LSQ
LEGISLATIVE STUDIES QUARTERLY
Published by the Department of Political Science at Washington University in St. Louis

AYSE ELDES
Princeton University
CHRISTIAN FONG
University of Michigan
KENNETH LOWANDE
University of Michigan

Information and Confrontation in Legislative Oversight

Committees can use oversight hearings to collect and communicate the information Congress needs to oversee the bureaucracy, but many worry that members instead focus on scoring political points by lambasting witnesses. We leverage the collective judgment of congressional staff to measure how exchanges between legislators and witnesses vary on two separable dimensions: information and confrontation. Contrary to the conventional wisdom that confrontation crowds out information, we show that members of the president's party that engage in less confrontational oversight and reveal no more or less information than their peers.

Oversight has been declared "an essential and appropriate auxiliary to the legislative function" by the Supreme Court. Its purpose, according to the Court, is plain from the fact that "[a] legislative body cannot legislate wisely or effectively in the absence of information" (273 US 135 [1927]). Congress "needs to ascertain the facts and identify and analyze the relevant issues. It needs to investigate" (Levin and Bean 2018). Information is supposed to be the goal.

But contemporary observers typically bemoan the work of Congress when it oversees the Executive Branch. Administrators say probes into their own work are invasive and time-consuming. The out-party says the investigations are a waste of Congress' agenda and resources. Both often regard oversight as politically motivated. Against the backdrop of this public conversation, studies show partisan patterns in legislative oversight across outcomes such as hearings (Aberbach 1990; Kriner and Schickler 2014; Kriner and Schwartz 2008; Lowande and Peck 2017; MacDonald and McGrath 2016; McGrath 2013), speech behavior (Bellodi 2021), and the testimony of witnesses (Ban et al. 2023).

Of course, the stylized fact that oversight increases or decreases in response to political incentives does not demonstrate it shirks its traditional purpose. It could be that the increased frequency of oversight introduces more information than otherwise would be released under unified government. On the contrary, the oversight undertaken during divided government may not

Author names are listed alphabetically.

DOI: 10.1111/lsq.12440

be designed to produce new information, but instead to produce opportunities for partisan messaging and grandstanding (Park 2021). To settle this, one has to know more about oversight than just when it occurs. Put differently, much is known about the frequency of inquiries, but little is known about their qualities.

We advance this area by introducing and measuring two separable dimensions of legislative oversight: confrontation and information. The first dimension, confrontation, describes the extent to which lines of questioning are intended to make witnesses embarrassed, angry, flustered, or uncomfortable. It makes the experience unpleasant for the relevant witnesses and draws contrasts between the witness and members. The second dimension, information, describes the degree to which the lines of questioning present or request facts and evidence (a la Esterling 2007). It contrasts with member statements that merely concern opinions, positions, or preferences. We argue these two qualities are not mutually exclusive. Oversight can be highly informational and highly confrontational simultaneously, but both qualities must be measured to relate patterns of oversight with its more normative purposes.[1]

We describe these qualities of legislative oversight in their fundamental unit: an exchange between legislator and witnesses. We examine oversight exchanges for legislators that take place in the two oversight committees over the 110th and 111th Congresses. Most studies collect long time series of countable oversight events and assess patterns. What we trade off in temporal scope, we gain in high-quality measurements based on the judgment of practitioners. Specifically, we recruited current and former congressional staff who organize oversight hearings; then, following Carlson and Montgomery (2017) and Park (2021), we aggregated their opinions using a pairwise comparison framework. There are two main innovations in our approach. First, we analyze oversight in strings of questions that, when presented out of context, might portray entirely different meanings. It is common for litigators, for example, to ask simple questions that set up more difficult and penetrating ones, which means considering individual instances of speech in unweighted isolation is misleading. Second, we leverage the judgment of practitioners with experience in oversight, rather than online workers. This approach required several methodological innovations broadly applicable for hard-to-recruit samples. To evaluate our experts, we also compare them to non-expert raters and show that they offer qualitatively different assessments of the texts.

Our approach leads to several important insights about the nature of legislative oversight. First, we show that far from being mutually exclusive, information and confrontation are distinct concepts and have a small, but statistically significant, positive correlation. Second, we identify some correlates of informational and confrontational oversight, and we find that they differ. For example, we use the presidential transition from George W. Bush to Barack Obama to replicate the finding that legislators' oversight becomes less vigorous when the President is a member of their party (Kriner and Schickler 2014; Kriner and Schwartz 2008), but we find that the presidential transition does not appear to have made oversight from either party any more or less

informational. We find that more extreme members engage in more informational oversight than their more moderate peers but, perhaps surprisingly, they are no more or less confrontational. These findings underscore the importance of treating information and confrontation as separable. Finally, we find that hearings with more confrontational exchanges are better attended by members than non-confrontational ones. This raises the possibility that confrontation itself may incentivize members to participate in oversight, thereby encouraging them to collect information to avoid embarrassing themselves in the hearings and exposing them to information revealed by other members.

We also reveal important stylized facts about the measuring qualities of legislative text. In companion surveys, we find evidence that the key dimension in question—information—is evaluated qualitatively differently by non-experts. Specifically, we find that non-experts rate information and confrontation as strongly, negatively correlated—the precise opposite of the experts. At the same time, we find no evidence that non-experts were simply lazy or less logically coherent. Both sets of raters spent similar time on the task and had similar minimum feedback arc sets, one measure of logical transitivity violations. We take these facts as evidence of what expertise buys with this approach—namely the ability to bring context that aids in judging the quality of information, whereas the confrontational nature or negative effect of an exchange is universally understood. This clarifies the research contexts where "crowd-sourced" measurements of legislative texts may be useful.

Our findings present a qualitatively different understanding of the politics of legislative oversight. In general, the findings suggest that partisan incentives do not defeat oversight's normative purposes. Put differently, partisan patterns in the frequency of oversight are consistent with the objective of introducing more information held in common by all legislators. This finding is consistent with theories of deliberation, which argue that moderate levels of disagreement lead to the introduction of falsifiable claims (e.g., Esterling 2011; Gutmann and Thompson 1996; Habermas 1984). Contrary to the conventional notion that confrontation-inducing partisanship damages the more substantive functions of a legislature, our study suggests they are often mutually reinforcing. Politicians who confront witnesses are often quite substantive and informational while doing so. Partisan disagreements may drive the tenor of these exchanges, but they may be no less useful, in terms of their utility for introducing public facts to Congress as a whole.

Partisanship and Oversight Quality

Partisan combat seems to be inseparable from legislative oversight. In the United States, the idea that legislative oversight is partisan actually predates the invention of political parties. As Kriner and Schickler (2014) note, in 1791, the first congressional inquest into a failed military operation allowed (future) Jeffersonians in the House of Representatives to blame officials appointed by President Washington. Kriner and Schickler go on to summarize the incentives behind cases like these:

While nominally defending the institutional prerogatives of Congress, the president's partisan opponents can use high-profile investigations in an orchestrated attempt to sour public perception of the president's party. (Kriner and Schickler 2014, 26)

Politicians sometimes say as much out loud. Representative Kevin McCarthy (R-CA) said publicly that the House Select Committee on Benghazi was a success because it damaged Hillary Clinton's prospects in the 2016 presidential election.[2] Senator Ron Johnson (R-WI), in the course of investigating former Vice President and then-2020 presidential candidate Joe Biden, said "what our investigations are uncovering […] will reveal that this is not somebody that we should be electing president of the United States."[3]

These political motives seem at odds with the idea that oversight should be driven by the desire for good government. Levin and Bean (2018), for example, equate the effectiveness of oversight itself with the degree to which it involves both parties, on the grounds that diverse viewpoints improve Congress' information-gathering. Legal scholars see oversight as a core function of the legislature. According to Weich (2019), for example, it is "not a game. […] Congress cannot carry out its constitutional duties without the power to investigate whether the laws it enacts are being faithfully executed and whether the money it appropriates is being properly spent." The key question is whether these normative conceptions about proper constitutional arrangements and the balance of power are compatible with the apparent partisan incentives.

Existing research, by-and-large, is not designed to answer this question. Instead, it has mostly confirmed empirical patterns in oversight that portend partisan motives. Ogul (1976) does this with a collection of case studies. Aberbach (1990) finds divided government is associated with increased committee oversight from 1961 to 1977. Kriner and Schwartz (2008) find split-party control increased the number and duration of investigatory hearings in the House and Senate held between 1946 and 2006, while Lowande and Peck (2017) find the same for the House from 1789 to 1948. Lee (2013) finds that partisan "team play" has become more prevalent in oversight of the national debt. An exception is McGrath (2013), who argues "ideological conflict may significantly influence oversight above and beyond specifically partisan concerns" (364). Another exception is oversight of foreign relations, which according to Fowler (2015) tends to be bipartisan—even if muted and on the decline. But more recently, others have shown similar partisan patterns beyond hearings, most notably in patterns of speech (Bellodi 2021; Park 2021) and in the presence of categories of witnesses (Ban et al. 2023).

Collectively, these findings suggest partisan incentives operate on members' decisions to conduct oversight. However, on their own, they are not meant to settle questions about the quality of these inquiries or their implications for

the separation of powers. Most often, studies speak only to patterns in countable instances of oversight. But oversight itself is heterogeneous, as any cursory examination of legislative hearings confirms. Most obviously, most oversight does not rise to the political salience of famous and often cited investigations, such as Harry S. Truman's investigations into defense appropriations during World War II. Relatedly, most do not uncover scandals or controversy that warrants consistent media coverage.

But more importantly, there are obvious behavioral differences in how members conduct themselves in the course of their oversight duties. Some members pitch "soft ball" questions to witnesses, feeding them queries easily answered, apparently designed for the purpose of portraying the witness (and themselves) in a positive light. Other members are more combative, designing their questioning for precisely the opposite. This variation, in and of itself, suggests that increased oversight might occasionally be a *positive* sign of partisan team play. MacDonald and McGrath (2016), for example, argue these competing motivations lead to divergent patterns in the frequency of oversight hearings depending upon the life cycle of a presidential administration. These aggregate patterns in committee behavior provide suggestive evidence. But ultimately, the motivations of individual members should be most apparent in their own behavior (Lowande 2018).

What this implies, in our view, is that some means of characterizing these differences in the behavior of individual legislators is needed. It cannot be limited to appearances, presence, or the volume of oversight, because oversight itself is neither uniformly negative nor positive in valence. In other words, we must describe how members' behavior differs to better understand whether partisanship is incentive-compatible with quality oversight.

Dimensions of Oversight

We argue that legislative oversight varies on two key dimensions that are separable but jointly relevant: information and confrontation. Though applied in this instance to legislative oversight, we think these index foundational concepts in the study of Congress and of American politics in general.

We define the first dimension, information, as the degree to which questions and comments directed at the witnesses either present or request facts, rather than opinions, positions, or preferences. This follows most closely with the work of Esterling (2007), who used falsifiability as a yardstick for whether congressional questions were "analytical." For example, a committee chair might engage in a factual exchange with a witness by asking "between today and the end of this administration does CMS plan to propose regulations that would cut Federal Medicaid payments to States for targeted case management services?" Alternatively, they may make a statement that contains a fact: "You recently announced that 100 new carriers would be brought on board in Chicago." In these cases, the question is directed at the witness and asks for factual information. The statement is a statement of fact, which the witness then responds to and corrects. This

differs from questions like "how important is it for the Congress to appropriate the necessary funds in 2008—in the fiscal year 2008 for the 2010 census?" or statements like "The difference between directing where our science goes and what we search and free speech is not a simple thing and is subject to direction by policy."

We see this dimension as the principal feature of normative arguments in favor of legislative oversight. It solicits facts that may help Congress to come to its own conclusions about possible legislative action, and it seeks to reduce the information asymmetry between legislators and executives. It is an endeavor for Congress to develop its own expertise. As work dating back to Gilligan and Krehbiel (1987) shows, even though soliciting the opinions of outside experts can be helpful for developing high-quality policies, Congress is in a much stronger position if it knows the facts for itself.

Our conception of information includes both questions that attempt to extract information from the witnesses and statements that offer information for the record, because both can provide the floor with a firmer informational foundation for policymaking. One legislator may prefer to extract information through questions at a hearing while another prefers to query the executive branch, stakeholders, and experts in private settings and then use a hearing to publicly convey their findings. We see no reason to privilege the former over the latter. Both reference the same legislative function. However, it is important to distinguish informational content in oversight questioning, as we have defined it, and the information itself. Ours is a study of legislators' attempts to reveal information, not the success or failure of those attempts, which would require a broader theory that incorporates the responsiveness of witnesses.

Confrontation is something else. We define this dimension as oversight intended to make the witness embarrassed, angry, flustered, or uncomfortable. This kind of orientation toward witnesses is immediately apparent from many exchanges. Take one example:

> Mr. Lynch: … I am hearing hedging, I am hearing some defenses about information not being available. This kid was 19 years old, 19 years old. He gets a $300 million contract, taxpayers' money from the United States of America. That is a disgrace. I don't hear that from the panelists. I am hearing defense of different individuals. Has anybody been fired for this? Can I ask the panel, anybody get their walking papers for what has happened here? Has anybody been fired?
>
> Mr. Parsons: No, sir. No one has been for instance fired.
>
> Mr. Lynch: I am sorry?
>
> Mr. Parsons: No one has been fired.
>
> Mr. Lynch: Well, that is a shame.

This kind of dialogue differs from neutral or even comforting exchanges found elsewhere. Mr. Lynch is upset and would like the audience to know it. He confronts the witnesses for their previous responses. He makes them repeat a particularly damaging fact for effect. His behavior attempts to draw contrasts between himself and the witness. He is good, they are bad, and it is an outrage. This kind of exchange is common and seems designed for purposes that differ entirely from the fact-based, informational content. It can benefit the questioner by producing favorable media coverage—either traditional news coverage, or by going viral on social media. It can also become an informal means of punishing the witness. As many practitioners point out, most failures and errors in the executive branch are not criminal—or even grounds for firing. That means a congressional hearing, with public airings of failures, is the only form of accountability available to legislators and constituents.

Each of these potential benefits points to a dimension of behavior with different conceptual character. Theoretically, we see confrontation as a dimension of oversight which—in contrast to the release of information—is zero-sum. By confronting (or not confronting) a witness, the member is taking a position on their relationship with that witness. They are communicating a preference or their level of agreement with the person in the chair across from them. Some members attack the witness, some are neutral, and others line up in defense. Most obviously, this is a dimension of spatial disagreement common to virtually every theory of policymaking.

Most importantly, these dimensions are distinct and not mutually exclusive. In the Lynch-Parsons exchange, for example, the questioner does ask for information. There are counterfactual scenarios with the same level of confrontation, but more or less information is requested. Similarly, Lynch could have posed the same question, with the same informational content, in ways that did not highlight a fact possibly unfavorable to the witness. In summary, we think these dimensions are distinct and expect oversight exchanges that take on all combinations of values for each. How frequently exchanges are, for example, high on both information and confrontation, is an empirical question.

These examples also help us distinguish our two-dimensional conception of member behavior in oversight from the concept of "grandstanding," most recently examined by Park (2021). Confrontation has a negative valence. In the oversight context, however, performative displays or "grandstanding" can orient the member toward the witness either positively or negatively. That is, members may be low in confrontation as they grandstand toward a friendly witness, or high in confrontation as they grandstand against a hostile one. Thus, because of the applied setting, we see our framework as non-competing and categorically distinct from Park (2021).

Ultimately, the utility of specifying these distinct dimensions is that they clarify the conditions under which partisan patterns in oversight are damaging to its more neutral goals. They offer a clearly defined set of interpretations for

existing evidence on oversight. We highlight one before going on to describe our measurement approach and data. A reasonable possibility is that information and confrontation are, on balance, substitutes, and that confrontational content is responsive to the same patterns present in oversight frequency. In other words, confrontational exchanges produce inherently fewer facts for public display, and because confrontation is incentivized by differences of opinion, it should be more prevalent among the party that does not control the target—in this case, the executive branch. This is the "worst case" scenario, from the perspective of observers, as it implies that the enhanced oversight does not produce the common benefits extolled in judicial proceedings and op-ed columns.

Alternatively, information and confrontation might be complementary or unrelated. Suppose, for example, that analogous to insights from deliberation theory, confrontation is associated with the introduction of more falsifiable claims (a la Esterling 2011; Habermas 1984). In this case, confrontational behavior on the part of a member might be indicative of interpersonal disagreement, which leads them to simultaneously be more informational. This presents somewhat rosier scenarios for oversight. It might be that confrontation is still tied to partisan incentives, but that this has little effect on information. Again, these are empirical questions, which may recast in interpretations of existing research on congressional oversight.

Measuring Qualities of Oversight

To study these qualities of oversight, we examine 5-min question-and-answer sessions allocated to committee members, which we call *partitions*. Though hearings often start with opening statements from the chair and ranking member, in general, each member of the committee gets 5 min to question the witnesses. Members have broad discretion in how to use their time. They might spend most of their time allowing witnesses to answer their questions, they might spend their whole time on a monologue and not ask any questions at all, or they might alternate between making their own remarks and asking questions of the witnesses.

These partitions give a complete picture of how the member uses their time. Some previous work has used utterances—that is, each uninterrupted chain of speech from a legislator. These have the advantage of being shorter. Because they are shorter, there is less opportunity for them to mix different latent traits, such as opinions and falsifiable facts. This makes them easier, faster, and less subjective to code. Their shorter length also makes it easier to use machine learning to generalize the decisions of human coders to a larger data set. This makes them well-suited to applications where the research design requires a very large sample.

But utterances have several disadvantages. First, utterances can omit crucial context for evaluating a legislator's speech. For example, "I'll look forward to your written responses to the questions I asked earlier" could be friendly or confrontational, depending on how adversarial the underlying questions were.

Second, utterances vary in length. Some members may make only a single utterance, either because they give a speech or because they ask a question and listen silently to the answers. Others may make many utterances, either because they repeatedly interrupt the witnesses, the witnesses or other members repeatedly interrupt them, or because they engage in a back-and-forth dialogue with the witnesses. Comparisons between an utterance that absorbs all of one member's time to one of several utterances a member makes during their 5 min could be misleading.

Most importantly, examining utterances in isolation ignores the conventional structure of witness questioning by members. Many members are lawyers, and many questions are prepared by legal counsel. It is common for members to set up particularly damaging or important questions with minor or seemingly unimportant ones. Treating each utterance as independent ignores this structure and assumes that all utterances are equally important within a string of questions.

Coding each entire 5-min period (partition) as a unit avoids these issues. These partitions contain all of the context necessary to evaluate the member's speech and each member is entitled to the same 5 min. If one member's speech is coded as more informational than another, we can be confident that this difference is attributable to substantive differences in the member's behavior and not stylistic differences in how the member divides their time into utterances. This is a crucial and necessary property for our application.

Before discussing our measures, it is also important to note that we are interested in *member* behavior, rather than the behavior of witnesses. Our approach minimizes the role of witnesses by pruning their responses from partitions. The role of witnesses in hearings is a subject well-treated by Ban et al. (2023). We study the preparation and execution questioning by of members of Congress, as distinct from the *reactions* or *responsiveness* of witnesses. The responsiveness of executive branch officials to question is an important topic of study in its own right, and necessarily beyond what we can address here.

Pairwise Comparisons

Focusing on two latent features of each text—how informational and how confrontational the partition is—presents a major challenge. There is no existing, commonly held scale. As Carlson and Montgomery (2017) argue, it is difficult to extract reliable cardinal measures of latent features like information and confrontation by asking about them directly. If we simply asked survey respondents to rate each partition on a scale from 0 to 100, respondents might have consequentially different understandings of what constitutes a 73.

Accordingly, we follow Carlson and Montgomery (2017) and Park (2021) in adopting a comparison-based framework. We ask survey respondents to

compare two partitions and rate which of the two is more informational and which of the two is more confrontational. These choices are distinct; a survey respondent is free to select one of the partitions as both more informational and more confrontational. Although a coder compares only two partitions at a time, aggregating incremental comparisons creates a continuous scale. This comparison-based framework relies on the much weaker assumption that different respondents will generally agree on which of the two partitions is more confrontational and which of the two is more informational. Ideally, the most accurate placement of a partition's amount of confrontation can be found by comparing it to every single line of questioning in the time frame and committees of interest.

To ensure that our coders worked within the same definition, we provided them with detailed instructions with examples (see Appendix A). Instructions to our coders state that a partition is informational insofar as the legislator presents or requests facts and evidence. A partition is confrontational insofar as the legislator attempts to make witnesses embarrassed, angry, flustered, or uncomfortable. For each coder, we also scheduled their coding and provided real-time clarification over the phone.

Expert Panel

Our second key innovation is the group of raters we call upon to code our texts. We fielded a survey to former staffers of the oversight committees, who we call "experts" as a shorthand. These experts' experiences make them extraordinarily qualified to code our two main variables. They understand the committee's role in the legislative process and the range of questions legislators ask, so they are particularly attuned to the degree to which texts present or request facts and evidence. They can accurately assess whether a member's behavior in a partition is intended to make a witness feel embarrassed, angry, flustered, or uncomfortable because they have themselves written questions designed to make witnesses feel embarrassed, angry, flustered, or uncomfortable.

We were able to recruit 10 experts to make 1565 comparisons. The details for precisely how the comparisons were generated are available in Appendix A. We also recruited an 11th expert to independently compare 306 pairs of partitions that had already been compared by the original experts. This expert's choices agreed with the original experts' 72% of the time on information and 86% of the time on confrontation (see Appendix G for more details). We find no evidence that experts were biased in favor of their own party—rating oversight by copartisans as more informational and less confrontational (see Appendix D).

Recruiting expert coders proved to be exceptionally difficult. We first obtained a list of 41 Democrat and Republican contacts with experience in the Senate and House of Representatives. Each of these was obtained by referral. We sent an initial invitation to each one with a $150 consultation honorarium and followed up with those who showed interest. Only 10

agreed to code; two volunteered without honorarium. This process took several months, from November 2021 to February 2022. The final sample is a mix of Democrats and Republicans from both chambers; however, there are more individuals who worked for Democrats in the Senate. Appendix B presents an anonymized summary of the characteristics of the experts we recruited.

Another notable feature of this approach is that it allows us to investigate whether there are any advantages associated with experts. We also fielded a survey to two other populations of coders: five research assistants and 69 students in an undergraduate course on the American legislative process. The research assistants coded 2773 comparisons. The students coded 1,133 comparisons. Like the experts, the research assistants and students did not get any context about the project or its goals besides the survey instructions.

Scaling and Bridges

It would be prohibitively expensive to compare each text to every other text. Scaling 183 partitions in this way (the number we ultimately scaled) would require 16,653 comparisons—more than 10 times as many as we were able to collect. Accordingly, most applications of Montgomery and Nyhan (2017) compare each text to a small sample of other texts.

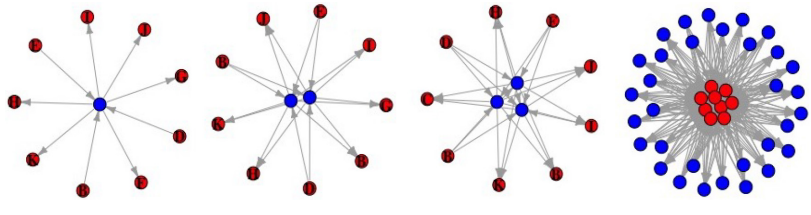
However, this poses its own problem: the researcher might end up with distinct clusters of texts with no way of comparing across clusters. Even if we know Partition A is more informational than Partition B and that Partition C is more informational than Partition D, we cannot say anything about whether Partition A is more or less informational than Partitions C and D. Perhaps Partition A is more informational than both, perhaps it is less informational than either, or perhaps it is more informational than D but less informational than C. It's not just that we cannot precisely rank A, B, C, and D. We have no information whatsoever about how A and B compare to C and D. To say something about how they compare, we would need some link between the two pairs, such as a comparison between Partitions A and D. But if the number of comparisons is very small relative to the number of possible comparisons, there is a good chance the researcher will end up with disconnected islands of texts and with no information about how texts on different islands compare to one another.

Previous researchers have addressed this problem by only gathering comparisons for a pre-defined subset of the texts. For example, Park (2021) takes a sample of 3000 texts and compares each of those texts to 20 other texts within the sample. However, this solution requires the researcher to know in advance how many texts their coders will be able to compare. We could not know ex ante how many experts we would be able to recruit, or how many partitions they would be willing to code. If we were optimistic and made our sample too large, we would end up with disconnected islands of texts.

Instead, we adopted a new approach that would ensure all texts could be placed on a common scale, even though we did not know in advance how many texts our coders would be able to compare. In our setup, coders compared each new text to the same set of nine bridging texts. These bridging texts place all new texts on a common scale. If we imagine (for the sake of simplicity) that the nine bridging texts are ordered from least informational to most informational, we might find that Text A is more informational than the first four bridging texts while Text B is more informational than the first eight bridging texts, which would allow us to conclude that Text B is more informational than Text A. Even if we found that Texts A and B were both more informational than the first four bridging texts but less informational than the other five, we would at least be able to characterize roughly how informational they are, even though we would not be in a position to offer a comparison between A and B. What's more, this method does not require the analyst to know how many texts will be coded in advance. Once the bridging texts are chosen, the coders can integrate as many new texts into the sample as they can. This procedure has the added benefit of making coders more productive. Once they have become familiar with the bridging texts, they can rapidly code new texts, because they do not need to re-read the bridging texts each time.

Figure 1 represents the procedure graphically. Each text is represented by a node in a graph, and each edge between nodes indicates a comparison. The edge points from the less informational text to the more informational text. The red nodes are the bridging texts. First, the coder takes a new text (the blue node) and compares it to each bridging text. Then, they repeat the process with a new text, over and over. None of the blue nodes have edges to one another because they are never directly compared, but the graph still conveys a great deal about how informational the blue nodes are relative to one another because they are all indirectly connected through the bridging texts. In the language of graph theory, this procedure creates a bipartite graph in which every non-bridging text is compared with every bridging text.

FIGURE 1
One Coder's Comparisons are Shown Above
Note: The leftmost diagram shows the coder's initial comparison of Text A, indicated in blue, to the nine red nodes representing the bridging texts. An arrow points from the less informational text to the more informational text. To the left is the diagram of the coder's second comparison denoted by another blue node. The rightmost diagram shows every text A that the coder compared with the red bridging set.



It is important to choose the bridging texts wisely. If all of the bridging texts have similar levels of confrontation, the coders' comparisons will sort the texts into those that are more confrontational to the bridging texts and those that are less confrontational, but they will not allow the analysis to make fine-grained comparisons. Accordingly, we read a sample of texts and identified nine that represented many different levels of information and confrontation. We then iteratively updated a handful of the bridging texts throughout the course of the experiment to obtain better coverage over the full range of information and confrontation.

Hearings Data

To answer our main research questions, we require repeated observations of the same legislators in different contexts. That, together with the difficulty and expense of recruiting experts, necessitates some trade-offs in the scope of our data. We draw our data from the hearings held in the House Committee on Oversight and Government Reform and the Senate Committee on Homeland Security and Government Affairs (along with their subcommittees) in the 110th and 111th Congresses (2007–2010). Although many congressional committees occasionally conduct oversight hearings, these two hold primary responsibility for investigation and oversight within their respective chambers, and most of their hearings are focused on oversight and investigation.

In total, during this period, these committees held 673 oversight hearings (see Appendix E for a precise definition of what constitutes an oversight hearing). We subsetted these hearings to the 492 that included at least one government witness. Some oversight hearings draw in testimony from outside experts, which we expect to follow a distinct strategic process that differs from the partisan combat found when there is at least one representative of the executive branch present. After selecting the partitions that would serve as bridges, we randomly sampled from these hearings for our coding exercise. Ultimately, we were able to measure information and confrontation in 183 partitions coded by experts, 303 coded by research assistants, and 168 coded by students in an undergraduate course.

The set of hearings appears to be a fairly representative cross-section of oversight hearings, with some weighting toward salient issues of the day. There are hearings that deal with aspects of the Iraq War, with more relating to Afghanistan and the broader war on terror. There are as many on the 2007–2008 financial crisis and associated bank failures. The sample also includes, for example, oversight of crop programs, environmental regulation, elections and voting systems, firearm regulation, D.C., defense contracts, border control, disease outbreaks, product recalls, and routine oversight of particular government agencies. Online Appendix C shows that the topics covered during this period were reasonably representative of the topics covered by oversight committee hearings between 1995 and 2020.

Estimation

Once we obtain the comparisons from our coders, we must assign a numeric score for how informational and confrontational each text is. We adopt the Bayesian approach of Carlson and Montgomery (2017). Let i and j index the partitions and k index the comparison. If the kth comparison is between i and j, let $y_k^{\text{info}} = i$ if the coder rates Partition i as more informational than Partition j and $y_k^{\text{info}} = j$ if the coder rates j as more informational than i. We suppose $Pr\left(y_k^{\text{info}} = i\right) = \frac{\exp\left(\alpha_i^{\text{info}} - \alpha_j^{\text{info}}\right)}{1+\exp\left(\alpha_i^{\text{info}} - \alpha_j^{\text{info}}\right)}$.[4] That is, α_i^{info} and α_j^{info} are latent parameters that reflect how informational Partitions i and j are, respectively. The more informational Partition i is, the higher α_i^{info} and the more likely it is to be chosen over other partitions. We define α_i^{conf} analogously for confrontation.

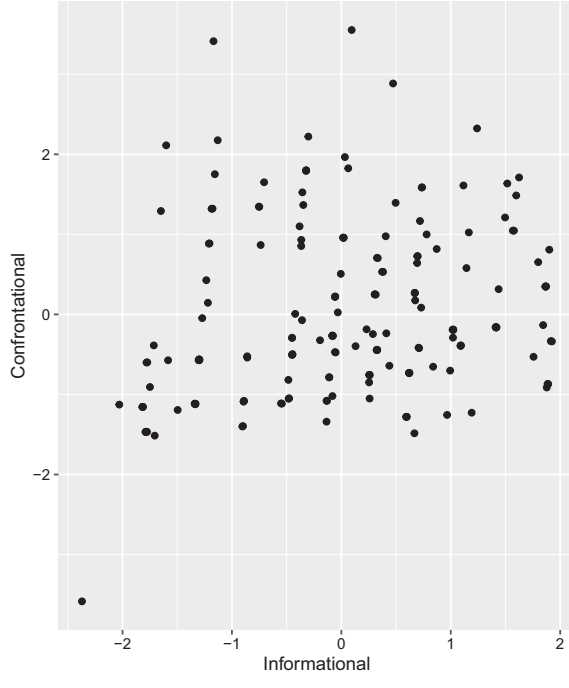
We fit this Bayesian model on the comparisons generated by our coders, with priors

$$\begin{pmatrix} \alpha_i^{\text{info}} \\ \alpha_i^{\text{conf}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix}\right)$$
$$\sigma \sim \textit{Uniform}(\text{-}1, 1)$$

σ is a hyperparameter that characterizes the correlation between information and confrontation within the partitions. If σ is close to 1, then high information implies high confrontation. If it is close to -1, then high information implies low confrontation. If it is close to 0, the two are uncorrelated. σ has a diffuse uniform prior, so it is posterior distribution comes from the observed correlation in the data.

We run four chains with a 5000 iteration burn-in and then sample 5000 draws from each chain, for a total of 20,000 draws from the posterior distribution of the α's and σ. The α's (how informational and confrontational each partition is) are not known with certainty; we can merely sample from their posterior distributions. If we simply plugged the posterior mean of each α into a standard linear regression, our standard errors would be too small, because they would not account for the fact that we don't know the α's for sure and that our estimates of α might change if we coded more data. We propagate this uncertainty into all of our regressions by bootstrapping. For each bootstrap iteration, we sample from the posterior distribution of α and then, conditional on those α's, resample complete observations from our data. In this way, our bootstrap procedure accounts for uncertainty from α as well as uncertainty from the regression model. The coefficients we report are the means of these bootstrapped draws and the 95% confidence intervals are from the empirical distribution implied by the bootstrapping procedure.

FIGURE 2
Information and Confrontation in Each Partition (Expert Coders)
Note: Each dot is the posterior mean of α_i^{info} and α_i^{conf} for a given partition, based on the experts' codings. There does not appear to be a strong negative correlation between the two.



Information and Confrontation in Oversight Hearings

Figure 2 summarizes the draws from the posterior distribution by plotting the posterior means of α_i^{info} and α_i^{conf} for each of the partition coded by the experts. α_i^{info} ranges from −2.37 to 1.92, with a mean of 0.02 and a standard deviation of 1.07. α_i^{conf} ranges from −3.58 to 3.55, with a mean of −0.00 and a standard deviation of 1.08.[5]

If information and confrontation are two poles of a single dimension, then confrontation and information should be negatively correlated, and the points in Figure 2 should lie on a line with a negative slope. That is not what we find. There does not appear to be a strong correlation between information and confrontation.

We can evaluate this claim more rigorously by examining the posterior distribution of σ, the hyperparameter that explicitly models the correlation between information and confrontation in the partitions. Table 1 shows that experts reported that partitions that are more confrontational also tend to be

TABLE 1
Relationship Between Information and Confrontation (σ). According to experts, partitions that tend to be more informational also tend to be more confrontational. The point estimates are the posterior mean for the hyperparameter σ and the 95% credible intervals are reported in parentheses below the point estimates. The p-values are for a two-tailed test

	Experts	Research Assistants	Students
Correlation Between Information and Confrontational (σ)	0.136*	−0.100*	−0.915***
	(−0.001, 0.272)	(−0.215, 0.014)	(−0.971, −0.829)
Observations	183	303	168

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

more informational. Although the correlation hyperparameter is positive, it is not so large to suggest that information and confrontation are synonymous, so the two appear to be separate dimensions.

Notably, this pattern only emerges from expert coders. Table 1 shows that if we perform the exact same exercise with research assistants or undergraduate students, the coefficient is negative and statistically significant. The coefficient for correlation research assistants is not too large, so the research assistants agree with the experts that information and confrontation are separate dimensions; they just disagree with whether they are positively or negatively correlated. The undergraduate students, however, reported a large negative correlation, which suggests that they uniquely believe information and confrontation are opposite poles of a single dimension.

This raises an important point on measurement. For reasons we have already discussed, we believe that the expertise of congressional staffers is important for our enterprise. They possess specialized knowledge that helps them assess the texts in the same way that the relevant political actors would. We would not have uncovered this positive relationship between information and confrontation without their assistance.

However, it is difficult to recruit experts for the survey, so they limit the sample size. This raises a bias-variance trade-off. Our study should be understood as an exercise in what would happen if we minimized bias; we recognize, however, that for other questions and research designs, the variance reduction associated with using research assistants or crowdsourcing could outweigh the bias they introduce.

Does the experts' belief that confrontation and information are separable—even positively correlated—make sense? To assess face validity, we describe one prototypical example that the experts found substantially more informational than the research assistants and one that the experts found substantially less informational. The first example is a medium information, high confrontation exchange between Henry Waxman and Alan Greenspan in the wake of the 2007 financial crisis (witness lines have been shortened):

> Chairman Waxman: Dr. Greenspan, Paul Krugman, the Princeton Professor of Economics who just won a Nobel Prize, wrote a column in 2006 as the subprime mortgage crisis started to emerge. He said, "If anyone is to blame for the current situation, it's Mr. Greenspan, who pooh-poohed warnings about an emerging bubble and did nothing to crack down on irresponsible lending." He obviously believes you deserve some of the blame for our current conditions. I would like your perspective. Do you have any personal responsibility for the financial crisis?
> Mr. Greenspan: Well, let me give you a little hist […] to be fairly major problems in predatory lending.
> Chairman Waxman: Well, he urged you to move with the power that you as chairman of the Fed, as both Treasury Department and HUD suggested, that you put in place regulations that would have curbed these emerging abuses in subprime lending. But you didn't listen to the Treasury Department or to Mr. Gramlich. Do you think that was a mistake on your part?
> Mr. Greenspan: Well, I questioned the facts of tha […] o a subcommittee of the Federal Reserve board—-
> Chairman Waxman: Dr. Greenspan, I am going to interrupt you. The question I had for you is you had an ideology. You had a belief that free, competitive–and this is shown–your statement, "I do have an ideology. My judgment is that free, competitive markets are by far the unrivaled way to organize economies. We have tried regulation, none meaningfully worked." That was your quote. You have the authority to prevent irresponsible lending practices that led to the subprime mortgage crisis. You were advised to do so by many others. Now, our whole economy is paying its price. You feel that your ideology pushed you to make decisions that you wish you had not made?

Waxman quotes a leading economist's explanation behind the crisis and puts the witness in an uncomfortable situation by asking him to respond. He also quotes Greenspan's statements from hearings in 1994, 1997, and 2002. Additionally, he references the Treasury and HUD's advice to Greenspan on preventing subprime lending. Waxman's questions follow well-researched statements and ask Greenspan to defend himself against causing an economic crisis that affected millions. Experts gave this exchange an average of 0.09 on information and 3.55 on confrontation, whereas research assistants gave it −1.42 on information and 3.34 on confrontation. The two groups have similar confrontation scores, but experts said the exchange was about average on information while the research assistants thought it was over a full standard deviation below average. Greenspan does ask some questions that solicit opinions rather than information. For example, "Do you think that was a mistake on your part?" seeks Greenspan's opinion on his accountability. However, Waxman asks opinion questions after

long recitations of relevant, falsifiable facts. The experts were able to identify this information amid the confrontation and dramatic opinion-based questions; the research assistants were not.

Waxman's line of questioning contrasts sharply with the following example is part of an exchange between Darrell Issa, Admiral Thad Allen, and Elaine Duke. Issa asks for their opinion on the potential for interagency collaboration and effectiveness of equipment.

> Mr. Issa: Well, following up on the sea fighter, you know, it was commissioned I guess now it is going on 2 years ago. It spent a lot of time dockside. How much more do you have to go through to find out what the advantages of this high-speed ferry, its air landing capability, fueling, etc., how much more is there before you know whether to build unit two?
> Admiral Allen: Well, if I could, I would like to g […] t be current. But happy to answer for the record.
> Mr. Issa: OK, I will put you on the spot where I can, though. How do you like it as a ship? How do you like it as a new category?
> Admiral Allen: I think it has some intriguing pote […] at. But those are the things we might talk about.
> Mr. Issa: I appreciate that. I certainly recognize that the air conditioned down below capability is very good. Switching to Ms. Duke, now, you work for the Secretary of Homeland Security, is that right?
> Ms. Duke: Yes.
> Mr. Issa: How do you interface–the Commandant made it clear that he feels he has the authority to have these liaisons and joint operations, but you work for a single cabinet officer who has a budget. What is it that you can do in your daily life, or can't do, that allows you to leverage other hundreds of contracts and contractors in the rest of the Federal system?
> Ms. Duke: Well, as the senior contracting person f […] s chaired by the OFPP Administrator, Paul Denett.
> Mr. Issa: OK. But does it have shortcomings? You know, today we are talking about whether or not there should have been a better integration of fleet Navy assets in this acquisition. Your organization was certainly part of the process of looking at your brethren in other procurements and saying, you know what, they have some expertise we should bring in to reduce the chances of exactly what has happened here happening. So what went wrong?
> Ms. Duke: Well, our focus from the contracting per […] ing able to bring the business deals to fruition.

Experts and research assistants scored this partition, similar scores on confrontation: −0.57 for experts and −0.45 for research assistants. Although Issa isn't making the witnesses uncomfortable or embarrassed, he also isn't praising them or agreeing with their opinions. The groups of coders differ in

their perception of information, though. Experts gave an average informational score of −1.30 while research assistants gave an average score of 0.67 to this partition. The non-experts may be distracted by jargon or the specificity in the opinion questions, causing them to rate this text as more informational than experts. However, this does not change that Issa is asking them to evaluate equipment, capabilities, and accountability. He does not seek responses that are falsifiable, and the statements he uses to preface his questions do not present much falsifiable information either.

Complete examples of varying levels of informational and confrontational partitions can be found in Appendix E.

Why Did the Experts Code Differently from Non-Experts?

Non-expert coders identified a qualitatively different relationship between information and confrontation than the experts did. One possibility is that the experts simply coded more carefully than the non-experts did. They might have been more interested in the texts and spent longer carefully assessing how informational and confrontational each partition was. The non-experts, bored by the texts, may have collapsed information and confrontation to make an unpleasant task easier. Alternatively, the non-experts might have coded less carefully because there were more demands on their time or because they are overconfident in their initial reactions.

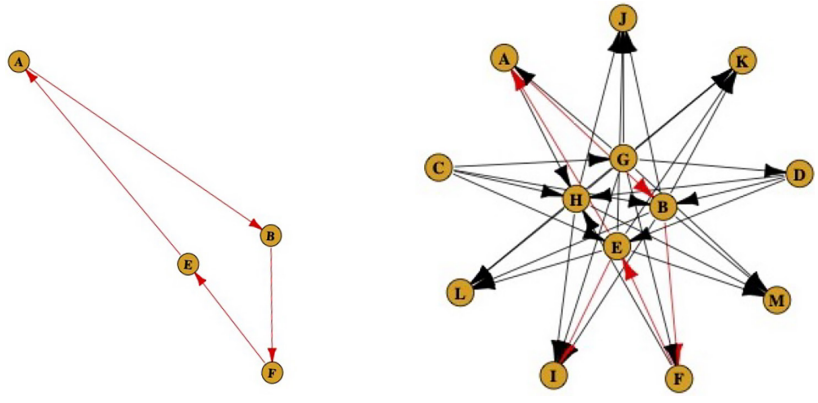
If the differences between the groups of coders stem from different levels of care in coding the texts, then we also ought to observe one group offering more logically consistent comparisons than the other. In particular, if one group was much more careful, then its comparisons should exhibit fewer transitivity violations. If a coder chose Partition B as more informational than Partition A and chose Partition C as more informational than Partition B, we expect them to choose Partition C as more informational than Partition A. If a coder chooses otherwise, there exists a transitivity violation in their comparisons.

To count the frequency of these transitivity violations, it is helpful to represent each coder's comparisons with a graph. Again, each text is represented by a node, and each directed edge represents a comparison. The edge points from the less informational (or depending on the graph, confrontational) text to the more informational text.

If there is a transitivity violation in the coder's choices, then there will be a cycle in the graph. Returning to the previous example, if Partition C was chosen over Partition B, Partition B was chosen over Partition A, and Partition A was chosen over Partition C, then there is an arrow that points from C to B, then another arrow that points from B to A, then a third arrow from A to C. We can trace a path that follows the directions of the arrow C to B to A and then back to C—a cycle.

There may be many such cycles in a given coder's graph, so to measure the transitivity violations of each graph in Figure 3, we calculated the length of its minimum feedback arc set. The minimum feedback arc set is the number of edges in the graph that must be removed so that the graph has no cycles.

FIGURE 3
Illustration of Transitivity Errors
Note: On the left is a directed acyclic graph. As indicated by the arrows, A is less confrontational than B, B is less confrontational than F, and F is less confrontational than E. Because an arrow points from E to A, A is both more and less confrontational than B. The logical error is a transitivity violation. On the right, this cycle is shown with red arrows among an expert's total comparisons. Without this violation, Text I would be more confrontational than A because it is more confrontational than E. Due to the cycle, Text I's relationship with A is unclear.



Greater lengths indicate more cycles in the original directed graph, which means the coder made more transitivity violations. If a graph's minimum feedback arc set length is 0, this means the graph is acyclic; the coder did not make logical mistakes in comparing the texts. If a graph's minimum feedback arc length was two, we would need to throw out two of that coder's comparisons to rationalize that coder's choices.

We found no significant difference in lengths of minimum feedback arc sets between experts and non-experts. Therefore, the differences in coding outcomes between the two groups do not appear to follow from one group thinking about their comparisons more carefully, and the experts are not more logically consistent than the research assistants or the students. Rather, experts appear to be looking for something qualitatively different when they assess how informational a line of questioning is. We have good reason to defer to these assessments of these experts, who have actually participated in the process, as to what counts as informational oversight.

Partisan Teams, Information, and Confrontation

Oversight is best characterized by the fact that it involves some *target*, often within the executive branch. Conventional understandings of oversight imply that confrontation tends to increase when the president is a member of

TABLE 2
Relationship Between Information, Confrontation, and Co-partisanship with the President.
Members' partitions are less confrontational when the president is a member of their party. The point estimates are the posterior means for the coefficients of the Bayesian model and the 95% credible intervals are reported in parentheses below the point estimates. The *p*-values are for one-tailed tests

	Informational	Confrontational
Member of President's Party	−0.029 (−0.425, 0.371)	−0.733*** (−1.164, −0.297)
Legislator FE	✓	✓
Observations	183	183

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

the out-party and that political teamsmanship crowds out information-oriented activities. Table 2 shows that this is not the case. Our data include a presidential transition from the Republican George W. Bush in the 110th Congress to the Democrat Barack Obama in the 111th Congress. This analysis regresses information and confrontation on whether the legislator associated with the partition is a member of the president's party. The specification includes legislator-level fixed effects, so this draws on within-legislator variation over the presidential transition. Confrontation decreases when the legislator's party controls the White House. However, co-partisanship with the president does not have a statistically significant effect on information. Legislators seek and reveal information at the same rate regardless of whether their party holds the presidency.

Although the presidential transition allows us to credibly estimate the effect of co-partisanship with the president with legislator fixed effects, the limited temporal scope of our data (2007–2010) prevents us from applying this strategy to study a broader set of interesting factors. Table 3, therefore, presents the results of a statement-level regression of confrontation and information on the legislator's characteristics. We regress our measures of confrontation and information on a vector legislator characteristics. Specifically, we are interested in features of legislators with less over-time variation that may nonetheless be associated with the qualities of members' oversight: measures of institutional standing (i.e., chamber, committee role, and seniority), political orientation (i.e., district competition and ideology), and background (i.e., sex and profession). These speak to fundamental questions about how legislators approach oversight—the extent to which it is predicated on stable differences across different legislators, political context, or individual capacity. Several broad themes emerge.

First, just like in the previous analysis, members of the president's party engage in less confrontational oversight, but they are no more or less informational. In short, this finding is strong enough to be robust to two different research designs—one with legislator fixed effects and another with explicit control variables.

TABLE 3
Predictors of Information and Confrontation.
The point estimates are the means for the coefficients over 5000 bootstrap draws, and the 95% confidence intervals are reported in parentheses below the point estimates. The *p*-values are for one-tailed tests

	Confrontational	Informational
President's Party	−0.614***	0.009
	(−1.050, −0.183)	(−0.418, 0.450)
Minority (Republican)	0.045	−0.453**
	(−0.445, 0.545)	(−0.956, 0.051)
Senate	−0.366	0.059
	(−0.990, 0.239)	(−0.553, 0.681)
Committee or Subcommittee Leader	−0.243	0.223
	(−0.797, 0.294)	(−0.312, 0.792)
Seniority	0.061**	0.037*
	(−0.003, 0.127)	(−0.018, 0.092)
Safe District	0.481	0.079
	(−0.174, 1.148)	(−0.581, 0.725)
|Nominate Score|	−0.489	0.323
	(−2.216, 1.138)	(−1.377, 1.938)
Male	−0.315	−0.133
	(−0.909, 0.290)	(−0.718, 0.463)
Lawyer	0.383	0.271
	(−0.123, 0.908)	(−0.250, 0.778)
Business Owner	0.659*	0.936**
	(−0.349, 1.632)	(0.060, 1.761)
Career Politician	0.127	0.474*
	(−0.468, 0.746)	(–0.125, 1.069)
Intercept	0.297	−0.521
	(−0.764, 1.326)	(−1.520, 0.501)
Observations	183	183

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Second, no variable has statistically significant and opposite signed effects on information and confrontation, but two have statistically significant effects with the same sign. More senior legislators engage in oversight that is both more informational and more confrontational, which suggests that both of these kinds of oversight might require costly preparation that is easier for more experienced legislators. Likewise, business owners are both more informational and more confrontational, which bolsters an already substantial literature that shows the importance of personal experiences on legislative action (Burden 2007; Washington 2008). These two findings of factors that are positively correlated with both information and confrontation underscore the importance of treating confrontation and information as two separate dimensions.

Third, this analysis replicates and elaborates upon a key result from Park (2021). Park finds that members of the minority party are more likely to grandstand. In Table 3, being a member of the minority party is negatively correlated with information but uncorrelated with confrontation. In the

context of oversight, the political reward for revealing information is lower for minority party members, who are unlikely to gain access to the floor agenda.

It is important to remember that these results are based on a small sample that includes little over-time variation. More data might uncover additional associations, and the associations we do find with this simple regression design might be attributable to selection bias, reverse causation, or aliasing. The results should therefore not be taken as definitive statements about the determinants of informational and confrontational oversight. Rather, they are intriguing suggestions that illustrate how it can be useful to distinguish between information and confrontation as separate dimensions rather than opposite ends of a single spectrum.

Does Confrontation Subsidize Information?

Finally, we consider the implications of this insight that information and confrontation are not mutually exclusive and may actually be positively correlated. This changes the normative status of confrontation in oversight. Far from crowding out information-oriented behavior, it may actually subsidize it. One possible mechanism is by drawing more legislators to hearings.

Legislators are generally free to skip hearings whenever they please. Preparing for a hearing is time-consuming, so legislators will not attend unless they will be rewarded for doing so. Some legislators might be drawn to attend for informational reasons—either to reveal information that could inform policymaking on the floor or to gather information that might be useful for crafting their own policy proposals. However, these sorts of rewards are most valuable to those with some leverage over policy or some base of expertise; they are probably less compelling for less-experienced, lower-capacity legislators.

However, the opportunity to attract media attention and take popular positions by excoriating unpopular witnesses may draw otherwise uninterested legislators to attend. Getting them to attend could have two informational benefits. First, confronting the witness without proper preparation is a dangerous exercise. The legislator might inadvertently say something that reveals their ignorance of the policy area and then goes viral, such as when Representative Steve King confronted Google CEO Sundar Pichai about content his granddaughter encountered on an iPhone, only for Pichai to reply that iPhones were made by a different company. To avoid these sorts of gaffes, legislators may acquire some information about the relevant policy area, even if their main goal is to confront the witnesses. Second, since confrontation does not necessarily crowd out information, if they listen to other legislators during the hearing, they might be exposed to the information being presented or extracted by the other hearing participants, which could lay the foundation for more information-oriented participation in future hearings.

To test whether confrontational hearings attract more legislators and whether informational hearings attract fewer, we conduct an analysis at the hearing level. We score how informational and confrontational a hearing is by taking the average of the scores of its partitions. Of the 492 hearings in our data set, there were 140 hearings for which the experts scored at least one partition.

In most cases, there was only one partition scored in the entire hearing, but since the experts coded a random sample of partitions, our approach gives as an unbiased, albeit noisy, measure of how informational and confrontational the entire hearing was. For each hearing, we regress the number of legislators who attended the hearing on how informational it was, how confrontational it was, whether it was a subcommittee hearing and whether it was a Senate hearing, and, to account for the fact that these last two control variables might be subadditive, the interaction of whether it was a subcommittee hearing and a Senate hearing.[6] Appendix H presents a robustness analysis that conducts a beta regression on the proportion of the committee or subcommittee that attended the hearing. As with all previous analyses, we account for uncertainty in our estimates of how informational and confrontational the hearings are by sampling from the posterior distribution of our partition scores for our bootstrapped standard errors.

Table 4 shows that legislators are indeed less likely to attend hearings that end up featuring highly informational partitions and are more likely to attend hearings that end up featuring confrontational partitions.[7] Of course, these results should not be interpreted to imply that information and confrontation at a hearing cause attendance; the exchanges between legislators and witnesses take place after legislators have already decided to attend. Rather, the topic of the hearing, the witnesses invited, and the political context make informational or confrontational approaches more attractive. Presumably, hearings on technical subjects with knowledgeable but obscure witnesses make information more attractive, and Table 4 suggests fewer legislators attend such hearings. Conversely, hearings on subjects that are interesting to the mass public with high-profile and unlikable witnesses make confrontation more attractive, and Table 4 suggests that more legislators attend these types of hearings.

This finding raises an interesting possibility. Perhaps some legislators attend highly informational hearings that they would otherwise skip because those hearings are also highly confrontational. Perhaps, either in preparing for the hearing or in listening to their fellow legislators, they get exposed to information, and perhaps that exposure makes them conduct more informational oversight in the future. In short, opportunities for confrontation could subsidize the provision of information.

This possibility relies on a number of assumptions. First, it assumes that the same legislators who shy away from informational hearings are attracted to confrontational hearings, and that the effects in the attendance analysis are not attributable to two distinct and non-overlapping groups of legislators. We could test this possibility by conducting a legislator-level analysis in which we interact legislator-level characteristics with how informational and confrontational the hearing is. If we find that certain kinds of legislators are repelled by information and drawn to confrontation, that would be consistent with the argument that confrontation subsidizes information. We conduct such an analysis in Appendix H, but, because our sample is so small, the confidence intervals are too wide to draw any conclusions.

TABLE 4
Information, Confrontation, and Attendance.
Hearings with confrontational partitions attract more committee members to attend. Hearings with informational partitions attract fewer committee members. The point estimates are the means for the coefficients over 5000 bootstrap draws and the 95% confidence intervals are reported in parentheses below the point estimates. The *p*-values are for one-tailed tests

	Number of Legislators Speaking at Hearing
Informational	−0.467*
	(−1.118, 0.129)
Confrontational	0.687**
	(0.074, 1.358)
Subcommittee Hearing	−9.086***
	(−11.522, −7.319)
Senate Hearing	−9.412***
	(−11.522, −7.319)
Senate Hearing × Subcommittee Hearing	7.393***
	(5.101, 9.753)
Observations	140

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Second, the conjecture assumes that drawing a legislator to a highly informational, highly confrontational hearing does not cause them to skip a hearing that would have been even more informational. If it did, then any information they gained from attending the one might not make up for the information they lost by skipping the other. Our sample is not suited to testing for this possibility, because it only tracks the oversight committee, and there might be spillovers across committees. However, with a wider data set, it would be testable. A researcher could exploit the fact that legislators are members of multiple committees. Suppose Legislator A is on Committees X and Y while Legislator B is on Committees Y and Z. If attending confrontational hearings crowds out attending other informational hearings, then if hearings in X are more confrontational than hearings in Z, we would expect Legislator A to be less likely to attend informational hearings in Y than Legislator B would be.

Third, the conjecture assumes that legislators learn from attending highly informational, highly confrontational hearings, either through preparation or through exposure to information from other legislators, and that what they learn leads them to conduct more informational oversight in the future. This could be tested by regressing the change in how informational a legislator's oversight is from one congress to the next on the number of highly informational, highly confrontational hearings they attended in the earlier congress. A positive relationship would support the conjecture. Unfortunately, our sample is too small to support such a design. It covers only two congresses, and there are not many legislators with scored partitions in both congresses.

Given the limitations of our sample, we can only raise the possibility that confrontation subsidizes information by getting members to participate who otherwise would not. A more complete test of this hypothesis would require a longer panel, coding a higher proportion of the partitions within each hearing, and expanding the data to include additional committees rather than just oversight committees.

Implications for Research on Oversight

Members of Congress routinely use hearings to confront bureaucrats. These confrontations can help politicians win reelection or cow troublesome bureaucrats into compliance, but they could conceivably distract legislators. Another key role of hearings is providing the floor with information that will help it craft well-informed, high-valence policy. Consistent with models of democratic deliberation, confrontation may actually facilitate attempts to acquire that information. We find that legislative oversight can be informational and confrontational at the same time. Legislators are more likely to attend hearings that provide ample opportunities to confront witnesses, but many confrontational hearings are also informational. Attending highly informational, highly confrontational hearings could encourage them to acquire some information to competently confront the witnesses and exposes them to information that other members reveal during their time.

When considered alongside existing empirical findings on legislative oversight, our study helps adjudicate between alternative scenarios with entirely different normative meanings. It has long been known that the frequency and apparent vigor of oversight changes depending upon the parties at the helm of Congress and the Presidency. This raised the possibility that the apparent surplus of oversight under divided government might simply be an expression of partisan teamsmanship that added little substance to debates in Congress. Our results suggest that even though the confrontational tenor of oversight responds in this way, that same oversight is no less informational. Put more succinctly, members do not simply abandon facts in service of politics.

To discover the often complementary relationship between information and confrontation, we needed to leverage experienced practitioners to evaluate texts. Paid research assistants and undergraduate students conflated confrontation with the absence of information. Insofar as the evaluations of the research assistants and undergraduate students are more representative of how the public evaluates congressional oversight, this raises the intriguing possibility that the public might systematically misperceive how informational congressional oversight is. If most of the oversight the general public encounters in the mass media is highly confrontational, they might conclude that it is not very informational, even if it is actually trying to provide Congress with important facts. As a result, the public could infer that the institution is more dysfunctional than it actually is.

Given the importance of this point, and the novel means by which it was uncovered, we close with a series of methodological recommendations that will aid in further research. Recruiting practitioners was time-consuming and expensive, so we had to restrict the scope of our analysis to a sample from two committees over two congresses. Even this analysis required significant innovations to existing coding protocols to squeeze as much value as possible from the practitioners' time (Carlson and Montgomery 2017; Park 2021).

Fortunately, the relationships we documented were so robust that we were able to reliably measure them even with a small sample. However, there are many other questions and research designs that would require a much wider span of data, including the designs we proposed to test about how opportunities for confrontation could increase informational oversight over the long run. It would be infeasible to get a significantly larger set of coded partitions from the practitioners, so how can researchers leverage our discoveries moving forward? One appealing option would be to use the practitioner-labeled partitions to train a machine learning classifier to predict information and confrontation for all other committees and all other congresses. However, we believe this will not work. Many of the partitions are quite long, and the cues for how informational a partition is are often subtle. Even with sophisticated natural language processing models, it will not be possible to train and validate an accurate supervised model with only a couple hundred labeled examples.

The other, more viable alternative is to train non-experts to code more like the practitioners. Through conversations with the practitioners and careful automated and manual analysis of their coding decisions, we may be able to articulate many of the features practitioners look for that others miss. If we could successfully train non-experts to look for those features, we would be able to massively expand the scope of our data. We plan to explore this possibility in future research.

Data Availability Statement. Data will be made available in a public repository upon acceptance/publication of the paper. The only limitation is that we cannot share the identities of our expert coders.

Ayse Eldes *is a PhD student in the Department of Politics at Princeton University.*

Christian Fong *is an assistant professor in the Department of Political Science at the University of Michigan.*

Kenneth Lowande *is an associate professor in the Department of Political Science at the University of Michigan.*

ENDNOTES

This project was funded by the Levin Center at Wayne State University Law School, the Sunwater Institute, the Center for Effective Lawmaking, and the College of Literature, Sciences, and the Arts at the University of Michigan. We are grateful for helpful comments from Luca Bellodi, Kevin Esterling, Jennifer Selin, and Julia Park. We are responsible for all remaining errors.

1. We discuss how these dimensions relate to other behaviors, like "grandstanding" (Park 2021), in the third section of this study.

2. Alex Moe and Carrie Dann. "McCarthy Links Benghazi Panel, Clinton's Sinking Poll Numbers," NBC News, Sept. 30, 2015.

3. Craig Gilbert. "A GOP colleague questions Sen. Ron Johnson's investigation of Joe Biden," Milwaukee Journal Sentinel, September 17, 2020.

4. Montgomery and Carlson have an additional parameter for how discerning each coder is, but because we have so few coders and so few comparisons in our application, we suppress this parameter and assume all experts are equally discriminating.

5. The summary statistics for partitions coded by the research assistants was similar: a range of −1.81 to 2.49 for information with a mean of 0.00 and a standard deviation of 1.00, and a range of −1.99 to 3.34 for confrontation with a mean of −0.00 and a standard deviation of 1.08.

6. We measure attendance by the number of legislators who spoke at the hearing. We assume a members do not often listen to other legislators without taking their own opportunity to speak.

7. Appendix F shows that we would not have found either of these results if we had used the data generated by the research assistants instead of the experts.

REFERENCES

Aberbach, Joel. 1990. *Keeping a Watchful Eye: The Politics of Congressional Oversight*. Washington, DC: Brookings Institution Press.

Ban, Pamela, Ju Yeon Park and Hye Young You. 2023. "How Are Politicians Informed? Witnesses and Information Provision in Congress." *American Political Science Review* 117:122–139.

Bellodi, Luca. 2021. *Is Partisanship Bad for Bureaucratic Accountability?* https://lucabellodi.com/material/partisanship_oversight.pdf

Burden, Barry C. 2007. *Personal Roots of Representation*. Princeton, NJ: Princeton University Press.

Carlson, David and Jacob M. Montgomery. 2017. "A Pairwise Comparison Framework for Fast, Flexible, and Reliable Human Coding of Political Texts." *American Political Science Review* 111:835–843.

Esterling, Kevin M. 2007. "Buying Expertise: Campaign Contributions and Attention to Policy Analysis in Congressional Committees." *American Political Science Review* 101:93–109.

Esterling, Kevin M. 2011. ""Deliberative Disagreement" in U.S. Health Policy Committee Hearings." *Legislative Studies Quarterly* 36:169–198.

Fowler, Linda L. 2015. *Watchdogs on the Hill: The Decline of Congressional Oversight of U.S. Foreign Relations*. Princeton, NJ: Princeton University Press.

Gilligan, Thomas W. and Keith Krehbiel. 1987. "Collective Decision Making and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures." *The Journal of Law, Economics, and Organization* 3: 287–335.

Gutmann, Amy and Dennis Thompson. 1996. *Democracy and Disagreement: Why Moral Conflict cannot be Avoided in Politics, and What Should be Done about It*. Princeton, NJ: Princeton University Press.

Habermas, Jurgen. 1984. *The Theory of Communicative Action, VolumeOne: Reason and the Rationalization of Society*. Boston, MA: Beacon Press.

Hitt, Matthew P., Craig Volden and Alan E. Wiseman. 2017. "Spatial Models of Legislative Effectiveness." *American Journal of Political Science* 61:575–590.

Kriner, Douglas L. and Eric Schickler. 2014. "Investigating the President: Committee Probes and Presidential Approval, 1953–2006." *The Journal of Politics* 76:521–534. https://doi.org/10.1017/S0022381613001448

Kriner, Douglas L. and Liam Schwartz. 2008. "Divided Government and Congressional Investigations." *Legislative Studies Quarterly* 33:295–321. https://doi.org/10.3162/036298008784310993/abstract

Lee, Frances E. 2013. "Presidents and Party Teams: The Politics of Debt Limits and Executive Oversight, 2001–2013." *Presidential Studies Quarterly* 43:775–791.

Levin, Carl and Elise J. Bean. 2018. "Defining congressional oversight and measuring its effectiveness." *Wayne Law Review* 64:1–22.

Lowande, Kenneth. 2018. "Who Polices the Administrative State?" *American Political Science Review* 112:874–890.

Lowande, Kenneth and Justin Peck. 2017. "Congressional Investigations and the Electoral Connection." *Journal of Law, Economics, and Organization* 33:1–27.

MacDonald, Jason A. and Robert J. McGrath. 2016. "Retrospective Congressional Oversight and the Dynamics of Legislative Influence over the Bureaucracy." *Legislative Studies Quarterly* 41:899–934. http://mcgrath.gmu.edu/wp-content/uploads/2015/08/MM-retrospectiveoversight.pdf

McGrath, Robert J. 2013. "Congressional Oversight Hearings and Policy Control." *Legislative Studies Quarterly* 38:349–376. https://doi.org/10.1111/lsq.12018

Montgomery, Jacob M. and Brendan Nyhan. 2017. "The Effects of Congressional Staff Netowrks in the US House of Representatives." *The Journal of Politics* 79:745–761.

Ogul, Morris S. 1976. *Congress Oversees the Bureaucracy: Studies in Legislative Supervision*. Pittsburgh, PA: University of Pittsburgh Press.

Park, Ju Yeon. 2021. "When Do Politicians Grandstand? Measuring Message Politics in Committee Hearings." *Journal of Politics* 83:214–228.

Ting, Michael M. 2011. "Organizational Capacity." *Journal of Law, Economics, and Organization* 27:245–271.

Washington, Ebonya L. 2008. "Female Socialization: How Daughters Affect Their Legislator Fathers." *American Economic Review* 98:311–332.

Weich, Ronald. 2019. "Five Myths About Congressional Oversight." *The Washington Post* 1. https://www.washingtonpost.com/outlook/five-myths/five-myths-about-congressional-oversight/2019/01/04/5b3b2a58-0f9c-11e9-831f-3aa2c2be4cbd_story.html

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Data S1.

Appendix A: Survey Design
Appendix B: Sample Recruitment
Appendix C: Topics Covered by Hearings
Appendix D: Partisan Bias Among Experts
Appendix E: What is an Oversight Hearing?
Appendix F: Replication with RAs Scores
Appendix G: Inter-Coder Reliability
Appendix H: Additional Attendance Analyses
Table A1. Count of Topics Covered by Hearings with At Least One Partition Scored by Expert Coders
Figure A1. Frequency of Topics in Sample Compared to 1995–2020
Figure A2. Evaluation of Partisan Bias in Experts' Comparisons
Table A2. Relationship Between Information and Confrontation (Research Assistants)
Table A3. Predictors of Information and Confrontation (Research Assistants)
Table A4. Information, Confrontation, and Attendance (Research Assistants)
Table A5. Beta Regression for Information, Confrontation, and Attendance
Table A6. Who Stays Away from Informational Hearings?