

What is Expressive Responding? Evidence from a Meta-Reanalysis

Matthew H. Graham*

August 29, 2024

*Preliminary draft for the annual meeting of the
American Political Science Association*

Research on expressive responding suggests that people often give survey responses that are more partisan than their underlying beliefs. How much more partisan? Does this reflect insincere cheerleading, congenial inferences based on biased recall and reasoning, or something else altogether? In a meta-reanalysis of 43 studies from 20 articles or preprints, treatments designed to reduce expressive responding shrink measured of partisan bias by about 25 percent, from 0.20 to 0.15 on a 0 to 1 scale. Treatment slightly deflates the correlation between partisan identity and measured beliefs, from 0.19 to 0.14, and slightly increases the between-question correlation between Democrats' and Republicans' beliefs, from 0.80 to 0.84. These findings suggests that expressive responding is a modest component of a modest component of beliefs. Contrary to expectations derived from existing accounts of expressive responding, I find no evidence that expressive responding is more common among people with stronger partisan identities or higher levels of education. An analysis of accuracy effects among independents provides some evidence for the congenial inference explanation. In sum, existing research shows that the beliefs people express in surveys are a little more partisan than their underlying perceptions of reality but does not clearly pin down a mechanism.

*Assistant Professor, Temple University, USA. mattgraham@temple.edu. Hai Binh Hoang and Joseph Murphy assisted with the data collection. Helpful comments on previous versions were provided by Alex Coppock, Eli Rau, Lilla Orr, Dan Hopkins, Marc Meredith, Matt Levendusky, Mike Sances, Yph Lelkes, Shanto Iyengar, Diana Mutz, Madeline Pfaff, seminar participants at Rutgers University, conference participants at the second annual meeting of the Polarization Research Lab, conference participants at the Center for the Study of Democratic Politics, Princeton University, and panel participants at the 79th annual conference of the American Association for Public Opinion Research. Thanks also to Erik Peterson, Andrea Robbett, and Gaurav Sood for sharing additional variables from their datasets, and to James Fahey, Maurice Jakesch, and Brendan Nyhan for answering questions about their data.

It is widely recognized that survey responses may not accurately reflect people's underlying beliefs. In recent years, particular attention has been paid to partisan expressive responding, the idea that surveys exaggerate the degree to which Democrats and Republicans differ in their beliefs and attitudes. Despite its popular currency, there is no consensus regarding how important expressive responding is or why exactly it happens. Some observers think that expressive responding has large and pervasive effects, dismissing apparent partisan divides over the economy or the election as partisan cheerleading. Others think that expressive responding has small effects and argue that beliefs that are affected by expressive responding may be more politically relevant than beliefs that are purged of it through some artificial technique.

This paper systematically re-analyzes the existing body of research on expressive responding. The goal, to the extent that the research basis allows, is to understand what exactly expressive responding is and how big of a deal is it on average. Based on a standardized dataset of 44 studies from 20 previously published articles and preprints, I conduct a meta-reanalysis of the entire expressive responding literature. Whereas traditional meta-analysis relies on the estimates published in the original papers, meta-reanalysis begins with the original studies' data ([Galos and Coppock 2023](#)). This allows the researcher to standardize the analytic approach and conduct analysis that was not originally reported.

The analysis begins by examining the magnitude of expressive responding's effect on beliefs. In existing studies, expressive responding is only a modest portion of partisan bias, which itself is only a modest component to what partisans believe. In experiments on expressive responding, the average partisan belief difference in the control group is equal to about 20 percent of the scale. Treatments that aim to reduce expressive responding reduce partisan bias to about 15 percent of the scale. Similarly, the control group correlation between beliefs and 7-point partisanship is about 0.19, compared with about 0.14 in the treatment groups. There is substantial evidence that treatment effects are heterogeneous across studies and question topics, suggesting that expressive responding may matter more in some contexts than in others.

The paper next turns to theoretical explanations for expressive responding. Some frame expressive responding as insincere “cheerleading,” wherein partisan respondents deliberately falsify their beliefs in order to send a partisan message. Most research acknowledges a second possible explanation, congenial inference (Bullock and Lenz 2019; Prior et al. 2015). Congenial inference holds that at baseline, peoples’ partisan identities motivate them to make biased on-the-spot inferences about the survey question. If a treatment designed to reduce expressive responding (such as paying for correct answers) heightens accuracy motivation, respondents will make more even-handed, less partisan inferences.

The meta-reanalysis creates three new opportunities to assess the plausibility of these accounts by enabling analyses that would be under-powered in any standalone study: heterogeneous effects by strength of partisanship, heterogeneous effects by educational attainment, and accuracy effects on independents who do not lean toward either party. These tests do not provide much support for conventional wisdom regarding the nature of expressive responding. As explained below, both accounts predict that effects will be larger among stronger partisans and the highly educated. Neither prediction bears out: strong partisans appear no more expressive than weak partisans or leaners, and the highly educated are no more expressive than the less-educated. The test of accuracy effects among independents pits the two explanations against one another. A small accuracy effect is observed, which is predicted by the congenial inference account but not the cheerleading account. Across the three tests, one theory goes 0-for-3 and one theory goes 1-for-3. Given that the third test is zero-sum (one must win and one must lose), this about as poorly as conventional wisdom could have fared.

Altogether, the results paint expressive responding as a poorly understood phenomenon that only explains a modest amount of the partisan bias observed in the average survey. However, it can be more important: the strong statistical evidence of heterogeneity in treatment effects is backed by clear cases of more substantial expressive responding (Schaffner and Luks 2018; Shino et al. 2022). This suggests that even though expressive responding is not a reason to dismiss

evidence of partisan bias out of hand, it also cannot easily be dismissed in any particular case. Researchers should continue to test for expressive responding in a wide range of topic areas, and should endeavor to include design features that help shed light on the mechanisms behind the phenomenon.

Possible Reasons for Expressive Responding

A broad, mechanism-agnostic definition of expressive responding is that people sometimes give survey responses that are more partisan than their underlying beliefs. Existing research on expressive responding highlights two possible reasons why this could occur. As a running example, consider Roy, a hypothetical survey respondent who prefers the Republican party. Roy thinks that the current president, Joe Biden, has generally done a poor job as president. He has some personal experience that suggests that the economy is doing pretty well but doesn't think that Biden deserves credit for this. If asked in a survey about the economy or some statistic about it, such as the unemployment rate, Roy may claim that the economy is doing poorly. Something about being asked to express their beliefs in a survey has created a disconnect between Roy's underlying beliefs about the economy and his response to the survey question.

One reason is cheerleading, i.e. an insincere or deliberate misreporting of one's beliefs. This could occur because of costs or benefits of expression that are internal or external to the respondent (Graham and Huber 2022).¹ Externally, respondents may think of their survey response as signaling to an audience what they or their group think (Connors 2023). Expressing a party-congenial belief may have expressive benefits, while expressing an uncongenial belief may have expressive costs. For example, Roy may be thinking of his response to questions about the economy as sending a signal about what Republicans think of Biden's job performance. His insincere claim to think that the unemployment rate is going up could be intended to send a signal that Republicans disapprove of Biden, or to avoid sending a signal that Republicans approve of his performance. Internal

¹In their review of list experiments, Blair et al. (2020) also cite possible reasons for insincere responding that are internal and external to respondents (self-image and self-presentation; see their Table 1).

costs and benefits are also possible. For example, Roy may find it more enjoyable to state responses that are more consistent with her self-image as a partisan, or less enjoyable to state responses that are inconsistent with it.

Another reason that expressive responding could occur is congenial inference (Bullock and Lenz 2019), which marries theories of the survey response and motivated cognition. Survey response theory holds that the beliefs measured in surveys are not generally preexisting beliefs that respondents extract directly from memory, like a file drawn from a drawer. Instead, responses are constructed on the spot based on whatever relevant information the respondent calls to mind (Zaller 1992; Tourangeau et al. 2000). Theories of motivated reasoning argue that identities like partisanship create directional motives that bias peoples' recall and inferences (Kunda 1990). For example, when posed with a question about how well the economy is doing, Roy could either rely on a partisan heuristic like "Biden is doing a bad job." If Roy were more motivated to make an accurate inference, he might instead rely on a non-partisan heuristic like "my sister had no problem finding work after she got laid off." In this way, heightened accuracy motivation could result in inferences that are more representative of peoples' underlying beliefs about the economy.

Both explanations suggest that expressive responding causes survey responses to be more partisan than people's underlying perceptions. However, the two explanations differ in terms of whether people sincerely believe their response. According to the cheerleading explanation, people are deciding not to report their true belief. According to the congenial inference explanation, people are reporting a true belief about a specific question—but one that does not faithfully reflect what's under the hood. If the congenial inference explanation is right, Roy might have made the same inference in a real-world argument about the economy, which he might use to argue in favor of his evaluation of Biden or his partisan preference. A researcher's interest could reasonably determine their view of whether congenial inference is a problem. Researchers who view survey responses as a projection of underlying perceptions should want to purge responses of both insincere misreporting and congenial inference, and do not need to worry so much about which

mechanism is actually at work. In contrast, researchers who want to measure what inference people would make if asked to do so in the real world may want to preserve congenial inference. In this sense, the importance of pinning down the precise mechanism behind expressive responding depends on the researcher's inferential target.

Techniques for Reducing Expressive Responding

Researchers use a variety of techniques for detecting and quantifying expressive responding. These techniques vary in terms of which reasons for expressive responding they target, as well as other possible pathways through which they could affect responses. This section reviews the techniques used in previous research and plausible reasons for their effects. Table 1 summarizes this discussion and links it to the previous section's example of Roy, a Republican who thinks the economy is doing well but says it is doing poorly.

Payment for correct answers (pay-for-correct). The most popular method for studying expressive responding, including the first two articles on the subject, is to pay randomly assigned respondents for correct answers (Bullock et al. 2015; Prior et al. 2015). Pay-for-correct experiments could either reduce insincerity or congenial inference. Bullock et al. (2015) focus on insincere misreporting, coining the term “partisan cheerleading” to describe their findings. They present a model in which payment increases incentives to be honest, outweighing expressive benefits. Prior et al. (2015) describing three reasons why people may give survey responses that are more partisan than “the information [they] hold in their memory”: the accessibility of information in memory, the on-the-spot inferences drawn from it, and insincere misreporting. The first two reasons fall into the category of congenial inference.

Payment for correct answers could plausibly affect responses through other pathways as well. I refer to these as *threats* because they would not be consistent with any reasonable definition of expressive responding. Payment could motivate respondents to look up the correct answers. Bullock et al. (2015) and Prior et al. (2015) make an effort to address this concern using a combi-

nation of question timers, “catch” questions that are hard to answer correctly without cheating, and self-reports. Payment for correct answers could also induce a form of demand effects, motivating respondents to try to anticipate the researcher’s beliefs about the correct answer rather than their own beliefs. This threat is plausibly circumvented by selecting questions with clear and indisputable referents. For example, given partisan controversy over what counts as the true unemployment rate (Vavreck 2014), question wording could specify the researcher’s interest in the official unemployment rate as defined by the Bureau of Labor Statistics. Researchers also sidestep demand effects by avoiding pay-for-correct experiments in cases when respondents’ preferred source of truth could diverge from the researcher’s, e.g. political rumors (Berinsky 2018).²

Encouragements, pledges, and requests. Researchers also use requests to be honest, accurate, or the like. As with payment for correct answers, honesty encouragement can either be framed as increasing incentives to be sincere (Bullock et al. 2015; Prior et al. 2015; Berinsky 2018) or as increasing accuracy motivation (Rathje et al. 2023). The appeal of honesty encouragement is its ability to credibly rule out threats to inference by assumption. Given the lack of financial incentives, it seems unlikely to encourage cheating. Because honesty encouragement does not tie anything to the researcher’s perception of what counts as correct, it also seems less likely to induce demand effects. However, research suggests that requests are weaker treatments than other strategies, in the sense that they eliminate less expressive responding (Bullock et al. 2015; Rathje et al. 2023; Graham and Yair 2024b).

²Also see Bullock and Lenz’s (2019) description of Nyhan’s unpublished list experiment and Graham and Yair’s (2024a) justification of the betting on the future strategy.

Table 1: Summary of Techniques for Studying Expressive Responding

Approach	Possible mechanism	Examples	
		Control (misreported)	Treatment (accurately reported)
<p><i>Pay-for-correct:</i> randomly assign monetary rewards for answering correctly.</p> <p>(Similar to honesty encouragement, requests, or pledges; see appendix.)</p>	<p><i>Outweigh internal costs/benefits of expression:</i> respondents value the money more than the good feeling from cheerleading.</p>	Roy thinks it decreased but saying this doesn't feel good given his view that Biden is doing a bad job. He picks "increased."	Roy thinks it decreased and wants to get the money. He picks "decreased."
	<p><i>Outweigh external costs/benefits of expression:</i> respondents value the money more than sending a signal by cheerleading.</p>	Roy thinks it decreased but thinks this will send a signal that Republicans think Biden is doing a good job. He picks "increased."	Roy thinks it decreased and wants to get the money. He picks "decreased."
	<p><i>Reduce congenial inference:</i> incentives motivate respondents to rely less on partisan considerations, causing them to form a more accurate belief.</p>	Roy thinks for a moment about how Biden is doing a bad job of managing the economy and picks "increased."	Roy thinks about how Biden is doing a bad job of managing the economy but also remembers that his company is hiring and his sister had no trouble finding work after losing her job. He picks "decreased."
<p><i>Ask the unasked question:</i> anticipate what people want to express and let them express it.</p>	<p><i>Shrink internal costs and benefits of expression:</i> the chance to express themselves reduces the expressive costs/benefits of stating an un/congenial belief.</p>	Roy thinks it decreased but saying this doesn't feel good given his view that Biden is doing a bad job. He picks "increased."	Roy thinks it decreased. It doesn't feel good to admit this, but the chance to say that Biden is doing a bad job makes it easier. He picks "decreased."
	<p><i>Shrink external costs and benefits of expression:</i> the chance to express themselves reduces the expressive costs/benefits of sending a signal that is in/consistent with the party line.</p>	Roy thinks it decreased but thinks this will send a signal that Republicans think Biden is doing a good job. He picks "increased."	Roy thinks it decreased. He doesn't love to admit this, but he already had a chance to send a message that Biden is doing a bad job. He picks "decreased."

(continued on next page)

<i>List experiments:</i> reduce expressive pressures by avoiding direct endorsement and ensuring anonymity.	<i>Shrink external costs imposed by interviewer or organization:</i> less fear that the interviewer will judge them or report them to the authorities.	Roy thinks it decreased but worries he will be shunned by his fellow Republicans if they find out. He picks “increased.”	Roy thinks it decreased and has no fear that anyone will find this out. He counts “unemployment decreased” as a statement he agrees with.
	<i>Shrink external costs/benefits created by audience for results:</i> less fear that responses will be used to cast group in an unfavorable light.	Roy thinks it decreased but thinks this will send a signal that Republicans think Biden is doing a good job. He picks “increased.”	Roy thinks it decreased and doesn’t realize that the list experiment will generate an estimate of what Republicans think. He counts “unemployment decreased” as a statement he agrees with.
	<i>Shrink internal costs and benefits of expression:</i> reduces worry that belief will be inconsistent with self-presentation as a partisan; less direct endorsement is less enjoyable.	Roy thinks it decreased but saying this doesn’t feel good given his view that Biden is doing a bad job. He picks “increased.”	Roy thinks it decreased. Counting this in a list doesn’t feel as bad as endorsing it directly, so he counts “unemployment decreased” as a statement he agrees with.
<i>Implausible beliefs:</i> ask about something nobody could possibly think, assume all endorsements are expressive. Non-experimental.	<i>Induce insincerity due to internal costs and benefits:</i> the party-congenial response is so implausible that sincere belief in it can be ruled out entirely.	Roy can see in the picture that Trump’s crowd is smaller but it doesn’t feel good to say something that suggests Obama is more popular. He says the smaller crowd is bigger.	(No treatment. Researchers assumes that the response was insincere.)
	<i>Induce insincerity due to external costs and benefits:</i> the party-congenial response is so implausible that sincere belief in it can be ruled out entirely.	Roy can see in the picture that Trump’s crowd is smaller but it feels like Republicans are being tricked into saying that Obama is more popular. He says the smaller crowd is bigger.	(No treatment. Researchers assumes that the response was insincere.)
<i>Administrative data:</i> ask about something with a more credible measure of ground truth.	<i>Reveal intentional misreporting due to internal / external benefits.</i>	Roy remembers that he voted by mail but it feels bad to admit that he defied the party line / he doesn’t want to send a signal that Republicans didn’t listen to Trump. He says he voted in person.	Voting records reveal that Roy voted by mail.
	<i>Reveal unintentional misreporting due to congenial inference.</i>	Roy doesn’t really remember how he voted last time but figures he wouldn’t have voted by mail given the politics. He says he voted in person.	Voting records reveal that Roy voted by mail.

Asking the unasked question. The theory of response substitution gives rise to another paradigm for studying expressive responding. According to the response substitution account, people sometimes respond insincerely to survey questions because they want to answer a different question than the one that was asked. Gal and Rucker (2011) give the example of a restaurant with terrible food and great service. If asked at the end of the meal about the quality of the service, one might say “terrible” in order to express their disapproval of the food. But if given a chance to rate the food first, one might rate the service more honestly. Researchers who apply this paradigm seek to anticipate and ask the “unasked question” that the respondent wants to answer (Yair and Huber 2020; Yair and Schaffner 2022). For example, Democrats and Republicans often claim that opinions they agree with are facts. This tendency is reduced when partisans are first given the opportunity to state their agreement or disagreement with the statement (Graham and Yair 2024b). Such interventions aim to increase sincerity without altering the motivational structure that could produce congenial inferences.

Although unasked question treatments can more credibly rule out congenial inference, they also raise a different set of threats to inference. One is priming. Answering the unasked question could change responses by priming respondents to call to mind considerations that they would not otherwise have thought about. Graham and Yair (2024b) argue that priming is unlikely in their case, as the instructions and question wording already made extensive reference to the agreeableness of the statements. Another threat is that asking the unasked question may clarify the intent of the original question. Under this interpretation, asking unasked questions signals to the respondent that the researchers intend them to exclude considerations that might otherwise have been deemed relevant. This concern is not typically addressed in the design; Yair and Huber (2020) and Graham and Yair (2024b) use post-hoc analysis to argue that it did not affect their results.

List experiments. Researchers also use list experiments, also known as the item count technique, to study expressive responding. List experiments were originally developed to increase sincere reporting of sensitive traits like illegal behavior (Miller 1984). Rather than ask respondents

to directly answer the question of interest, researchers present respondents with a list of statements and ask how many apply to the respondent. For a random subset of respondents, the statement of is omitted from the list. The difference in the number of statements endorsed between the two lists provides an estimate of the prevalence of the trait of interest.

List experiments have been used to study a wide range of beliefs and traits. Conventional wisdom regarding when how they work is divined from examining cases in which there is a statistically significant difference between the list estimate and a direct question. Among these cases are beliefs and traits that would threaten peoples' self-image or self-presentation (Blair et al. 2020), which respectively are internal and external sources of expressive costs and benefits. For example, Heide-Jørgensen (2023) finds that conservatives are more likely to endorse progressive taxation in list experiments than with direct questions.

Implausible beliefs. This strategy takes advantage of situations in which one belief is so unlikely to be sincerely held that it can be ruled out by assumption. Schaffner and Luks (2018) use the approach to study a controversy over the relative size of the crowd at Barack Obama and Donald Trump's presidential inaugurations. It rained the day of Trump's inauguration, dampening attendance. Trump claimed that his crowd had been larger. Schaffner and Luks showed survey respondents photos of the two crowds and asked which was larger. About 15 percent of Republicans picked the obviously smaller crowd, defying the plain visual evidence to express an implausible but party-congenial belief. By comparison, just 4 percent of Democrats picked the same. This provides strong evidence of insincere misreporting. Other possible reasons for choosing the incorrect answer seem unlikely to explain partisan bias.³

External validation. This strategy takes advantage of situations in which surveys ask about behaviors that can be verified externally, e.g. in administrative records. After mail-in voting be-

³For example, some may not be paying careful attention to the survey, and others may think that if the researcher is asking about something so obvious, there must be a trick. It is possible for careless responding to cause illusory between-group differences when one group is small (Fan et al. 2002; Cimpian 2017). In this case, however, the comparison is between two groups that make up similarly large shares of the population.

came politicized during the 2020 presidential election, [Shino et al. \(2022\)](#) fielded a survey that asked registered voters in Florida how they usually voted. Among voters who were confirmed in public records to have voted by mail during each of the previous four elections (the 2016 primary, 2016 general, 2018 primary, and 2018 general), Republicans were more likely to misreport that they usually voted in-person. This pattern of misreporting echoed attacks on mail-in voting by Republican politicians. This provides strong evidence of misreporting. In principle, self-reports could be inconsistent with administrative records because people forget what they did and use partisan heuristics to fill in the blanks. In this case, however, this concern is mitigated by the focus on respondents who voted the same way in four consecutive elections. A more general limitation is the small number of behaviors that can be validated externally.

Meta-Reanalysis

To gain more leverage on the issues raised above, I conduct a meta-reanalysis of 52 treatments in 43 studies from 20 articles or preprints. Meta-reanalysis differs from traditional meta-analysis in its use of the raw data ([Galos and Coppock 2023](#)). I collected data from each study's publication file or original author, harmonized the data structure across studies, specified common scoring rules and estimation strategies, separately calculated the quantities of interest for each study, and used random effects meta-analysis to aggregate the study-by-study estimates. Two benefits of this approach are to (1) force a detailed examination of the original estimation strategies and their relationship to one another, and (2) permit analysis that was not part of the original studies, including analysis of heterogeneous effects that are not usually examined in existing research.

Inclusion Criteria

The meta-reanalysis endeavors to include all studies that (1) make individual-level data available, (2) use an experimental treatment or other approach that appears in research on expressive responding, (3) examine partisan or ideological bias, and (4) allow a comparison between two sets

of responses that are expected to be more- or less-affected by expressive responding. The most notable exclusion is studies the implausible beliefs strategy. Although some partisans' willingness to give survey responses that contradict plain visual evidence is clearly suggestive of expressive responding, the approach does not provide an estimate of how much of the observed partisan bias is due to expressive responding.

Scoring Rules

Existing research quantifies expressive responding in two main ways: partisan differences and congenial responding. Handling these differences was an essential part of the data and estimation strategies. In both cases, researchers begin by preprocessing the data based on conception of what responses are more congenial to one party of the other. For example, to quantify partisan differences, [Bullock et al. \(2015\)](#) coded all responses so that “the most Republican response to each question (either the largest or smallest response) is coded as 0 [and] the most Democratic response is coded as 1,” then estimated the average difference between Democrats and Republicans. To estimate party-congenial responding, [Prior et al. \(2015\)](#) score each response as a congenial error, approximately correct, or an uncongenial error, then subtract the proportion of congenial and uncongenial errors. This procedure is equivalent to scoring the responses on a 1, 0, or -1 scale and taking the average.

Though different studies' use of different quantities looks at first blush like thorny territory, measures of partisan difference and congenial responding turn out to be closely related to one another. Across the 52 estimates that appear below in [Figure 2](#), the correlation between the estimated effects on partisan differences and congenial responding is 0.86. To provide a deeper sense of the relationship between the two measures, [Appendix \[INCOMPLETE\]](#) examines them analytically. Effects on both measures can be expressed as averages of the conditional average treatment effects among Democrats and Republicans. Estimates of effects on partisan differences weight the two parties equally, while effects on congenial responding weight each party according to its proportion

of the sample. Estimated effects on congenial responding are sensitive to chance differences in the proportion of Democrats and Republicans who are randomly assigned to each treatment condition, while estimated effects on partisan differences are not affected by this source of variance. These features help explain why the two methods of quantification can differ from study to study but average out to similar magnitudes in the meta-analysis.

One advantage of the response congeniality measure is that it can be used in studies that only include one party. This applies to three of the 52 studies ([Fahey 2022](#); [Graham and Yair 2024a](#), Study 1; [Nyhan 2020](#)). These studies are included in the congenial responding-based estimates but excluded from the partisan difference estimates.

Data

With the help of research assistants, I obtained data from 20 previously published articles about expressive responding through a combination of public replication files and personal communications. In the current draft, 40 of the 43 studies were conducted in the United States. The other three were conducted in Israel ([Graham and Yair 2024b](#), Studies 3 and 4; [Yair and Schaffner 2022](#), Study 1).⁴ I harmonized the structure of the datasets so that each row of the dataset includes one respondent's answer to one survey question. For example, a study with 10 questions has 10 rows per respondent. In total, the data contain about 260,000 responses from about 57,000 unique respondents.

To place responses on a common scale, I began by rescaling responses to each question to range from 0 to 1, where 0 is the low point on the scale as the authors defined it and 1 is the high point. I then used information gleaned from the original authors' data and documentation to place each response on the two response scales described above: a partisan scale from least to most Republican, as well as a congeniality scale from least to most congenial to the respondent's

⁴This draft refers to the left party as "Democrats" and the right party as "Republicans," but the terminology will probably need to change. Alternatively, the scope could be limited to U.S. studies.

partisanship. This is possible for all studies because, as noted above, all studies embed a conception of partisan congeniality somewhere in their original approach. In all cases, I follow the original author’s stipulations about which end of the response scale is congenial to which party.

Estimation

Each part of the analysis follows the same two steps. I begin by estimating treatment effects separately for each study. In all cases but the list experiments, this can be done using linear regression. To estimate the effect on partisan differences, I used OLS to estimate

$$YR_{ik} = \beta_0 + \beta_1 \text{Republican}_i + \beta_2 \text{Treatment}_i + \beta_3 \text{Republican}_i \times \text{Treatment}_i + \epsilon_{ik}$$

where i indexes respondents, k indexes questions, YR_{ik} is an answer to a question coded from 0 (most Democratic) to 1 (most Republican), Republican_i is a Republican indicator, and Treatment_i is a treatment indicator. The key parameters are β_1 , the partisan difference in the control group, and β_3 , the treatment effect on partisan differences.

To estimate the effect on congenial responding, I used OLS to estimate

$$YC_{ik} = \beta_0 + \beta_1 \text{Treatment}_i + \epsilon_{ik}$$

where YC_{ik} is a survey response coded from -1 (least party-congenial) to 1 (most party-congenial) and all other terms are defined above. The key parameters are β_0 , which is congenial responding in the control group, and β_1 , the treatment effect on congenial responding. Because each observation is a respondent \times question pair, I cluster standard errors by respondent.

After collecting the study-by-study estimates, I aggregate across studies by conducting a meta-analysis. I begin this step by extracting the relevant estimates and standard errors. This table of estimates has one estimate and standard error per treatment condition, meaning that studies with multiple treatment conditions (e.g., [Prior et al. 2015](#), Study 2) contribute more than one estimate.

Using these estimates, I conduct a random effects meta-analysis using the `meta` package in R (Schwarzer et al. 2015). To account for the fact that some estimates share the same control group, I cluster the meta-analysis at the level of the study.

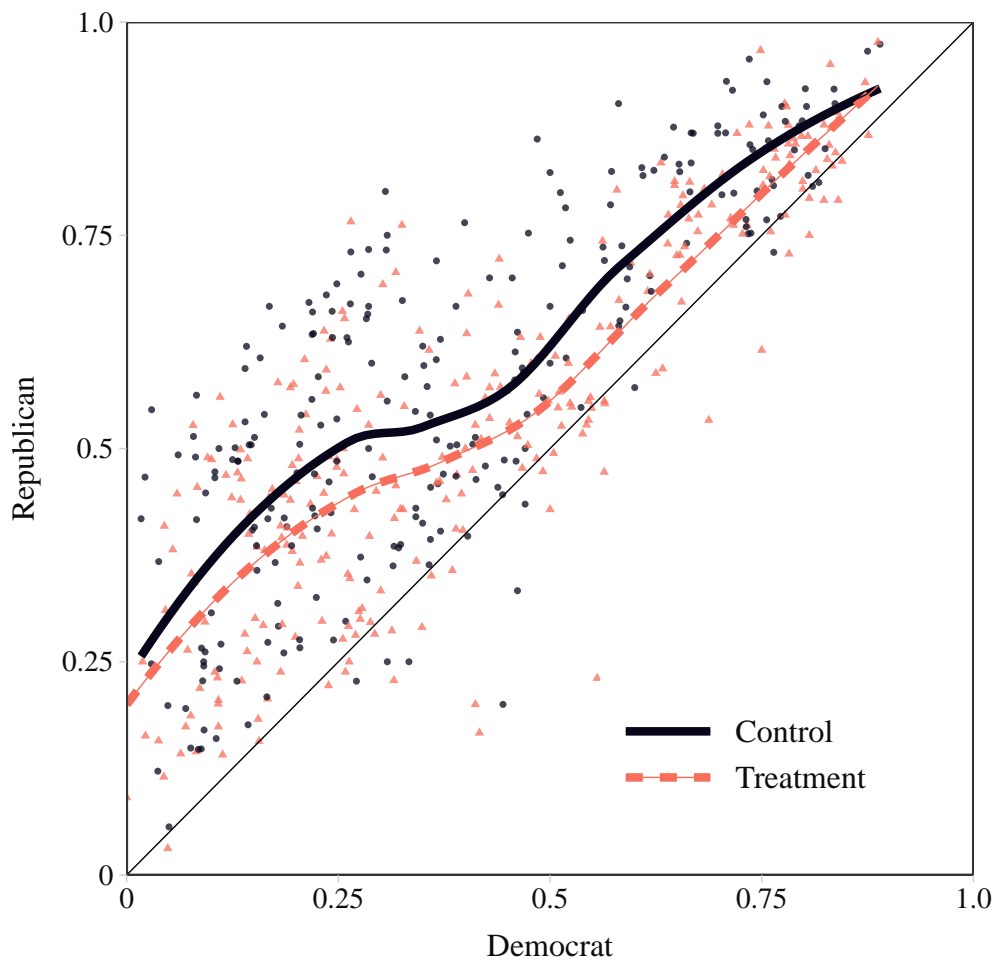
Effects on Partisan Bias

To place partisan bias in context, I begin by comparing the average Democrat and Republican's response to each question. Across the studies, there are 254 study \times question pairs and 538 study \times question \times treatment combinations. For each combination, Figure 1 plots the average response among Democrats and Republicans on a 0 to 1 scale from most Democratic to most Republican. The dark purple dots and purple smoothed line represent the control means for each question, and the orange triangles and dashed smoothed line represent the treatment group means. The thin black 45-degree line represents an absence of partisan bias. Almost all of the dots appear above the 45-degree line, indicating that partisan bias is observed on almost all questions. The dashed line is modestly closer to the 45-degree line, indicating that treatment tends to reduce partisan bias by a modest amount.

The question-by-question estimates suggest that even within a universe of questions that is intentionally selected based on the suspicion of partisan bias, there is also a strong common component to peoples' perceptions of the political world. Across the 254 questions, the correlation between Democrats' and Republicans' beliefs is 0.80 in the control conditions and 0.84 in the treatment conditions.

To aggregate across questions, I begin by estimating treatment effects within each study, using both measures of partisan bias discussed above. The full set of estimates appears in Figure 2. Across the 49 estimates of the effect on partisan differences, 40 (82 percent) are negative and 19 (39 percent) are statistically significant. Across the 52 response congeniality estimates, 42 (81 percent) are negative and 17 (33 percent) are statistically significant. The correlation between the two sets of estimates is 0.86. The largest estimates are in the -0.1 to -0.15 range, while the smallest

Figure 1: Average Response by Question, Party, and Treatment Condition



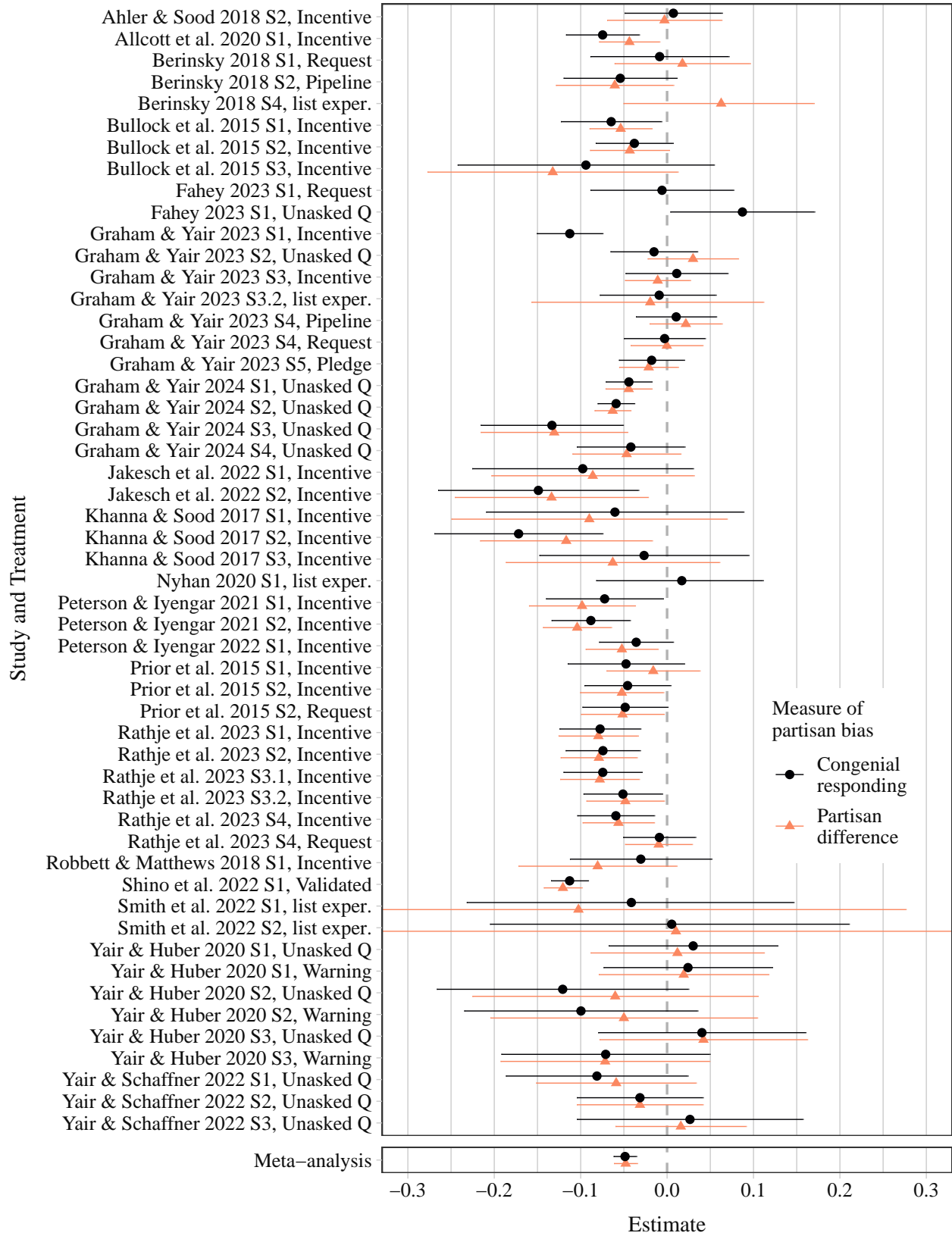
Note: Each dot represents a question from a study about expressive responding. Responses are coded on a 0 to 1 scale from most Democratic to most Republican. The x-axis is the average response for a Democrat; the y-axis, for a Republican. Thin 45 degree line corresponds to no partisan difference. Almost all points are above the 45-degree line, indicating that on average, Republicans give more Republican responses than do Democrats (i.e., partisan bias). The averages in the treatment group (dashed line) are closer to the 45-degree line than the average response in the control groups (solid line), indicating a reduction in partisan bias.

estimates are close to 0.

To aggregate across studies, I conducted a random effects meta-analysis. In the control groups, average partisan bias is 0.186 using both measures. In the treatment conditions, partisan differences shrink by about -0.048 (s.e. = 0.007) and congenial responding is reduced by about -0.049 (s.e. = 0.007). Both estimates are equal to about 25 percent of the baseline.

When making generalizations based on the meta-analytic estimates, two points of caution are worth noting. First, expressive responding may have larger effects in some contexts than in others. The estimated standard deviation of the true effect sizes (τ) is about 0.030 and there is strong statistical evidence for heterogeneous treatment effects ($Q \approx 150, p < 0.0001$). Second, there is some evidence in existing research that some treatments are stronger than others (Bullock et al. 2015; Graham and Yair 2024b). Given this, the true amount of expressive responding is likely to be larger than the average estimate.

Figure 2: Effects on Partisan Bias



Note: Figure displays treatment effects on two measures of partisan bias, partisan differences and congenial responding (equations X and X). Dots are point estimates. Horizontal bars are 95 percent confidence intervals. Standard errors clustered by respondent.

Heterogeneity by Strength of Partisan Identity

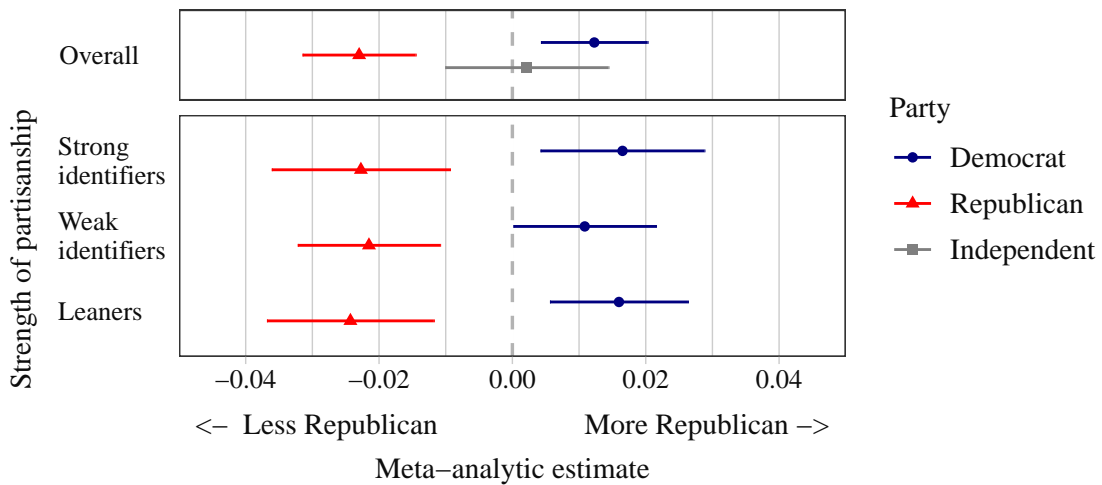
A natural first step in a deeper look at expressive responding is to examine heterogeneous effects by strength of partisan identity. The congenial inference account offers a clearer basis to make predictions along these lines. Because directional motivations are thought to be rooted in identities (Leeper and Slothuus 2014; Taber et al. 2009, 139), people with stronger partisan identities should have a greater tendency to engage in expressive responding. The insincerity account's predictions with respect to partisan identity are less clear, as it does not specify any particular source of insincerity beyond partisanship. Despite this, it seems likely that partisan cheerleading is more fun for people with stronger partisan identities.

Given this, examining heterogeneous effects by partisan identity has potential to provide value in the following ways. First, the presence of such heterogeneity would add to the credibility of both the cheerleading and congenial inference accounts. Second, a lack of such heterogeneity would be bad news for both accounts—but particularly for congenial inference accounts, which draw on theories that squarely focus on identification with a party as the source of partisan motivation. Third, a lack of such heterogeneity would suggest a need to consider other possible mechanisms for the observed effects that have weaker connections to identity.

To examine the relationship between partisan identity strength and expressive responding, I collected the 7-point partisanship scale in each study and coded it to range from 1 to 7, where 1 is a strong Democrat and 7 is a strong Republican. This was available in 16 of the 20 articles and 32 of the 40 U.S. studies. For this set of studies I estimated the average treatment effect for each level of the 7-point partisan identity scale (Figure 3), as well as differences in treatment effects between the two categories (Table 2).

The treatment weakens the association between partisanship and measured beliefs. First consider Figure ??, which displays the average response in each partisan group on a scale from most Democratic to most Republican. In the treatment group, the mean for Democrats is a few

Figure 3: Effects by 7-Point Partisan Identity



percentage points higher, meaning that Democrats’ responses become a little more Republican. Similarly, Republicans’ responses become a bit less Republican. The correlation between the 7-point partisan identity scale and the Republican belief scale declines from 0.19 to 0.14.

Next I estimate the treatment effect within each partisan strength group (Table 2). Little evidence emerges that expressive responding is more pronounced among strong partisans. When partisan bias is measured in terms of partisan differences, the effects range from -0.032 to -0.044 (Table 2),⁵ with slightly larger estimates among stronger partisans. When partisan bias is measured in terms of congenial responding, the estimates range from -0.038 to -0.045. When the effect sizes are compared to the baseline differences observed in the control group, there is no evidence that stronger partisanship predicts larger effects. The effects among leaners and strong partisans are less than 20 percent of the baseline, while the effects among weak partisans equal 25 to 30 percent of the baseline.

In addition to being small in absolute terms and relative to the baseline difference, the effects are small relative to the variation in belief strength that exists among survey-takers. Within each group on the 7-point partisanship scale, the effect size is 3 to 6 percent of the control group standard

⁵Once data collection is complete, these results will be updated with a statistical test of the differences between the groups.

Table 2: Effects by Strength of Partisanship

Identity strength	Congenial responding		Partisan difference	
	Baseline	Treatment effect	Baseline	Treatment effect
All partisans	0.197 (0.030)	-0.041 (0.007)	0.198 (0.022)	-0.039 (0.007)
Leaners	0.183 (0.030)	-0.038 (0.010)	0.173 (0.022)	-0.032 (0.009)
Weak identifiers	0.151 (0.037)	-0.045 (0.007)	0.166 (0.019)	-0.040 (0.007)
Strong identifiers	0.242 (0.027)	-0.041 (0.010)	0.247 (0.027)	-0.044 (0.009)

Note: Cell entries are estimates from a random effects meta-analysis using the 26 studies with a measure of 7-point partisan identity. Standard errors in parentheses.

deviation within the same partisan category (and about 0.1 percent of a standard deviation for independents). The effects on partisan bias are equal to 5 to 10 percent of the control group standard deviation among the entire sample.

Heterogeneity by Educational Attainment

Educational attainment provides another opportunity to check predictions derived from the two theories. Both accounts depend on a respondent’s awareness of which response options are more or less favorable to one’s partisan team. For example, in order to cheerlead or make a congenial inference about the unemployment rate, one must know that high unemployment rates are bad and that observers of politics are likely to credit or blame the president’s party for whatever happens to unemployment. In other domains, more-educated people have a better sense of “what goes with what” (Bishop 1976). A few studies have used heterogeneity by educational attainment as evidence for their favored mechanism for expressive responding (Schaffner and Luks 2018; Shino et al. 2022). In addition, theories of motivated reasoning give a prominent role to cognitive resources: one who is more skilled at reasoning and critical thinking is likely to be more capable

of engaging in the mental acrobatics required to make a congenial inference (Taber et al. 2009, 139).⁶ Given this, examining heterogeneous effects by educational attainment provides an opportunity to check a prediction that is consistent with both accounts. If more-educated people are more prone to expressive responding, both are more plausible. If not, an alternative explanation may be necessary.

To examine the relationship between educational attainment and expressive responding, I collected the educational attainment measure for each study that made it available. This amounted to 18 of the 20 articles and 38 of the 43 studies. I then identified the most detailed coding scheme that was compatible with all of the studies. The four-level coding scheme classifies respondents as having a high school diploma or less, some college or an associate's degree, a bachelor's degree, and a graduate degree.

Within each educational attainment category, I estimated the treatment effects on congenial responding and partisan differences. The results appear in Table 3. For comparability, I also estimated the treatment effect across all levels of education among only those respondents who reported an education level. For this subset of studies and respondents, the estimated effects in the top row of Table 3 are the equivalent of the bottom row of Figure 2.

At baseline (in the control groups), there is not much evidence that the level of partisan bias varies by educational attainment. Using the congenial responding measure, estimated bias falls between 0.187 and 0.215 in all four education groups. Using the partisan difference measure, baseline partisan bias ranges from 0.161 to 0.217, with the point estimates monotonically decreasing in education. The estimated difference between the congenial responding and partisan difference measures is rarely more than one standard error.

I find no evidence that a higher level of education predicts a greater tendency to engage in

⁶Tappin, Pennycook, & Rand (2021) question the link between cognitive resources and biased reasoning. Whereas they frame this as evidence that cognitive resources may not matter for motivated reasoning, I view this as evidence that motivated reasoning may not be a good explanation for the phenomenon they study.

Table 3: Effects by Educational Attainment

Educational attainment	Congenial responding		Partisan difference	
	Baseline	Treatment effect	Baseline	Treatment effect
All levels	0.196 (0.026)	-0.050 (0.007)	0.191 (0.020)	-0.051 (0.007)
High school or less	0.209 (0.028)	-0.063 (0.017)	0.214 (0.026)	-0.058 (0.016)
Some college or associate's	0.215 (0.030)	-0.050 (0.010)	0.217 (0.026)	-0.051 (0.011)
Bachelor's degree	0.195 (0.026)	-0.052 (0.010)	0.184 (0.017)	-0.052 (0.010)
Graduate degree	0.187 (0.030)	-0.042 (0.009)	0.161 (0.017)	-0.042 (0.009)

Note: Cell entries are estimates from a random effects meta-analysis using the 28 studies with a measure of educational attainment. Standard errors in parentheses.

expressive responding. For all education groups and measures of bias, the estimated effect on partisan bias is between 25 and 35 percent of the baseline. Using both measures, the largest point estimate is among the lowest level of educational attainment, high school or less. In this group, the estimated effect on congenial responding was -0.063 (s.e. = 0.017), and the estimated effect on partisan differences was -0.058 (s.e. = 0.016). The smallest point estimate according to both measures was among the highest level of educational attainment, a graduate degree. The estimated effects on this group were both -0.042 (s.e. = 0.009). The middle two education categories fell in between these two, with point estimates ranging from 0.050 to 0.052 (s.e. 0.010 to 0.011).

Accuracy Effects among Independents

The previous sections highlight the challenge of distinguishing the cheerleading and congenial inference accounts, which generate similar predictions about how effects vary by partisan identity and educational attainment. One potential route is to examine how the two accounts view the motivation to be accurate. The congenial inference account is built on motivated reasoning

theory, which views pay-for-correct treatments as increasing accuracy motivation. In contrast, the cheerleading account does not include any mechanism that should increase accurate responding. This difference in predictions regarding the role of accuracy motivation can be leveraged by examining effects on pure independents, who have no partisan motivations at play. To the extent that the motivational account is correct, a treatment that provides incentives to be accurate should increase accurate responding among independents. To the extent that the cheerleading account is correct, independents should not be responsive one way or the other.

A limitation on this strategy is the fact that not all studies of expressive responding concern topics with objectively correct answers. In total, 18 of the 20 articles focus on questions with objectively correct answers.⁷ Among these, 24 studies contain data on political independents, excluding respondents who lean toward a party. Moreover, features of the substantive topic affect researchers' decisions about what treatments to use. The two articles without correct answers both use unasked question treatments. The articles with correct answers use are more likely to use financial incentives and requests. Conveniently, however, all of the studies that use strategies for which congenial inference is a plausible mechanism (pay-for-correct, requests and encouragement, administrative data) also have correct answers.

To examine accuracy effects on independents, I used the original authors' definitions of the correct answer to code all responses on an accuracy scale from 0 to 1, where 1 is the most accurate answer possible and 0 is the least accurate answer possible.⁸ Based on the discussion above, I split the results according to whether congenial inference is a plausible mechanism for the observed treatment effects. If accuracy effects are observed on the treatments where congenial inference is plausible, the test constitutes evidence in favor of the congenial inference account. If no accuracy effects are observed on the treatments where congenial inference was deemed implausible, the test supports the decision to rule out this mechanism.

⁷The exceptions are [Yair and Huber \(2020\)](#) and [Yair and Schaffner \(2022\)](#).

⁸That is, I transformed all responses according to the formula $1 - |x_i - \theta| / \max(|\max(X) - \theta|, |\min(X) - \theta|)$ where x_i is respondent i 's response, θ is the correct answer, and X is a vector of all responses to the question. This yields a 0-1 scale where 1 is the most accurate possible response and 0 is the least accurate possible response.

Table 4: Accuracy Effects on Independents

Treatments	Estimate	SE	95% CI
All treatments	0.014	0.007	0.001, 0.028
Congenial inference plausible	0.016	0.008	0.000, 0.032
Incentives	0.017	0.009	0.000, 0.035
Requests and pledges	0.013	0.016	-0.018, 0.044
Congenial inference implausible	0.007	0.013	-0.018, 0.032
Unasked question	0.001	0.013	-0.025, 0.027
List experiments	0.106	0.055	-0.001, 0.214

Table 4 presents meta-analytic estimates of the effect on independents. Across all treatment conditions, the effect on accuracy was about 0.014 (s.e. = 0.007, $p = 0.04$). Among the treatment conditions for which congenial inference was deemed a plausible mechanism, the accuracy effect was 0.016 and was also statistically significant (s.e. = 0.008, $p = 0.05$), providing evidence that the treatment heightened accuracy motivation. Among the remaining treatment conditions, the estimated accuracy effect smaller. However, the estimate was sufficiently imprecise (s.e. = 0.013) that accuracy effects of a similar magnitude cannot be ruled out.

These results are mostly encouraging for the congenial inference account, which predicts that treatments increase respondents' motivations to form accurate beliefs. The main caveat is the fact that the effect size is plausibly in the same range for treatments on which congenial inference is not predicted.

Discussion

In what sense does existing research suggest that partisans don't believe what they say in surveys? Across existing research, the net effect of reducing expressive responding was to reduce partisan differences by about one-fourth, from 20 percent of the scale to 15 percent of the scale. Given that many or most treatments are not strong enough to eliminate all expressive responding, this may be an under-estimate of expressive responding's total effect. Even so, partisan bias is far

from the only force shaping peoples' beliefs. At baseline, the correlation between partisan identity and beliefs is 0.19, falling to 0.14 in the treatment groups. From question to question, Democrats' and Republicans' responses are strongly predictive of one another, with correlations of 0.80 and 0.84 in the control and treatment groups. Given all this, the 30,000 foot view is that expressive responding is a modest component of measured beliefs about politics. Its effects are larger in some situations than in others, but on the whole, its effects do not appear significant enough to fundamentally warp our big-picture view of public opinion.

The findings did not not clearly support either of the leading explanations for expressive responding, cheerleading and congenial inference. The two tests in which the two accounts generated the same predictions—heterogeneous effects by partisan identity and educational attainment—did not support either account. The analysis of accuracy effects on independents provided some evidence that incentive and request treatments boost accuracy motivation, a key ingredient of the congenial inference explanation. Against the backdrop of considerable uncertainty about the mechanism in foundational articles on the topic, the limited support for either explanation suggests a need for caution when describing the nature of the observed effects. Although it is clear that survey responses are sometimes at least somewhat more partisan than people's underlying perceptions, we do not know much about the underlying mechanisms.

A path forward may be to move away from the expressive responding label, treating cheerleading and congenial inference as distinct phenomena that are handled by distinct interventions. The challenge, of course, is that both pathways are plausible for some popular interventions. One could begin to untie this knot by combining approaches that are hypothesized to work differently within the same experiment. For example, to better-understand the effects of pay-for-correct treatments (where congenial inference is plausible), they could be included in the same experiment as the unasked question or implausible belief approach (where congenial inference is unlikely). A challenge in forming such combinations is that the substantive topic of the study is often relevant to ruling out threats to inference. In particular, researchers tend to favor non-financial treatments

when there is no objectively correct answer (e.g., physical attractiveness) or researchers and respondents may not share the same point of reference (e.g., politicized controversies). Though some non-financial strategies are more clearly targeted at cheerleading, they are generally limited by fairly specific scope conditions, e.g. an obvious unasked question, plain visual evidence that renders some beliefs implausible, or the existence of external verification. In turn, cases that support the use of multiple strategies may then face questions about their external validity.

The process of assembling the data highlights replication practices that would make individual studies more useful to the collective scientific endeavor. When possible, replication files should include more background variables than are necessary to reproduce their main results. They should also include data on respondents who were excluded from the main analysis, even if this exclusion was pre-registered or otherwise merited in the context of the original study. Including “extras” like this has potential to create new points of meta-analytic leverage that may not have been feasible to examine in the original study.

References

- Berinsky, Adam J. 2018. "Telling the Truth about Believing the Lies? Evidence for the Limited Prevalence of Expressive Survey Responding." *The Journal of Politics* 80(1):211–224.
- Bishop, George F. 1976. "The Effect of Education on Ideological Consistency." *Public Opinion Quarterly* 40(3):337–48.
- Blair, Graeme, Alexander Coppock and Margaret Moor. 2020. "When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments." *American Political Science Review* 114(4):1297–1315.
- Bullock, John G, Alan S Gerber, Seth J Hill and Gregory A Huber. 2015. "Partisan Bias in Factual Beliefs about Politics." *Quarterly Journal of Political Science* 10:1–60.
- Bullock, John G and Gabriel Lenz. 2019. "Partisan Bias in Surveys." *Annual Review of Political Science* 22:325–342.
- Cimpian, Joseph R. 2017. "Classification Errors and Bias Regarding Research on Sexual Minority Youths." *Educational Researcher* 46(9):517–529.
- Connors, Elizabeth C. 2023. "Social Desirability and Affective Polarization." *Public Opinion Quarterly* 87(4):911–934.
- Fahey, James J. 2022. "The Big Lie: Expressive Responding and Conspiratorial Beliefs in the United States." *Journal of Experimental Political Science* (forthcoming).
- Fan, Xitao, Brent C. Miller, Mathew Christensen, Kyung Eun Park, Harold D. Grotevant, Manfred van Dulmen, Nora Dunbar and Bruce Bayley. 2002. "Questionnaire and interview inconsistencies exaggerated differences between adopted and non-adopted adolescents in a national sample." *Adoption Quarterly* 6(2):7–27.
- Gal, David and Derek D Rucker. 2011. "Answering the Unasked Question: Response Substitution in Consumer Surveys." *Journal of Marketing Research (JMR)* 48(1):185–195.
- Galos, Diana and Alexander Coppock. 2023. "Gender composition predicts gender bias: A meta-reanalysis of hiring discrimination audit experiments." *Science Advances* 9(18).
- Graham, Matthew H. and Gregory A. Huber. 2022. *The Expressive Value of Answering Survey Questions*. Oxford University Press pp. 83–112.
- Graham, Matthew H. and Omer Yair. 2024a. "Expressive Responding and Belief in 2020 Election Fraud." *Political Behavior* 46:1349–74.
- Graham, Matthew H. and Omer Yair. 2024b. "Less Partisan but No More Competent: Expressive Responding and Fact-Opinion Discernment." *Public Opinion Quarterly* (forthcoming).
- Heide-Jørgensen, Tobias. 2023. "Does Social Desirability Bias Distort Survey Analyses of Ideology and Self-Interest? Evidence from a List Experiment on Progressive Taxation." *Public Opinion Quarterly* (first view).

- Kunda, Ziva. 1990. "The case for motivated reasoning." *Psychological Bulletin* 108(3):480–498.
- Leeper, Thomas J. and Rune Slothuus. 2014. "Political parties, motivated reasoning, and public opinion formation." *Political Psychology* 35(SUPPL.1):129–156.
- Miller, Judith D. 1984. "A New Survey Technique for Studying Deviant Behavior."
- Nyhan, Brendan. 2020. "Race and the Obama Muslim Myth." *Time-Sharing Experiments in the Social Sciences Archive* (March 9, 2020).
- Prior, Markus, Gaurav Sood and Kabir Khanna. 2015. "You Cannot be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions." *Quarterly Journal of Political Science* 10:489–518.
- Rathje, Steve, Jon Roozenbeek, Jay J. Van Bavel and Sander van der Linden. 2023. "Accuracy and Social Motivations Shape Judgements of (Mis)Information." *Nature Human Behavior* (first view).
- Schaffner, Brian F. and Samantha Luks. 2018. "Misinformation or Expressive Responding?" *Public Opinion Quarterly* 82(1):135–147.
- Schwarzer, Guido, James R. Carpenter and Gerta Rücker. 2015. *Meta-Analysis with R*. Springer International Publishing.
- Shino, Enrijeta, Daniel A. Smith and Laura Uribe. 2022. "Lying for Trump? Elite Cue-Taking and Expressive Responding on Vote Method." *Public Opinion Quarterly* 86(4):837–861.
- Taber, Charles S., Damon Cann and Simona Kucsova. 2009. "The Motivated Processing of Political Arguments." *Political Behavior* 31(2):137–155.
- Tourangeau, Robert, Lance J. Rips and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge University Press.
- Vavreck, Lynn. 2014. *Want a better forecast? Measure the campaign not just the economy*. Vol. 47 Cambridge University Press pp. 345–347.
- Yair, Omer and Brian F. Schaffner. 2022. "Response Decoupling and Partisans' Evaluations of Politicians' Transgressions." *Political Behavior* .
- Yair, Omer and Gregory A Huber. 2020. "How robust is evidence of perceptual partisan bias in survey responses? A new approach for studying expressive responding." *Public Opinion Quarterly* 84(2):469–92.
- Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge University Press.

Appendix

Supplemental Figures

This section plots the study-by-study estimates that entered the analysis of heterogeneous effects by strength of partisan identity and educational attainment.

Figure 4: Conditional Average Treatment Effects by Strength of Partisanship

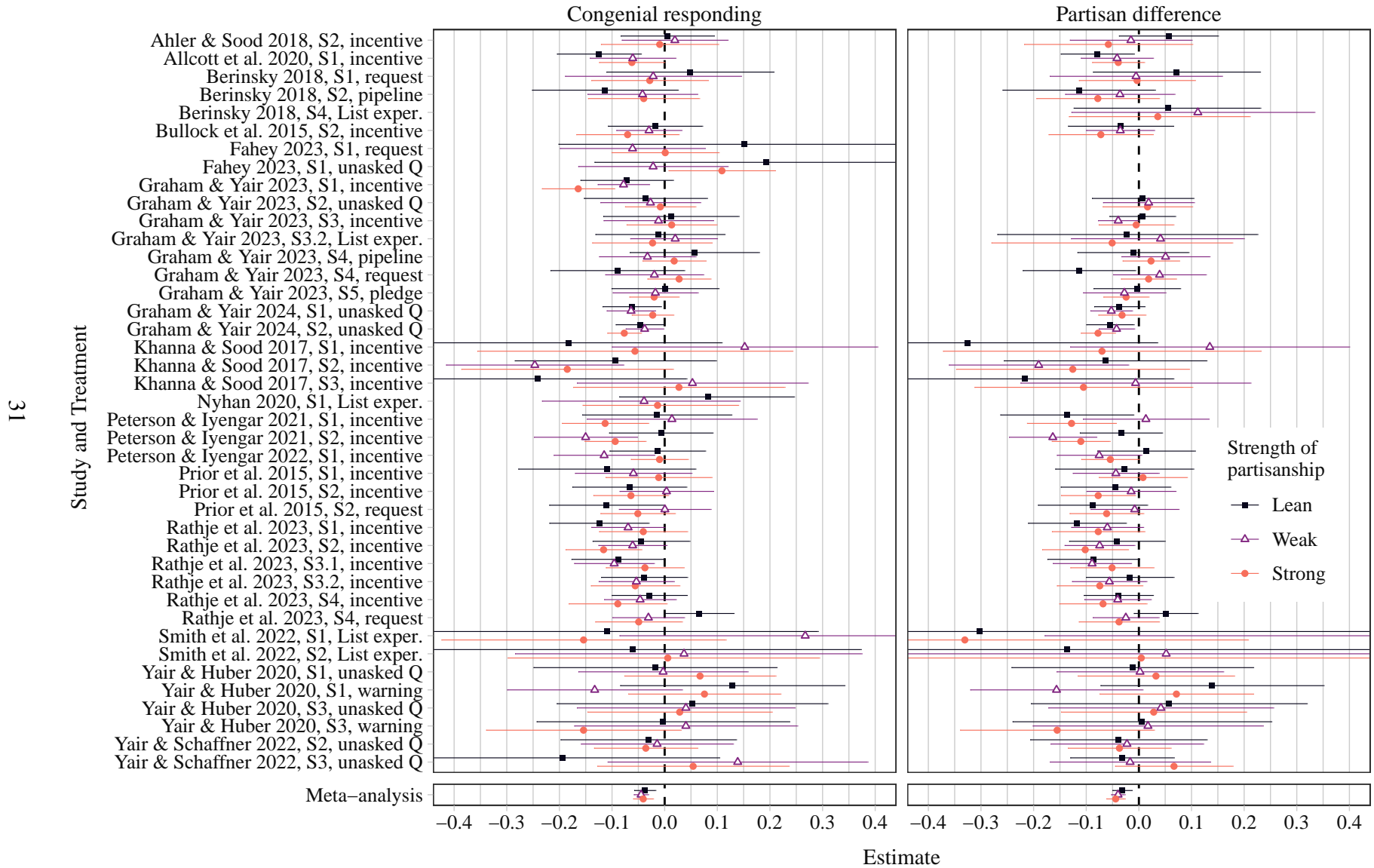
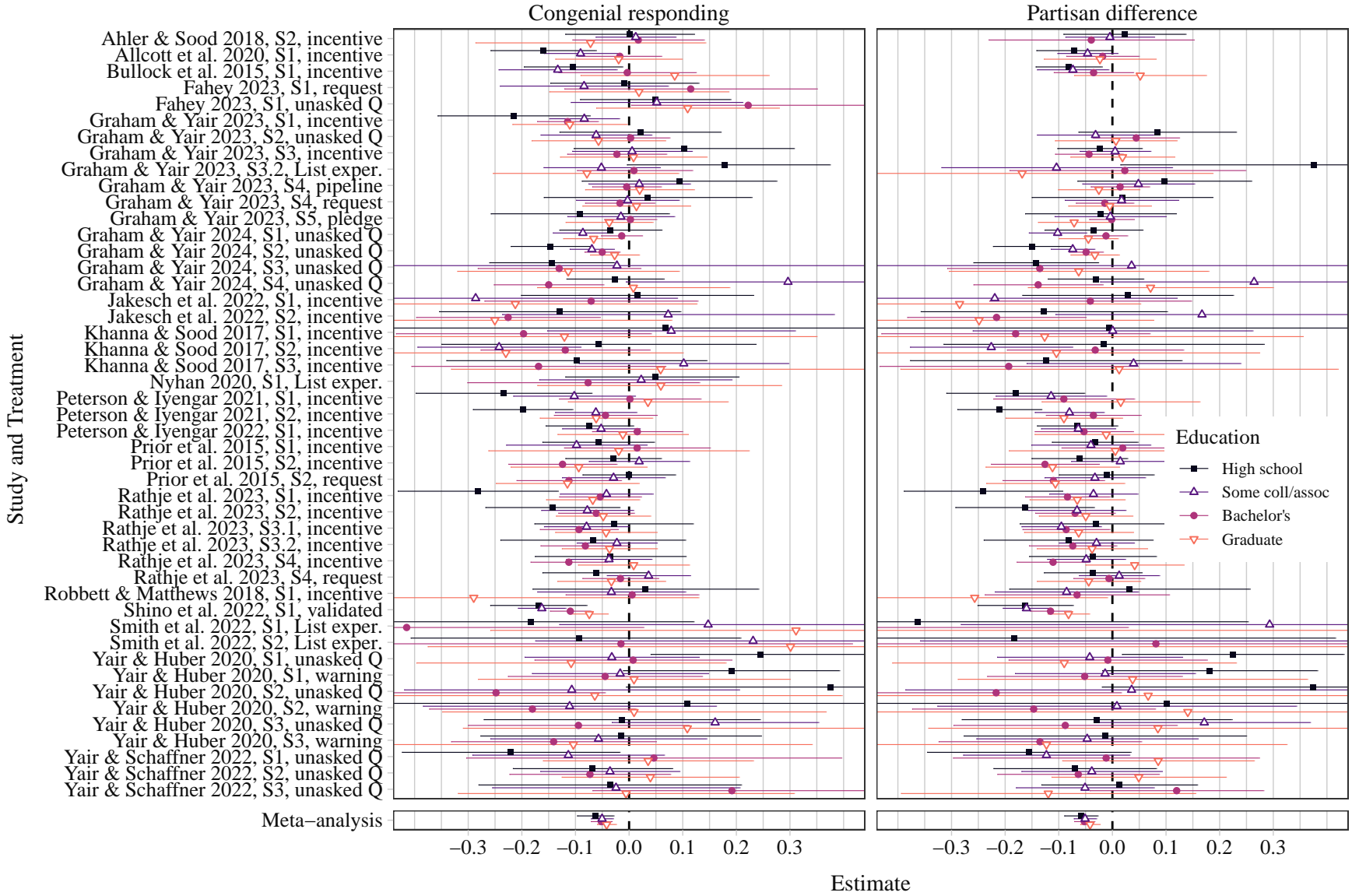


Figure 5: Conditional Average Treatment Effects by Educational Attainment



Additional Results

Accuracy Effects by Question Congeniality

Another situation in which the cheerleading and motivational accounts offer differing predictions emerges from questions whose correct answers are unambiguously congenial to one's party (that is, the correct answer would also be good for your party if true). I believe that the argument in this section is consistent with an extended version of the [Bullock et al. \(2015\)](#) model (Appendix A) that treats the belief as an endogenous function of motivations rather than a constant, but for now I rely on a verbal explanation.

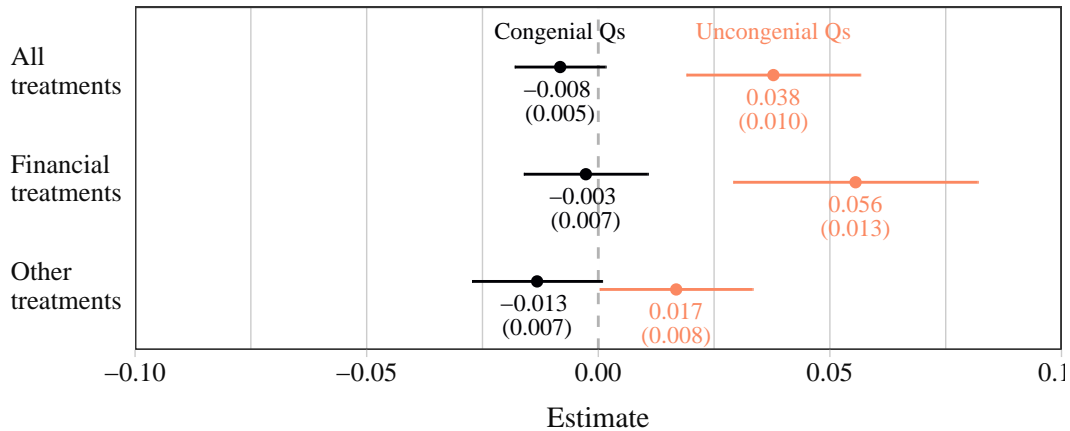
On congenial questions, the motivational account predicts that partisan and accuracy motivations push in the same direction. Consequently, a treatment that alters the mix of motivations should not have much effect on accuracy. A reduction in partisan motivation may pull respondents away from the correct answer, but the concurrent increase in accuracy motivation should pull them back toward it. In contrast, the insincerity account views respondents as choosing whether to honestly report their sincere belief or the partisan response. When the sincere belief is less accurate than the partisan response, heightened motivation to be honest should make the response less accurate on average. In this way, a treatment that works primarily by heightening sincerity should be expected to result in less accurate responses to congenial questions. By contrast, a treatment that works primarily by altering the mix of partisan and accuracy motivations in the formation of a belief should not be expected to have much effect one way or another.

The two accounts make the same prediction about uncongenial questions, whose correct answer is inconvenient for the respondent's party. According to the insincerity account, expressive responding hides knowledge of such answers; a treatment that counteracts expressive responding encourages respondents to reveal their hidden knowledge. Though the motivational account again considers both partisan motivation *and* accuracy motivation, these should be expected to work in the same direction. Just as in the insincerity account, the reduction in partisan motivation encourages more correct answers. Accuracy motivation also encourages more correct answers. In sum:

- The insincerity account predicts that treatment effects on partisan bias come from both congenial and uncongenial questions. In response to treatment, respondents should give less accurate answers to congenial questions and more accurate answers to uncongenial questions.
- The motivational account predicts that treatment effects on partisan bias are driven by uncongenial questions. In response to treatment, there should be little effect on congenial questions, where the decline in accuracy due to reduced partisan motivation is offset by the increase in accuracy motivation. Most of the effect should be concentrated in uncongenial questions, where reduced partisan motivation and increased accuracy motivation should push responses in the same direction.

To test these expectations, I examine effects on the subset of questions whose correct answers lie on the extreme ends of the response scale. In these cases, the correct answer is unambiguously

Figure 6: Accuracy Effects by Question Congeniality



congenial to one party, and uncongenial to the other party, in the sense that the most partisan answer is also the most (un)congenial answer. By contrast, when the correct answer is somewhere in the middle of the scale, there is no clear prediction about whether partisan motivation will make responses more or less accurate, and vice versa.

To make these criteria more concrete, consider the following example. At the time of this writing, the unemployment rate in the United States is about 4 percent, with Democrats controlling the presidency. This gives Democrats a partisan incentive to say that the unemployment rate is low. If the low point on the response scale were 4 percent, the most partisan answer that a Democrat can give is also the correct answer, making the question clearly congenial to Democrats. This means that partisan and accuracy motivations would both serve to pull Democrats toward the correct answer, and a change in the relative balance of partisan and accuracy motivation would not have much effect. However, if the low point on the response scale were 0 percent, the most partisan answer would no longer be correct. In this case, partisan motivation could make Democrats more or less accurate, and accuracy motivation could make Democrats more or less partisan. In this way, the lack of complete alignment between the most accurate and most partisan responses prevents one from setting clear expectations about how the two sources of motivation should affect responses.

Based on these definitions, I estimated the effects on accuracy separately for congenial and uncongenial questions. The results appear in Figure 6. As predicted by both theories, there is strong evidence of effects on uncongenial questions. However, there is only suggestive evidence of an effect on congenial questions. Across all treatment conditions, the meta-analytic estimate suggests that accuracy declined by an estimated 0.008. This is one-fourth the magnitude of the effect on congenial questions, and is statistically insignificant (s.e. = 0.005). Evidence for an effect on congenial questions is weaker still among financial treatments (estimate = -0.003, s.e. = 0.007). In contrast, there is suggestive evidence that non-incentive treatments reduce accuracy on congenial questions (estimate = -0.013, s.e. = 0.007).

These results are broadly consistent with the Motivation section's arguments regarding which mechanisms can be ruled in and out on theoretical grounds. For financial treatments, the motivational account could not be dismissed theoretically. The results for these treatments are also more

supportive of the motivational account: the effects are concentrated among uncongenial questions, just as one would predict if decreasing partisan motivation and increasing partisan motivation had opposite effects on congenial questions. In contrast, for the non-incentive treatments, the cheer-leading account was more plausible than the motivational account on theoretical grounds. For that set of treatments, there are about equal effects on congenial and uncongenial questions.