 5.0 correct answers (Prior and Lupia, 2008, p. 175).

[7] In Prior (2007), incentives reduced partisan gaps in responses to four items. Results on a fifth item were mixed. Results were null for two other items. There was no partisan gap in the control group for three further items, and results for the remaining four items were not reported.

from offering partisan-tinged responses and (b) differ in their underlying knowledge of the truth. We use this model to understand the effect of incentives on a respondent's tendency to answer questions in a manner that reflects either her partisan affinity or her beliefs about the truth. We also show that our model can be used to understand the extent to which partisan differences arise because people are uncertain about the truth.

2 A Theory of Expressive Survey Response

To explore the role that insincere "cheerleading" plays in the partisan polarization of survey responses, and to motivate our experimental design, we present in the Appendix a formal model of responses to factual questions in the presence and absence of financial incentives. As in our experiments, incentives take two forms: respondents may be paid for offering the correct answer or for admitting that they don't know the correct answer. We present here a summary of results from the model.

The first results show that incentives for correct responses reduce partisan divergence under three conditions: (1) participants would give inaccurate, partisan-tinged responses in the absence of incentives; (2) the value of the incentive is greater than the value of partisan cheerleading; and (3) the same strong beliefs about the correct answer are held by members of both parties.[8] The intuition for this result is straightforward: giving a response that one does not believe but that portrays one's party in a favorable light is more costly when it entails giving up the chance to earn a reward for answering correctly. Therefore, under the three conditions listed earlier, a researcher can reduce partisan divergence and elicit responses more informative of people's true beliefs by offering incentives to answer correctly.

[8]If members of different parties have different underlying beliefs about the truth, there is no strong reason to expect that responses in the presence of incentives will be less divergent than in the absence of those incentives. Additionally, it may be that only members of one party change their responses in the presence of incentives, in which case divergence will be reduced only if members of that party move in the direction of the other party's responses.

The third condition — if incentives for correct responses are to reduce partisan divergence, members of different parties must share the same belief about the truth — requires elaboration. This condition is an implication of our model, not an assumption that underpins it. There are surely cases in which members of different parties hold different beliefs about the truth. In these cases, paying them to answer truthfully will not cause their survey responses to converge. On the other hand, to the extent that payments for correct answers do cause partisans' survey responses to converge, we can infer that partisans' beliefs about the correct answers are more similar than they seem to be under normal survey conditions.

An alternative interpretation of partisan convergence when people are paid for correct answers does not imply that they "know" the correct answers with much confidence. Instead, it suggests that partisan differences arise because of "congenial inference": when trying to answer a question under ordinary conditions, partisans are especially likely to call to mind those considerations that put their own party in a favorable light, and they infer the correct answer to the question at hand from this congenial set of considerations (e.g., Zaller, 1992, Chapter 5). But payment for correct answers heightens the desire to provide a correct answer. In turn, respondents who are paid for correct answers undertake a more even-handed (and perhaps more effortful) search of their memory for relevant considerations. They make more accurate inferences on the basis of this different set of considerations — even though they were not at all sure of the correct answer before the question was asked. In this paper, we are agnostic about which mechanism better explains the effects of payment for correct answers, but both mechanisms reveal that conventional survey responses do not fully characterize individuals' beliefs about political facts.

In addition to identifying the conditions under which incentives promote partisan convergence, our model highlights a little-appreciated explanation for divergent factual responses: even when partisans are paid for correct responses, their answers may diverge because they are unsure of the correct response and therefore default to an expressive response. To see how uncertainty can increase partisan divergence, note that the expected value of an uncertain respondent's best guess is discounted by her uncertainty. If she is sufficiently uncertain, the expected value of her best guess may be smaller than the expected

value of partisan cheerleading. At the extreme, if there are two answers to a question and she is completely uncertain about which response is correct, in expectation she earns the incentive for a correct response half the time for offering either response, and she therefore has no reason to deviate from her preferred partisan response. This will be true even if the incentives are very large.

In light of this ambiguity, we extend the model by incorporating incentives for admitting one's lack of knowledge. When respondents are paid for both correct and "don't know" answers, our analysis shows that the proportion of respondents choosing "don't know" is increasing in the proportion who (1) place low value on partisan cheerleading relative to the incentive for choosing "don't know," *and* (2) are so unsure of the correct answer that they are better off choosing "don't know" than any other option. This is so because one can earn the incentive for a "don't know" response with certainty (by choosing "don't know"), whereas the incentive for a correct response is earned only if the respondent chooses the response that is correct. Overall, incentives for "don't know" responses allow us to understand the proportion of partisan divergence that arises because respondents default to expressive responding when they are unsure of the correct answer.

Our model implies that an experiment in which subjects receive incentives for correct and "don't know" responses to factual questions can identify the presence of partisan cheerleading. We now describe two experiments that meet these conditions.

3 Experiment 1: Effects of Incentives for Correct Responses on Partisan Divergence

Our first experiment was fielded on the Cooperative Congressional Election Study in October 2008. CCES subjects are part of a nationally representative opt-in sample. In our experiment, 626 participants were randomly assigned to the control group ($N = 312$) or the treatment group ($N = 314$). We restrict our analysis to the 419 participants who identified as either Democrats or Republicans.[9]

[9]In our analysis, Democrats are those who responded "Democrat" to the first question in the standard two-question measure of party identification. Republicans are those who responded "Republican." We discuss the behavior of partisan "leaners" later, and we present question wording for both the experiments, along with further

We told control-group subjects that they would be asked questions about politics, that they would have 20 seconds to answer each question, and that their scores would not be shared with anyone. Treated subjects received the same instructions and were told that answering correctly would increase their chance of winning a prize:

> For each question that you answer correctly, your name will be entered in a drawing for a $200 Amazon.com gift certificate. For example, if you answer 10 questions correctly you will be entered 10 times. The average chance of winning is about 1 in 100, but if you answer many questions correctly, your chance of winning will be much higher.

After receiving their instructions, all subjects were asked the 12 factual questions shown in Table 1.[10] The first 10 items had closed (i.e., multiple-choice) response options and were similar to questions for which other research has found partisan differences. No "don't know" option was offered. Each question referred to a potentially salient partisan issue. The last two "placebo" questions were open-ended and required participants to enter numerical responses. We fielded the placebo questions, which were about obscure historical facts, to ascertain whether participants were using their allotted 20 seconds to look up answers using outside references. Using these questions, we find little evidence that participants "cheated": rates of correct responding were below 3% and statistically indistinguishable between the control and payment conditions.

This experiment allows us to understand whether some partisan divergence in responses to factual questions arises because of the expressive benefit of providing partisan responses. Specifically, we can learn about the role of expressive benefits by comparing partisan divergence in the treatment and control conditions. If divergence is lower in the treatment group, it suggests that, for some respondents, our incentives are of greater value than partisan cheerleading. Given the modest size of the incentives offered, we view the estimates that we obtain

information about the construction of the sample, in the Online Appendix.

[10]We note that in the presence of ambiguity about which responses is correct, incentives should have weaker effects. For our purposes, what matters is not which answer is correct, but simply that partisans of different stripes have common beliefs about which answer is most likely to be correct.

Table 1: Experiment 1: question wording and baseline partisan differences in scale scores.

Question	Question wording	Response options	Control group, mean Democratic response	Control group, mean Republican response	Control group difference in scale scores, Democrats − Republicans	p-value of difference of party means, one-tailed test	N
Iraq casualties, 2007 versus 2008	Was the number of U.S. soldiers killed in Iraq in the first half of 2008 lower, about the same, or higher than the number who were killed in the second half of 2007?	*Lower (0),* About the same (0.5), Higher (1)	0.416	0.177	0.239	0.000	212
Bush inflation change	Compared to January 2001, when President Bush first took office, has the level of inflation in the country increased, stayed the same, or decreased?	*Increased (1),* Stayed about the same (0.5), Decreased (0)	0.894	0.694	0.201	0.000	207

(*Continued*)

Table 1: (*Continued*)

Question	Question wording	Response options	Control group, mean Democratic response	Control group, mean Republican response	Control group difference in scale scores, Democrats − Republicans	*p*-value of difference of party means, one-tailed test	*N*
Bush unemployment change	Compared to January 2001, when President Bush first took office, has the level of unemployment in the country increased, stayed the same, or decreased?	*Increased (1),* Stayed about the same (0.5), Decreased (0)	0.766	0.598	0.168	0.002	208
Estimated Bush approval	About what percentage of *Americans* approve of the way that George W. Bush is handling his job as President?	*20% (1),* 30% (0.75), 40% (0.5), 50% (0.25), 60% (0)	0.909	0.817	0.092	0.000	216

(*Continued*)

Table 1: (*Continued*)

Question	Question wording	Response options	Control group, mean Democratic response	Control group, mean Republican response	Control group difference in scale scores, Democrats − Republicans	*p*-value of difference of party means, one-tailed test	*N*
Iraq total casualties	About how many U.S. soldiers have been killed in Iraq since the invasion in March 2003?	*4,000 (0),* 8,000 (0.25), 12,000 (0.5), 16,000 (0.75), 20,000 (1)	0.200	0.114	0.087	0.013	210
Estimated Bush approval among Republicans	About what percentage of *Republicans* approve of the way that George W. Bush is handling his job as President?	40% (1), 50% (0.75), *60% (0.5),* 70% (0.25), 80% (0)	0.794	0.724	0.070	0.039	211
Obama age	How old is Barack Obama?	37 (0), 42 (0.33), *47 (0.66),* 52 (1)	0.558	0.508	0.050	0.055	213
McCain age	How old is John McCain?	62 (0), 67 (0.33), *72 (0.66),* 77 (1)	0.681	0.637	0.044	0.035	215

(*Continued*)

Table 1: (*Continued*)

Question	Question wording	Response options	Control group, mean Democratic response	Control group, mean Republican response	Control group difference in scale scores, Democrats − Republicans	*p*-value of difference of party means, one-tailed test	*N*
Afghanistan casualties, 2007 versus 2008	Was the number of U.S. soldiers killed in Afghanistan in the first half of 2008 lower, about the same, or higher than the number who were killed in the second half of 2007?	Lower (0), *About the same* (0.5), Higher (1)	0.608	0.598	0.010	0.430	208
Bush deficit change	Compared to January 2001, when President Bush first took office, has the federal budget deficit in the country increased, stayed the same, or decreased?	*Increased (1),* Stayed about the same (0.5), Decreased (0)	0.938	0.944	−0.006	0.589	212

(*Continued*)

Table 1: (*Continued*)

Question	Question wording	Response options	Control group, mean Democratic response	Control group, mean Republican response	Control group difference in scale scores, Democrats − Republicans	p-value of difference of party means, one-tailed test	N
Placebo: price of gold in 1980	What was the price of gold, in dollars per ounce, on January 18, 1980?	In dollars, 0=0, 1,000=1, Correct is *between $800 and $900*	0.791	0.680		N/A	
Placebo: Bangladeshi independence	In what year did Bangladesh become independent of Pakistan?	In years, 1800=0, 2000=1, Correct is *1971*	0.151	0.185			

Note: Questions are ordered by size of partisan gap in control-group responses, with placebo questions at the bottom. All responses are scaled from 0 to 1; 1 is the most Democratic response. In the "response options" column, the correct response options are italicized. Placebo questions were open-ended and were recoded to range from 0 to 1.

Source: 2008 CCES.

from treatment–control comparisons as lower bounds on the extent of expressive partisan responding in this experiment.

To measure partisan divergence, we create scale scores by coding responses to each question to range linearly from 0 to 1. These scores are the dependent variables in our analyses. The most Republican response to each question (either the largest or smallest response) is coded as 0; the most Democratic response is coded as 1. For example, when we ask about the change in unemployment under President Bush, the response "decreased" is coded as 0 because it portrays a Republican president most positively, "stayed about the same" is coded as 0.5, and "increased" is coded as 1 because it portrays the president most negatively. If partisans are answering in a manner consistent with their partisanship, Democrats should offer "larger" responses than Republicans.

Table 1 shows the average partisan difference in scale score, by question, for those in the control group. The questions in Table 1 are ordered by the size of these control-group partisan gaps. For 9 of the 10 questions, the gaps are consistent with our expectations about patterns of partisan responding.[11] Eight of the differences are significant at $p < 0.10$ (one-tailed). The gaps for these eight items vary substantially in size, with the largest gaps appearing for questions about casualties in Iraq and Bush's economic performance. Because our theory of expressive responding is about the effects of incentives on partisan differences, we focus on these eight items, that is, the items to which partisanship makes a difference under ordinary survey conditions. (In the online Appendix, we analyze our data while including responses to all questions, including those for which we do not find partisan gaps.)

What effect do incentives for correct responses have on observed partisan divergence? To measure the effects, we estimate a model in which we predict scale score R for individual i and question j:

$$R_{ij} = b_0 + b_1\text{Democrat}_i + b_2\text{PayCorrect}_i + b_3(\text{Democrat}_i \times \text{PayCorrect}_i) + \text{Question}_j + e_i,$$

where *Democrat* equals 1 for Democratic participants and 0 for Republicans, *PayCorrect* equals 1 for those assigned to the incentive

[11]The exception is the question about the change in the deficit under George W. Bush. For both Democrats and Republicans, 92% of respondents correctly reported the deficit had increased.

condition, and *Question* is a vector of question-specific fixed effects. The coefficient b_1 is therefore the average party difference in scale scores in the control condition, while $b_1 + b_3$ is the average party difference in the incentive condition. Prior research suggests $b_1 > 0$, while our theoretical model predicts that b_3 will be negative if partisans offer partisan-tinged responses in the absence of incentives, share common and sufficiently strong beliefs about the truth, and give less weight to partisan responding than to the expected value of the incentive.

OLS estimates, with standard errors clustered at the respondent level, appear in Table 2. Pooling across the eight questions for which we observe statistically significant partisan gaps in the control condition, column (1) provides estimates of the average effect of incentives on responses. The 0.118 ($p < 0.001$) coefficient for *Democrat* (b_1) is the average gap between Democrats and Republicans in the control condition. The -0.065 ($p < 0.001$) coefficient for *Democrat* $\times$ *PayCorrect* (b_3) means that this gap is reduced to 0.053 ($0.118 - 0.065$), or by 55%, when incentives are offered. In column (2), we add demographic controls; the results are nearly unchanged.[12]

In Table A.1 of the Appendix, we repeat the analysis for each question individually. The estimate for b_3 is negative in all eight cases. While most of these individual-question estimates are not statistically significant — perhaps because the impact of sampling variability is heightened when we examine individual questions — the estimates are large, accounting for between 13% and 100% of the partisan gap between Democrats and Republicans. These estimates are especially noteworthy for the questions about the most salient issues in the 2008 campaign: the Iraq War and Bush's performance on unemployment. On these matters, incentives reduced partisan gaps by between 33% and 73%. Importantly, these questions about war and unemployment were not salient only in 2008: they speak to the issues that political scientists often use when they link objective conditions to election outcomes (e.g., Hibbs, 2000).

[12]We have also repeated our analysis excluding the Bush approval item, which is the item for which we find our largest estimate of b_3. In this case, we continue to find a negative and statistically significant coefficient for b_3 in the pooled analysis ($-0.06, p < 0.01$). Our analysis excludes cases in which participants didn't provide a response, which occurs 3% of the time in both treatment and control conditions. Replacing nonresponses to each question with party averages for each question produces substantively similar results.

Table 2: Experiment 1: effect of payment for correct responses on partisan differences in scale scores.

	(1)	(2)	(3)
Democrat (b_1)	0.118	0.105	0.082
	[0.015]***	[0.016]***	[0.022]***
Political interest × Democrat			0.059
			[0.030]**
Payment for correct response × Democrat (b_3)	−0.065	−0.059	−0.057
	[0.022]***	[0.022]***	[0.037]
Payment for correct response × Political interest × Democrat			−0.023
			[0.046]
Payment for correct response	0.038	0.031	0.045
	[0.016]**	[0.016]*	[0.029]
Payment for correct response × Political interest			−0.005
			[0.035]
Knowledge (0–1)		0.013	
		[0.015]	
White		0.017	
		[0.024]	
Hispanic		0.040	
		[0.028]	
Other race		0.051	
		[0.030]*	
Female		0.016	
		[0.012]	
Age (in years)		0.001	
		[0.002]	
Age^2/100		−0.001	
		[0.002]	
Region: Northeast		0.043	
		[0.017]***	
Region: Midwest		0.042	
		[0.016]***	

(*Continued*)

Table 2: (*Continued*)

	(1)	(2)	(3)
Region: South		0.014	
		[0.014]	
Income (1 = < $10,000; 14 = > $150,000; 15 = RF/Missing)		0.005	
		[0.002]**	
Income missing		−0.046	
		[0.024]*	
Education (1 = no high school; 6 = graduate degree)		0.000	
		[0.006]	
Education: No high school		0.006	
		[0.024]	
Education: Some college		0.019	
		[0.014]	
Education: 2-year college		0.032	
		[0.026]	
Education: 4-year college		−0.003	
		[0.019]	
Married or in a domestic partnership		−0.007	
		[0.013]	
Religious attendance (1–6)		−0.002	
		[0.004]	
Political interest (0,1)			−0.034
			[0.021]
Constant	0.239	0.160	0.261
	[0.021]***	[0.059]***	[0.024]***
Observations	3321	3299	3305
R^2	0.398	0.407	0.400

Note: The dependent variable is the mean scale score for the eight questions on which we observed control-group partisan gaps of $p < 0.10$. It ranges from 0 to 1. The analysis includes only Democrats and Republicans from the control and pay-for-correct-response conditions. Cell entries are OLS coefficients with robust standard errors, clustered by respondent. Question fixed effects not reported. *significant at 10%; **significant at 5%; ***significant at 1% (two-tailed tests).

Source: 2008 CCES.

These results show that even modest incentives can substantially reduce partisan divergence in factual assessments. For example, in this experiment, participants are told that answering correctly will improve their chances of earning a $200 gift certificate, and that the baseline chance of winning was around 1 out of 100. If they estimate that answering all questions correctly would double their chances of winning this prize, the expected value of answering any given question correctly is approximately 17 cents.[13] In turn, the finding that incentives reduced partisan gaps by more than 50% means that more than half of the party gap may be generated by participants for whom partisan responding to any given question is worth less than 17 cents.

Of course, the effects of incentives are unlikely to be equal across all of the people in our data set. We focus on two characteristics across which variation might be expected: political interest and strength of partisanship. So far as interest is concerned, partisans who are most interested in politics may be most likely to engage in partisan cheerleading under ordinary survey conditions. In this case, they may be *more* affected than low-interest respondents by incentives for correct response. Another possibility, however, is that highly interested partisans are most likely to sincerely hold different factual beliefs about politics (e.g., Abramowitz and Saunders, 2008; Taber and Lodge, 2000). If they do, they may be *less* affected by incentives. The estimates presented in column (3) of Table 2 show that both accounts are informative. In the control group, partisan gaps are larger among high-interest respondents, that is, those who report being "very much interested" in politics and current events. The average partisan gap is 0.14 for high-interest respondents and 0.08 for all others (whom we label "low-interest respondents"). The treatment reduces partisan gaps more for high- than for low-interest respondents — but only to an insignificant extent (−0.08 versus −0.06),

[13]Suppose that respondents believe that (a) they will answer 6 of our 12 questions correctly if they simply respond in a partisan manner, and (b) answering 6 questions correctly will give them a 1-in-100 chance of winning $200. If they also believe that answering all 12 questions correctly will double their chances to 2 in 100, then the expected value of answering all 12 questions correctly, relative to the "baseline" of answering 6 correctly, is $[(\$200 \times 2/100) - (\$200 \times 1/100)]/12$ questions = $0.167 per question. These calculations are speculative, because we did not verify how subjects interpreted the instructions. In our second experiment, the calculations are more straightforward, because subjects were given specific rewards on a question-by-question basis rather than entries in a lottery.

and high-interest respondents in the treatment group remain more polarized than their low-interest counterparts. (The treatment-group partisan gaps are 0.06 for high-interest respondents and 0.03 for low-interest respondents). Thus, highly interested people are initially more polarized, and their slightly greater responsiveness to incentives is not enough to overcome their initially greater polarization. Political interest is associated with polarization, but it does not significantly moderate the effects of incentives.[14]

The analyses that we report earlier exclude partisan "leaners" who may identify with a party less strongly than other partisans. In the Online Appendix, we present parallel analyses that include leaners. The results are similar: partisan leaners appear to behave like those who identify more strongly with the major American political parties.

Treatment-effect heterogeneity aside, the main finding of Experiment 1 is that small incentives for correct answers reduce partisan gaps in responses to factual questions by about 55%. Of course, Experiment 1 cannot tell us why 45% of the partisan gap remains. Following our model, the people responsible for this gap may sincerely disagree about which response is correct. Or they may agree about the correct response but value partisan cheerleading more than giving a correct answer. Or they may be so uncertain about which response is correct that incentives for correct responses cannot offset the expressive value of partisan responding. To evaluate these explanations, we turn to our second experiment.

4 Experiment 2: Effects of Incentives for Correct and "Don't Know" Responses on Partisan Divergence

We fielded our second experiment in 2012 using subjects recruited from Amazon.com's Mechanical Turk marketplace (Berinsky *et al.*, 2012). Subjects were required to pass a two-question attention screener and were then randomly assigned to a control group ($N = 156$) or to one

[14]Sixty-five percent of our CCES subjects report being "very much interested" in politics and current events. By contrast, the corresponding percentage among partisans in the 2008 ANES is 38%. That said, the overrepresentation of the interested in the 2008 CCES does not seem to affect the results. See the Online Appendix for a discussion of this point.

of three treatment groups, two of which we examine here.[15] In the first treatment group, participants were paid for each correct response ($N = 534$). In the second treatment group, participants were paid for each correct response and each "don't know" response ($N = 660$). Later, we restrict our analysis to the 795 individuals in these three groups who identified as either Democrats or Republicans.[16]

There are two major differences between this experiment and Experiment 1. First, and of greatest importance theoretically, we introduce a new condition here, in which we offer subjects a "don't know" response option and incentives for both correct and "don't know" responses. Therefore, unlike Experiment 1, Experiment 2 permits us to assess the extent to which partisan divergence that persists in the face of incentives for correct responses reflects self-aware ignorance, rather than partisan cheerleading or sincere differences in beliefs. Second, in both treatment conditions, we pay subjects for each correct response (instead of entering them into a lottery, as in Experiment 1), and we vary the amount offered for correct responses across participants. In the treatment that includes payment for "don't know" responses, we also vary the amount offered for that response across participants. These randomizations allow us to assess the degree to which partisan divergence is affected by the size of incentives.[17]

As before, we gave subjects 20 seconds to answer each question to limit opportunities for consultation of outside information sources. In all conditions, participants were initially asked five questions that were selected at random from a larger list that we describe later. All questions had a closed response format without a "don't know" option. Subjects

[15]In the third treatment, we paid participants a flat fee to answer questions post-treatment, just as we did in the control group. However, in this condition, we also allowed respondents to offer "don't know" answers. 14.8% of responses in this condition were "don't know."

[16]We fielded a one-item replication of this experiment on the 2012 CCES. The item was an economic retrospection item similar to those that have been used in the past to document partisan divergence (e.g., Bartels, 2002). The results were similar. See the Online Appendix for a discussion.

[17]As we discuss in the online appendix, one additional difference is that we used a graphical input device — a "slider" — to gather responses for this experiment. The advantage of this input device is that it allows subjects to provide responses continuously across the entire range of possible responses instead of requiring them to select one response from a small set of predefined options.

then received instructions that indicated how they would be paid for answers to the subsequent questions. They were then asked seven more questions: two new questions followed by the same five questions that they had previously been asked. (See the Online Appendix for details.) This design feature addresses one potential objection to our analysis of Experiment 1, which is that we use the control group in that experiment both to identify questions for which party gaps arise and as a baseline against which to evaluate the treatment group. In this experiment, by contrast, we use pre-treatment responses from all subjects to identify items for which partisan divergence arises, and we then compare post-assignment responses across treatment and control conditions.[18]

In the control condition, participants were paid a flat \$0.50 bonus to answer those seven post-treatment questions. In the pay-for-correct (PC) condition, participants were informed that they would be paid for each correct response. The amount offered for each correct response was randomly assigned to be \$0.10 (at probability $p = 0.25$), \$0.25 ($p = 0.25$), \$0.50 ($p = 0.25$), \$0.75 ($p = 0.15$), and \$1.00 ($p = 0.10$). (These amounts varied only across subjects, not within subjects across questions.) Finally, in the pay-for-correct-and-"don't know" (PCDK) condition, participants were again informed they would be paid for each correct response, and the amount offered for each correct response was assigned as in the prior treatment. Participants in this condition were also given "don't know" response options, and if they selected "don't know," they were randomly assigned to receive a fraction of the amount offered for a correct response: 20% of the payment for a correct response ($p = 1/3$), 25% ($p = 1/3$), and 33% ($p = 1/3$).

We list the 12 questions that we fielded in this experiment in Table 3, which also shows the correct response and the range of the response options that we offered. The correct responses varied across the entire range of potential answers: they were not concentrated at either end of the scale or in the middle. The effects of incentives therefore cannot be attributed to a tendency among treated subjects to offer middle-of-the-scale responses. The direction of partisan responding also

[18]In the Online Appendix, we also show that if we leverage this pre–post design by conducting a within-person analysis, we find results similar to those that we obtain when we focus only on post-assignment comparisons across conditions.

Table 3: Experiment 2: question wording and baseline partisan differences in scale scores.

Question	Question wording	Range of response line	Correct response	Mean pre-treatment Democratic response	Mean pre-treatment Republican response	Pre-treatment difference in scale scores, Democrats − Republicans	p-value of difference of party means, one-tailed test	N
Obama unemployment	From January 2009, when President Obama first took office, to February 2012, how had the unemployment rate in the country changed?	–2 (Unemployment decreased) to 4% (Unemployment increased)	Increased by 0.5%	0.552	0.378	0.174	0.000	389
Bush II unemployment	From January 2001, when President Bush first took office, to January 2009, when President Bush left office, how had the unemployment rate in the country changed?	–2 (Unemployment decreased) to 4% (Unemployment increased)	Increased by 3.6%	0.715	0.583	0.132	0.000	383

(*Continued*)

Table 3: (*Continued*)

Question	Question wording	Range of response line	Correct response	Mean pre-treatment Democratic response	Mean pre-treatment Republican response	Pre-treatment difference in scale scores, Democrats − Republicans	*p*-value of difference of party means, one-tailed test	*N*
Defense spending	For every dollar the federal government spent in fiscal year 2011, about how much went to the Department of Defense (US Military)?	3–27 cents	19.4 cents	0.731	0.631	0.101	0.000	355
Obama vote in 2008	In the 2008 Presidential Election, Barack Obama defeated his Republican challenger John McCain. In the nation as a whole, of all the votes cast for Obama and McCain, what percentage went to Obama?	50–62%	53.70%	0.544	0.444	0.100	0.001	366

(*Continued*)

Table 3: (*Continued*)

Question	Question wording	Range of response line	Correct response	Mean pre-treatment Democratic response	Mean pre-treatment Republican response	Pre-treatment difference in scale scores, Democrats − Republicans	*p*-value of difference of party means, one-tailed test	*N*
Iraq deaths: percent black	Approximately 12–13% of the US population is Black. What percentage of US Soldiers killed in Iraq since the invasion in 2003 are Black?	9–21%	9.90%	0.430	0.344	0.085	0.006	373
Medicaid spending	Medicaid is a jointly funded, Federal-State health insurance program for low-income and needy people. For every dollar the federal government spent in fiscal year 2011, about how much went to Medicaid?	3–27 cents	7.5 cents	0.577	0.502	0.075	0.013	343

(*Continued*)

Table 3: (*Continued*)

Question	Question wording	Range of response line	Correct response	Mean pre-treatment Democratic response	Mean pre-treatment Republican response	Pre-treatment difference in scale scores, Democrats − Republicans	p-value of difference of party means, one-tailed test	N
TARP: percent paid back	The Treasury Department initiated TARP (the first bailout) during the financial crisis of 2008. TARP involved loans to banks, insurance companies, and auto companies. Of the $414 billion spent, what percentage had been repaid, as of March 15, 2012?	1 (less repaid) to 100 (more repaid)	69.56%	0.391	0.324	0.068	0.027	349
Global warming	According to NASA, by how much did annual average global temperatures, in degrees Fahrenheit, differ in 2010 from the average annual global temperature between 1951 and 1980?	−1 (temperatures cooler) to 2 (temperatures warmer)	Increased by 1.1 degrees	0.685	0.640	0.045	0.013	382

(*Continued*)

Table 3: (*Continued*)

Question	Question wording	Range of response line	Correct response	Mean pre-treatment Democratic response	Mean pre-treatment Republican response	Pre-treatment difference in scale scores, Democrats − Republicans	*p*-value of difference of party means, one-tailed test	*N*
Iraq deaths	About how many U.S. soldiers were killed in Iraq between the invasion in 2003 and the withdrawal of troops in December 2011?	1000–7000	4,486	0.549	0.504	0.044	0.072	382
Debt service spending	The Treasury Department finances U.S. Government debt by selling bonds and other financial products. For every dollar the federal government spent in fiscal year 2011, about how much went to pay interest on those Treasury securities?	3–27 cents	6.2 cents	0.501	0.458	0.043	0.095	360

(*Continued*)

Table 3: (*Continued*)

Question	Question wording	Range of response line	Correct response	Mean pre-treatment Democratic response	Mean pre-treatment Republican response	Pre-treatment difference in scale scores, Democrats − Republicans	p-value of difference of party means, one-tailed test	N
Foreign-born population	According to the Census Bureau, in 2010 what percentage of the total population of the United States was born outside of the United States (foreign-born)?	1–100%	12.92%	0.785	0.772	0.013	0.239	388
Placebo: Mantle home runs 1961	In 1961, Roger Maris broke Babe Ruth's record for most home runs hit in a major league baseball season. He hit 61 home runs that year. How many home runs did his Yankees teammate Mickey Mantle hit in that year?	36–60	54	0.339	0.319		N/A	

Note: Questions are ordered by size of partisan gap in pre-treatment responses, with placebo question at the bottom. All responses scaled 0 to 1; 1 is the most Democratic response.

Source: Mechanical Turk, March–April 2012.

varied: sometimes, responses at the higher end of the scale favored the Democratic Party; sometimes, they favored the Republican Party. As before, we fielded a placebo question to assess whether participants were consulting outside references, and we found little evidence of this behavior.[19] (See the Online Appendix.)

As with Experiment 1, we recoded all responses to range from 0 to 1, with 0 corresponding to the response that portrayed Republicans most favorably and 1 corresponding to the response that portrayed Democrats most favorably.[20] Table 3 reports, for each non-placebo question, the observed pre-treatment difference in mean scale scores between Democrat and Republican participants. (Recall that each participant was asked five pre-treatment questions.) We find statistically significant ($p < 0.10$, one-tailed) partisan gaps for 10 of the 11 questions, with the largest gaps for questions about unemployment under Bush and Obama, and the smallest gaps for a question about the proportion of the population that is foreign-born. Our subsequent analysis is restricted to these 10 questions, that is, the questions to which partisanship makes a difference under ordinary survey conditions.[21] (Including all items produces similar results; see the Appendix.)

[19]In this experiment, subjects were explicitly asked, after they had completed the entire experiment, whether they had consulted any outside resources for an answer. (We told them that their pay would be unaffected by their answers to this question.) In the control condition, 1% of respondents reported consulting an outside reference, compared to 4% who reported doing so when paid $1.00 for a correct response. In the Online Appendix, we show that excluding all responses from any respondent who reported looking up the answer to any question produces highly similar results.

[20]We coded one end of the (continuous) input range at 0 and the other end at 1. Empirically, subjects use the entire scale range for all 10 questions. Our scaling implies that identical movements on the scale response range (e.g., 1 additional point of unemployment) are equivalent across the entire scale range.

[21]In pooled models, we assume movements across the scale range are on average the same across questions. As the units and endpoints of each question are different, this is a simplification for ease of presentation. While this is not a necessary assumption for our data analysis, we do not have strong *ex ante* theoretical reasons for presuming a different functional relationship for each question. We present a question-by-question analysis, which does not use this approximation, in Table A.2 of the Appendix.

4.1 *The Effect of Incentives for Correct and "Don't Know" Responses*

We begin by reporting the effect of the treatments on the frequency of selecting "don't know." Our model suggests that the rate at which participants select "don't know" when offered a payment for doing so indicates the degree to which they understand that they don't know the correct responses. In particular, if participants are sufficiently uncertain about the correct response and preferences for expressive partisan responding are not too large, then choosing "don't know" when paid to do so will yield greater expected utility than either expressive or sincere responses.

Pooling across the 10 questions for which we found pre-treatment partisan gaps, we find that 48% of responses in the PCDK condition are "don't know." That is, nearly half of participants forgo a response that would allow them to support their party or give them a chance to earn the larger payment that we offered for a correct response. Recall that for "don't know" responses, participants were randomly assigned to receive 20%, 25%, or 33% of the payment that they received for correct responses. Across these conditions, "don't know" responses were given 46%, 47%, and 50% of the time, respectively. These percentages are ordered as the theoretical model predicts, but only the difference between the 20% and 33% conditions approaches statistical significance ($p < 0.07$, one-tailed).[22]

This pattern — frequent "don't know" responses when subjects are paid to give that response, even when they are also offered more for correct responses — implies that many participants are so uncertain about the correct answers that they expect to earn more by selecting "don't know." In this experiment, uniformly distributed blind guesses will be correct about 17% of the time. Subjects who are completely unsure of the correct answers can therefore receive, in expectation, 17%

[22]One concern is that respondents may choose "don't know" simply because it allows them to avoid thinking about the question altogether. In footnote 15, we show that when offered a "don't know" option without payment, only 15% of responses were "don't know," a much lower rate than in this condition. Of note, as our model shows, choosing "don't know" when also offered a payment for a correct response is optimal only if the respondent is uncertain enough about the correct answer that it makes sense to give up the chance to guess and potentially earn a much larger amount.

of the payment that we offer for correct answers just by guessing blindly. Yet, when we paid subjects just 20% of the correct-answer payment for "don't know" responses, 46% chose to say "don't know" rather than to guess. We therefore infer that many respondents are highly unsure of which response is correct and give low weight to partisan responding.

As in the previous section, we study the effect of the treatments on party polarization by examining whether post-treatment partisan gaps differ between the control and treatment conditions. Our analysis initially takes the following form:

$$\begin{aligned} R_{ij} = {} & b_0 + b_1 \text{Democrat}_i + b_2 \text{PayCorrect}_i + b_3 \text{PayCorrectDK}_i \\ & + b_4(\text{PayCorrect}_i \times \text{Democrat}_i) \\ & + b_5(\text{PayCorrectDK}_i \times \text{Democrat}_i) + \text{Question}_j + e_i, \end{aligned}$$

where Democrat = 1 for Democratic participants and 0 for Republicans, PayCorrect = 1 for those assigned to the PC condition, PayCorrectDK = 1 for those assigned to the PCDK condition, and Question is a vector of question-specific fixed effects.[23] In this specification, b_1 is the amount of partisan divergence in the control condition, while $b_1 + b_4$ is the gap in the PC condition, and $b_1 + b_5$ is the gap in the PCDK condition.[24]

Our model predicts that $b_1 > 0$, $b_4 < 0$, and $b_5 < 0$. That is, both treatments will reduce partisan divergence relative to the control condition. Additionally, our theoretical model suggests that some

[23]We have multiple observations from the same respondent, which is why we cluster our standard errors by respondent. To test whether this clustering is sufficient to account for the correlated nature of multiple responses by the same respondent, we have also collapsed the data (to one observation per respondent) and estimated an otherwise identical specification. The results are highly similar, and we present them in the Online Appendix.

[24]To incorporate "don't know" responses into our analysis of partisan divergence, we must decide where to place those responses on the 0–1 scale that we use to analyze other responses. Because participants who admit that they don't know thereby forgo the opportunity to express support for their party, we treat these responses as being non-polarized. That is, we assign both Democrats and Republicans who choose "don't know" to the same position on the 0–1 scale. Specifically, we assign "don't know" responses for a given question to the average pre-treatment response that participants offered to that question. In practice, the specific value makes little difference to our analyses; the important point is that Democrats and Republicans are assigned to the same position on the scale if they say "don't know." If everyone chose "don't know," we would therefore find no differences between the parties.

partisans who will not respond to incentives for correct responses will nonetheless respond to incentives for "don't know" responses. For this reason, we also predict $b_5 < b_4$ (a larger reduction of partisan differences in the PCDK condition than in the PC condition).

The first column of Table 4 reports OLS estimates of the equation. (Parallel analysis for each individual question appears in Table A.2 of the Appendix.) The estimate of b_1 is 0.145 ($p < 0.01$), which means that, on average, control-group Democrats and Republicans differ by about 15% of the range of the scale. The estimate of b_4 is -0.087 ($p < 0.01$), so the total partisan gap in the PC condition is 0.058 ($0.145 - 0.087$). In other words, only 40% of the previously observed party gap remains when participants are paid small amounts for correct responses. Despite the differences between Experiments 1 and 2 in subject pools, questions, and other respects, this effect is similar to the effect that we find in Experiment 1. And like Experiment 1, this experiment shows that analyses of ordinary survey responses are likely to overstate the true extent of partisan polarization.[25]

This experiment also allows us to estimate the effect of incentives for "don't know" responses on polarization. The estimate of b_5 is -0.117 ($p < 0.01$), so the total partisan gap in the PCDK condition is 0.028 ($0.145 - 0.117$), or 80% smaller than the control-condition gap and about 50% smaller than the PC–condition gap. (These differences are significant at $p < 0.01$ and $p < 0.05$, respectively.) In practical terms, whereas the control-group difference between Democrats and Republicans was about 15% of the range of the scale, it shrinks to 3% of the range when we offer incentives for both correct and "don't know" responses.

In column (2), we estimate a Tobit specification because our response scales were bounded and unable to accommodate extreme responses. The estimates are similar to those shown in column (1). Indications of statistical significance do not change.

In column (3), we leverage the variation in incentive size to assess more fully the effect of differences in correct and "don't know" payments

[25] As in Experiment 1, question-by-question analysis yields less precise estimates and reveals heterogeneity across topics. Incentives have their largest effects on responses to questions about unemployment under Obama and the racial composition of Iraq War casualties. They also have large effects on basic retrospective assessments, reducing average partisan divergence by 41% and 72% in responses to questions about unemployment under Bush and Obama, respectively. (See Table A.2.)

Table 4: Experiment 2: effect of payment for correct responses on partisan differences in scale scores.

	(1) OLS	(2) Tobit	(3) OLS
Democrat (b_1)	0.145	0.152	0.145
	[0.028]***	[0.029]***	[0.028]***
Payment for correct response × Democrat (b_4)	−0.087	−0.091	
	[0.030]***	[0.032]***	
Payment for correct response and DK × Democrat (b_5)	−0.117	−0.123	
	[0.029]***	[0.030]***	
Payment for correct response	0.018	0.018	
	[0.025]	[0.026]	
Payment for correct response and DK	0.049	0.052	
	[0.024]**	[0.025]**	
Amount correct = \$0.10 × Dem.			−0.082
			[0.033]**
Amount correct = \$0.25 × Dem.			−0.092
			[0.033]***
Amount correct = \$0.50 × Dem.			−0.096
			[0.033]***
Amount correct = \$0.75 × Dem.			−0.061
			[0.036]*
Amount correct = \$1.00 × Dem.			−0.116
			[0.036]***
(Proportional payment for DK = 0.20) × Democrat			−0.031
			[0.018]*
(Proportional payment for DK = 0.25) × Democrat			−0.016
			[0.020]
(Proportional payment for DK = 0.33) × Democrat			−0.041
			[0.020]**
Amount correct = \$0.10			0.010
			[0.027]

(*Continued*)

Table 4: (*Continued*)

	(1) OLS	(2) Tobit	(3) OLS
Amount correct = \$0.25			0.028
			[0.027]
Amount correct = \$0.50			0.020
			[0.027]
Amount correct = \$0.75			0.005
			[0.029]
Amount correct = \$1.00			0.042
			[0.029]
Proportional payment for DK = 0.20			0.023
			[0.013]*
Proportional payment for DK = 0.25			0.030
			[0.017]*
Proportional payment for DK = 0.33			0.034
			[0.016]**
Constant	0.614	0.617	0.614
	[0.026]***	[0.026]***	[0.026]***
Observations	4608	4608	4608
R^2	0.179	N/A	0.181
F-test, 'Pay Correct × Dem.' > 'Pay Correct and DK × Dem.'	0.020	0.020	N/A

Note: The dependent variable is the mean scale score for the 10 questions on which we observed pre-treatment partisan gaps of $p < 0.10$. It ranges from 0 to 1. The analysis includes only Democrats and Republicans. Cell entries are coefficients with robust standard errors, clustered by respondent. Question fixed effects are not reported. *Significant at 10%; **significant at 5%; ***significant at 1% (two-tailed tests).

Source: Mechanical Turk, March–April 2012.

on observed divergence. Our specification includes indicators for each level of payment, each interacted with partisanship. The specification is highly flexible because it does not make assumptions about the functional form that relates incentive size to responses (e.g., a linear interaction between incentive size and responses).

Under this specification, the estimated 0.145 ($p < 0.05$) coefficient for *Democrat* is the average difference between Democrats and Republicans in the control condition. As expected, all five interactions between the amount paid for a correct response and *Democrat* are negative and statistically significant at $p < 0.10$, which means that party gaps are smaller when participants are offered incentives for correct responses. With one exception, larger payments are associated with smaller partisan gaps. For example, we estimate that partisan gaps are 56% smaller in the $0.10 payment condition than in the control group and 80% smaller in the $1.00 payment condition. The difference between the two coefficients (*Amount correct = $0.10 × Democrat* and *Amount correct = $1.00 × Democrat*) is marginally significant ($p < 0.10$, one-tailed test).

The third column of Table 4 also reports the effects of variation in the amount paid for "don't know" responses. All of the interactions between the fractional payment amounts and partisanship are in the expected negative direction, meaning that payments for "don't know" responses further reduce partisan gaps. For payments that are 20% or 33% as large as the payments for correct responses, the estimates are statistically significant at $p < 0.10$ (two-tailed), and the pooled estimate of the effect of "don't know" payments is significant at $p < 0.05$. To interpret these coefficients, one can fix the payment for a correct response at $0.10, in which case the estimated partisan gap is 0.063 ($0.145 - 0.082$, $p < 0.01$). Adding the "don't know" payment is estimated to reduce this party gap by between 0.02 (a 25% reduction for a "don't know" payment of $0.025) and 0.04 (a 65% reduction for a payment of $0.033).

The ordering of the effects for the proportional payments is non-monotonic. The largest reduction in partisan divergence is associated with the 33% payment for "don't know" responses, the next-largest reduction is associated with the 20% payment, and the smallest reduction is associated with the 25% payment. None of these estimates are statistically distinguishable from one another, perhaps reflecting the relatively small sample sizes in each condition. At the same time, the estimates imply that the combination of a $1.00 payment for a correct response and a $0.33 payment for a "don't know" response will eliminate *the entire* gap between Democrats and Republicans in responses to partisan factual questions.[26]

[26]This calculation is $0.145 - 0.116 - 0.041$, which is actually slightly smaller than 0.

Taken as a whole, these results have two implications. First, as in Experiment 1, modest incentives for correct responses substantially reduce partisan gaps, which is consistent with these gaps being due partly to expressive responding rather than to sincere differences in beliefs. Second, at least half of the partisan divergence that remains in the presence of incentives for correct responses alone appears to arise because people know that they do not know the correct response but continue to engage in expressive responding. On average, payments for correct responses in this experiment reduce partisan gaps by 60%. Adding "don't know" payments reduces partisan gaps by an additional 20 percentage points, leaving only 20% of the original gap. This result implies that fully half of the remaining gap arose because participants were unaware of the correct response and understood their lack of knowledge. Indeed, the relatively high rate of "don't know" response (about 48%) reveals that a surprising number of respondents were aware that they lacked clear knowledge of partisanship-relevant facts.

5 Expressive Survey Response and the Relationship Between Facts and Votes

Our experiments speak most directly to the role that partisan cheerleading plays in responses to factual questions about politics. But they also speak to the relationship between factual assessments and the political choices that people make. In particular, they suggest that efforts to understand the relationship between facts and votes with survey responses are likely to be biased in the absence of efforts to account for partisan cheerleading. To make this concern clear, we use Experiment 1 to assess the correlation between factual assessments and candidate preference in 2008. By comparing the correlations in the control and treatment conditions, we can understand whether the use of survey measures of economic perceptions to predict vote choice — a common practice in the literature on retrospective economic voting (e.g., Duch and Stevenson, 2006) — leads to biased conclusions when those measures are affected by partisan cheerleading.

With the data from Experiment 1, we estimate

$$\begin{aligned}\text{PresVote}_i = {} & b_0 + b_1\text{FactualAssessments}_i + b_2\text{PayCorrect}_i \\ & + b_3(\text{PayCorrect}_i \times \text{FactualAssessments}_i) + e_i,\end{aligned}$$

where PresVote = 1 indicates an intended vote for Obama and PresVote = 0 indicates an intended vote for McCain. (We exclude from the analyses those who aren't registered, prefer other candidates, or report that they won't vote.) FactualAssessments is the mean of the eight items that we included in our earlier analysis of the experiment, with each item coded so that 1 is the most Democratic response and 0 is the most Republican response. PayCorrect is an indicator for assignment to the pay-for-correct-response condition. Existing research suggests that $b_1 > 0$: statements of factual beliefs that favor the Democratic Party are associated with voting for the Democratic candidate. But if those statements are affected by cheerleading under ordinary survey conditions, then the association should be weaker in the treatment condition, implying $b_3 < 0$.

We present OLS estimates with clustered standard errors in Table 5.[27] Per these estimates, a one-standard-deviation increase (0.124) in the factual assessments scale is associated with a 22-percentage-point increase in the probability of voting for Obama ($p < 0.01$). Among those assigned to the treatment group, however, the negative estimate for b_3 means that this effect is reduced. For those subjects, the same shift in the assessments scale increases the probability of voting for Obama by 13 percentage points, a decrease of more than 40% ($p < 0.05$) in the association between those assessments and vote choice. This finding suggests that the observed correlation between normal (unincentivized) survey reports of factual assessments and voting is exaggerated by partisan cheerleading.

We are not suggesting that partisanship does not shape vote choice. However, the clear implication of our experiments is that standard survey measures of factual beliefs are affected by expressive responding. It is therefore difficult to use those measures to test the claim that partisanship works by shaping factual beliefs. When incentives are used to measure factual assessments more accurately, the apparent role of factual assessments in vote choice is reduced.

[27] In this sample, the mean *FactualAssessments* score is 0.59 and 50% of respondents prefer Obama. Probit results are substantively similar.

Table 5: Experiment 1: association of factual assessments with vote choice.

	Vote for Democratic presidential candidate
Average factual assessments scale score (b_1; 0 = most Republican, 1 = most Democratic)	1.770 [0.222]***
Payment for correct response (b_2)	0.418 [0.224]*
Payment for correct response × Average factual assessments scale score (b_3)	−0.741 [0.367]**
Constant	−0.548 [0.135]***
Observations	373
R^2	0.130

Note: The dependent variable is coded 1 for subjects who expressed an intention to vote for the Democratic candidate (Barack Obama), 0 for those who expressed an intention to vote for the Republican candidate (John McCain). The analysis includes only those Democrats and Republicans who expressed an intention to vote for one of the major-party candidates. "Payment for correct response" is coded 0 or 1. "Average factual assessments scale score" is computed by averaging across the eight non-placebo questions for which we found partisan gaps in the control condition. *Significant at 10%; **significant at 5%; ***significant at 1%.

Source: 2008 CCES.

6 Discussion and Conclusion

Differences between Democrats and Republicans in statements about factual matters are a hallmark of American politics. How should those differences be interpreted? One view is that they reveal perceptual biases. That is, Democrats and Republicans answer questions differently because they perceive "separate realities" (e.g., Kull *et al.*, 2004). Another possibility, highlighted in this paper, is that differences in survey responses arise because surveys offer partisans low-cost opportunities to express their partisan affinities.

To explore the distinction between beliefs and expressive statements made in surveys, we have presented a model of survey response that accounts for the possibility of expressive partisan responding. Our model shows that, if respondents have this sort of knowledge, incentives for correct responses can be used to distinguish sincere from insincere parti-

san responding. It also shows that incentives — no matter how large — may fail to reduce partisan responding. However, by providing incentives for both correct and "don't know" responses, one can estimate the proportion of partisan responding that arises either because of partisan cheerleading or because of uncertainty about the correct answers.

Guided by the model, we designed and fielded two novel experiments. In the first experiment, some participants were paid for correct answers to factual questions. The payments reduced observed partisan gaps by about 55%. In the second experiment, we also paid some participants for "don't know" responses. Payments for correct responses reduced partisan gaps by 60%. Payments for both correct and "don't know" responses reduced them by an additional 20%, yielding gaps that were 80% smaller than those that we observed in the absence of payments. Taken together, these results from experiments with small incentives provide lower-bound estimates of the extent to which partisan divergence arises because of expressive partisan returns and self-aware ignorance of the truth.

Why do we observe partisan responding in the first place? We have suggested that it follows from a conscious desire to offer a partisanship-consistent message. But it may also arise unconsciously. Survey respondents may not think seriously about correct answers under ordinary survey conditions, but incentives may reduce partisan gaps by causing respondents to think more carefully about correct answers (e.g., Kahan *et al.*, 2015; Kuklinski *et al.*, 2001, pp. 419–420). In either case, the takeaway is the same: conventional survey measures overstate partisan differences.[28]

The article most closely related to ours is Prior *et al.* (2015), which also appears in this issue. One basic difference is that Prior *et al.* focus on the accuracy of answers to factual questions about politics, while we focus on partisan differences in responses to those questions. That is, Prior *et al.* examine the extent to which payments or unpaid appeals

[28]We also designed Experiment 1 to test whether merely enhancing accuracy motivations would reduce partisan gaps. Specifically, we fielded an additional treatment, not discussed earlier, in which some respondents were told that their answers would be scored. This condition is similar to the "accuracy appeal" condition in Study 2 of Prior *et al.* (2015). But unlike those authors, we did not find that this treatment made much difference to partisans' responses, perhaps due to imprecision in our estimates.

for accurate responses reduce respondents' factual errors in surveys. By contrast, we examine the extent to which payments reduce differences in responses between Democrats and Republicans, and we do not focus on whether respondents answer correctly. Despite this difference, the basic results are complementary: ordinary surveys seem to exaggerate both the differences between partisans and the extent to which they are misinformed.

Two other differences between this article and Prior *et al.* (2015) merit attention, and we hope that they will guide future research. First, while we asked partisans about a range of issues in our two studies, Prior *et al.* focused on economic issues. Both articles show that ordinary surveys have been overstating partisan bias on a set of economic issues. However, our work shows that this pattern also extends to important issues beyond the economy, including evaluations of foreign affairs.

Second, we present a model of survey response which allows for the possibility that respondents know that they do not know the correct answers to factual questions. This model shows that, if respondents have this sort of knowledge, increasing the incentive to be accurate alone will not reduce partisan bias in survey responses. However, the experimental manipulation that we undertake, in which individuals are paid for "don't know" responses, permits us to gauge how many respondents recognize their own lack of knowledge about basic political matters. We find that a surprisingly large proportion of subjects appear willing to admit their ignorance by choosing "don't know" for a small financial incentive, despite the fact that this means forgoing the chance to express one's partisan feelings or to earn a larger reward by choosing the correct response. Furthermore, paying respondents to admit their own ignorance further reduces partisan divergence beyond what is achieved by only encouraging accuracy. We see this finding and its implications as particularly deserving of further study.

6.1 Implications of Our Findings

The main implication of our findings is that partisan differences in responses to factual questions may not imply partisan differences in beliefs. Instead, some portion of partisan polarization in survey responses about facts — perhaps a very large portion — is affective and insincere. Our

results thus call into question the common assumption that what people say in surveys reflects their beliefs. Of course, this assumption has often been called into question for sensitive topics. But our results suggest that a broader range of survey responses should be subject to scrutiny.

In light of this concern, efforts to assess the dynamics of public opinion should grapple with the possibility that over-time changes in partisans' expressed attitudes do not reflect changes in real beliefs. Instead changes in survey responses may reflect changes in the social returns to cheerleading (see Iyengar *et al.*, 2012) or in the degree to which different responses are understood to convey support for one's party. For example, elections may make more salient the need to support one's party, explaining why party polarization is more pronounced during campaigns (Iyengar *et al.*, 2012), just as "sorting" (Levendusky, 2009) may arise because holding particular policy positions may come to be associated with public support of one's party.

Our results may also help to resolve the tension between partisans' divergent assessments of objective conditions in surveys and the power of those conditions to explain aggregate election outcomes (e.g., Bartels and Zaller, 2001; Hibbs, 2000). We show not only that partisans do not fully believe their own survey responses, but also that they appear to be aware of their own ignorance. This self-awareness may make it easier to inform them of the facts and in turn change their votes. It may also help to explain why even some simple informational interventions appear to have relatively large effects on voting (e.g., Ferraz and Finan, 2008; Kendall *et al.*, 2015). And if these interventions have large effects on vote choice, then partisan patterns in voting may reflect, in large part, a self-aware lack of information rather than some persistent unwillingness to tie electoral sanctions to performance.

While our experiments are confined to factual questions, our argument applies to a wider range of questions. Our model suggests that, in the absence of a motivation to answer "partisan" questions accurately, partisan divergence should be large. These factors are also likely to apply to nonfactual matters. In particular, when survey reports of attitudes have expressive value, they may be inaccurate measures of true attitudes. And survey reports of vote intention may also be systematically biased by expressive responding.

We have focused on factual statements because our experimental design requires objectively verifiable responses. But other approaches

that do not rest on payments for objectively correct answers may also create pressures to be objective. For example, forward-looking judgments can be tied to incentives that are paid based on the realization of future events. And creative studies in psychology show that enhancing accuracy motivations can reduce partisan divergence even for questions that lack objectively correct answers (Campbell and Kay, 2014; Waytz *et al.*, 2014). In other words, partisans may be exaggerating not only their statements of factual belief but also their attitudinal statements.

Another area for subsequent research is the potential heterogeneity of treatment effects. We asked questions about many different policy areas, and we found variation across questions, in both the degree of partisan divergence that exists in the absence of incentives and the degree to which incentives reduce that divergence. Further exploration of this variation will be useful. For example, are certain policy topics perceived as more important, leading partisans to feel that they must stay "on message" when answering questions about those topics? In Experiment 2, we find the largest baseline partisan gaps for questions about economic performance, which is a key issue in almost all presidential campaigns. (See Table A.2.) Did partisans feel that straying from their team's message on these questions would be particularly damning? (Interestingly, despite the large initial gaps, incentives for correct and "don't know" responses reduced partisan divergence for these items by about as much as they did for other items.)

Similarly, which sorts of people are most likely to engage in expressive responding, and how do those people respond to incentives for correct responses? In our discussion of Experiment 1, we find that strength of partisanship does not seem to moderate expressive responding. Political interest does moderate expressive responding — as expected, more interested partisans are more polarized — but neither strength of partisanship nor political interest changes the effects of incentives. That said, these results are tentative, and a comprehensive examination of heterogeneity across subjects awaits future research.

Additionally, the imprecision of our estimates about the effects of increasing incentive size, for example, means that it would be valuable to conduct additional experiments with larger samples. The apparent effect of increasing incentive size also implies that it would be desirable to ascertain whether even larger incentives can further reduce apparent bias.

Our main contributions are a model of expressive survey response and two experiments that distinguish cheerleading behavior from sincere partisan divergence. We find that small financial inducements for correct responses can substantially reduce partisan divergence, and that these reductions are even larger when inducements are also provided for "don't know" answers. In light of these results, survey responses that indicate partisan polarization with respect to factual matters should not be taken at face value. Analysts of public opinion should consider the possibility that the appearance of polarization in American politics is, to some extent, an artifact of survey measurement rather than evidence of real and deeply held differences in assessments of facts.

Appendix: A Model of Expressive Survey Response

We begin with a model in which respondents derive utility from their survey responses in three ways: by offering answers that cast their party in a favorable light, by expressing their sincere beliefs, and by earning financial rewards. For now, we set aside the possibility that people can choose to say "don't know." For simplicity, we focus on the case in which there are two survey responses, r_1 and r_2. Individuals, indexed by the subscript i, are either Democrats ($T = D$) or Republicans ($T = R$). Individuals differ in their taste for partisan cheerleading and their beliefs about the truth.

Turning first to expressive benefits, individual i's taste for partisan cheerleading is denoted by the parameter c_i, for cheerleading, which ranges from 0 (no taste for it) to any positive number. Beliefs about the truth are described by the function $p_i(r_j)$, which is the probability that i believes response r_j, $j = 1$ or 2, is correct. In this example, we assume that response r_1 portrays Democrats most favorably, that response r_2 portrays Republicans most favorably, and that these assumptions are shared by respondents from both parties. Specifically, the expressive function $e(T, r_j)$ maps an individual's partisanship T to the personal benefit of offering response r_j, and is defined as $e(T = D, r_1) = e(T = R, r_2) = 1$ and $e(T = D, r_2) = e(T = R, r_1) = 0$. That is, Democrats and Republicans receive an expressive partisan utility boost from offering the response that portrays their party in a favorable light, and they receive no partisan utility from offering the response that is inconsistent with their partisan leanings.

The utility associated with providing a sincere response is measured by the "honesty" function $h_i(r_j)$. For simplicity, we assume $h_i(r_j) = p_i(r_j)$, that is, the honesty value of offering response r_j is the probability that the respondent believes it is true. Finally, some respondents may also receive an incentive, $I > 0$, which is the additional reward for a correct response. We assume utility is linear in I.

These assumptions allow us to describe a respondent's expected utility for offering response r_j as the sum of three terms. We omit the individual subscript i for clarity:

$$\mathrm{EU}(r_j|.) = h(r_j) + I \times p(r_j) + c \times e(T, r_j). \tag{A.1}$$

The first term is simply the honesty value of response r_j. The second term is the additional value of providing response j in the presence of incentive I (realized with the probability that response is correct). The third term is the partisan value of offering response r_j weighted by the respondent's value of expressive partisan responding, c. Using the assumption that $h()$ is equivalent to $p()$, we rewrite (A.1) as:

$$\mathrm{EU}(r_j|.) = (1 + I) \times p(r_j) + c \times e(T, r_j), \tag{A.2}$$

which is the form of the expected utility we focus on here. A respondent will offer the response r_j from (r_1, r_2) that maximizes (A.2).

To make the exposition as clear as possible, we suppose that the respondent is a Democrat ($T = D$). The analysis for the Republican partisan mirrors that for the Democratic partisan and is omitted. Recall that r_1 is the partisan Democratic response, and so $e(D, r_1) = 1$ and $e(D, r_2) = 0$.

First, consider how our model predicts that partisans will respond to a survey in the absence of incentives for correct responses. In this case, equation (A.2) reduces to

$$\mathrm{EU}(r_j|.) = p(r_j) + c \times e(T, r_j). \tag{A.3}$$

Using (A.3), the utility from reporting response r_1 is $p(r_1) + c$, and the utility from reporting r_2 is $p(r_2) = 1 - p(r_1)$. Therefore, the Democrat will report r_1 whenever $c \geq c* = 1 - 2p(r_1)$.

As c is weakly positive, whenever $p(r_1) > 0.5$ (i.e., the Democrat believes response r_1 is at least as likely to be correct as r_2), the Democrat

will offer the partisan response r_1 even in the absence of expressive returns (i.e., even if $c = 0$). By contrast, as $p(r_1)$ grows small (i.e., as the Democrat becomes increasingly likely to believe the pro-Republican response is correct), larger values of c are required to cause her to offer r_1. To produce a response of r_1, the partisan expressive return must be larger to offset the greater cost of providing an answer that is likely to be untrue.

This relationship is displayed graphically in Figure A.1(a), which shows that for each value of $p(r_1)$ there is a value of expressive partisan responding such that, for those Democrats with c at least this large, r_1 will be their survey response. Democrats offering r_1 are therefore composed of two groups. The first group consists of those who believe that r_1 is more likely to be correct than r_2; this group is represented by the right-hand side of the panel, for which $p(r_1) > 0.5$. The second group consists of those who believe that r_2 is more likely to be correct, but for whom that belief is offset by a larger return from offering an expressive partisan response. This group is represented by the upper segment of the left-hand side of the panel, which is labeled "insincere choice of r_1."

To link expressive returns to polarization of partisan responses, consider Panels (b) and (c). Panel (b) shows the response pattern for Republicans, which is a mirror image of Panel (a). And Panel (c) displays both partisan response patterns at once. It shows that in the presence of expressive returns, Democrats and Republicans *who share common beliefs about the truth* (are at the same position on the horizontal axis) *can nonetheless offer polarized survey responses if their value of expressive partisan responding is large enough.* When beliefs about the truth are shared, polarization is most prevalent when beliefs are most uncertain, that is, when $p(r_1) = p(r_2) = 0.5$. Polarization will also arise, even in the absence of returns to expressive partisan responding (i.e., when $c = 0$), if Democrats and Republicans hold different beliefs about the truth.

We next consider what happens when incentives are offered for correct responses, that is, when $I > 0$. From Equation (A.2), for a given value of I, there is a unique $c*' = (1 + I)(1 - 2p(r_1))$ such that all Democrats with an expressive responding parameter greater than $c*'$ will offer r_1. As before, incentives have no effect on the responses of Democrats who believe that response r_1 is correct (i.e., $p(r_1) > 0.5$).

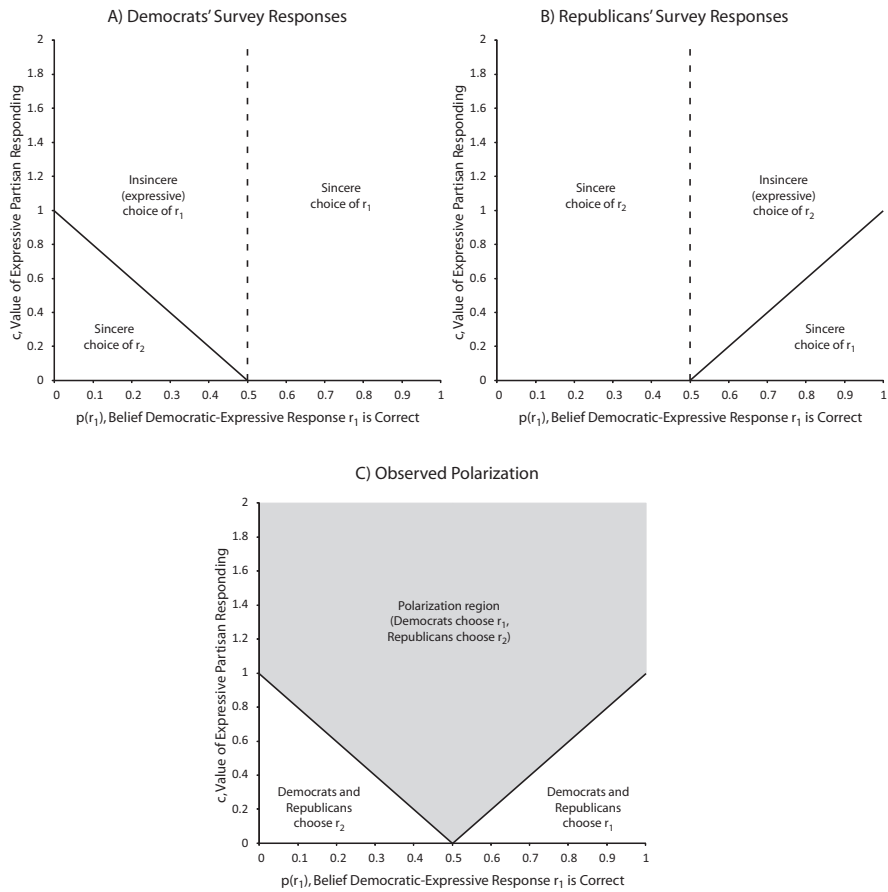

Figure A.1: Patterns of survey response in the absence of incentives by value of expressive partisan responding and beliefs about correct responses.

Note: Panel (a) displays Democrats' survey responses in the absence of incentives for different levels of returns to expressive partisan responding and beliefs about whether response r_1 is correct. Panel (b) displays responses for the same parameters for Republicans. Finally, the grey area in Panel (c) is the range of parameters for which Democrats and Republicans offer different survey responses despite common beliefs about which response is correct.

But for Democrats who believe response r_2 is more likely to be correct, a larger return to cheerleading is now required to offset the earnings that are likely to be lost by offering response r_1. Formally, $c*' = c* + (I \times (1 - 2p(r_1)))$. This relationship is shown in Panel (a) of Figure A.2. (For simplicity, we assume throughout Figure A.2 that $I = 1$.)

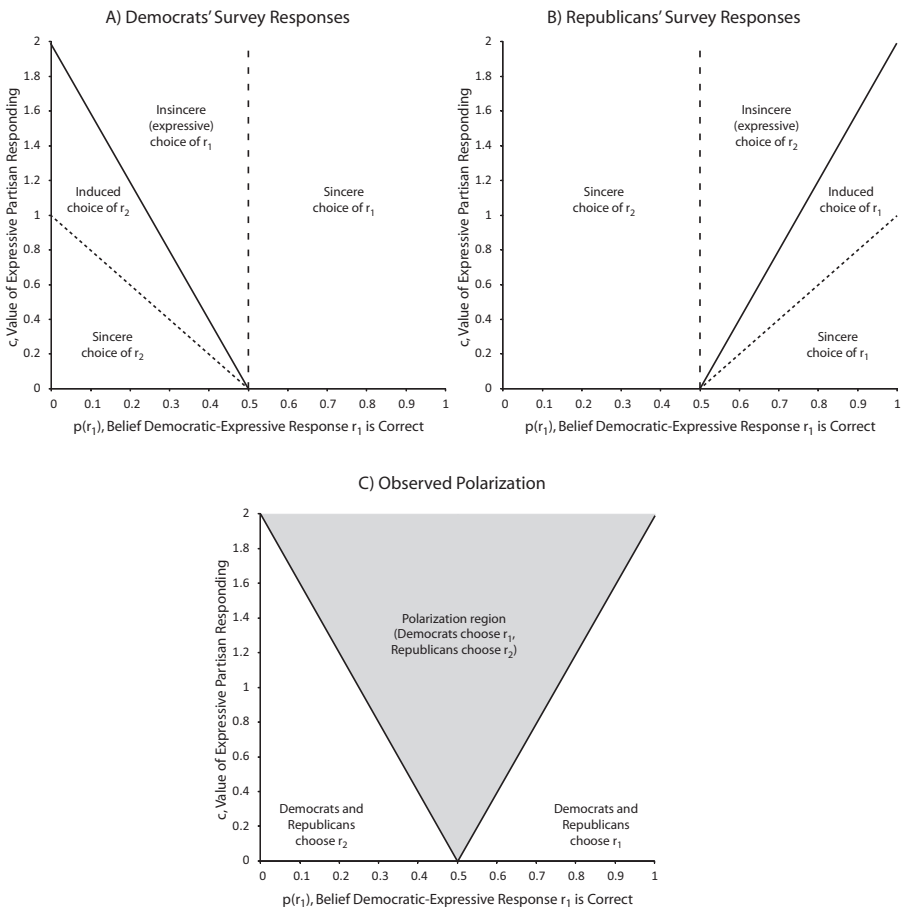

Figure A.2: Patterns of survey response given incentives for correct responses ($I = 1$) by value of expressive partisan responding and beliefs about correct responses.

Note: Panel (a) displays Democrats' survey responses given incentives $I = 1$ for correct responses for different levels of returns to expressive partisan responding and beliefs about whether response r_1 is correct. Panel (b) displays responses for the same parameters for Republicans. Finally, the grey area in panel (c) is the range of parameters for which Democrats and Republicans offer different survey responses despite common beliefs about which response is correct.

Comparison of Panel (a) in Figure A.1 and Panel (a) in Figure A.2 draws out a basic but important result: incentives for correct responses reduce expressive partisan responding by causing some of those who know that response r_1 is less likely to be true to offer response r_2 instead.

In Panel (a) of Figure A.2, these respondents are represented by the region that is labeled "induced choice of r_2."

Figure A.2 draws out a second important result: when a Democrat believes that r_2 is more likely to be correct, the additional value of expressive returns (c) that is required to make her offer response r_1 increases in her belief that r_2 is correct. Formally, $c*' - c*$ is increasing in $p(r_2)$. To see this result graphically, note that the vertical gap between the dashed and solid lines increases as one approaches the left side of the x-axis. This gap increases because the difference between $c*'$ and $c*$ is a function of $p(r_1)$. In other words, for those who are more uncertain ($p(r_1)$ is closer to 0.5), incentives have smaller effects. The intuition for this result is that a person who chooses the answer she thinks is most likely to be correct only earns the incentive for a correct response if that answer is in fact correct, which she expects to occur with the probability that she believes that response is correct. If a person believes r_1 is correct with probability 0.75, she earns the incentive I with probability 0.75 if she chooses r_1 and 0.25 if she chooses r_2. At the extreme, an individual who believes that r_1 and r_2 are equally likely to be true — that is, she knows that she does not know the truth — continues to offer r_1 regardless of incentives for correct responses because she won't (in expectation) do better by giving up the certain benefit of a partisan response because she earns the incentive I, in expectation, half the time for either response.

To illustrate the effect of incentives on polarization, Panel (b) of Figure A.2 shows the effect of incentives for Republican partisans, and Panel (c) displays both partisan response patterns at once. Comparison of Panel (c) in Figure A.1 to Panel (c) in Figure A.2 shows that increasing incentives decreases polarization. In particular, incentives reduce the frequency with which Democrats and Republicans who share common beliefs about the truth offer different survey responses, apart from the case in which $p(r_1) = p(r_2) = 0.5$.

This exposition leads us to two conclusions. First, incentives for correct answers reduce partisan divergence in the presence of shared beliefs about the truth. Second, partisan divergence may persist in the face of incentives. It is clear that if partisan groups have different sincere beliefs about which response is most likely to be true, paying respondents for correct responses will not reduce polarization. However, although it may seem intuitive that persistent partisan divergence in the

presence of incentives for correct responses *implies* underlying differences in beliefs about the truth, our analysis suggests partisan divergence may nonetheless persist for two other reasons. First, the taste for expressive partisan cheerleading (c) may be large. Second, even if that taste is small, individuals may be uncertain about the truth. In that case, they will offer partisan responses even in the face of large incentives for correct responding.

We have considered respondents who must provide either a partisan-consistent or a partisan-inconsistent response. But giving respondents the option to decline to provide a response may reduce observed polarization. To explore this possibility, we consider a model with an additional response option: "don't know."

A.1 Incorporating "Don't Know" Responses

To incorporate a "don't know" response option, we must specify the utility that a respondent receives from selecting "don't know." For simplicity, we assume that a "don't know" response (r_{dk}) yields some fixed positive psychological benefit $V_{dk} > 0$ plus whatever financial incentive is offered for giving that response (I_{dk}). (The results here are robust to allowing negative values of V_{dk}.) Specified this way, $U(r_{dk}) = V_{dk} + I_{dk}$. One can think of V_{dk} as the honesty value of choosing "don't know" relative to an incorrect response. As before, the individual is offered an incentive I for providing a correct response.

When will a respondent choose "don't know"? Note that the value of "don't know" is unaffected by c or $p()$, so a respondent chooses "don't know" when the values of c and $p()$ make both r_1 and r_2 less attractive than "don't know." Critically, one can earn the incentive I_{dk} with certainty by choosing "don't know", unlike the incentive for a correct response which is realized only if the chosen response is revealed after the fact to be correct, which occurs with the belief $p(r_j)$. Ceteris paribus, therefore, increasing uncertainty ($r_j = 0.5$) will make the "don't know" option more attractive. Recall from the previous analysis (illustrated in Panel (a) of Figure A.2) that a Democrat's selection of r_1 or r_2 depends on whether c is greater or less than $c*' = (1 + I)(1 - 2p(r_1))$.

Consider first a Democrat who would otherwise choose the "Republican" response, r_2. Her expected utility for choosing this response is

$(1+I)\times(1-p(r_1))$. This utility is greater than the utility associated with selecting "don't know" when $p(r_1) < p*(r_1) = 1-(V_{dk}+I_{dk})/(1+I)$. This $p*(r_1)$ is the lowest probability that the Democratic response (r_1) is correct for which the Democrat will select "don't know" rather than the Republican response. When $p(r_1)$ is below this critical value, the Democrat prefers to report the Republican response. Note that this critical value of $p*(r_1)$ is unaffected by the expressive value of partisan responding c, because the return to r_2 is unaffected by c.

Figure A.3 illustrates this logic. For presentation, we assume that $I = 1$, $I_{dk} = 0.75$, and $V_{dk} = 0.5$.[29] The value of $p*(r_1)$ is thus $1-(0.5+0.75)/(1+1) = 0.375$. Graphically, this solution is represented in Panel A by the leftmost line that defines the "induced don't know" region. Substantively, the point is that when $p(r_1)$ exceeds the critical value $p*(r_1)$, all cases in which the Democrat would have offered the Republican response are replaced by "don't know" answers.

We next examine how a Democrat who otherwise would have chosen the "Democratic" response, r_1, behaves in the presence of incentives for "don't know." We have already shown that if $c = c*'$, the Democrat is indifferent between the Democratic and the Republican responses, and that if $p(r_1) = p*(r_1)$, she is also indifferent between those responses and "don't know." However, as $p(r_1)$ rises above $p*(r_1)$, the expected return from choosing the "Democratic" response increases. This means that as the Democratic response becomes more likely to be true, smaller returns to expressive responding are required to keep the Democratic response more attractive than "don't know." In Panel (a) of Figure A.3, this condition is illustrated by the downward-sloping line that defines the top of the region labeled "induced don't know." Formally, $c = c*'' = (V_{dk}+I_{dk})/(p(r_1)(1+I))$ is the critical value, such that when $c > c*''$ (and $c > c*'$), the Democrat chooses the Democratic response over "don't know."

Parallel analysis for Republicans appears in Panel (b) of Figure A.3. For both Democrats and Republicans, the subjects who offer "don't

[29] We choose a relatively high level of I_{dk} because Figure A.3 illustrates the logic of our model when there are only two survey responses (in addition to "don't know"). Given only two responses, even complete uncertainty means that one is, in expectation, correct half of the time. In a model with more response options, the value of I_{dk} necessary to sustain don't know responses would be smaller. For example, one could also allow the value of I_{dk} to be negative.

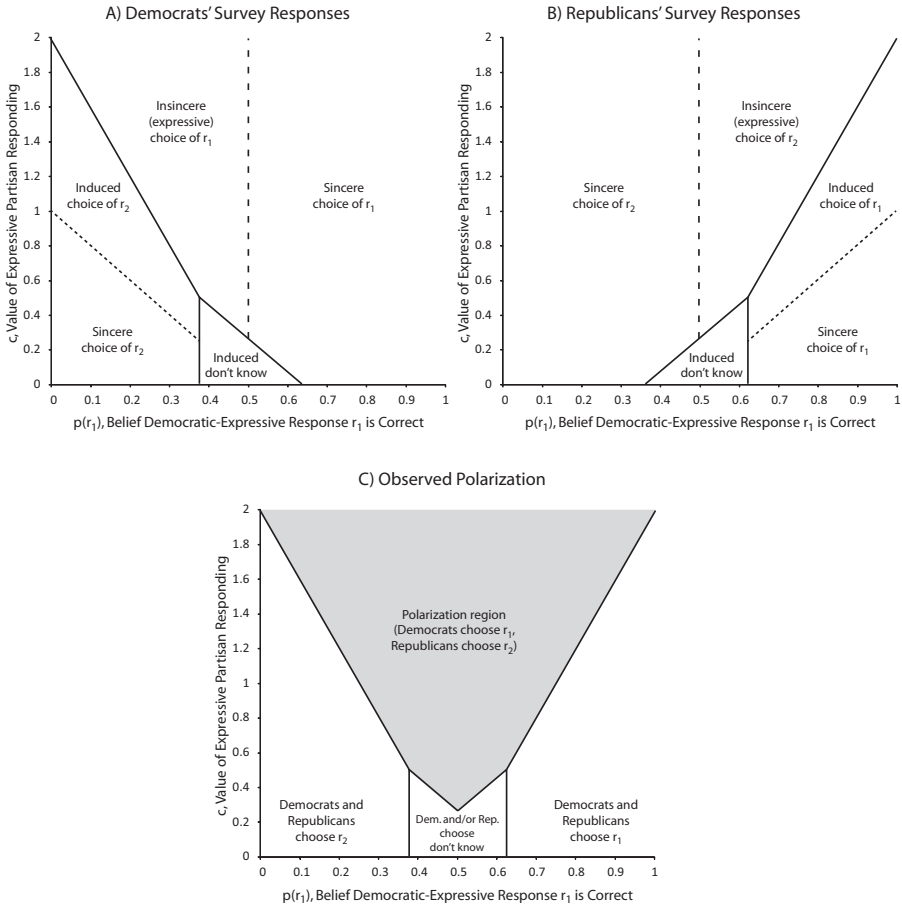

Figure A.3: Patterns of survey response given incentives for correct ($I = 1$) and don't know ($I_{dk} = 0.75$) Responses by value of expressive partisan responding and beliefs about correct responses.

Note: Panel (a) displays Democrats' survey responses in the absence of incentives for different levels of returns to expressive partisan responding and beliefs about whether response r_1 is correct. Panel (b) displays responses for the same parameters for Republicans. Finally, the grey area in Panel (c) is the range of parameters for which Democrats and Republicans offer different survey responses despite common beliefs about which response is correct.

know" responses are drawn from those who are most uncertain about which answer is correct, that is, from subjects for whom $p(r_1)$ is close to 0.5. Our analysis above establishes that it is this uncertainty that makes incentives for correct answers least likely to affect survey responses. Accordingly, for these uncertain respondents, the "sure thing" of a "don't know" payment is a more effective inducement than the smaller probability of earning a potentially larger payment for a correct response.

Combining these analyses, as we do in Panel (c), and comparing that plot to panel (c) of Figure A.2 allows us to assess the effect on observed polarization of offering incentives for both correct and "don't know" responses. Relative to simply offering incentives for correct responses, adding incentives for "don't know" responses decreases the frequency with which Democrats and Republicans who share common but weak beliefs about the correct response ($p(r_j)$ is not close to 1 for any j) provide divergent (non-"don't know") survey responses.

Table A.1: Experiment 1: Effect of payment for correct responses on partisan divergence in scale scores by question.

	(1) Iraq 07 to 08 change casualties	(2) Bush inflation change	(3) Bush unemployment change	(4) Est. bush approval	(5) Iraq total casualties	(6) Est. Bush approval among reps.	(7) Obama age	(8) McCain age
Democrat (1=Yes, 0=Republican)	0.239 [0.052]***	0.201 [0.044]***	0.168 [0.056]***	0.092 [0.023]***	0.087 [0.038]**	0.070 [0.039]*	0.050 [0.031]	0.044 [0.025]*
Payment for correct response* Democrat	−0.078 [0.077]	−0.026 [0.061]	−0.074 [0.079]	−0.100 [0.034]***	−0.064 [0.054]	−0.072 [0.055]	−0.048 [0.045]	−0.053 [0.033]
Payment for correct response	0.043 [0.051]	0.059 [0.052]	0.091 [0.058]	0.018 [0.024]	0.051 [0.036]	0.026 [0.039]	0.005 [0.034]	0.010 [0.024]
Constant	0.177 [0.033]***	0.694 [0.036]***	0.598 [0.042]***	0.818 [0.016]***	0.114 [0.024]***	0.724 [0.029]***	0.508 [0.023]***	0.637 [0.019]***
Observations	415	409	407	421	412	416	419	422
R-squared	0.064	0.093	0.032	0.044	0.014	0.008	0.008	0.012
Percentage of partisan gap eliminated by payment for correct response	32.5%	12.9%	44.4%	108.7%	73.3%	103.4%	95.2%	119.1%

Note: Includes only Democrats and Republicans. Cases included are from control and paid for correct response condition. OLS Coefficients with robust standard errors. *Significant at 10%; **significant at 5%; ***significant at 1% (two-tailed tests).

Source: 2008 CCES.

Table A.2: Experiment 2: Effect of payment for correct and don't know responses on partisan divergence in scale scores by question.

	(1) Obama unemployment	(2) Bush II unemployment	(3) Defense spending	(4) Obama vote 08	(5) Iraq deaths % black	(6) Medicaid spending	(7) TARP % paid back	(8) Global warming amount	(9) Iraq deaths	(10) Debt service spending
Democrat (1=Yes,	0.293	0.239	0.118	0.126	0.219	0.136	0.107	0.133	0.051	0.010
0=Republican)	[0.065]***	[0.068]***	[0.085]	[0.086]	[0.081]***	[0.086]	[0.091]	[0.057]**	[0.072]	[0.089]
Payment correct *	−0.210	−0.097	−0.021	−0.091	−0.207	−0.088	−0.059	−0.107	−0.043	0.062
democrat	[0.078]***	[0.077]	[0.093]	[0.095]	[0.092]**	[0.096]	[0.100]	[0.063]*	[0.081]	[0.099]
Payment DK and	−0.184	−0.202	−0.092	−0.056	−0.209	−0.139	−0.091	−0.093	−0.052	−0.035
correct * Democrat	[0.073]**	[0.073]***	[0.088]	[0.092]	[0.086]**	[0.089]	[0.096]	[0.060]	[0.077]	[0.092]
Payment for correct	0.021	−0.049	−0.053	−0.028	0.113	0.081	−0.019	0.058	−0.009	0.042
response	[0.057]	[0.072]	[0.080]	[0.080]	[0.069]	[0.079]	[0.073]	[0.055]	[0.064]	[0.082]
Payment for DK	−0.019	0.079	0.059	−0.031	0.158	0.067	0.053	0.038	0.013	0.039
and correct response	[0.054]	[0.068]	[0.076]	[0.078]	[0.064]**	[0.073]	[0.071]	[0.053]	[0.062]	[0.076]
Constant	0.401	0.586	0.630	0.467	0.241	0.489	0.346	0.605	0.522	0.490
	[0.048]***	[0.066]***	[0.073]***	[0.073]***	[0.060]***	[0.071]***	[0.066]***	[0.050]***	[0.057]***	[0.074]***
Observations	444	485	446	457	470	442	452	466	479	467
R-squared	0.077	0.099	0.050	0.029	0.023	0.022	0.017	0.028	0.005	0.029
F-test, Pay Correct * Dem. > Pay DK and Correct * Dem.	0.310	0.010	0.060	0.250	0.490	0.150	0.280	0.340	0.420	0.030
Percentage of partisan gap eliminated by payment for correct response	71.6%	40.6%	17.8%	72.1%	94.2%	64.8%	55.3%	80.5%	83.3%	
Increases percentage of partisan gap eliminated by payment for DK and correct response	62.7%	84.5%	78.2%	44.7%	95.1%	101.9%	84.8%	70.3%	101.0%	335.7%

Note: Includes only Democrats and Republicans. Comparison of post-treatment responses in control, pay correct, and pay correct and don't know conditions. OLS Coefficients with robust standard errors. *Significant at 10%; **significant at 5%; ***significant at 1% (two-tailed tests). **Sources:** 2012 Mechanical Turk study.

References

Abramowitz, A. I. and K. L. Saunders (2008), "Is Polarization a Myth?", *The Journal of Politics*, 70(April), 542–55.

Ansolabehere, S., M. Meredith, and E. Snowberg (2013), "Asking about Numbers: Why and How", *Political Analysis*, 21(January), 48–69.

Bartels, L. M. (2002), "Beyond the Running Tally: Partisan Bias in Political Perceptions", *Political Behavior*, 24(June), 117–50.

Bartels, L. M. and J. Zaller (2001), "Presidential Vote Models: A Recount", *PS: Political Science and Politics*, 34(March), 9–20.

Berinsky, A. J. (2005), *Silent Voices: Public Opinion and Participation in America*, Princeton, NJ: Princeton University Press.

Berinsky, A. J. (2015), "Rumors and Health Care Reform: Experiments in Political Misinformation", *British Journal of Political Science*.

Berinsky, A. J., G. S. Lenz, and G. A. Huber (2012), "Using Mechanical Turk As a Subject Recruitment Tool for Experimental Research", *Political Analysis*, 20(3).

Bishop, G. F., R. W. Oldendick, and A. Tuchfarber (1984), "What Must My Interest in Politics Be If I Just Told You 'I Don't Know'?"", *Public Opinion Quarterly*, 48(2), 510–9.

Brennan, G. and L. Lomasky (1997), *Democracy and Decision: The Pure Theory of Electoral Preference*, New York: Cambridge University Press.

Campbell, A., P. E. Converse, W. E. Miller, and D. E. Stokes (1960), *The American Voter*, Chicago: University of Chicago Press.

Campbell, T. H. and A. C. Kay (2014), "Solution Aversion: On the Relation Between Ideology and Motivated Disbelief", *Journal of Personality and Social Psychology*, 107(November), 809–24.

Cohen, G. L. (2003), "Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs", *Journal of Personality and Social Psychology*, 85(November), 808–22.

Conover, P. J., S. Feldman, and K. Knight (1986), "Judging Inflation and Unemployment: The Origins of Retrospective Evaluations", *Journal of Politics*, 48(3), 565–88.

Conover, P. J., S. Feldman, and K. Knight (1987), "The Personal and Political Underpinnings of Economic Forecasts", *American Journal of Political Science*, 31(August), 559–83.

Duch, R. M. and R. Stevenson (2006), "Assessing the Magnitude of the Economic Vote over Time and across Nations", *Electoral Studies*, 25, 528–47.

Ferraz, C. and F. Finan (2008), "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes", *The Quarterly Journal of Economics*, 123(2), 703–45.

Fiorina, M. P. (1981), *Retrospective Voting in American National Elections*, New Haven, CT: Yale University Press.

Gaines, B. J., J. H. Kuklinski, P. J. Quirk, B. Peyton, and J. Verkuilen (2007), "Same Facts, Different Interpretations: Partisan Motivation and Opinion on Iraq", *Journal of Politics*, 69(November), 957–74.

Gerber, A. S., G. A. Huber, and E. Washington (2010), "Party Affiliation, Partisanship, and Political Beliefs: A Field Experiment", *American Political Science Review*, 104(November), 720–44.

Green, D., B. Palmquist, and E. Schickler (2002), *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*, New Haven, CT: Yale University Press.

Hamlin, A. and C. Jennings (2011), "Expressive Political Behaviour: Foundations, Scope and Implications", *British Journal of Political Science*, 41(3), 645–70.

Healy, A. and N. Malhotra (2009), "Myopic Voters and Natural Disaster Policy", *American Political Science Review*, 103(3), 387–406.

Hibbs, D. A. (2000), "Bread and Peace Voting in U.S. Presidential Elections", *Public Choice*, 104(1–2), 149–80.

Iyengar, S., G. Sood, and Y. Lelkes (2012), "Affect, Not Ideology: A Social Identity Perspective on Polarization", *Public Opinion Quarterly*, 76(Fall), 405–31.

Jacobson, G. C. (2006), *A Divider, Not a Uniter: George W. Bush and the American People*, Upper Saddle River, NJ: Pearson.

Jerit, J. and J. Barabas (2012), "Partisan Perceptual Bias and the Information Environment", *Journal of Politics*, 74, 672–84.

Kahan, D. M., H. Jenkins-Smith, T. Tarantola, C. L. Silva, and D. Braman (2015), "Geoengineering and Climate Change Polarization: Testing a Two-Channel Model of Science Communication", *The Annals of the American Academy of Political and Social Science*, 658(1), 192–222.

Kendall, C., T. Nannicini, and F. Trebbi (2015), "How Do Voters Respond to Information? Evidence from a Randomized Campaign", *American Economic Review*, 105(January), 322–53.

Kuklinski, J. H., M. D. Cobb, and M. Gilens (1997), "Racial Attitudes and the 'New South'", *Journal of Politics*, 59(May), 323–49.

Kuklinski, J. H., P. J. Quirk, J. Jerit, and R. F. Rich (2001), "The Political Environment and Citizen Competence", *American Journal of Political Science*, 45(April), 410–24.

Kull, S., C. Ramsay, S. Subias, and E. Lewis (2004), "The Separate Realities of Bush and Kerry Supporters", Program on International Policy Attitudes. Available at: http://www.pipa.org/OnlineReports/Iraq/IraqRealities_Oct04/IraqRealitiesOct04rpt.pdf (accessed on 20/04/2015).

Lau, R. R., D. O. Sears, and T. Jessor (1990), "Fact or Artifact Revisited: Survey Instrument Effects and Pocketbook Politics", *Political Behavior*, 12(3), 217–42.

Lavine, H. G., C. D. Johnston, and M. R. Steenbergen (2012), *The Ambivalent Partisan: How Critical Loyalty Promotes Democracy*, Oxford University Press.

Levendusky, M. (2009), *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*, Chicago, IL: University of Chicago Press.

Luskin, R. C. and J. G. Bullock (2011), "'Don't Know' Means 'Don't Know': DK Responses and the Public's Level of Political Knowledge", *Journal of Politics*, 73(April), 547–57.

Morton, R. C. and K. C. Williams (2010), *Experimental Political Science and the Study of Causality: From Nature to the Lab.* New York: Cambridge University Press.

Palmer, H. D. and R. M. Duch (2001), "Do Surveys Provide Representative or Whimsical Assessments of the Economy?", *Political Analysis*, 9(1), 58–77.

Prior, M. (2007), "Is Partisan Bias in Perceptions of Objective Conditions Real? The Effect of an Accuracy Incentive on the Stated Beliefs of Partisans", Presented at the Annual Conference of the Midwest Political Science Association, Chicago.

Prior, M. and A. Lupia (2008), "Money, Time, and Political Knowledge: Distinguishing Quick Recall and Political Learning Skills", *American Journal of Political Science*, 52(January), 169–83.

Prior, M., G. Sood, and K. Khanna (2015), *The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions*, Manuscript: Princeton University, Princeton, NJ.

Schuessler, A. A. (2000), *A Logic of Expressive Choice*, Princeton, NJ: Princeton University Press.

Sears, D. O. and R. R. Lau (1983), "Inducing Apparently Self-Interested Political Preferences", *American Journal of Political Science*, 27 (May), 223–52.

Shapiro, R. Y. and Y. Bloch-Elkon (2008), "Do the Facts Speak for Themselves? Partisan Disagreement as a Challenge to Democratic Competence", *Critical Review*, 20(1), 115–39.

Taber, C. and M. Lodge (2000), "Three Steps Toward a Theory of Motivated Political Reasoning", in *Elements of Reason*, ed. A. Lupia, M. D. McCubbins, and S. L. Popkin, New York: Cambridge University Press.

Waytz, A., L. L. Young, and J. Ginges (2014), "Motive Attribution Asymmetry for Love vs. Hate Drives in Intractable Conflicts", *Proceedings of the National Academy of the Sciences of the United States of America*, 111(44), 15687–92.

Wilcox, N. and C. Wlezien (1993), "The Contamination of Responses to Survey Items: Economic Perceptions and Political Judgments", *Political Analysis*, 5(1), 181–213.

Zaller, J. R. (1992), *The Nature and Origins of Mass Opinion*, New York: Cambridge University Press.