

Partisan Bias in Factual Beliefs about Politics

John G. Bullock, Alan S. Gerber, and Gregory A. Huber

Yale University

March 29, 2011

Latest version at <http://bullock.research.yale.edu/facts.pdf>

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 use two experiments and a model of responses to “partisan” factual questions. All subjects in our experiments were asked factual questions about politics, but some received financial incentives to answer correctly. Partisan gaps in correct responding are reduced only moderately when incentives are offered, which constitutes some of the strongest evidence to date that such patterns reflect sincere differences in factual beliefs.

*Please send comments to john.bullock@yale.edu, alan.gerber@yale.edu, or gregory.huber@yale.edu. David Doherty and Conor Dowling helped us to set up the first experiment. Celia Paris helped with the analysis. Bethany Albertson, Matias Bargsted, Ryan Enos, Anthony Fowler, Seth Hill, Ben Lauderdale, Tse-min Lin, Robert Luskin, Brendan Nyhan, Jason Reifler, Daron Shaw, and Alex Theodoridis made many useful comments. We thank all of them.

One year ago, Harris Interactive surveyed U.S. adults to determine their beliefs about Barack Obama. Majorities of Republicans claimed to believe that Obama is a Muslim and that he wants to “turn over the sovereignty of the United States to a one world government.” Forty-five percent also claimed to believe that he was not born in the United States. Far fewer Democrats claimed to hold the same beliefs (Harris Interactive 2010). Subsequent polls suggested that Harris may have overstated the proportions of Americans holding such beliefs, but even in these polls, the gap between Democrats and Republicans was substantial (e.g., Pew Research Center 2010).

Such poll results may seem extreme, but they exemplify a common pattern in American public opinion: people’s tendency to give “partisan” answers to survey questions that may not reflect their beliefs. The pattern would be less noteworthy if it were confined to questions about political attitudes or tastes. But it exists even in responses to factual questions about politics. These questions may be about economic trends during the tenure of the incumbent President (Bartels 2002, 133-38). Or about the President’s foreign policy stands (PIPA/Knowledge Networks 2004, 13-17). Or about whether Iraq possessed weapons of mass destruction at the time of the U.S. invasion (ABC News/The Washington Post 2004, 2005), or about whether the United States has discovered such weapons (Harris Interactive 2006; see also Duelfer 2004, 6).

Systematic errors in responses to factual questions about politics are not mere curiosities. They call into question the ability of voters to judge politicians’ performance in office. For example, anti-tax voters who believe that an incumbent has raised taxes when he has instead lowered them will be unable to reward him for his stance on taxes (Cooper 2010). The errors also call into question a widespread assumption about legislators’ duties: should legislators heed constituents’ views if those views are rooted in false beliefs?

These dilemmas aside, many constituents may not believe their own incorrect statements about factual matters. Their statements may instead be “expressive,” intended to make themselves feel good by promoting one political party or denigrating another. This possibility arises for many factual questions on which partisan differences have been observed. But little research

has attempted to determine whether responses to such questions are sincere. Making such determinations is the point of this paper.

To determine the extent to which responses to “partisan” factual questions are sincere, we conducted two experiments. All subjects in both experiments were asked questions about politics, but some were given a financial incentive to answer correctly. Partisan gaps in correct responding proved largely invariant to incentives, suggesting that most partisans sincerely believe their answers—even when those answers are starkly at odds with the truth.

We begin by reviewing theory and evidence about partisan response patterns in answers to factual questions. The next sections introduce experiments and discuss what they reveal about sincere responding and partisan cheerleading in response to charged factual questions about politics. The final section considers normative implications and avenues for future research.

Theory and Prior Evidence

The most common model of responses to factual questions is the “Rasch model” (Rasch 1960, 1966), which holds that a person’s probability of answering a question correctly depends on only his level of knowledge and the question’s difficulty. Rasch’s model rarely appears directly in political science, but Lord and Novick (1968, 426-49) show that sums of correct answers to questions with binary (correct/incorrect) scoring are sufficient statistics for knowledge under the Rasch model. The Rasch model thus underpins most scholarship on political knowledge, which relies heavily on summing correct answers to factual questions (e.g., Delli Carpini and Keeter 1996, esp. 296-98).

The Rasch model is noteworthy for what it leaves out. There is no place in it for partisanship or person-to-person variation in the desire to answer questions accurately. In most cases, the omission of these factors may be unimportant. For example, a workhorse of political knowledge studies like “How much of a majority is required for the U.S. Senate and House to override a presidential veto?” affords little opportunity to express a partisan attitude by giving an incorrect answer. The desire to answer accurately is also unimportant: one is likely to either

know the answer or not know it, so increasing one's motivation to respond accurately is unlikely to affect one's response.¹ For such items, the Rasch model may do a good job of capturing the qualities that lead people to answer questions correctly.

But many items do not fit this description. Bartels (2002, 133-38) describes some prominent examples: retrospective questions about economic trends. In the 1988 American National Election Study, subjects were asked whether, "compared to 1980, the level of unemployment in the country" had increased, decreased, or stayed about the same. It had decreased: the average unemployment rate was 7.0% in 1980, 5.4% in 1988 (U.S. Census Bureau 1990, 378). Subjects were asked a similar question about inflation, which had also declined. Such questions afford an opportunity for partisan attitude expression: the ANES interview was conducted in the context of the 1988 Presidential campaign, and Democrats who knew the correct answers may nevertheless have seized the opportunity to criticize Ronald Reagan, even if doing so required answering incorrectly. Indeed, nearly 50% of strong Democrats responded that unemployment and inflation had increased "since 1980" (Bartels 2002, 135). Republicans were much more likely to answer correctly. The opposite patterns appeared in 2000 among ANES subjects who answered questions about crime rates and federal deficit trends during Bill Clinton's terms in office (Bartels 2002, 137).

A small but growing body of literature is devoted to identifying partisan gaps in correct responding to factual questions (Bartels 2002; Prior 2007; Shapiro and Bloch-Elkon 2008). But little of this research attempts to determine the extent to which the gaps reflect sincere differences in knowledge and the extent to which they instead reflect partisan cheerleading. To be sure, many social scientists have wrestled with the problem of inauthentic survey responses (e.g., Berinsky 2005; Kuklinski, Cobb, and Gilens 1997; Noelle-Neumann 1993). But their concern lies far more with social desirability biases than with the problems that may be caused by "expressive benefits" in survey responses. And the methods used to overcome problems associated with

¹ See Luskin and Bullock (2011, 4-5) for a discussion of "partial knowledge."

social desirability or sensitive topics—for example, the “list experiment” (Kuklinski, Cobb, and Gilens 1997)—do not seem applicable to the problem of eliciting sincere responses when people derive “expressive benefits” from answering insincerely.

Instead, experimenters have long turned to incentives to induce honest responses or rational behavior. In a review of experiments involving incentives, Morton and Williams (2010, 358-61) argue that incentives are often successful at reducing the size and frequency of decision-making errors. But almost all of the studies that they survey are apolitical and do not involve tests of knowledge. Prior and Lupia (2008) do study the effects of financial incentives on responses to factual questions about politics, and they find that the effects are weak.² But they do not examine partisan patterns in responses to factual questions, leaving open the possibility that incentives are effective in combating partisan responses to factual questions even if they are ineffective on average.

To date, only one study has examined the effects of incentives on partisan response patterns to factual questions about politics: Prior (2007). Subjects in the study were asked to answer 14 factual questions about politics; some were randomly assigned to receive \$1 for each correct answer. The results were mixed, but they suggested that \$1 incentives can reduce the gap in partisans’ responses to such questions.³ Prior examined differences in the average answers given by Democrats and Republicans; he did not focus on whether incentives made partisans more likely to respond correctly.

² 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, 175). Subjects who were paid \$1 for each correct answer and given a day to look up the answers—rather than one minute per question, as in the other conditions—averaged 5.6 correct answers (Prior and Lupia 2008, 177).

³ Incentives reduced the partisan gap 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.

None of this research directly considers the extent to which answers to factual questions are sincere or the extent to which they represent cheerleading. The problem is that data alone cannot tell us about the proportion of “true believers” or cheerleaders in any sample. These are unobservable quantities: we can elicit responses from subjects and we can pay them to give correct answers, but we cannot peer inside their minds. To estimate these unobservable proportions, we need an organized set of assumptions—a model.

Model of Sincere Responses and Partisan Cheerleading

Assume an experiment in which all subjects are asked factual questions about politics. Some subjects are randomly assigned to a treatment that induces sincere responding. Let $y_{ij} \in \{0, 1\}$ be i 's answer to item j . It is correct ($y_{ij} = 1$) or incorrect ($y_{ij} = 0$). Let $x_i \in \{0, 1\}$ equal 1 if i knows the correct answer to j , 0 if he does not. Of course, some people will guess when they do not know the answer: we say that $g_{ij} \in \{0, 1\}$ indicates whether i correctly guesses the answer to j .

Assume that all of the people whose answers we study are Democrats or Republicans. We make this assumption for notational convenience, and it can easily be relaxed. Let $Dem_i, Rep_i \in \{0, 1\}$ indicate whether i is a Democrat or a Republican. Let $p_{ij} \in \{0, 1\}$ equal 1 if i is a *partisan responder*: unless he is paid, he answers j in the way that casts his party in the best light or the other party in the worst light. If $p_{ij} = 0$, i is a *sincere responder*: his response is never dictated by his partisanship. And let $\delta_{ij} \in \{0, 1\}$ equal 1 when the correct answer favors i 's party, 0 otherwise. We assume that there is an *advantaged party* and a *disadvantaged party* for each question, i.e., $\delta_{ij} = 1$ for all members of one party and $\delta_{ij} = 0$ for all members of the other.

Finally, let $t_i \in \{0, 1\}$ equal 1 if i is assigned to the treatment condition, 0 if he is not. By random assignment, $t_i \perp x_{ij}, p_{ij}, \delta_{ij}$.

Our model of correct response is

$$y_{ij} = (1 - t_i) \left[p_{ij} \delta_{ij} + (1 - p_{ij}) x_{ij} + (1 - p_{ij})(1 - x_{ij}) g_{ij} \right] + t_i \left[x_{ij} + (1 - x_{ij}) g_{ij} \right] \quad (1)$$

$$g_{ij} \stackrel{\text{iid}}{\sim} \text{Bernoulli} \left(\frac{1}{J} \right)$$

By this model, subject i answers question j correctly if he (a) is untreated, sincere, and knows the answer, (b) is untreated and a partisan responder but is answering a question whose correct answer favors his party, (c) is treated and knows the answer, or (d) guesses correctly. Otherwise, he answers incorrectly.

In the model, all guessers are ignorant, i.e., people guess only when $x_{ij} = 0$. The assumption that $g_{ij} \stackrel{\text{iid}}{\sim} \text{Bernoulli}(1/J)$ implies that guessers distribute their guesses evenly over all possible response options. (For example, a guesser confronting a true/false question has a 50% chance of answering correctly, and increasing the number of response options always decreases the guesser's chance of answering correctly.) This assumption is consistent with Luskin and Bullock (2005), who investigate guessing in response to political knowledge questions.

This model captures three important aspects of political reality that are not captured by most studies of political knowledge. First, it acknowledges that responses to factual questions about politics are affected not only by knowledge but by guessing and partisan responding. Second, it allows that the influence of partisanship varies across issues and people, such that one may give a partisan response to a question about (say) health care but a sincere answer to a question about immigration, while another member of the same party may answer the health-care question sincerely and the immigration question in a partisan way. Third, it allows that partisan responding can cause people to answer correctly as well as incorrectly. For example, an ignorant Democrat asked about the budget deficit under Bill Clinton may answer correctly in spite of his ignorance because he wants to cast Clinton in a favorable light, and because the answer that achieves this end happens to be the correct answer.

The most critical assumption in the model is that the treatment completely eliminates partisan responding: no matter how partisan they are, treated subjects answer correctly if they

know the answer and guess if they do not. The assumption is strong, and in the following sections, we test it and consider the implications of violating it. A second, subtler assumption is that members of both parties share the same understanding of “favorable” and “unfavorable” outcomes and thus agree on which party is favored or disfavored by any particular political outcome. Of course, this assumption will not always hold. The existence of such cases does not invalidate the model, but it does restrict its scope.

The model’s main virtue is that it gives us a principled way of estimating political knowledge in spite of problems caused by guessing and partisan responding. Specifically, it implies that the proportion of people who know the answer to any question j is

$$E_i[x_{ij}] = E_i[y_{ij} | t_i = 1] - E_i[(1 - y_{ij}) | t_i = 1] \left[\frac{1}{J - 1} \right]. \quad (2)$$

We can estimate this proportion separately for Democrats and Republicans. In turn, we can estimate the true knowledge gap between the parties for any question of interest and compare it to the gap that we might observe under ordinary conditions—for example, a control group or a political poll.⁴

Experiment 1

Subjects were asked five questions about politics. They were given 20 seconds to answer each question. Some subjects were assigned to receive an entry in a lottery for each correct answer. Other subjects were offered no incentive to answer correctly.

⁴By adding further assumptions, we might also estimate other quantities of interest. For example, we might estimate the proportion of people in each party who are partisan responders for a given question. To date, our efforts along these lines have produced implausible estimates, suggesting that the additional assumptions are flawed. We therefore focus in this draft on estimating the actual knowledge gap between the parties and comparing it to the ordinary (i.e., control-group) gap in correct responding.

Participants, Design, and Procedure

Experiment 1 was embedded within the pre-election wave of the 2008 Cooperative Congressional Election Study, an Internet survey of U.S. citizens that was fielded in October 2008. The CCES sample closely matched the population of adult U.S. citizens on age, race, gender, education, income, and other demographic variables (Ansolabehere 2009, 7-17). Six hundred and twenty-six members of the sample were randomly assigned to participate in our experiment. These subjects were randomly assigned to a control group ($N = 312$) or a treatment group ($N = 314$). A randomization check indicated that assignment to these groups was independent of observed pre-treatment covariates: see Table A2 of the appendix.⁵

Control-group subjects were told that they would see a series of 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 also told

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 the instructions, all subjects were asked five closed-ended factual questions. Two were about U.S. fatalities in Iraq, two were about the ages of Barack Obama and John McCain, and one was about the level of national unemployment “compared to January 2001, when President Bush first took office.” Table 1 reports the text of each question.⁶

⁵ Our experiment also included a “score” condition to which 267 other subjects were assigned. Subjects in this condition were not given any financial incentive to answer questions correctly, but they were told that we were tracking their performance. Our data suggest that this treatment induces correct responding to a much lesser extent than the prize treatment. For simplicity, we omit further discussion of the score condition, focusing instead on the contrast between subjects in the control and prize conditions.

⁶ We also asked all subjects about Bush’s job approval ratings and about the deficit and inflation during the Bush presidency. We omit the inflation question from our analysis because it

Questions	Control-group partisan gap favors
Compared to January 2001, when President Bush first took office, has the level of unemployment in the country <i>increased</i> , stayed about the same, or decreased?	Democrats
About how many U.S. soldiers have been killed in Iraq since the invasion in March 2003? [Response options: <i>4,000</i> , 8,000, 12,000, 16,000, 20,000.]	Republicans
Was the number of U.S. soldiers killed in Iraq in the first half of 2008 <i>lower</i> , about the same, or higher than the number who were killed in the second half of 2007?	Republicans
How old is Barack Obama? [Response options: 37, 42, <i>47</i> , 52.]	Democrats
How old is John McCain? [Response options: 62, 67, 72, <i>77</i> .]	Republicans

Table 1: Factual Questions in the Experiment. Each question was put to every subject in the experiment. Questions were presented in the order listed here. Italicized response options were coded as correct; other options were coded as incorrect. The second column indicates which party answered correctly more often in the control group.

All subjects were asked about the year of Bangladeshi independence, the political offices held by George von L. Meyer, and the price of gold on January 18, 1980. We presumed that almost no one would be able to answer these questions correctly without looking up the answers, and we included them to determine how many subjects were in fact looking up the answers. Fewer than 1% of subjects in any condition answered any of these questions correctly, suggesting that almost no one was doing research before answering our questions.

A common concern about factual questions in a political survey is that they may be preceded by items that prime subjects' partisanship or emotions and thereby distort subjects'

was ambiguous: whether an answer was correct or incorrect depended on which Bureau of Labor Statistics definition of inflation one used. (See <ftp://ftp.bls.gov/pub/special.requests/cpi/cpiiai.txt>.) We omit the other items because they failed to generate partisan gaps in correct responding in the control group. Our analysis concerns questions that generate such gaps under ordinary conditions; we have little to say about items that do not meet this criterion.

answers to the factual questions. For example, factual questions about economic trends in the 1988 ANES were immediately preceded by items that asked subjects about their emotional reactions to Ronald Reagan (Bartels 2002, 134n30). Such items may lead subjects to answer subsequent factual questions as they would not otherwise do. But no such items appeared in the CCES: our factual items were immediately preceded by questions about the subjects' personalities and perceptions of ballot secrecy. Subjects' party identification was measured at the beginning of the interview with a standard seven-category branching item; more than 40 items intervened between the party-ID question and the beginning of our experiment. Most importantly, our model explicitly accounts for the possibility that people may respond in a "partisan way" to factual questions—either because they have been primed or for other reasons.

Results

Figure 1 reports the average proportions of correct answers in each condition. It shows that the treatment had little overall effect. In each condition, subjects answered 64% of questions correctly and the median number of correct responses was 3. The middle panels of Figure 1 show that the treatment had a weak effect on Democrats (raising correct-response rates from 63% to 64%) and none on Republicans (66% correct in both conditions). The largest effect appears among independents, who were substantially *less* likely to answer correctly when offered an incentive. Control-group independents answered 61% of questions correctly, against only 53% in the prize condition. The difference is statistically insignificant ($p = .26$, two-tailed), but it is quite contrary to our expectations, and we investigate it below.⁷

The average results in Figure 1 reflect responses to all items in our study. But following our model, we expect the treatment to have positive effects only on those questions that members of the party were less likely to answer correctly in the control group. Figure 2 reports the average

⁷ When we also consider the questions for which we observed no control-group partisan gap—questions about Bush's approval rating and changes in the deficit under Bush—the gap for independents widens to 10% and $p = .10$, two-tailed.

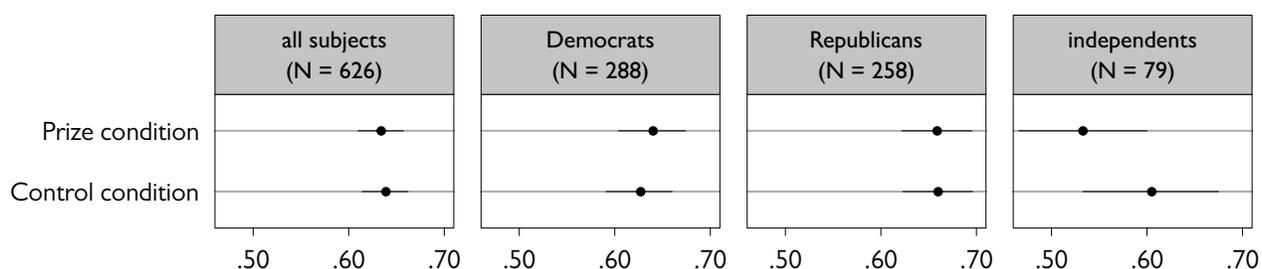


Figure 1: Proportions of Subjects Answering Correctly. Each panel plots the average proportion of correct answers in the control and prize conditions. Black lines are 95% confidence intervals.

The results show that the treatment had little overall effect on partisans. Contrary to our expectations, the largest effects were found among independents: incentives to answer correctly made independents more likely to answer *incorrectly*.

percentages correct in each condition when we consider only these items: for Democrats, we consider the item about John McCain’s age and the two items about Iraq; for Republicans, we consider the items about unemployment and Obama’s age. For both Democrats and Republicans, the results are in the expected direction: treated subjects are more likely than control-group subjects to answer correctly. Control-group Democrats answer correctly 60% of the time while their treated counterparts answer correctly 64% of the time ($p = .16$). Among Republicans, the percentages are 50% and 55% ($p = .13$).

PARTISAN DIFFERENCES IN CORRECT RESPONDING

Our chief interest lies with partisan gaps in correct responding and in knowledge. Figure 3 is instructive on this topic. Its left-hand panel reports, for each item, the partisan gaps (Democratic % correct – Republican % correct) in the control and treatment conditions. It also reports the partisan gaps in actual knowledge that we estimate via Equation 2. There is wide variation in all of these gaps from item to item, but the most telling parts of the panel are the summary rows at the bottom. Averaging across all five items, they show that the mean treatment-group gap (9%) is scarcely half the size of the mean control-group gap (16%; for the difference, $p = .02$). But the

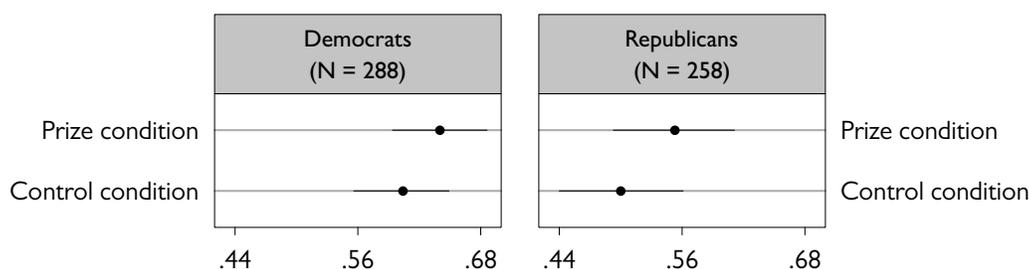


Figure 2: Proportions of Subjects Answering Correctly: “Observed Partisan Bias” Items Only. Each panel plots the average proportion of correct answers in the control and prize conditions. Black lines are 95% confidence intervals.

For each party, the averages are computed for only those items that the party was less likely to answer correctly in the control group. Thus, the averages for Democrats in the left-hand panel are taken over the two questions about Iraq and the question about John McCain’s age. And the averages for Republicans in the right-hand panel are taken over the questions about unemployment and Obama’s age. (See Table 1 for the text of each question.)

For both Democrats and Republicans, the results are in the expected order: treated subjects are more likely to answer correctly than their control-group counterparts. Among Democrats, 60% of control-group subjects and 64% of treated subjects answer correctly ($p = .16$). Among Republicans, 50% of control-group subjects and 55% of treated subjects answer correctly ($p = .13$).

mean estimated knowledge gap, 13%, is only insignificantly different from the control-group gap in correct responding ($p = .30$).

The second panel of Figure 3 casts the connection between responding and knowledge in a slightly different light. It does so by reporting

$$1 - \frac{\text{estimated partisan gap in knowledge}}{\text{control-group partisan gap in correct responding}} \tag{3}$$

for each item. This quantity equals 1 when there is no estimated knowledge gap between the parties. It equals 0 when the estimated knowledge gap is the same as the observed gap in correct responding. And in practice, Figure 3 shows that it is typically closer to 0 than to 1. Averaging over the five items, it is only .21 ($p = .16$). Further inspection of Figure 3 shows that the corresponding ratio for the treatment group,

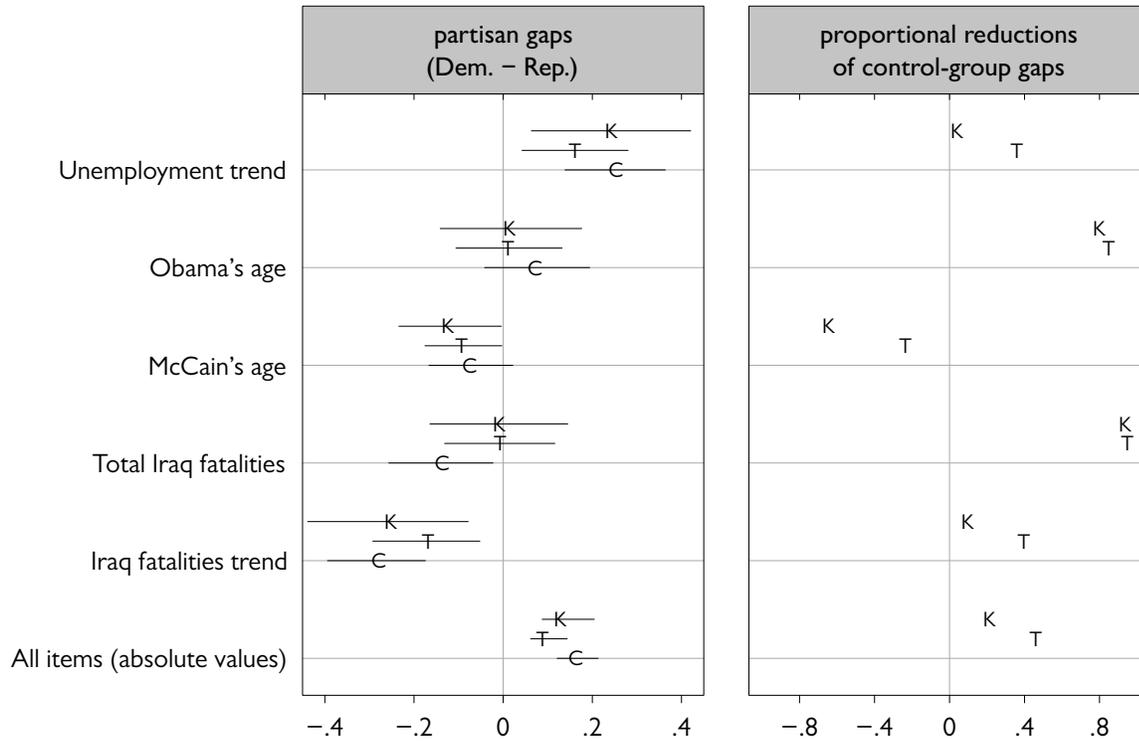


Figure 3: Partisan Gaps in Correct Responding. The left-hand panel plots the difference between the proportions of correct answers given by Democrats and Republicans. Differences are plotted for the control- and treatment-group gaps in correct responding (C and T). Differences are also plotted for the gaps in actual knowledge between the parties (K). These differences are estimated via Equation 2. In all cases, positive values indicate that Democrats answered correctly more often (or knew more) than Republicans. Negative values indicate that Democrats answered correctly less often (or knew less) than Republicans. Black lines are 95% confidence intervals. The exact wording of each item is given in Table 1.

The right-hand panel represents the treatment-group gap in correct responding and the estimated knowledge gap on the scale of the control-group gap in correct responding. (See expressions (3) and (4).) The plotted quantities equal 1 when they are equal to the control-group gap. The second-to-bottom row in the panel shows that, on average, the estimated knowledge gap was only 21% smaller than the control-group gap in correct responding.

$$1 - \frac{\text{treatment-group partisan gap in correct responding}}{\text{control-group partisan gap in correct responding}}, \quad (4)$$

is larger for all five items. The treatment-group results reflect diminished partisan responding (because treated subjects were paid to give correct answers), but they do not adjust for guessing. Comparison of terms (3) and (4) for any particular item therefore suggests that the control group gap is better than the treatment group at reflecting the actual gap in knowledge. The general lesson is that trying to estimate a gap in knowledge by accounting for partisanship but not guessing may be worse than trying to estimate a gap in knowledge without accounting for either confound.

Experiment 1 therefore suggest that ordinary gaps in correct responding owe more to real partisan differences in knowledge than to “partisan cheerleading.” But the results are far from dispositive. In particular, Experiment 1 contained only five items, and a wider range of items might have yielded different results. Experiment 2 addresses this concern.

Experiment 2

Subjects were asked 10 questions about politics. They were given 35 seconds to answer each question. Some subjects were randomly assigned to receive \$0.50, \$1, \$1.50, or \$2 for each correct answer. The order of questions in the battery and the order of response options for particular questions were also randomized.

Participants, Design, and Procedure

One thousand, five hundred and fifty-two American adults were recruited through Amazon Mechanical Turk to participate in a study of “what people know about politics.” The study was

estimated to take eight minutes. All subjects were paid \$0.75 for agreeing to participate. The study was fielded from September 2, 2010 through October 18, 2011.⁸

Four hundred and seventy-two subjects were randomly assigned to a control group. One thousand and eighty subjects were randomly assigned with equal probability to receive \$0.50, \$1, \$1.50, or \$2 for each correct answer to a set of questions. They were given 35 seconds to answer each question. The appendix includes the text of the instructions given to subjects (page A13), the distribution of subjects across parties and conditions (Table A3), and information about a randomization check which indicates that assignment to the treatments was independent of observed pre-treatment covariates (Table A4).

After receiving the instructions, all subjects were asked ten factual questions. Six were about changes in unemployment, inflation, and the federal budget deficit under Presidents Obama, Clinton, and Reagan. Two others were about American's recent wars: one about weapons of mass destruction in Iraq, another about changes ordered by Obama to the number of U.S. troops in Afghanistan. The remaining questions were about Obama's position on gay marriage and the proportion of people who would be affected by a federal estate tax. Except for the estate-tax question, all questions were closed-ended. The text of each question is given in Table 2.⁹

As in Experiment 1, all subjects were also asked about the year of Bangladeshi independence, the political offices held by George von L. Meyer, and the price of gold on January 18, 1980. We presumed that almost no one would be able to answer these questions

⁸ An additional 138 subjects were recruited but failed a prescreening task intended to filter out survey-taking software programs and subjects who were not paying minimal attention. The task asked subjects to read two short news articles; subjects who spent fewer than eight seconds doing so were not invited to take the study. The prescreening task was given before treatment was assigned and it therefore does not affect the randomizations in our analysis. The text of the prescreening articles appears in the appendix; see Berinsky and Lenz (2011) for more on the use of prescreening tasks in survey experiments.

⁹ We also asked all subjects about the total value of President Obama's assets, including his property and his investments. There was no partisan gap in correct responding to this item in the control group, and we therefore omit it from our analysis.

Questions	Control-group partisan gap favors
<p>Which of the following is closest to President Obama’s position? [Response options: The Constitution should be amended to explicitly permit gay marriage, Gay marriage should be allowed, <i>Gay marriage should not be allowed, but other kinds of legal partnership should be encouraged</i>, Gay marriage should not be allowed, The Constitution should be amended to explicitly prohibit gay marriage]</p>	Democrats
<p>“Weapons of mass destruction” (WMD) are chemical, biological, or nuclear weapons that can kill many people at once. Just before the U.S. invasion of Iraq in 2003, Iraq had . . . [working WMD; no working WMD, but an active program for developing them; <i>neither WMD nor an active program for developing them</i>]</p>	Democrats
<p>In 2011, there will be a federal estate tax—a tax on money that people leave to others when they die. If current law does not change, about what percentage of people who die in 2011 will leave enough to others for the federal estate tax to kick in? [Subjects were asked to type an answer between 0 and 100.]</p>	Democrats
<p>Since taking office in January 2009, how has President Obama changed the number of U.S. soldiers in <i>Afghanistan</i>? [reduced troop levels, no significant changes, <i>increased troop levels</i>]</p>	Republicans
<p>Compared to January 2009, when President Obama first took office, how has the level of <i>unemployment</i> in the country changed?’ [decreased, stayed the same, <i>increased</i>]</p>	Republicans
<p>Compared to January 2009, when President Obama first took office, how has the <i>federal budget deficit</i> changed? [decreased, stayed about the same, <i>increased</i>]</p>	Republicans
<p>Consider the presidency of Ronald Reagan, which began in January 1981 and ended in January 1989. Between the start and the end of his presidency, how did the level of <i>inflation</i> in the country change? [<i>decreased</i>, stayed about the same, increased]</p>	Republicans
<p>Consider the presidency of Ronald Reagan, which began in January 1981 and ended in January 1989. Between the start and the end of his presidency, how did the <i>federal budget deficit</i> in the country change? [decreased, stayed about the same, <i>increased</i>]</p>	Democrats

Questions	Control-group partisan gap favors
Consider the presidency of Bill Clinton, which began in January 1993 and ended in January 2001. Between the start and the end of his presidency, how did the level of <i>inflation</i> in the country change? [<i>decreased</i> , stayed about the same, increased]	Democrats
Consider the presidency of Bill Clinton, which began in January 1993 and ended in January 2001. Between the start and the end of his presidency, how did the <i>federal budget deficit</i> in the country change? [<i>decreased</i> , stayed about the same, increased]	Democrats

Table 2: Factual Questions in Experiment 2. Each question was put to every subject in the experiment. Questions were presented in the order listed here. Italicized response options were coded as correct; other options were coded as incorrect. Entries in the “partisan gap favors” column indicate which party answered correctly more often in the control group.

correctly without looking up the answers, and we included them to determine how many subjects were in fact looking up the answers. Fewer than 1% of subjects answered the question about gold correctly, but 7% of subjects correctly answered the question about Meyer, and 9% correctly answered the question about Bangladesh. 13% of Democrats and 12% of Republicans answered at least one of these questions correctly ($p = .77$). But having measured these responses, we can control for them in our analyses. Most importantly, for our purposes, correct responding to these questions does not differ significantly across parties, suggesting that research effort does not differ significantly across parties, either.

Results

Figure 4 reports the average proportions of correct answers in each payment condition. The bottom two rows of each panel show that paid subjects answer correctly 3% more often than unpaid subjects ($p = .01$). The same 3% result holds among Democrats alone ($p = .07$), Republicans alone ($p = .10$), and independents alone ($p = .12$).

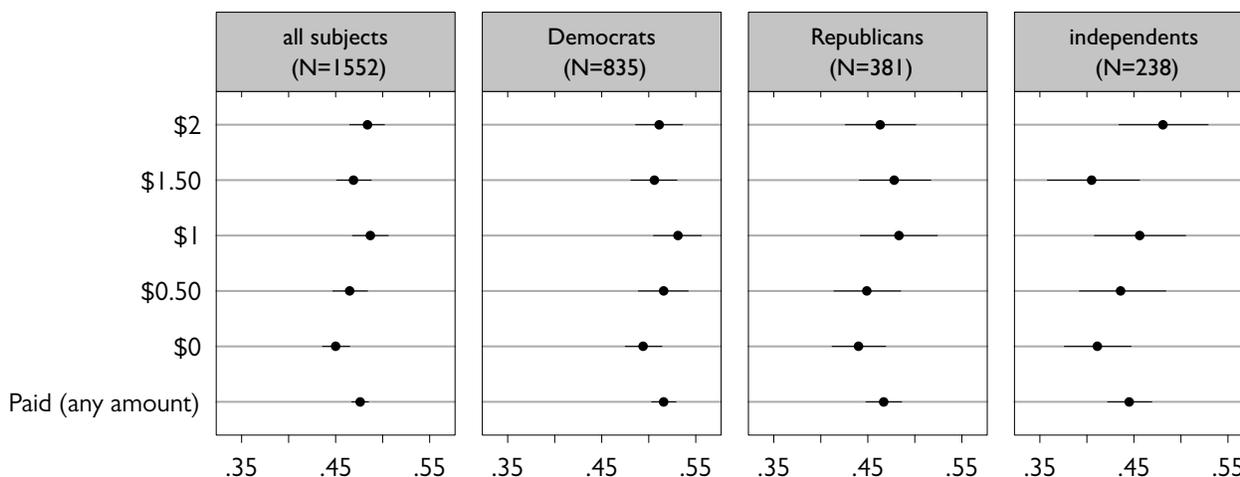


Figure 4: Proportions of Subjects Answering Correctly in Experiment 2. Each panel plots the average proportion of correct answers in the control and treatment conditions. Black lines are 95% confidence intervals. Some subjects did not answer the party ID question; these subjects’ answers are reflected in the first panel.

As in Experiment 1, these overall results may obscure important effects among Democrats and Republicans by averaging over items on which no partisan bias should be expected. We refine the analysis with Figure 5, which reports the average percentages correct in each payment condition when we consider only these items. For Democrats, we consider the questions about troop levels in Afghanistan, the deficit and unemployment under Obama, and inflation under Reagan. For Republicans, we consider the questions about the estate tax, Obama’s stand on gay marriage, weapons of mass destruction in Iraq, changes in the deficit under Reagan and Clinton, and changes in inflation under Clinton.

Two patterns stand out in Figure 5. The first is that incentives have a substantial six-percentage-point effect on correct-response rates among members of each party when we narrow our analysis to the items on which partisan bias should be expected. Democrats answered such questions correctly 50% of the time when unpaid, 56% when paid some amount of money per correct answer ($p = .001$). The corresponding percentages for Republicans are 28% and 34% ($p = .006$).

The second and more striking result is that, conditional on being paid, the *amount* paid per correct answer matters to Republicans but not to Democrats. The right-hand panel of Figure 5 reveals a clear linear trend: the more that Republicans are paid, the more likely they are to answer correctly. Correct-response rates rise from 32% among Republicans who are paid 50 cents per correct answer to 37% among those who are paid \$2 per correct answer. This difference is only marginally statistically significant ($p = .14$), perhaps owing to the relatively small number of Republicans in the study. But the regression estimates in Table 3 offer stronger evidence.

Table 3 reports OLS estimates from four models:

$$\text{sum of correct answers} = \beta_0 + \beta_1(\text{paid}) + \beta_2(\text{amount paid}) + \epsilon, \quad (5)$$

$$\text{sum of correct answers} = \beta_0 + \beta_1(\text{paid}) + \beta_2(\text{amount paid}) + \beta\mathbf{X} + \epsilon, \quad (6)$$

$$\begin{aligned} \text{sum of correct answers} = \beta_0 + \beta_3(\text{paid } \$0.50) + \beta_4(\text{paid } \$1) + \beta_5(\text{paid } \$1.50) \\ + \beta_6(\text{paid } \$2) + \epsilon, \text{ and} \end{aligned} \quad (7)$$

$$\begin{aligned} \text{sum of correct answers} = \beta_0 + \beta_3(\text{paid } \$0.50) + \beta_4(\text{paid } \$1) + \beta_5(\text{paid } \$1.50) \\ + \beta_6(\text{paid } \$2) + \beta\mathbf{X} + \epsilon. \end{aligned} \quad (8)$$

Each model is estimated twice: once for Democrats and once for Republicans. “Sum of correct answers” ranges from 0 to 4 for Democrats, 0 to 6 for Republicans. “Paid” equals 0 if subjects were in the control group, 1 if they were paid any positive amount per correct answer. “Amount paid” ranges from 0 to 2 in increments of .5. “Paid \$0.50,” “paid \$1,” “paid \$1.50,” and “paid \$2” are scored 0 or 1 for each subject depending on the payment condition to which he was assigned. \mathbf{X} is a matrix of control variables that includes general political knowledge (the subject’s score on a battery of seven general political knowledge items), a dummy variable indicating whether the subject answered at least one of the “placebo” questions correctly, and dummy variables measuring sex, level of education, age, age squared, and region of residence (Northeast, Midwest, South, and West, as defined by the Census Bureau).

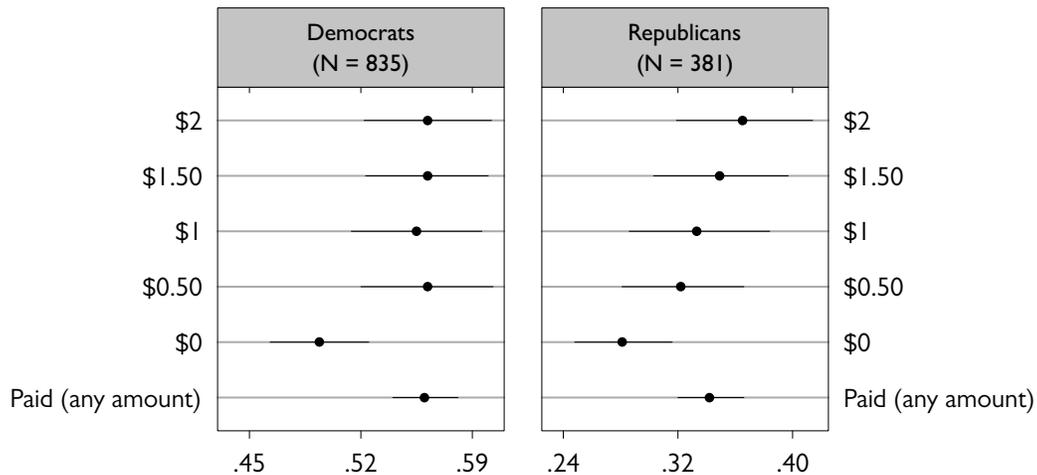


Figure 5: Proportions of Subjects Answering Correctly: “Observed Partisan Bias” Items Only. Each panel plots the average proportion of correct answers in the control and treatment conditions. Black lines are 95% confidence intervals.

The averages in each panel are computed only for items on which we expected to observe a partisan response bias. Thus, the averages for Democrats in the left-hand panel are taken over the questions about troop levels in Afghanistan, the deficit and unemployment under Obama, and inflation under Reagan. And the averages for Republicans in the right-hand panel are taken over the questions about the estate tax, Obama’s stand on gay marriage, weapons of mass destruction in Iraq, changes in the deficit under Reagan and Clinton, and changes in inflation under Clinton. (See Table 2 for the text of each question.)

For both Democrats and Republicans, the effect of being paid is as expected: it induces correct responding. Among Democrats, 49% of unpaid subjects and 56% of paid subjects answer correctly ($p = .001$). Among Republicans, 28% of control-group subjects and 34% of treated subjects answer correctly ($p = .01$). But disaggregating these results by amount of payment per correct answer reveals a sharp partisan difference: Republicans’ performance improved with every 50-cent increment in the amount offered, while Democrats’ performance did not vary according to the amount paid (so long as some amount was paid). See Table 3 for further analysis.

	Democratic subjects				Republican subjects											
Intercept	1.98	.07	.97	.29	1.98	.07	.98	.29	1.68	.13	1.02	.61	1.68	.13	.94	.61
Paid	.26	.13	.25	.12					.16	.24	.13	.22				
Amount paid	.00	.08	.00	.07					.17	.15	.18	.14				
Paid \$0.50					.27	.11	.26	.11					.25	.20	.30	.19
Paid \$1.00					.24	.11	.27	.10					.32	.22	.13	.21
Paid \$1.50					.27	.11	.20	.10					.41	.21	.39	.20
Paid \$2.00					.27	.11	.29	.10					.51	.21	.53	.20
General political knowledge			1.00	.13			1.00	.13			1.41	.28			1.44	.29
Answered placebo correctly			-.05	.10			-.06	.10			.26	.21			.27	.21
Female			-.36	.07			-.37	.07			-.56	.14			-.56	.14
Some college			.02	.12			.02	.12			.26	.21			.24	.21
B.A.			.11	.13			.12	.13			.33	.22			.33	.22
Postgraduate degree			.14	.14			.14	.14			.57	.26			.57	.26
Age			.01	.01			.01	.01			.00	.03			.01	.03
Age ²			.00	.00			.00	.00			.00	.00			.00	.00
Midwest			.29	.10			.30	.10			-.29	.22			-.31	.22
South			.16	.10			.16	.10			-.21	.21			-.21	.21
West			.29	.10			.28	.10			-.21	.24			-.23	.24
R ²	.01		.18		.01		.18		.02		.18		.02		.19	
Standard error of regression	1.06		.97		1.06		.97		1.36		1.26		1.37		1.27	
Number of observations	835		830		835		830		381		377		381		377	

Table 3: Experiment 2 Knowledge Scores. Each column reports OLS estimates and standard errors. In the “Democratic subjects” regressions, the outcome is the sum of correct answers to the four questions that control-group Democrats were less likely to answer correctly than control-group Republicans. In the “Republican subjects” regressions, the outcome is the sum of correct answers to the six questions that control-group Republicans were less likely to answer correctly than control-group Democrats. See Table 2 for the text of each question and page 20 for descriptions of the predictors.

These regressions provide stronger evidence for a linear effect of amount of payment on Republican correct-response rates. In our preferred specification, Equation 6, increasing the amount that Republicans are paid per correct answer by \$1 increases the mean number of correct responses by .18, or 3% of the range of the scale ($p = .10$). The Table 3 also strongly suggest that, conditional on being paid, the amount that Democrats are paid does not make a difference to correct response rates. Democrats receiving fifty cents per correct answer answered correctly 56.3% of the time; those paid \$2 per correct answer answered correctly 56.2% of the time. The results in Table 3 also confirm that the amount that Democrats are paid makes approximately no difference to their answers: the estimated effect of amount of payment is .00 under specifications 5 and 6, and under specifications 7 and 8, there is neither a substantive nor a statistically significant difference between any of the different payment conditions.

This difference between Democrats and Republicans is not just interesting in itself: it suggests differences in our ability to draw inferences about the extent to which survey responses are influenced by “partisan cheerleading.” For Democrats, the results suggest that incentives greater than \$2 per correct answer would be no more effective at inducing truthful responses. We may therefore be able to speak with relative precision about the extent to which Democrats’ answers to factual questions are sincere and the extent to which they are informed by partisan cheerleading. But the results for Republicans suggest that we cannot be as precise when evaluating their responses. We return to this idea below.

PARTISAN DIFFERENCES IN CORRECT RESPONDING

Table 3 cannot speak to the way in which incentives affect the partisan gap in correct responding, but Figure 6 does. It shows that incentives reduced the magnitude of the partisan gap in correct responding on nine of the ten questions in Experiment 2. The exception is the question about Obama’s stand on gay marriage, and here, the effect of incentives is trivial: among control-group subjects, the Democratic correct-response rate was higher than the Republican correct-response rate by 9.4%; among paid subjects, the Democratic advantage rose to 9.5%.

But—and more importantly—Figure 6 also suggests that actual knowledge gaps between the parties correspond rather closely to the control-group gaps in correct responding. Averaging over all ten items, the mean absolute control-group gap in correct responding between the parties is 20%. The estimated absolute knowledge gap, 13%, is smaller ($p = .16$), but only by about one third. As in Experiment 1, these results suggest that the partisan gaps that appear in response to ordinary survey questions—represented here by the gaps in responses to control-group questions—may be surprisingly good indicators of actual gaps in knowledge.

Figure 6 considers a single set of “treated” subjects—those who were offered any amount of money for correct answers. One may therefore wonder whether it elides differences between the different treatment levels. For example, perhaps the \$2-per-correct-answer treatment is especially successful at eliminating partisan responding. If this is so—and Figure 5 suggests that it may be—Figure 6 overstates the actual gap in knowledge between the parties. But analyses in the appendix suggest that focusing on those who received the maximum treatment (\$2 per correct answer) makes only a modest difference. In particular, Figure A9 shows that the estimated mean knowledge gap in this condition is still more than 60% of the mean control-group gap.

Conclusion

Partisan differences in responses to factual questions about politics are attracting increasing attention from both political scientists and the news media. But observed partisan gaps appear chiefly for questions on which the correct answer casts a party in a positive or negative light. This invites the suspicion that partisan gaps in correct responding are due to partisan motivations rather than sincere differences in beliefs. Little research has spoken to this possibility.

Our experiment is consistent with the suggestion that partisan gaps are due chiefly to sincere differences of belief between members of different parties. Telling subjects that their chances of winning a \$200 prize would be increased by answering correctly had real but modest effects: it increased the percentages of partisans responding correctly by four to five percentage points on questions for which partisanship might have deterred subjects from answering correctly.

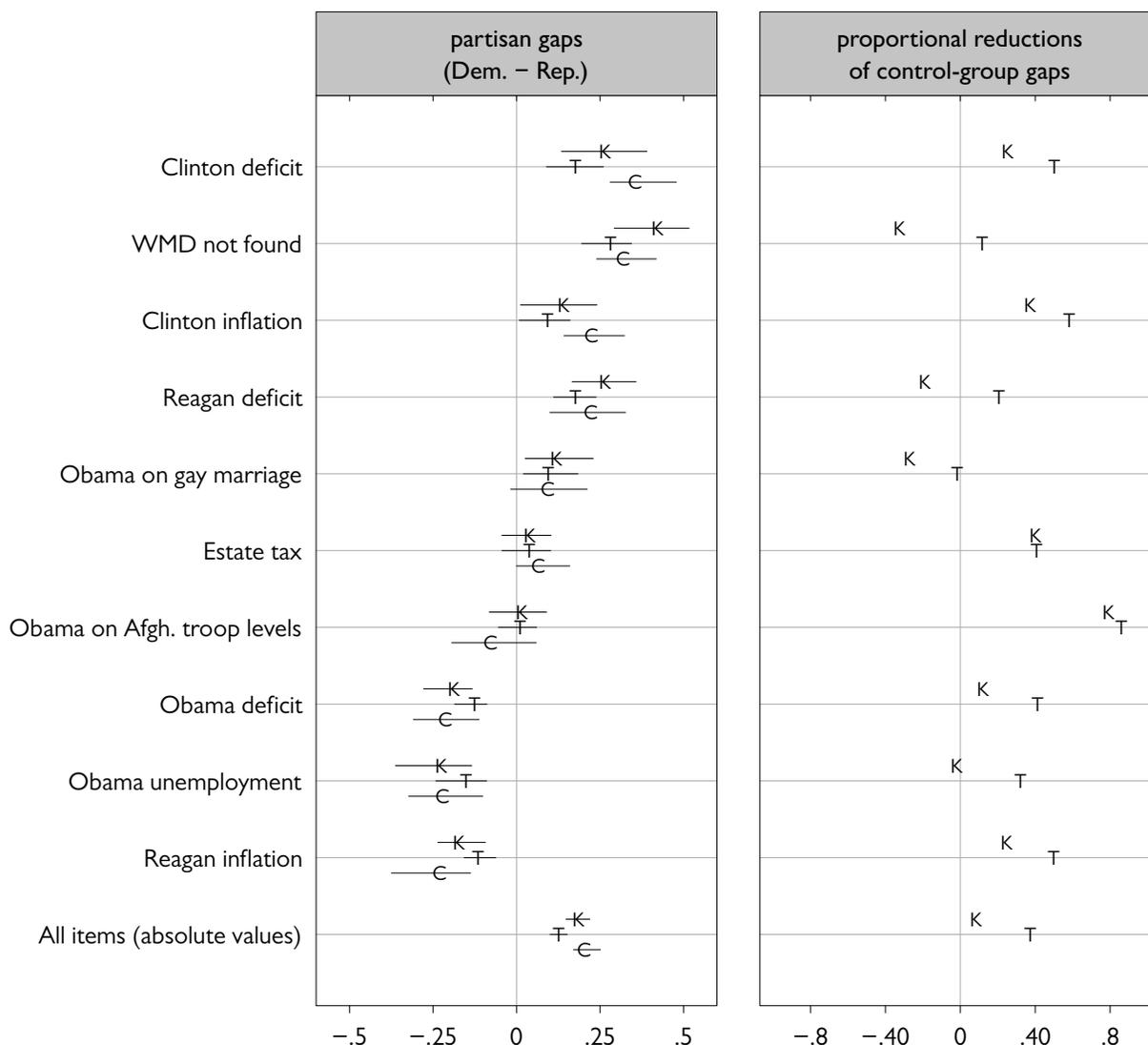


Figure 6: Partisan Gaps in Correct Responding in Experiment 2. The left-hand panel plots the difference between the proportions of correct answers given by Democrats and Republicans. Differences are plotted for the control- and treatment-group gaps in correct responding (C and T). Differences are also plotted for the gaps in actual knowledge between the parties (K). These differences are estimated via Equation 2. In all cases, positive values indicate that Democrats answered correctly more often (or knew more) than Republicans. Negative values indicate that Democrats answered correctly less often (or knew less) than Republicans. Black lines are 95% confidence intervals. The exact wording of each item is given in Table 1.

For the purposes of this figure, the “treatment group” includes all subjects who were offered money in exchange for correct responding. See the appendix for a related figure that accounts for only those subjects who received the maximum treatment, \$2 per correct answer.

The right-hand panel represents the treatment-group gap in correct responding and the estimated knowledge gap on the scale of the control-group gap in correct responding. (See expressions (3) and (4).) The plotted quantities equal 1 when they are equal to the control-group gap. The second-to-bottom row in the panel shows that, on average, the estimated knowledge gap was only 35% smaller than the control-group gap in correct responding.

(See Figure 2.) Offering subjects fixed amounts of money for correct answers had slightly greater effects, but even at their most effective, they reduced the average partisan gap in correct responding by only 50%.

Most importantly, estimated gaps in knowledge between the parties are, in general, close to the observed gaps in correct responding that we obtain from our control groups. This is surprising. Our control-group analyses, like analyses of almost all ordinary surveys, make no effort to account for the influences of either guessing or partisan cheerleading. The results in both experiments suggest that naive control-group estimates of knowledge are inflated by partisan responding. But guessing works in the opposite direction, causing the control-group estimates to understate the gaps in knowledge between the parties. Accounting for both of partisan cheerleading and guessing thus produces estimates of knowledge that are quite close to the control-group gaps in correct response.

These results are far from dispositive, and they suggest other experiments that might be conducted. Stronger incentives might be offered: perhaps the incentives that we offered were simply too weak to diminish the partisan gaps that we observed. Other questions might be asked: the partisan gaps that have attracted recent media attention are typically very large, and the gaps for such questions may be especially susceptible to our treatments. And the motivation to answer correctly or incorrectly might be manipulated in other ways, e.g., by priming accuracy or honesty for some subjects before asking them factual questions about politics. If such manipulations increased accuracy among partisans, it would suggest that their errors in control conditions and ordinary surveys are not entirely due to sincere differences in beliefs.

References

- ABC News/The Washington Post. 2004. ABC News/Washington Post Weapons of Mass Destruction Poll, February 2004 [computer file] (Study 4034). ICPSR version. Horsham, PA: Taylor Nelson Sofres Intersearch [producer], 2004. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2004.
- ABC News/The Washington Post. 2005. ABC News/Washington Post Poll, March 2005 [computer file] (Study 4319-v1). ICPSR version. Horsham, PA: Taylor Nelson Sofres Intersearch [producer], 2005. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005.
- Ansolabehere, Stephen. 2009. "Guide to the 2008 Cooperative Congressional Election Survey." http://web.mit.edu/polisci/portl/cces/material/CCES_Guide_2008_Rough_Draft_v2.pdf (accessed March 26, 2010).
- Bartels, Larry M. 2002. "Beyond the Running Tally: Partisan Bias in Political Perceptions." *Political Behavior* 24 (June): 117-50.
- Berinsky, Adam J. 2005. *Silent Voices: Public Opinion and Participation in America*. Princeton, NJ: Princeton University Press.
- Cooper, Michael. 2010 October 18. "From Obama, the Tax Cut Nobody Heard Of." *New York Times*: A1.
- Cragg, John G., and Russell S. Uhler. 1970. "The Demand for Automobiles." *Canadian Journal of Economics* 3 (August): 386-406.
- Delli Carpini, Michael X., and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters*. New Haven, CT: Yale University Press.
- Duelfer, Charles. 2004. "Comprehensive Report of the Special Advisor to the Director of Central Intelligence on Iraq's Weapons of Mass Destruction." <http://www.gpoaccess.gov/duelfer/index.html> (accessed October 1, 2009).
- Harris Interactive. 2006 July 21. "Belief that Iraq Had Weapons of Mass Destruction Has Increased Substantially." <http://news.harrisinteractive.com/profiles/investor/ResLibraryView>.

[asp?ResLibraryID=34134&GoTopage=4&Category=1777&BzID=1963&t=8](http://www.harrisinteractive.com/ResLibraryView.asp?ResLibraryID=34134&GoTopage=4&Category=1777&BzID=1963&t=8).

Harris Interactive. 2010 March 24. "'Wingnuts' and President Obama." [http://news.](http://news.harrisinteractive.com/profiles/investor/ResLibraryView.asp?ResLibraryID=37050&Category=1777&BzID=1963)

[harrisinteractive.com/profiles/investor/ResLibraryView.asp?ResLibraryID=37050&Category=1777&BzID=1963](http://news.harrisinteractive.com/profiles/investor/ResLibraryView.asp?ResLibraryID=37050&Category=1777&BzID=1963).

Kuklinski, James H., Michael D. Cobb, and Martin Gilens. 1997. "Racial Attitudes and the 'New South'." *Journal of Politics* 59 (May): 323-49.

Lord, Frederic M., and Melvin R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Luskin, Robert C., and John G. Bullock. 2005. "Measuring Political Knowledge." University of Texas at Austin. Typescript.

Luskin, Robert C., and John G. Bullock. 2011. "'Don't Know' Means 'Don't Know': DK Responses and the Public's Level of Political Knowledge." *Journal of Politics*.

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

Noelle-Neumann, Elisabeth. 1993. *The Spiral of Silence: Public Opinion, Our Social Skin*. Chicago: University of Chicago Press.

Pew Research Center for People & the Press. 2010. "Growing Number of Americans Say Obama is a Muslim." http://pewforum.org/uploadedFiles/Topics/Issues/Politics_and_Elections/growingnumber-full-report.pdf.

PIPA/Knowledge Networks. 2004. "Separate Realities of Bush and Kerry Supporters." Questionnaire. <http://www.worldpublicopinion.org/pipa/pdf/oct04/IraqRealities%20Oct04%20quaire.pdf> (accessed April 16, 2010).

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

Rasch, Georg. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks Paedagogische Institut.

Rasch, George. 1966. "An Item Analysis which Takes Individual Differences into Account." *British Journal of Mathematical and Statistical Psychology* 19: 45-57.

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

U.S. Census Bureau. 1990. *Statistical Abstract of the United States: 1990*. 110th ed. Washington, D.C.: U.S. Government Printing Office.

Online Appendix to “Partisan Bias in Factual Beliefs about Politics”

John G. Bullock, Alan S. Gerber, and Gregory A. Huber

March 30, 2011

2008 CCES Experiment	A2
Instructions to Subjects	A2
Sample Size by Condition and Party ID	A3
Randomization Check	A4
Item-by-Item Results	A5
Overall Partisan Differences in Correct Responding	A7
Strong Partisans: Overall and Item-by-Item Results	A8
Low-Income Subjects: Overall and Item-by-Item Results	A10
2010 Mechanical Turk Experiment	A12
Prescreening Task	A12
Instructions to Subjects	A13
Sample Size by Condition and Party ID	A14
Randomization Check	A15
Overall Results	A16
Item-by-Item Results	A17

2008 CCES: Instructions to Subjects

Subjects were assigned with equal probability to the control group or the “prize” group. All subjects read

The next 13 questions will ask about your knowledge of politics. You will have 20 seconds to read and answer each question. After 20 seconds, you will automatically be forwarded to the next question.

Next, control-group subjects read

Your performance and score will not be shared with anyone else.

while “prize”-group subjects read

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. Your performance and score will not be shared with anyone else.

All subjects then read

Please answer every question. If you do not know the answer to a question, **please give your best guess.** The questions will begin in 45 seconds, or when you click the button at the bottom of the screen.

Experiment 1: Sample Size by Condition and Party ID

	Dem.	Ind.	Rep.	No Answer	Total
Control condition	148	37	126	1	312
Prize condition	140	42	132	0	314
Total	288	79	258	1	626

Table A1: Sample Size by Condition and Party ID in Experiment 1. Democrats are those who identified with the Democratic Party on a standard seven-point measure of party ID, including those who “lean” toward the Democratic Party. Republicans were coded in the same way.

Experiment 1: Randomization Check

Intercept	.56	.58
Age	.00	.01
High school	-.58	.43
Some college	-.83	.44
Two-year college	-.98	.54
Four-year college	-.79	.45
Postgraduate degree	-.84	.49
Female	.25	.17
Democrat	.13	.21
Republican	.10	.21
Registered voter	-.21	.33
Asian	-1.74	1.12
Black	.20	.36
Hispanic	-.34	.34
Middle Eastern	14.82	623.93
Mixed race	.68	.89
Native American	-.21	.69
Other	-.22	.60
Midwest	.12	.24
South	.15	.23
West	.21	.25
Log likelihood	-424	
Likelihood ratio test	18.72	($p = .54$)
Cragg and Uhler (1970) pseudo- R^2	.03	
Number of observations	625	

Table A2: Randomization Check for Experiment 1. Cell entries are parameter estimates and standard errors from a logistic regression in which assignment to treatment is regressed on pre-treatment variables. The baseline education category is “no high school.” The baseline racial category is “white.” The baseline region is the Northeast.

The “Likelihood ratio test” entry is a chi-square statistic for a comparison against an intercept-only model. It is statistically insignificant and the pseudo- R^2 value is low, suggesting that random assignment to the treatment conditions worked as intended.

Experiment 1: Item-by-Item Results

Average results can mask item-by-item variation in responsiveness to the treatments. Figure A1, which reports the percentages correct for each item, suggests little variation of this type in our experiment. The left-hand panel of the figure reports these percentages for all subjects, and it shows that the treatments made little overall difference in correct responding to any item. For three items, correct responses were most likely in the control group; for the other six items, correct responses were more likely in at least one of the treatment groups. The largest effects were found for the questions about Bush's approval rating among all Americans (92% of control-group subjects answered correctly against 87% of treated subjects, $p = .02$), Bush's approval rating among Republicans (18% against 23%, $p = .02$), and total American fatalities in Iraq (66% against 62%, $p = .11$). All other differences were smaller and did not approach statistical significance.

A similar message is conveyed by the middle panels of Figure A1, which report percentages of Democratic and Republican subjects answering correctly. As expected, partisans generally answered correctly more often when offered an incentive: for members of each party, correct responses were highest among treated subjects for seven of the nine questions. But the differences were generally slight. Treated Democrats were more likely to correctly answer the questions about Bush's approval rating among Republicans (16% vs. 24%, $p = .06$) and the trend in Iraq fatalities (46% vs. 52%, $p = .15$, but less likely to correctly answer the question about Obama's age (63% vs. 57%, $p = .06$). Treated Republicans were more likely to correctly answer questions about unemployment and inflation trends during the George W. Bush presidency (44% vs. 55%, $p = .04$ and 52% vs 60%, $p = .09$). Surprisingly, they were also less likely to correctly answer the question about total Iraqi fatalities (74% vs. 62%, $p = .02$). All other treatment effects were smaller and did not approach statistical significance.

One might expect that the treatment reduces the size of the partisan gap by inducing correct responses among those who were motivated by partisanship to give an incorrect answer. Inspection of Figure A1 shows that this is indeed why the treatment narrowed the partisan gap for

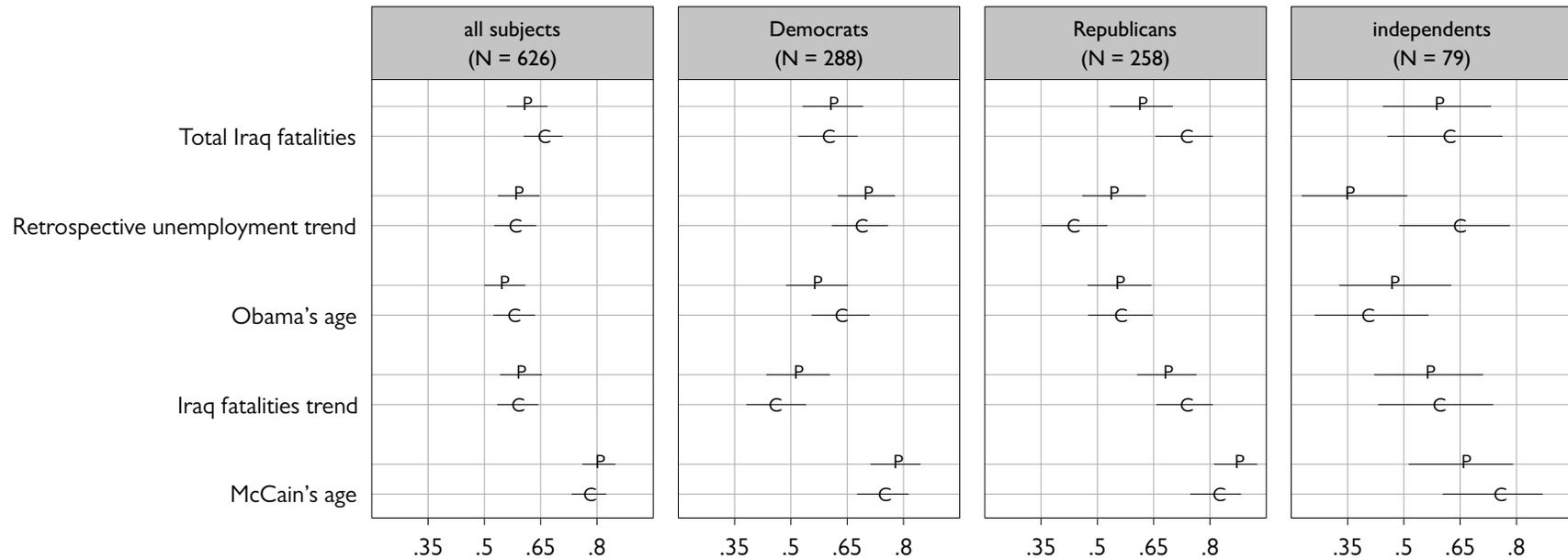


Figure A1: Proportions of Subjects Answering Correctly: Item-by-Item Results. Each panel plots the proportions of subjects answering correctly in the control (C) and prize (P) conditions. Black lines are 95% confidence intervals. The “retrospective trend” questions refer to trends during George W. Bush’s terms as President. The “fatalities” question refers only to deaths of U.S. soldiers. The exact wording of each item is given in Table 1.

Items are ordered according to the maximum of the absolute average treatment effects among Democrats and Republicans. Thus, “Total Iraq fatalities” appears at the top of each panel (treated Republicans were 12 percentage points less likely to answer correctly), and “Retrospective deficit trend” appears at the bottom of each panel (treated Republicans were only three percentage points more likely to answer correctly, and the effect was smaller among Democrats).

retrospective economic evaluations: it had little effect on Democrats but made Republicans more willing to give the correct, negative response. But different mechanisms appear to be at work for the questions about Iraq. For the question about total U.S. fatalities in Iraq, Democrats were the subjects whom we expected to answer incorrectly, and they were therefore the subjects whom we expected to be affected by the incentives. That is not what happened. Instead, Democrats were unaffected by the treatment, but as noted above, Republicans were made substantially *less* likely to answer correctly. And for the question about the trend in Iraq fatalities, Figure A1 shows that the treatment made Democrats more likely to answer correctly while simultaneously making Republicans less likely to answer correctly. Neither of these within-party effects is statistically significant, but they collectively produce an 11-percentage-point reduction in the partisan gap ($p = .08$).

Experiment 1: Overall Partisan Differences in Correct Responding

The left-hand panel of Figure A2 presents the overall differences: the average, taken over all items, of the partisan gap in correct responses. Formally, the panel plots

$$y_c = \left(\sum_i |D_{ci} - R_{ci}| \right) / 9,$$

where c indexes conditions (prize, control), i indexes the nine items in the experiment, and D_{ci} and R_{ci} are the percentages of Democrats and Republicans in condition c who correctly answer item i . The average partisan difference is 13% in the control condition, 8% in the prize condition ($p = .04$). This difference is in the expected direction: incentives reduce the average partisan gap in correct responding.

The middle and right-hand panels of Figure A2 show that the overall effect is generated primarily among weak partisans. For them, the average partisan gap is 15% in the control condition, 8% in the prize condition ($p = .03$). The corresponding percentages for strong partisans are 13% and 10% ($p = .17$). Figure 3, which shows item-by-item results, reveals that

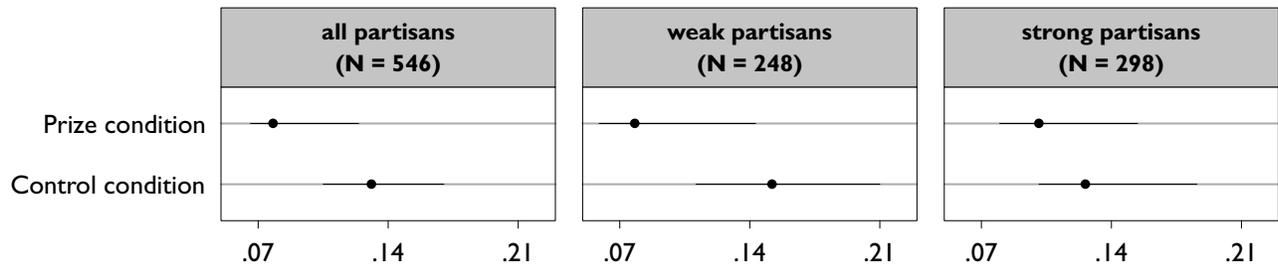


Figure A2: Overall Partisan Differences in Correct Responding. Each panel plots the average, taken over all nine questions, of absolute differences between the proportion of correct answers by Democrats and the proportion of correct answers by Republicans. Black lines are bootstrapped 95% confidence intervals.

The first panel shows that, as expected, the treatments reduced partisan differences in correct responding. The second and third panels show that this overall effect was generated primarily among weak partisans. (“Weak partisans” includes “leaning” subjects who initially identify as independents. It also includes subjects who initially identify themselves as partisans but subsequently refuse to call themselves “strong” partisans. See page ??.)

treatment reduced the gap in correct responding among weak partisans for every item. But among strong partisans, the treatment reduced the partisan gap for only five items, increasing it for the other four. Our experiment does not shed light on whether strong partisans were actually more likely to believe their false claims or whether the incentives offered were simply insufficient to overwhelm their stronger partisan motives.

Experiment 1: Strong Partisans: Overall and Item-by-Item Results

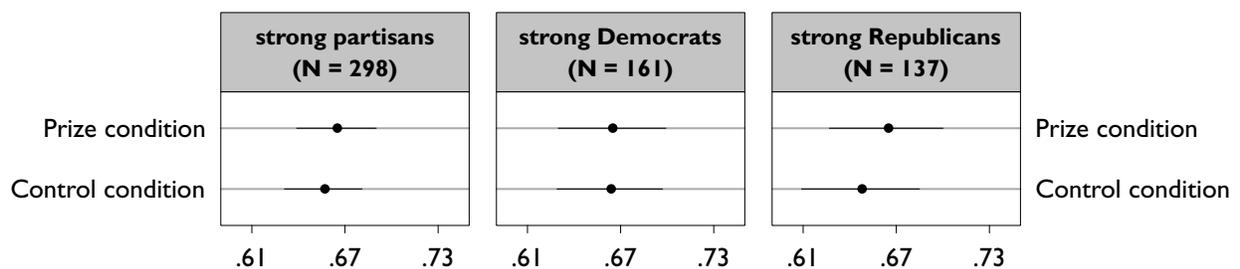


Figure A3: Proportions of Subjects Answering Correctly: Strong Partisans Only. Each panel plots the average proportion of correct answers given by strong partisans in the control and prize conditions. Black lines are 95% confidence intervals.

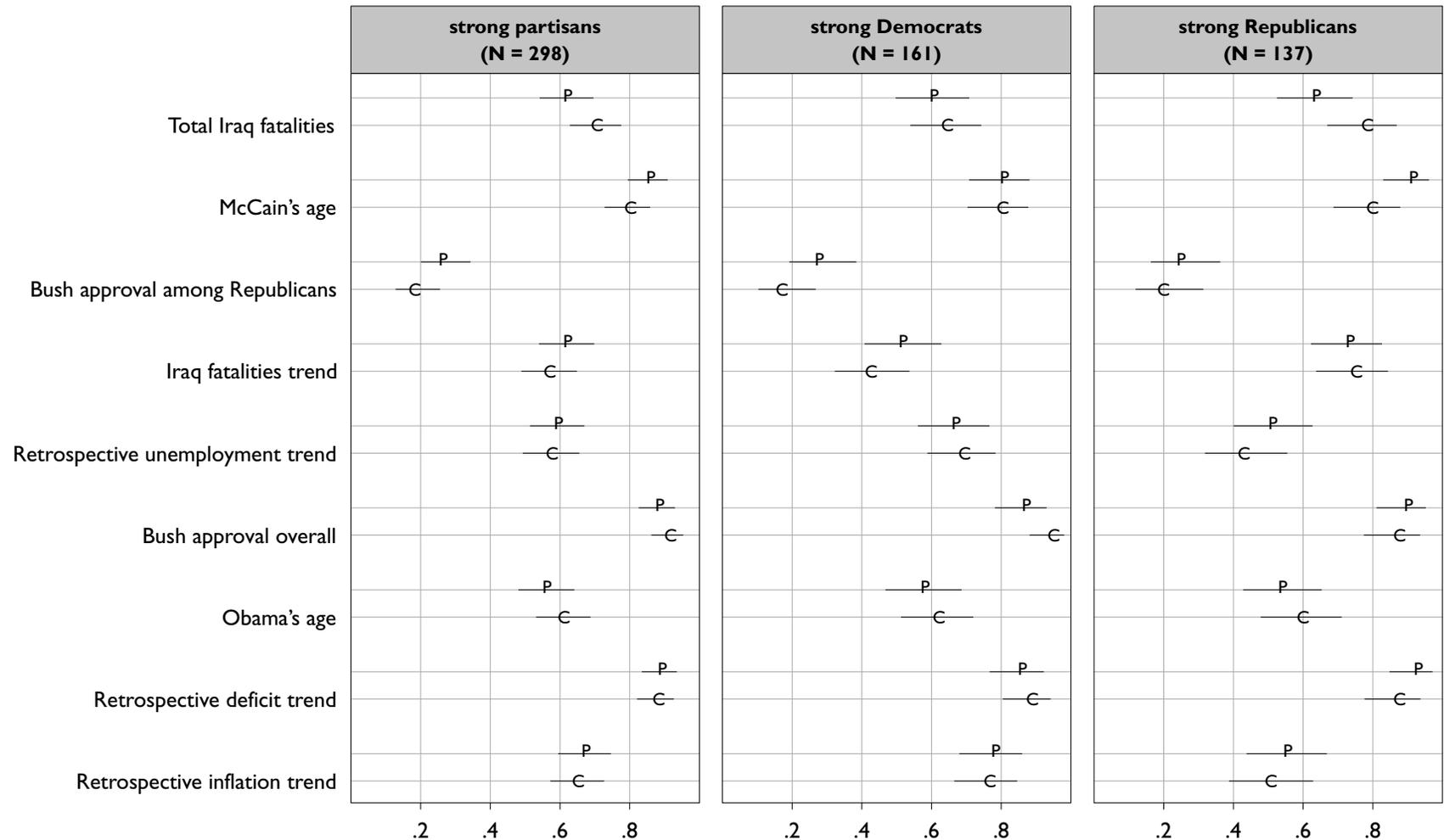


Figure A4: Proportions of Subjects Answering Correctly in Experiment 1: Strong Partisans Only. Each panel plots the proportions of strong partisans answering correctly in the control (C) and prize (P) conditions. Strong partisans are those who identified themselves as “strong Republicans” or “strong Democrats” in response to a party ID question with seven response options. Black lines are 95% confidence intervals. The “retrospective trend” questions refer to trends during George W. Bush’s terms as President. The “fatalities” questions refer only to deaths of U.S. soldiers. The exact wording of each item is given in Table 1.

Items are ordered according to the maximum of the absolute average treatment effects among strong Democrats and strong Republicans. Thus, “total Iraq fatalities” appears at the top of each panel (treated strong Republicans were 15 percentage points less likely than their control-group counterparts to answer correctly) and “retrospective inflation trend” appears at the bottom of each panel (treated strong Republicans were only five percentage points more likely to answer correctly, and the effect was smaller among Democrats).

Experiment 1: Low-Income Subjects: Overall and Item-by-Item Results

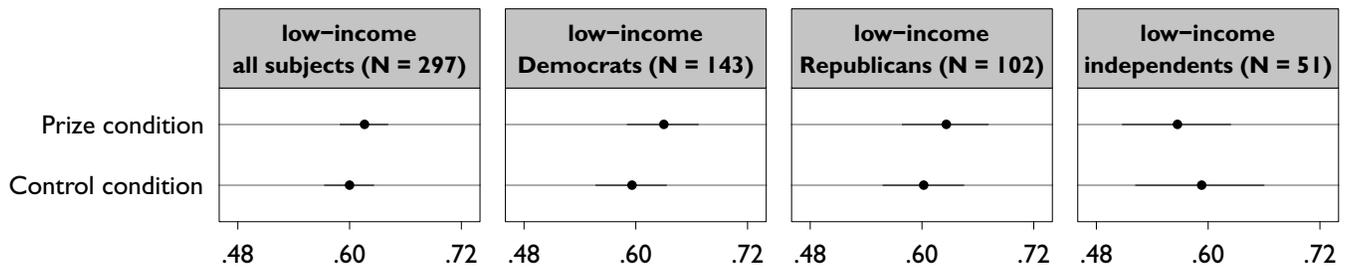


Figure A5: Proportions of Subjects Answering Correctly in Experiment 1: Low-Income Subjects Only. Each panel plots the average proportion of correct answers given in the control and prize conditions by subjects who report income under \$50,000. Black lines are 95% confidence intervals.

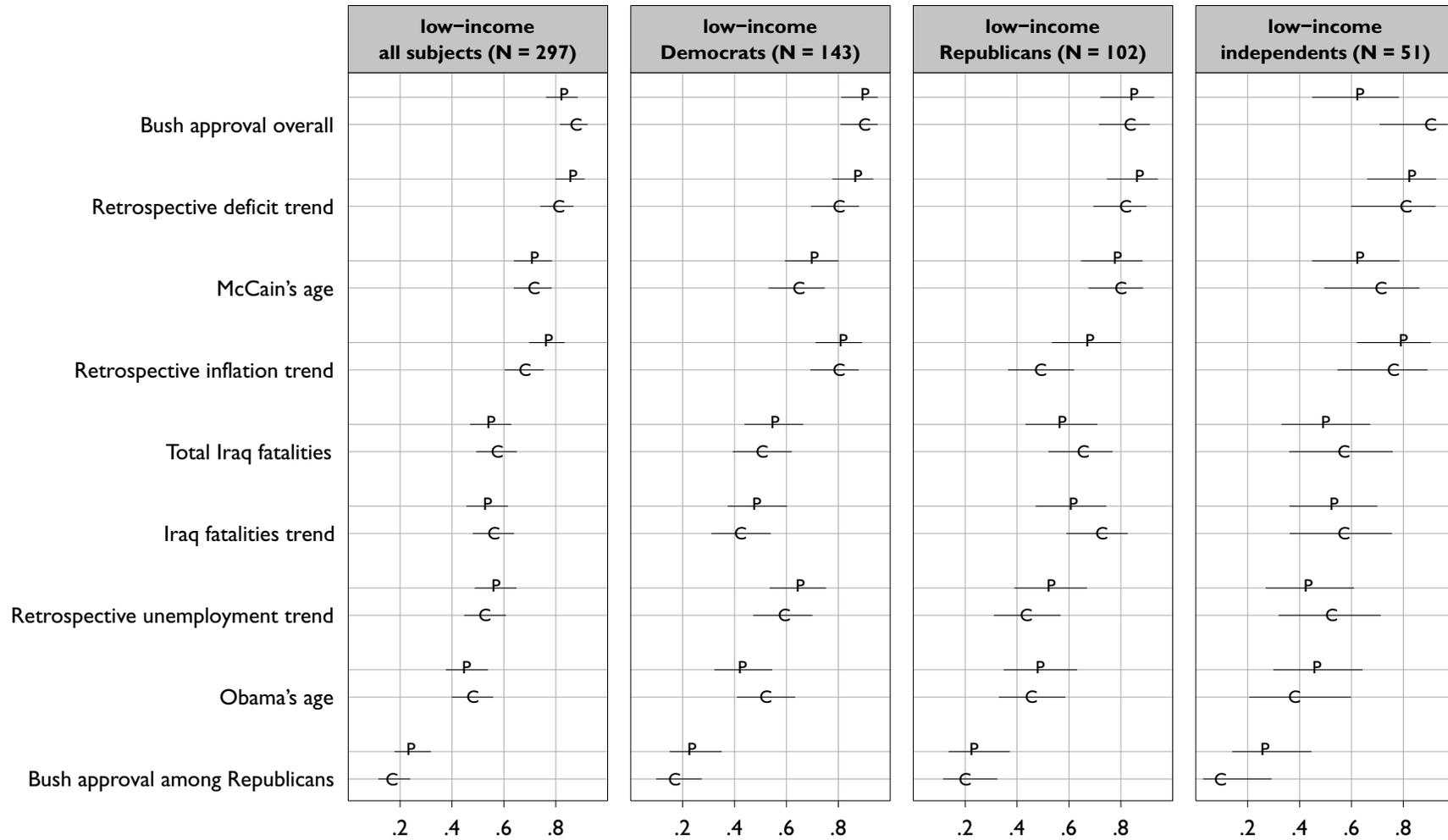


Figure A6: Proportions of Subjects Answering Correctly: Low-Income Subjects Only. Each panel plots the proportions of correct answers in the control (C) and prize (P) conditions among subjects who report annual income under of \$50,000. Black lines are 95% confidence intervals. The “retrospective trend” questions refer to trends during George W. Bush’s terms as President. The “fatalities” questions refer only to deaths of U.S. soldiers. The exact wording of each item is given in Table 1.

Items are ordered according to the proportion of all subjects answering correctly in the control condition.

Experiment 2: Prescreening Task

Before assignment to experimental condition, all subjects were given the following task:

News Articles

Here are two stories that were recently reported in Nebraska newspapers.

When you have finished reading them, please proceed to the next screen.

BELLEVUE, Neb. – The Bellevue City Planning Commission planned to hear a proposal on Thursday that would allow residents to raise chickens in their yards. It would allow people to keep chickens, but regulate the size of pens and the distance to a neighbor’s property. It would also impose fines for noise or health violations.

OMAHA, Neb. – The Omaha City Planning Department said a developer still has plans to build a hotel on a vacant lot near 12th and Jackson streets in the city’s Old Market district. The site has remained empty for more than a year and financing for the project has not been finalized. City officials said the developer does want to build a Residence Inn on the site and recently submitted some new, cost-effective designs.

The 138 subjects who spent fewer than eight seconds viewing this information were not invited to participate in Experiment 2.

Experiment 2: Instructions to Subjects

Subjects assigned to receive a reward for each correct answer received the following instructions, where *amount per correct answer* was the financial reward, e.g., \$0.50:

The next 14 questions are about politics and public affairs in America. You will have 35 seconds to answer each question. After 35 seconds, your screen will automatically advance to the next question.

For each correct answer that you give to the next 14 questions, you will receive an extra [*amount per correct answer*]. For example, if you answer six of the questions correctly, you will receive an extra $\$[6 \times \textit{amount per correct answer}]$, and if you answer all 14 questions correctly, you will receive an extra $\$[14 \times \textit{amount per correct answer}]$. This prize money is in addition to the money that you receive just for participating in this study. The prize money will be sent as a bonus payment within four weeks.

Please do not look up the answers. We are trying to understand what people know about politics, not what they can look up.

Many people have trouble answering questions like these. So if you can't think of the answer to a question, please don't worry. Just skip the question.

Control-group subjects received the same instructions except for the second paragraph, which was omitted.

Experiment 2: Sample Size by Condition and Party ID

	Dem.	Ind.	Rep.	No Answer	Total
\$0 per correct answer (control)	255	73	114	30	472
\$0.50 per correct answer	132	44	75	21	272
\$1.00 per correct answer	145	41	59	17	262
\$1.50 per correct answer	154	38	65	20	277
\$2.00 per correct answer	149	42	68	10	269
Total	835	238	381	98	1552

Table A3: Sample Size by Condition and Party ID in Experiment 2. Democrats are those who identified with the Democratic Party on a standard seven-point measure of party ID, including those who “lean” toward the Democratic Party. Republicans were coded in the same way.

Experiment 2: Randomization Check

Intercept	.82	.10
Female	.02	.04
Some college	.08	.06
Four-year college	.05	.07
Postgraduate degree	.01	.08
Age	.00	.00
Age ²	.00	.00
Midwest	-.06	.06
South	.04	.06
West	-.01	.06
Democrat	.04	.06
Republican	.02	.06
R^2	.01	
Standard error of regression	.75	
F	.70	($p = .71$)
Number of observations	1446	

Table A4: Randomization Check for Experiment 2. Cell entries are OLS estimates and standard errors from a regression in which assignment to treatment is regressed on pre-treatment variables. The outcome is the payment condition to which subjects were assigned (\$0, \$0.50, \$1, \$1.50, or \$2). The baseline education category is “no education beyond high school.” The baseline racial category is “white.” The baseline region is the Northeast. One hundred and six observations were omitted from the analysis due to missing data on the pretreatment covariates.

The F statistic is statistically insignificant, suggesting that random assignment to the treatment conditions was uncorrelated with the observed pretreatment covariates.

2010 Mechanical Turk Experiment: Overall Results

	Democratic subjects	Republican subjects
Intercept	1.34 .42	1.63 .80
Amount paid	.01 .07	.18 .14
General political knowledge	.86 .16	1.31 .37
Answered placebo correctly	-.08 .12	.25 .23
Female	-.29 .09	-.52 .18
Some college	-.10 .14	.17 .27
B.A.	.04 .15	.21 .29
Advanced degree	.01 .18	.57 .34
Age	.01 .02	-.02 .04
Age ²	.00 .00	.00 .00
Midwest	.31 .12	-.20 .29
South	.23 .11	-.04 .27
West	.33 .12	-.23 .31
R^2	.14	.15
Standard error of regression	.97	1.33
Number of observations	577	264

Table A5: Experiment 2 Knowledge Scores: Alternative Specifications. Each column reports OLS estimates and standard errors. In the “Democratic subjects” regressions, the outcome is the sum of correct answers to the four questions that control-group Democrats were less likely to answer correctly than control-group Republicans. In the “Republican subjects” regressions, the outcome is the sum of correct answers to the six questions that control-group Republicans were less likely to answer correctly than control-group Democrats. See Table 2 for the text of each question. **See page XX for descriptions of the predictors.**

The regressions reported here are closely related to the regressions in Table 3. The sole difference is that these regressions are estimated exclusively on data from subjects who were paid to answer questions correctly. Because we condition on payment amount in the Table 3 regressions, restricting the sample in this way should make little difference to our estimates. And indeed, comparison of these results to those in Table 3 show that there are no substantive differences.

2010 Mechanical Turk Experiment: Item-by-Item Results

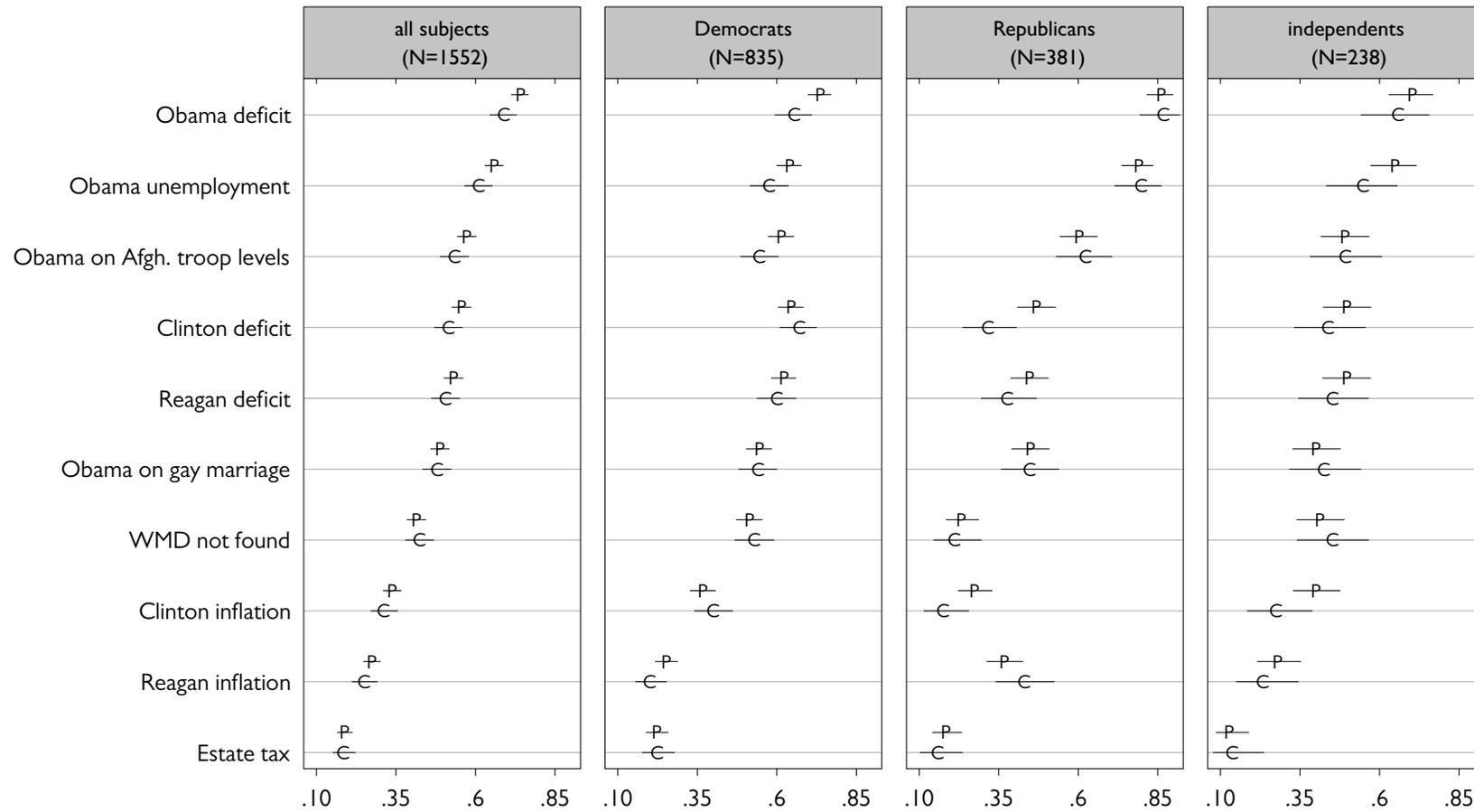


Figure A7: Experiment 2 Proportions of Subjects Answering Correctly: Results by Item. Each panel plots the proportions of subjects answering correctly in the control (C) and prize (P) conditions. Black lines are 95% confidence intervals. The “retrospective trend” questions refer to trends during George W. Bush’s terms as President. The “fatalities” question refers only to deaths of U.S. soldiers. The exact wording of each item is given in Table 1.

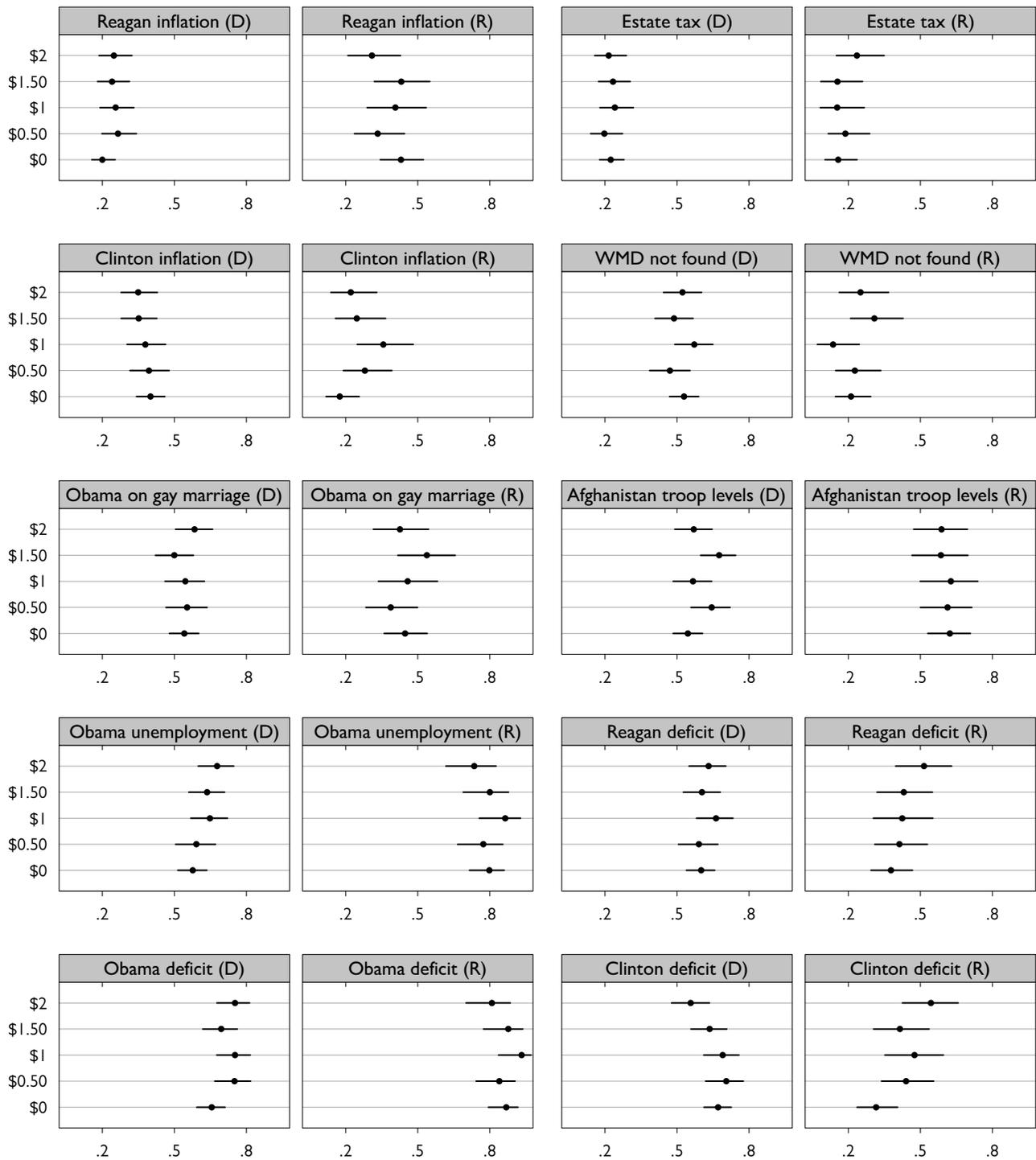


Figure A8: Experiment 2 Proportions of Subjects Answering Correctly: Results by Item and by Amount Paid. Each panel plots the proportions of subjects answering correctly in each payment condition. The first and third columns display proportions from Democratic subjects. The second and fourth columns display proportions from Republican subjects. Black lines are 95% confidence intervals. The exact wording of each item is given in Table 1.

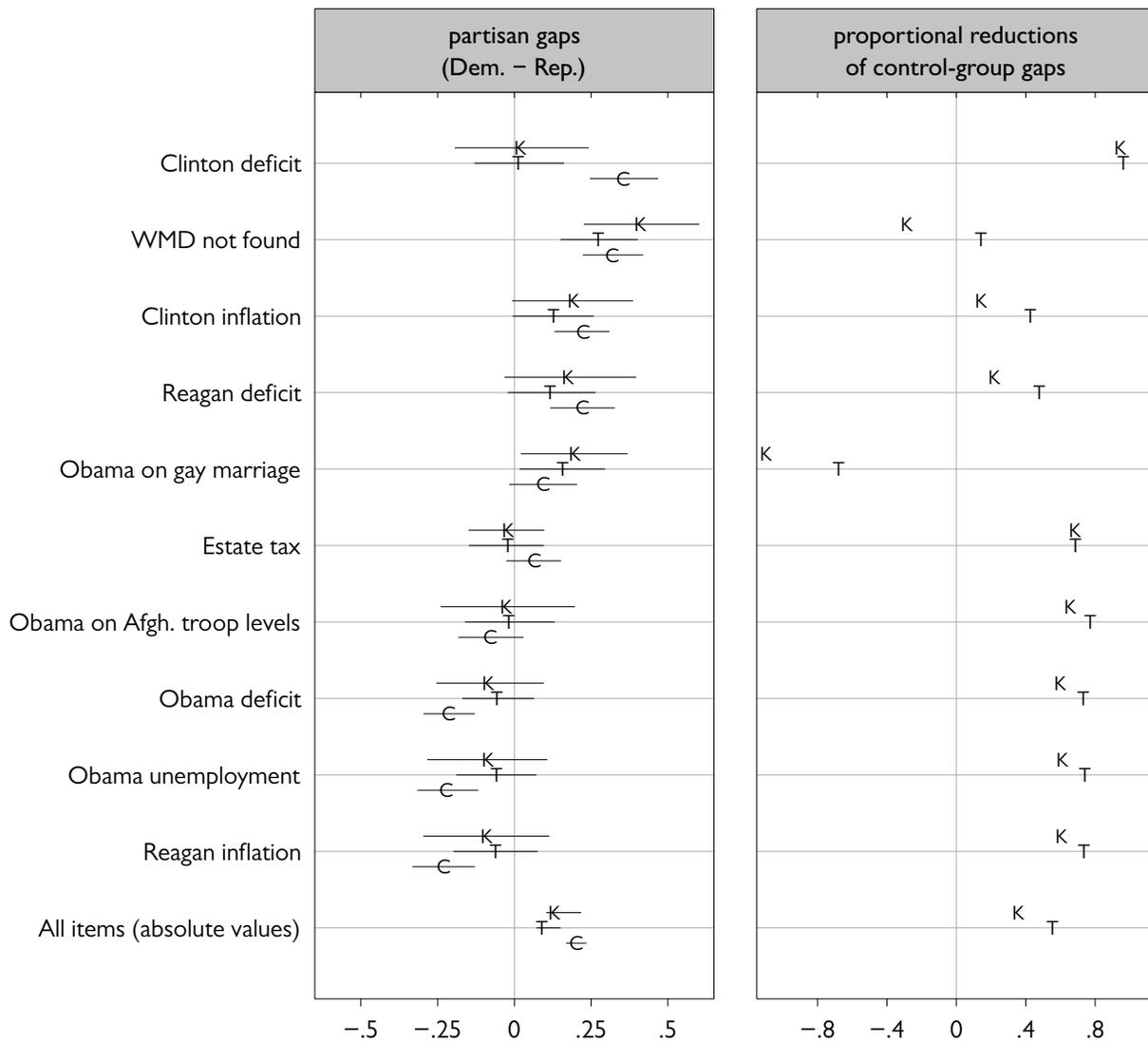


Figure A9: Partisan Gaps in Correct Responding in Experiment 2 (“maximum treatment” version). This figure is like Figure 6. The difference is that this figure compares control-group subjects to subjects who were offered \$2 per correct answer (the strongest incentive condition in the experiment), whereas Figure 6 compares control-group subjects to subjects who were offered any positive reward for answering correctly.

The left-hand panel plots the difference between the proportions of correct answers given by Democrats and Republicans. Differences are plotted for the control (C) and prize (P) conditions. The prize condition includes all subjects who were paid to offer correct answers. Positive values indicate that Democrats answered correctly more often than Republicans. Negative values indicate that Democrats answered correctly less often than Republicans. Black lines are 95% confidence intervals. The exact wording of each item is given in Table 2 (page 18).