

Congressional Oversight Overseas*

Rachel Yu[†]

September 8, 2026

Abstract

Congressional oversight requires information about policy implementation and executive activities, but legislators often struggle to obtain this information, particularly when information asymmetries with the executive branch are most acute. I argue that tools of direct observation are useful for legislators to gather firsthand information about policy implementation, which can shape legislative behavior in downstream oversight venues. Focusing on foreign policy oversight, I study congressional travel abroad in the form of congressional delegations (CODELs), which enable legislators to directly observe overseas operations and question relevant stakeholders on the ground. Using data on CODELs and House Foreign Affairs Committee hearings from 1999-2006, I evaluate how CODELs shape and inform downstream, public oversight activity within congressional hearings. I find that while CODELs do not mobilize hearing participation, they can inform how legislators engage in downstream oversight. Members who travel abroad are more likely to ground their hearing remarks in firsthand observations and use language that is more detailed and relevant to foreign policy issues. These findings have implications not only for how well-informed legislators are when they engage in oversight, but also for how Congress can reduce its reliance on executive branch accounts of its own performance.

*Many thanks to Jim Curry, Emily Ommundsen, and Jason M. Roberts for graciously sharing their data on congressional travel and to Susan Stanley, Aleena Amir, Erick Guzman, and Elaine Yang for excellent research assistance. For helpful comments and discussions, thanks to Pamela Ban, Michael Bednarczuk, Lawrence Broz, Ben Kinnard, Michael Joseph, David Lake, Ben Noble, Sean Theriault, Yujia Wan, Alex Zhao, and audiences at Yale CSAP, MPSA and the Junior Americanist Workshop. This material is based upon work supported by the NSF Graduate Research Fellowship Program under Grant No. DGE-2038238.

[†]Ph.D. Candidate, University of California, San Diego; Email: rquyu@ucsd.edu; Web: <https://rachelqyu.github.io/>

1 Introduction

The US withdrawal from Afghanistan in August 2021 involved the evacuation of approximately 124,000 people, a loss of \$7.1 billion in military equipment, and hundreds of casualties.¹ Despite reporting requirements and the creation of a dedicated oversight body, Congress struggled to get timely and desired information from the executive branch about related policy decisions and their outcomes. In the days after the fall of Kabul, Rep. James Comer expressed this frustration, saying “since the White House either cherry-picks what questions to answer or completely fails to respond to questions, Republicans have had to find answers elsewhere.”² To find such answers, some members turned to using congressional delegations (CODELs), or official travel abroad, as an independent channel of firsthand information. During the withdrawal, Reps. Seth Moulton and Peter Meijer conducted a bipartisan visit to the Kabul airport, explaining that the CODEL was intended to “gather more information” and “provide oversight on the executive branch.”³

This example highlights the importance of Congress’s ability to gather information about policy implementation and outcomes as a part of the oversight process (McCubbins and Schwartz 1984; Levin and Bean 2018; Selin and Moore 2023). Much of the work on how Congress is informed on matters related to oversight focuses on information provided in the context of congressional hearings (Krehbiel 1991; Quirk 2005; Ban, Park, and You 2022). Given their importance to policy implementation, scholars have focused on the information transmitted by bureaucrats (Eldes, Fong, and Lowande 2023; Bellodi 2023; Ban, Park, and You 2025). Beyond hearings, others have emphasized the importance of information conveyed through informal mechanisms of oversight, such as direct communications between legislators and bureaucrats (Mills, Kalaf-Hughes, and MacDonald 2016; Lowande 2018; Ritchie 2023; Ritchie and You 2019; Lowande and Augustine Potter 2021).

These mechanisms, however often rely on reported information that is sourced by and transmitted through an intermediary source. As illustrated above, reported information is not always readily available and can often be filtered through the interpretations and incentives of that source. Thus,

1. https://www.stateoig.gov/uploads/report/report_pdf_file/aud-geer-24-21.pdf; <https://soaa.org/sigars-verdict-a-fghanistan/>; <https://foreignaffairs.house.gov/getting-answers-on-afghanistan-withdrawal>. Last accessed: August 6, 2026.

2. <https://oversight.house.gov/release/comer-republicans-will-gain-answers-on-botched-afghanistan-withdrawal-with-or-without-white-house/>. Last accessed: August 6, 2026.

3. <https://www.npr.org/2021/08/24/1030838666/2-congressmen-make-secret-trip-to-kabul-amid-u-s-withdrawal-chaos>. Last accessed: April 17, 2026.

legislators often have incentives use direct observation and collect their own oversight relevant information (Aberbach 1990). Mechanisms of direct observation, such as site visits (Garvey 2025) and inspections (Levin and Bean 2018), are important because they allow legislators to assess ground realities and gather information about policy conditions firsthand. Direct observation can still be subject of intermediary influences, but this approach offers legislators access to information closer to the point of policy implementation and potentially less mediated by the interpretations and interests of intervening actors.

In this paper, I focus on a key mechanism of oversight within the context of foreign policy, where Congress faces relatively large information asymmetries (Schlesinger 1973). I argue that CODELs, allow legislators to gather oversight relevant information directly. Through meetings with relevant stakeholders and observation of program implementation (Scoville 2013), CODELs provide legislators with direct access to the site of policy implementation. Such firsthand exposure is particularly valuable for contextualizing policy outcomes and understanding the parameters that define foreign policy success. Because CODELs equip legislators with this contextual grounding, the observations gained often inform and shape downstream oversight activity.

To illustrate this, I focus on how CODELs shape legislative behavior within relevant, subsequent congressional hearings. Tracing the full scope of CODEL interactions with foreign policy oversight is difficult given that much of the oversight activity is unobservable, but hearings provide an empirical window into this broader process. Where CODELs supply firsthand knowledge from abroad, hearings provide a forum to act on, interrogate, and disseminate that information domestically. Given this, I expect CODEL participation to complement hearing participation, with members who participated in CODELs to be more likely to engage in relevant hearings. Because CODELs are uniquely helpful for gathering firsthand observations, I also expect CODEL participants to be more likely to use discourse that reflects firsthand, policy relevant, and specific observations gained through travel. In other words, deeper policy and contextual understanding of foreign policy that is gained through travel should be reflected in statements that engage with the specific entities and observations a legislator could have observed.

To evaluate how CODELs inform foreign policy oversight, I consider how member-level CODEL participation shapes hearing participation and discourse within House Foreign Affairs Committee (HFAC) hearings. Disclosure data on CODEL participation allows me to examine when and where

a legislator traveled to.⁴ To get within-hearing measures, I collect all hearing transcripts from the HFAC from 1999–2006. First, I create a set of measures for hearing participation which capture whether and how often a legislator speaks and asks questions within a hearing. Second, I create two text measures of hearing discourse which are used to assess whether legislators who travel convey information that relies on firsthand observations and is more specific to foreign policy.

The first discourse measure uses a dictionary approach to capture the extent to which legislators make explicit references to firsthand observations gained through travel. The second measure captures hearing statements that are more foreign policy specific, or richer in specific references to specific entities that are directly relevant to foreign policy. Building on existing methods to measure latent traits within committee hearings (Carlson and Montgomery 2017; Park 2021), I employ and validated a large language model to code a sample of statements and then trained and optimized multiple machine learning models to predict a score of foreign policy specificity for all legislator statements. After validating this measure substantively and statistically, I combine the data on CODEL participation with measures of hearing participation and discourse, resulting in a dataset at the hearing-legislator level.

Using a within-legislator-region design, which holds constant the legislator’s innate proclivity to participate in CODELs and engage in hearings focused on a given region, I find that HFAC members who participated in a relevant CODEL in the prior month are not significantly more likely to participate in subsequent hearings. This result suggests that CODEL participation does not change whether legislators choose to speak up about a region within a hearing. When I disaggregate by committee context, CODEL participation is a positive and significant predictor of subcommittee hearing participation but not full committee hearing participation. Subcommittees are typically used for information provision and policy discourse, while full committee hearings are often used for political staging. This suggests that travel matters more in settings where information is actively being disseminated, and that the relevant question is not whether CODELs bring legislators into the hearing room, but what information they bring to the room once there.

I turn to this question by examining the effect of CODEL participation on hearing discourse. The results show that, compared to not traveling to a particular region, a legislator who travels

4. I am grateful to Jim Curry, Emily Ommundsen, and Jason Roberts for generously sharing their data on congressional travel.

on a CODEL to a particular region is (1) more likely to reference observations from a firsthand travel experience and (2), conditional on speaking, more likely to use language that is richer in foreign policy specificity in a subsequent, relevant hearing. To further clarify how CODELs shape hearing discourse, I provide a case study that leverages the COVID-19 pandemic as a plausible exogenous shock to CODEL travel. Focusing on Africa subcommittee member, Karen Bass, I trace her behavior in similar hearings before and during the pandemic to illustrate how her discourse changes with the absence of CODEL access. Altogether, these findings demonstrate that CODEL participation can inform what legislators say in the broader foreign policy oversight process.

This paper advances our understanding of how Congress gathers information in the face of persistent information asymmetries with the executive branch. Tools that improve members' independent access to information matter not just for members' individual competency, but also for how reliant Congress is as a whole on the executive as a source of policy-relevant information. I show that CODELs, as a method of direct observation, help supply legislators with information that the executive would otherwise dominate. This has implications not only for how well-equipped legislators are to effectively engage in oversight, but for Congress's ability, as an institution, to check the executive branch.

Second, where existing scholarship has read the decline of hearings as a weakening of foreign policy oversight capacity (Fowler 2015; Drezner and Fowler 2025), I show that hearing quality varies with upstream investments in information acquisition. Because hearings capture only a formal, public stage of the broader oversight process (Selin and Moore 2023), focusing on hearings in isolation risks an incomplete picture of foreign policy oversight. This study suggests that the path to effective oversight might depend on how oversight is performed at different stages of the process and how these stages interact with each other. In particular, CODELs can be used to inform and shape legislative behavior and policy discourse within venues of oversight such as hearings.

Finally, this paper also offers a different lens on congressional travel itself. Much of the existing literature treats CODELs as relational tools or personal junkets (Alduncin, Parker, and Theriault 2017; Curry and Roberts 2022; McGee and Theriault 2022), while I show that they also function as instruments of oversight.

2 Information, Oversight, and Direct Observation

Congressional oversight is defined as congressional efforts to “detect and remedy executive-branch violations of legislative goals” (McCubbins and Schwartz 1984, 165). Central to oversight is Congress’s ability to gather information about policy implementation and outcomes in order to detect potential “policy drift”. This is similarly emphasized by Selin and Moore (2023, 4) who conceptualize oversight as “the process of reviewing, monitoring, and supervising implementation of public policy by the executive branch”. In the context of oversight, gathering information about policy implementation and outcomes is essential in detecting whether agencies deviate from Congress’s intended policy outcomes. As Eldes, Fong, and Lowande (2023, 6) notes, one of oversight’s primary functions is an “endeavor for Congress to develop its own expertise”.

Much of the empirical work on congressional oversight and its informative qualities has primarily, and oftentimes solely, focused on legislative oversight hearings. Hearings are a salient and public venue of information collection and dissemination for lawmakers (Quirk 2005; Ban, Park, and You 2022; Eldes, Fong, and Lowande 2023) and its occurrence is often used as a measure for congressional oversight (Fowler 2015). Furthermore, scholars have found that empirical patterns in oversight hearings are largely driven by partisan and ideological disagreement between Congress and the bureaucracy (Kriner and Schwartz 2008; Parker and Dull 2009; McGrath 2013; Kriner and Schickler 2016; Lowande and Peck 2016; MacDonald and McGrath 2016).

Hearings, however, only capture the formal, public stage of the broader oversight process. The information gathering stage of oversight is often conducted through both formal and informal channels (Aberbach 1990), with the latter tending to be less visible to the public (Selin and Moore 2023). For example, Congress has access to independent agencies that are explicitly tasked with providing information that is related to oversight (e.g., CRS, GAO, IGs) (Levin and Bean 2018). Members themselves can also gather information through direct contact with bureaucrats (Mills, Kalaf-Hughes, and MacDonald 2016; Lowande 2018; Ritchie 2023; Ritchie and You 2019), through a range of substantive and procedural legislative tools (Lowande and Augustine Potter 2021), and through interest groups and civil society organizations (Hall and Deardorff 2006).

A common feature across these mechanisms is that they remain mostly reliant on accounts that are collected and reported by an intermediary, whether a bureaucrat, a reporting agency, or an inter-

est group. In this paper, I focus on another mechanism of oversight: direct observation. Legislators can use tools of direct observation to gather information from the site of policy implementation. Importantly, when legislators visit the sites where policy is implemented, they acquire independent knowledge of policy conditions firsthand without solely relying on official or third-party accounts (Levin and Bean 2018; Garvey 2025). The claim is not that direct observation is unmediated, but that it shortens the chain between legislators and the source of policy implementation.

While direct observation as an oversight mechanism applies broadly, these mechanisms are particularly helpful when sources of reported information are unavailable or unreliable. Foreign policy is a policy domain in which information asymmetries between Congress and the executive branch are large. The Executive branch controls various institutions and agencies designed to collect foreign policy information, giving it a large informational advantage over Congress within this policy domain. Canes-Wrone, Howell, and Lewis (2008, 5) describes this dynamic, writing that “because the White House and executive branch collect most information regarding foreign policy, presidents can better tailor the presentation of facts to suit their interests on these issues”. This offers a useful setting for examining how and why legislators employ mechanisms of direct observation.

I argue that legislators use congressional travel abroad, in the form of congressional delegations or CODELs, to conduct congressional oversight over foreign policy. CODELs are publicly funded trips abroad containing at least one Member of Congress⁵. CODELs allow legislators to acquire targeted and relevant information that is an important component of conducting oversight over foreign policy (Scoville 2013) which can shape and inform downstream foreign policy oversight.

3 Congressional Travel and Foreign Policy Oversight

How do CODELs serve as a tool for foreign policy oversight? I theorize that there are several challenges legislators face when conducting foreign policy oversight and that congressional travel provides key advantages in overcoming these challenges.

First, given the sensitive nature of foreign and national security, legislators may not be able to

5. This is different from privately funded travel that are typically sponsored by non-governmental organizations (i.e., gift travel (Rosenson 2009)) or foreign governments (i.e., FGDA or MECEA trips).

access as much and/or detailed information due to special classification restrictions (Joseph and Poznansky 2024). For example, all members of Congress can access generally classified information by virtue of their offices, but they are not automatically entitled to “Sensitive Compartmented Information” (SCI) access. Access to this information is granted on a “need to know basis” in which Congress and the executive branch differ on what constitutes a “need to know”. Typically, however, the executive branch, which controls access to information, has more influence in defining “need to know” (Hathaway et al. 2021). On occasion the legislative branch has expressed distrust of executive branch information, prompting efforts to seek their own information through travel (Scoville 2013). Travel provides an avenue through which legislators can gather missing information firsthand.

Second, evaluating foreign policy requires understanding the domestic bureaucratic and regulatory context of the country in which it is implemented. The success of any policy depends not only on what actions are permissible and feasible in a foreign setting, but also on which policy outcomes are most relevant for assessing effectiveness. Such evaluation, therefore, requires familiarity with local conditions and the limits within which foreign policy can operate. By enabling legislators to meet directly with executive branch officials stationed abroad as well as with foreign officials, CODELs provide access to individuals who possess broader networks within their host countries and more context-specific insights relevant to oversight. While legislators can gather general information about foreign policy from news outlets or interest groups, CODELs facilitate the acquisition of more detailed and firsthand knowledge. This is especially important for policies involving countries that are smaller, receive limited media or interest group attention, or are less accessible to civil society due to security concerns or domestic political constraints. As Scoville (2013, 343) notes, “these kinds of investigatory acts are helpful to the legislative process because they enable members of the House and Senate to ask their own questions, learn about foreign conditions through firsthand observation rather than third-party reports, and consequently form their own impressions about appropriate action”.

Third, challenges to foreign policy oversight are not solely limited to access to policy relevant information. When conducting foreign policy oversight, legislators must convince their colleagues that an issue warrants the chamber’s scarce time and resources. In making decisions over policy, legislators often rely on informational cues from credible and more knowledgeable colleagues

(Curry 2019; Box-Steffensmeier, Christenson, and Craig 2019). CODELs allow legislators to gather more targeted and specialized information than their non-traveling peers. The expertise acquisition is particularly helpful in building momentum for further oversight action, especially in a policy domain such as foreign policy. For legislators, “having a reputation for [foreign] policy expertise, along with the perception that MCs care more about accomplishing something... can go a long way in helping them to establish a degree of personal influence ” (Carter and Scott 2004, 35). In other words, legislators can use travel in an attempt to demonstrate to colleagues that a foreign policy issue is serious enough to merit increased oversight.

An illustrative example that highlights the informative qualities of such trips comes from a series CODELs to oversee congressionally appropriated military aid and reimbursements to Pakistan between 2007 and 2008⁶. During this period, 24 different House members participated in several bipartisan CODELs to Pakistan and Afghanistan to evaluate the use of congressionally appropriated funds allocated to the Pakistani government for military expenditures (U.S. House Subcommittee on National Security 2008).⁷

Oversight reports detailing the activities of these CODELs confirm that members met with senior officials from the US, NATO, and Afghan and Pakistani political and military officials. CODEL members gathered information that contextualizes the performance of US military aid in the region. For example, one report highlighted that the support of Pakistani military operations in the Federally Administered Tribal Areas of Western Pakistan had the unintended consequence of damaging the local, political administration of the region. These sentiments were explicitly expressed by leaders of the Awami National Party to the CODEL. Anecdotal evidence also suggests that CODEL members gained key pieces of information about how aid policy was being administered abroad. The reports referenced a conversation in which “a senior United States military official told a Subcommittee-sponsored Congressional delegation that he believed approximately 40 percent to 50 percent of the Coalition Support Funds have not reached the military components...and instead have been used for other Pakistani government priorities” (18).

6. The goal of this example is to reveal insights about the mechanisms that link CODELs with oversight activities.

7. Using data released by the house I identified all trip level disclosures to Afghanistan and Pakistan during the 110th Congress and cross-referenced the dates and members with those mentioned in subsequent hearings and reports.

4 Empirical Context: Congressional Hearings

To understand whether and how CODELs shape the broader foreign policy oversight process, I consider the extent to which legislators' behavior in congressional hearings is altered as a result of participating in relevant CODELs. Hearings are well-suited for assessing the informative effects of CODELs for a few reasons. First, where CODELs allow legislators to *gather* firsthand information from foreign contexts, hearings provide a public venue to *share* these findings. Oversight constitutes a series of actions designed to both gather information and act upon potential deviations from legislative goals (Aberbach 1990). The importance of CODELs as an oversight tool therefore depends on its capacity to shape the broader foreign policy oversight process, which should be evidenced in downstream venues such as hearings (Selin and Moore 2023).

Second, hearings are more likely than other legislative venues to feature oversight relevant discourse. In contrast to other forms of legislative communications, such as floor speeches or press releases, what legislators say in committees is more likely to be directed at policy-relevant stakeholders rather than constituents (Ban and Kaslovsky 2024; Grimmer and Powell 2013). While grandstanding (Park 2021) still occurs, information-seeking remains a prevalent part of committee hearings, even among ideologically extreme members and presidential out-partisans (Eldes, Fong, and Lowande 2023).

Third, hearings provide a relatively difficult case to observe the informative effects of CODELs. Tracing the full scope of CODEL effects is difficult because much of what follows a CODEL remains unobservable. Who legislators meet with and each CODEL's explicit purpose are not systematically disclosed⁸. Conversations with former and current congressional staffers reveal that CODELs often shape oversight in informal and less observable ways. One current staffer recounted how their member used a CODEL to press embassy officials about canceled foreign assistance programs, with the staffer tasked with following up State Department bureaucrats until a satisfactory response was received.⁹ A former committee staffer similarly noted that "traveling for oversight can help gain data points of how different embassies are otherwise structured or how they func-

8. Ex-post disclosure requirements do not require members to disclose the purpose of their travel (Scoville 2013)

9. Interview with current personal staffer, January 7, 2026. Interviews with former and current congressional staffers. Qualitative interviews were carried out in the Spring & Fall of 2025 and the Winter of 2026. Interviewees include former and current congressional staff members who have experience with a foreign policy related portfolio and members of the State Department who have worked with CODELs. All interviews were exempted by UCSD's IRB (IRB #812103).

tion, different priorities, and it just helps develop themes and how operations work or how priorities are.”¹⁰ This is not to suggest that hearings are the only venue through which CODELs shape oversight; rather, hearings offer a systematic and observable window into a process that is otherwise difficult to trace.

Legislators who participated in a CODEL are equipped with direct and firsthand observations. Hearings provide legislators a venue to disseminate these observations. Legislators returning from a CODEL have both greater familiarity with the issue at hand, a deeper understanding of policy specifics, and a stronger motivation to pursue them. I focus here on two observable implications of this theory: compared to their non-traveling counterparts, (1) legislators who participate on a CODEL should be more likely to participate in a relevant congressional hearing and (2) more likely to provide information that reflects specific, firsthand observations gained through travel.

5 Data and Descriptive Patterns

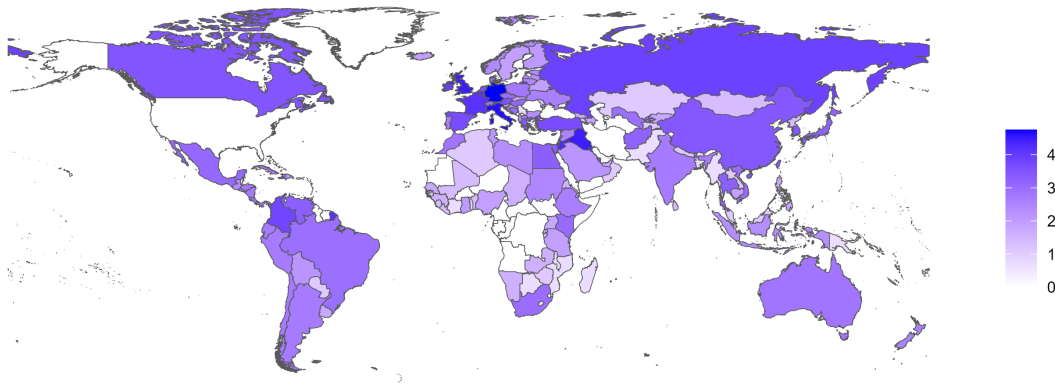
To understand how CODEL participation affects when and how legislators engage in foreign policy hearings, I use two different analyses. First, I merge CODEL participation data with novel data collected from hearing transcripts to assess whether participating on a CODEL affects how much a legislator participates within a relevant hearing. Second, I leverage new text measures that capture how specific and foreign policy relevant legislator speech is to assess whether CODEL participation affects the informative qualities of legislator engagement within hearings.

5.1 CODEL Participation

First, I use disclosure data reported at the legislator level, which includes the destination and travel dates, to identify whether a legislator participated in a CODEL (Curry and Roberts 2022). There are a few important characteristics of CODELs that are worth noting. First, by focusing on CODELs exclusively, I theorize about and evaluate the effects of publicly funded travel abroad. Privately funded trips and trips funded by foreign governments are excluded from this analysis. These trips are less likely to serve a purely informational role and often designed to serve the policy interests of their funders (Rosenson 2009). In addition, in order for a trip to be designated as

10. Interview with former committee staffer, May 1, 2025.

Figure 1: Logged Total Number of CODELs by Country Destination (1999-2006)



a CODEL, it must also be approved by congressional leadership. This gives congressional leadership and the majority party discretion over what constitutes travel for official business¹¹. In fact, Curry and Roberts (2022, 11) explains that CODELs are organized by committee chairs and usually include members of each party. There are some caveats to this; members who are considered to be less friendly are less likely to be invited on trips, as members are less inclined to spend time with disagreeable legislators. Newer members are typically encouraged to come on CODELs in order to learn more about a policy domain.

Descriptively, CODELs also seem to be conducted to destinations with relevance for foreign policy oversight. Legislators often use CODELs to travel to a wide range of destination countries, including to countries of geopolitical or foreign policy relevance (Alduncin et al. 2014; Scoville 2013). Within this time frame, the common destinations include countries such as Iraq, Jordan, Russia, and China (see Figure 1). To provide additional evidence of this, I descriptively examine the predictors of CODEL destinations. Significant predictors of CODEL destinations include whether the destination country received greater executive branch attention through the State of the Union addresses, executive branch travel, and higher levels of security assistance and foreign

11. The State Department works with Congressional members on organizing CODELs but does not have overriding authority over the purpose and destination of these trips. Scoville (2013, 339) writes that “ultimately... delegations themselves decide where to go, whom to visit, and what to say. The State Department does not claim power to prohibit CODELs from traveling abroad”.

aid. This suggests that legislators are not traveling arbitrarily; rather, CODEL destinations reflect the same foreign policy priorities that motivate oversight attention more broadly. For a fuller discussion of this, see Appendix A1.

I aggregate legislator level travel to the regional level because legislators often visit multiple countries within a region on around the same time issues pertaining to that region (e.g., Afghanistan and Pakistan to assess military aid). I categorize regions using the US State Department regional bureaus, which directly match to subcommittees within the HFAC.¹² Within each region, I sum the number of CODELs and total travel days conducted by each legislator to each region by month.

5.2 Hearing Transcripts and Participation

Next, I collect, clean, and process all published hearing transcripts from the House Foreign Affairs Committee (HFAC) from 1999-2006.¹³ There are a couple of reasons as to why I limit the analysis to the HFAC. First, the HFAC is an ideal venue to discuss foreign policy oversight given its focus on foreign policy issues (Carter and Scott 2004). While committees like Armed Services and Intelligence are also important for foreign policy issues, these committees are more likely to also overlap with domestic policy issues (e.g., military personnel, surveillance). Second, given their policy jurisdiction, HFAC committee members are relatively more likely to participate on CODELs (Scoville 2013). By focusing within the HFAC, I can better connect CODEL participation to relevant, downstream oversight activity.

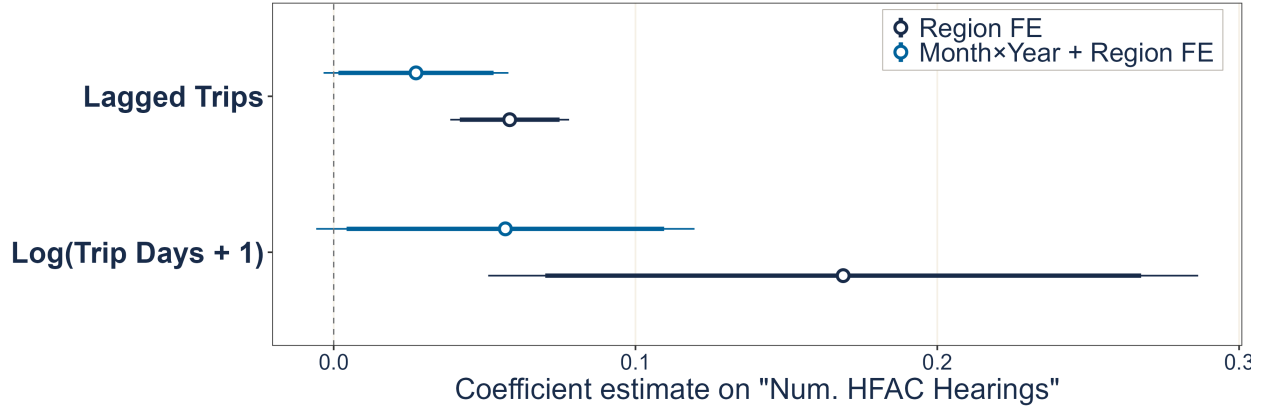
One limitation of using these data sources is that they do not capture closed-door or unpublished hearings. Drezner and Fowler (2025, 4) find that closed hearings were particularly prevalent in the Senate national security committees. House committees, however, are use closed door hearings relatively sparingly. Thus, I expect this limitation to be less consequential for the analyses that follow. In total, I collect 249 HFAC transcripts, with 153 subcommittee hearings (61%) and 96 full committee hearings (39%).

For CODELs to inform legislative oversight, a necessary precondition is that relevant hearings follow CODEL activity. I present some aggregate, descriptive trends of CODELs and foreign

12. These regions include: Africa (AF), East Asia and Pacific (EAP), Near East/ Middle East and North Africa (NEA), South and Central Asia (SCA), and the Western Hemisphere (WHA).

13. <https://commdocs.house.gov/committees/intlrel/>. Last accessed: March 6, 2026. See Appendix A2.

Figure 2: CODELs and Hearing Occurrence



This analysis captures hearings from the HFAC from 1999–2006. This plot shows the coefficient estimates for two measures of CODEL activity in the prior month (*Lagged Trips* and *Log(Trip Days + 1)*) on the number of relevant HFAC hearings in the current month, estimated at the region-month level.

policy oversight hearings to assess whether legislators have a hearing venue to disseminate oversight information gathered through CODELs. I regress the number of HFAC hearings in region and month on the number of CODELs directed at that region in the prior month. Figure 2 shows that CODELs to a particular region in a given month are positively associated with foreign policy oversight hearings about that particular region in the following month. In the baseline specification with region fixed effects, each additional lagged CODEL trip is associated with a 0.06 increase in the number of hearings ($p < 0.01$), roughly 7% of the sample mean (0.83 hearings). These associations attenuate once month-year fixed effects are introduced, suggesting part of the relationship is shared with broader temporal patterns in congressional activity, though the coefficients remain positive across all specifications. For a fuller discussion of this analysis, see Appendix A3.

From each hearing transcript, I construct several measures of hearing participation. I first code whether a member spoke at all during a hearing (*Spoke*), and then the intensity of participation through three count measures: the number of speaking instances (*Total Speaking Instances*), total words spoken (*Total Words*), the number of questions asked, operationalized as speaking instances containing a question mark (*Total Questions*), and the number of times a legislator addressed a bureaucratic witness (*Bureaucrat Address*).¹⁴ *Bureaucrat Address* counts the number of times a

14. Speaking instances of fewer than twenty words are omitted to exclude short procedural remarks such as “Without objection”. I also omit committee members with the same last name (e.g., “Smith”).

legislator speaks immediately before a bureaucrat witness.¹⁵

5.3 Text Measures of Hearing Discourse

To assess the relationship between CODEL participation and *what* legislators say within hearings, I construct two novel text measures.

First, I use a dictionary approach to measure legislator references to first person experiences that are relevant to foreign travel. This is a blunt measure that captures whether a legislator refers to a previous travel experience in a statement that is made during a hearing. For a statement to be coded as having a first person reference to travel, the statement must explicitly mention a travel experience in the past tense (i.e., the travel has already happened) and use first-person language. To make sure these references are relevant for foreign travel, I omit statements that reference domestic travel destinations.¹⁶ These statement patterns used in this approach are listed in Table 1.

Table 1: First-Person Experiential Travel Statement Patterns

Pattern Type	Pattern Structure
Subject-Verb (When)	When I/we traveled/travelled/visited/toured
Subject-Verb (While)	While I/we traveled/travelled/visited/toured
Possessive (Opportunity/Chance)	I/We had an opportunity/chance to visit
Possessive (My/Our)	My/Our trip/visit
Possessive (On)	On my/our trip/visit
Possessive (During)	During my/our trip/visit

A statement is coded as referencing a firsthand experience if it contains one of the following sentence structure listed in the table. Both singular and plural first-person forms are included. Statements that allude to domestic travel are explicitly excluded.

For each legislator within a hearing, create a binary measure of whether a legislator made a mention to a first person travel experience and sum the number of relevant statements they make. Overall, mentions of first person travel experiences are rare. They comprise approximately 2% of all speaking instances (n = 145). Among all hearings, 46.5% of hearings had at least one member

15. For this measure, I omit legislator statements made directly before the first speaking turn of each witness, as these typically corresponds to their opening statement.

16. A statement is excluded if it references travel to any of these keywords: “Administration”, “Congress”, “Senate”, “House”, “Cabinet”, “Army”, “Navy”, “Marines”, “Air Force”, “Guard”, “Corps”, “Committee”, “Subcommittee”, “Service”, “Department”, “Agency”, “Bureau”, “Office”, “Government”, “Court”, as these could be in reference to domestic travel experiences. (E.g., “I visited the Armed Services Committee”.)

use a first-person experiential reference ($n = 114$). On average, a legislator has a baseline likelihood of 2% to make an explicit reference to first person travel within a given hearing.

Oftentimes, however, the information that legislators gather on CODELs will shape what they say in hearings in more subtle ways. Legislators who travel are exposed to firsthand observations. In particular, specific people, places, organizations, and facts on the ground, that their non-traveling counterparts lack. This accumulated knowledge should surface in how they frame questions and statements within a hearing, even when they do not explicitly reference their travel. To capture this, I develop a text measure for “foreign policy specificity” by leveraging LLMs to conduct pairwise comparisons with the purpose of predicting and scaling a measure to capture the relative specificity of a legislator’s statement.

Conducting pairwise comparisons has been used to measure latent characteristics of political text (Carlson and Montgomery 2017). Notably, scholars have used this method to capture traits such as grandstanding (Park 2021), confrontation (Eldes, Fong, and Lowande 2023), and analytical information (Esterling and Park 2024; Ban, Park, and You 2025; Eldes, Fong, and Lowande 2023) in congressional committees. Following recent work using large language models (LLMs) to scale pairwise comparison methods (DiGiuseppe and Flynn 2025; Halterman and Keith 2026), I combine human-coded validation with LLM-assisted scaling to help produce a specificity score for each statement in my corpus.

First developed a codebook defining foreign policy specificity and adapted it for both human coders and LLMs. Broadly, a statement is more foreign policy specific when it *contains concrete, detailed references to people, places, quantities, policies, events, or institutions that are directly relevant to US. foreign policy* (see full codebook in Appendix C1.) Three trained research assistants compared 650 unique statement pairs and were asked to choose which statement was more foreign policy specific.¹⁷ Fifty pairs are coded by all three, to establish inter-coder reliability and consistency of codebook interpretation, achieving Cohen’s Kappa values of 0.63–0.96.

I then benchmarked two open-source LLMs¹⁸ (Llama 3.1 8B and Mistral 7B) using the behavioral tests proposed by Halterman and Keith (2026), which assess label accuracy, codebook

17. All coders were instructed to prioritize statements that were both relevant to foreign policy and contained specific references. For example, if one statement was more specific but not relevant to foreign policy, the foreign policy relevant statement should be prioritized. Additional edge cases are detailed in the codebook (see Appendix C1).

18. I favored open-source models over proprietary alternatives (e.g., Claude, ChatGPT) to preserve reproducibility.

comprehension, out-of-sample classification, and order-invariance. Llama 3.1 8B performed best and was selected for the full scale coding. To validate the LLM’s coding performance against human benchmarks, I ran it on the same 650 RA coded pairs. Agreement between the LLM and individual RAs ranged from 0.45 - 0.63 (74 - 82%), and on the 50 pairs coded by all four coders, the LLM matched the majority vote 86% of the time. Bootstrapped ($n = 1000$) Krippendorff’s alpha estimates showed substantial overlap between human-only codings ($\alpha = 0.751$, 95% CI: [0.581, 0.886]) and human-plus-LLM codings ($\alpha = 0.651$, 95% CI [0.487, 0.780]), suggesting the LLM’s addition did not significantly alter coding reliability compared to the trained human coders (see Appendix C2).

Next, I sampled 600 statements from the full corpus of 18,395 hearing statements, generated 3,000 pairwise comparisons with each statement appearing 10 times and had the LLM code all pairs. I fit a Bradley-Terry model to these comparisons to derive a specificity score for each of the 600 statements.¹⁹ This model produced a score for each sample statement which is plotted in the left-hand panel of Figure 3.

Finally, I used these scores to train a machine learning model to extend the scores to the full corpus. I randomly selected 500 statements for training and 100 held out for validation and ran eight machine learning models.²⁰ The final model is constructed using an ensemble Bayesian model (EBMA) averaging technique to optimize prediction performance (Montgomery, Hollenbach, and Ward 2012). This final model was used to predict a score for each statement in the corpus. The final foreign policy specificity score is scaled from 0-100 and plotted in the right-hand panel of Figure 3.

To statistically validate this measure, I assess how the final model predicts the LLM coded

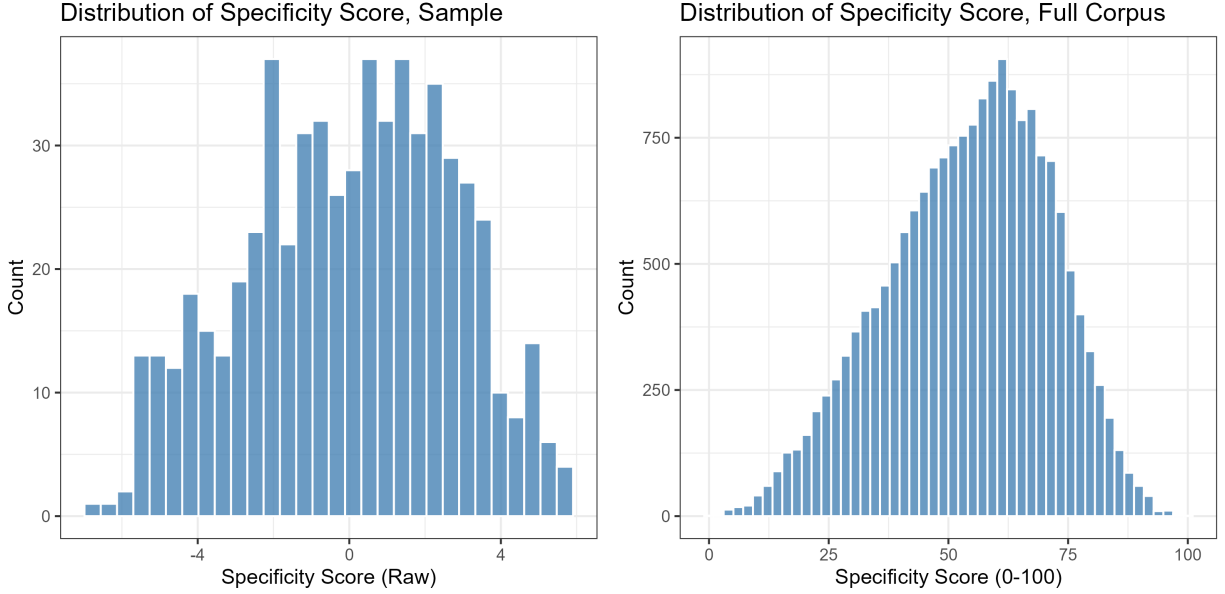
19. Formally, the Bradley Terry model predicts:

$$Pr(y_k = j) = \frac{\exp(a_j - a_i)}{1 + \exp(a_j - a_i)} \quad (1)$$

with priors of $\alpha \sim N(0, 1)$ and $\sigma \sim trN(0, 3)$. The model estimates the foreign policy specificity of each statement, a_j . As this method does not rely on crowd-sourced pairwise comparisons, I treat LLM as a single “worker” per the advice of DiGiuseppe and Flynn (2025). Priors are chosen to match similar models for latent traits in committee transcripts (Park 2021).

20. I first constructed two document-level matrices: a term-document frequency matrix and a doc2vec matrix. For each matrix, I fit four models: support vector machine, Kernlab’s support vector machine, LASSO, and Gradient Boosting Machine. Because the doc2vec matrix struggled with statements longer than approximately 950 words ($n = 753$) these were split into sub-segments for embedding and their vectors averaged.

Figure 3: Specificity Score Distribution



Distribution of the foreign policy specificity score among sample statements (left-hand panel) and the entire corpus (right-hand panel).

labels of the validation set. The Pearson correlation between between the LLM coded labels and the predicted labels is 0.76 and the normalized root means squared error is 0.16²¹ (see Appendix Figure C2). This is similar to other prediction exercises within the context of committee hearings (Ban, Park, and You 2025).

To qualitatively validate the measure, I present the highest, lowest, and average legislator specificity score in Table 2. The highest scoring statement makes references to specific policy and events in Libya and displays an understanding of a crisis that falls within the domain of foreign policy. The statement that achieved an average score off approximately 55 makes references to foreign countries and people (e.g., Vietnam/ Nixon) but does not refer to any specific foreign policy issue and uses these references to ask a question that is generally about military strength. The lowest scoring statement reflects a procedural statement.²²

21. I normalize the RMSE using the range of my raw score predictions. $NRMSE = \frac{RMSE}{y_{max} - y_{min}} = \frac{1.79}{11} = 0.16$.

22. I omit all statements that are less than 20 words to address to exclude as many procedural statements as possible. Longer procedural statements, however, are kept as some legislator statements of substance are of a similar length. I run additional analyses in later sections to account for procedural statements.

Table 2: Sample Legislator Statements with Different Specificity Scores

Score	Text
90.47	Thank you, Mr. Chairman. I would like to take this opportunity to commend President Bush and his Administration for their commitment and the dedication to rid the world of the threats posed by dictators such as Libya's Gadhafi. U.S. policy and determination appears to have prompted Libya to modify its position regarding its support for terrorism, its non-conventional weapons program and its deplorable human rights record. However, given the repeated refusals by Libya to accept responsibility for the Pan Am 103 attack and its continual denials prior to December of last year of its involvement in other threatening activities, can we really trust the Libyan regime? Should the U.S. be in any hurry to ease sanctions and other restrictions on Libya until there is a greater body of evidence that Libya means what it says and it will follow through on its promises. There has to be more cooperation from the Libyans on getting all of the facts as required by the U.N. resolution lifting sanctions on Libya . . .
55.02	And because I am very concerned and I am somebody whose politics are sort of Nixonian in that sense of seeing how we conducted Vietnamization and what the underpinning problems were after we left Vietnam, even though from a military standpoint we left a solid military able to defend itself, but there were a number of problems both here in Congress and there, would you please give us your view of how important the differences again not spending too much time on Vietnam but the difference between us leaving a government with a strong military and us leaving a government that is reflective of the people and an appropriately strong military?
0	Thank you. Mr. Crowley, you are welcome to join the Committee, and if you have any questions, you are recognized for 5 minutes.

5.4 Hearing-legislator Dyads

To test the effect of CODEL participation on hearing participation and discourse, I construct a panel at the hearing-(sub)committee member level.²³ Because hearings are held at the committee and subcommittee level, all HFAC committee members are not always expected to attend every subcommittee meeting. Thus, for every committee hearing in the panel, there is an observation for every HFAC committee member. For every subcommittee hearing, there is an observation for every relevant subcommittee member. Within each hearing, this makes the comparison between members who traveled versus did not travel among legislators who could be expected to be at the hearing.

To connect CODEL participation to the different measures of hearing participation and discourse at the member level, I match member-level CODEL travel to hearings based on whether the hearing's title references the CODEL destination region or any countries within that region (Jones

23. Note that all foreign policy scores associated with non-legislator statements are omitted.

et al. 2023). For each member-hearing pair, I merge the number of CODELs and trip days the member took to the relevant region in the month prior to the hearing. The resulting panel consists of 7,306 member-hearing pairs drawn from 249 hearings and 83 distinct HFAC committee members over the study period.

6 CODEL Participation and Hearing Participation

To estimate the effect of CODEL participation on the probability that a legislator speaks during a subsequent, relevant hearing, I estimate the following ordinary-least-squares regression:

$$\text{Spoke}_{ih} = \beta \text{Traveled in Last Month}_{ir} + X_{it} + \alpha_{ir} + \gamma_t + \varepsilon_i \quad (2)$$

where Spoke_{ih} is a binary indicator equal to 1 if legislator i speaks during hearing h . The main independent variable, $\text{Traveled in Last Month}_{ir}$, indicates whether the legislator participated in a CODEL to the relevant region r in the month immediately preceding the hearing. One challenge to estimating this effect is separating the role legislator-level interest and expertise play in driving both travel and hearing participation. To address this, I use legislator-region fixed effects (α_{ir}) to control for legislator characteristics, geographic region characteristics, and the time invariant component of a legislator's interest and expertise in a specific region that might drive both CODEL and hearing participation. With the inclusion of congress fixed effects (γ_t) to account for time-variant confounders, this results in a within-legislator-region specification that compares whether a legislator participates more in a hearing following a recent CODEL to that region, compared to hearings about the same region where a legislator had not traveled before.

In addition, I incorporate legislator-level controls (X_{it}) to capture institutional and individual factors that may independently affect both the propensity to travel and participate in hearings. Whether the hearing was a subcommittee hearing is included because legislators are more likely to use committee hearings for political staging and grandstanding than subcommittee hearings, which are more likely to feature information provision (Ban, Park, and You 2025). I also control for legislative effectiveness (Volden and Wiseman 2014) and the number of years a legislator has been on a committee as these reflect a legislator's institutional capacity and incentives to engage

in foreign policy oversight.²⁴ Following prior work suggesting that members more predisposed to bipartisan engagement are more likely to travel together (Curry and Roberts 2022; Alduncin, Parker, and Theriault 2017; McGee and Theriault 2022), I include ideological distance from the chamber median as a proxy for bipartisan orientation. Finally, I control for the legislator’s previous vote share, as electorally vulnerable members may be less willing to devote time to foreign policy (Carter and Scott 2004). Standard errors are clustered by legislator (ϵ_i).²⁵

Table 3: CODEL Travel and Hearing Participation

Dependent Variable:	Spoke (0/1)			
	All Hearings		Committee	Subcommittee
Traveled in Prev. Month	0.0427 (0.0280)		-0.0154 (0.0284)	0.0943* (0.0458)
Logged Trip Days		0.0344 [†] (0.0174)		
Subcommittee	0.0061 (0.0196)	0.0058 (0.0196)		
Committee Seniority	0.1256*** (0.0220)	0.1260*** (0.0219)	0.0759* (0.0292)	0.1471*** (0.0296)
Dist. from Median	-0.0086 (0.3029)	-0.0078 (0.3025)	-0.2250 (0.4328)	0.4824 (0.4266)
LES	0.0047 (0.0100)	0.0046 (0.0100)	-0.0009 (0.0155)	-0.0054 (0.0150)
Prev. Vote Share	-0.0023** (0.0009)	-0.0023** (0.0009)	-0.0024 [†] (0.0013)	0.0003 (0.0017)
Legislator-Region FE	Yes	Yes	Yes	Yes
Congress FE	Yes	Yes	Yes	Yes
Observations	7,070	7,070	4,490	2,481
Dependent variable mean	0.22207	0.22207	0.19488	0.27489
Within Adjusted R ²	0.00238	0.00271	0.00113	0.00439
Std. Dev. (dep. var.)	0.41566	0.41566	0.39615	0.44655

The unit of observation is hearing-legislator. Standard errors clustered by legislator. ** $p < 0.01$; * $p < 0.05$; [†] $p < 0.1$

The results are shown in Columns 1 and 2 of Table 3. I find a member who traveled on a CODEL to a relevant region in the previous month is not significantly more likely to participate

24. Other institutional factors that might predict travel and hearing participation, such as majority party status, are absorbed by the inclusion of both legislator-region and congress fixed effects.

25. I also run this specification with clustered standard errors at the legislator-region level. See Appendix B4.

in a subsequent hearing compared to when she did not travel on a CODEL. Neither measure for CODEL travel in the previous month achieves traditional levels of statistical significance.

There is reason to suspect, however, that the effect of travel on hearing participation differs by institutional setting. Full committee hearings are structurally higher-visibility and receive higher levels of attention than subcommittee hearings. This makes it more likely that legislators will use committee hearings for political staging and messaging that is less relevant for information provision (Park 2017, 2021). Subcommittees, by comparison, may function more as workhorse venues. These hearings are more likely to draw a narrower, more specialized audience, and members are afforded more policy relevant resources (Hall 1987).

I rerun the main analysis on subsets of the data by whether the hearing was held in a committee or subcommittee. These results are shown in columns 3 and 4 of Table 3. Travel does not have a significant effect on whether a legislator speaks in a full committee hearing. However, I find that at the subcommittee level and holding fixed the geographic region of interest, a member who travels on a CODEL in the prior month is associated with a 9 percentage point increase in the probability of speaking at a hearing compared to when she did not travel. Given a member's average rate of participation within a subcommittee hearing, this is equivalent to a 33% increase relative to the subcommittee mean participation rate (27%) and roughly 0.2 standard deviations of the outcome. These results are substantively similar when using an alternative specification that interacts *Traveled in Prev. Month* with *Subcommittee* (see Appendix Figure B1). While the effect is moderate, this suggests that the effect of CODEL travel on whether a legislator participates in a relevant hearing is apparent only when the hearing is held in a setting that is relatively more likely to be used for information provision and policy-oriented discussion.

In Table 4, I also estimate the effect of CODEL travel on other measures of hearing participation. All in all, CODEL travel does not have a significant effect on the number of speaking instances a legislator takes, the number of question she asks, or number of statements she directs at a bureaucratic witness. Interestingly, there is a positive and significant relationship between CODEL travel and the number of words a legislator uses within a hearing. This suggests that legislators are more likely to speak for longer and more verbosely as a result of traveling to a relevant region. This also suggests that CODEL travel may have a stronger effect on what a legislator says in a hearing rather than how much they participate in a hearing.

Table 4: CODEL Travel and Intensity of Hearing Participation

Dependent Variables:	Speaking Instances	Total Words	Total Questions	Total Bureaucrat Addresses
Traveled in Prev. Month	0.0728 (0.0709)	0.3928* (0.1941)	0.0612 (0.0689)	0.0366 (0.0540)
Controls	Yes	Yes	Yes	Yes
Legislator-region FE	Yes	Yes	Yes	Yes
Congress FE	Yes	Yes	Yes	Yes
Observations	6,889	6,889	6,889	6,889
Dependent variable mean	0.41108	1.4335	0.35263	0.23933
Within Adjusted R ²	0.00702	0.00445	0.00626	0.00711
Std. Dev. (dep. var.)	0.87028	2.7322	0.80575	0.61126

Controls include: Subcommittee, Committee Seniority, Distance from the Median, LES, and Previous Vote Share. All dependent variables are logged. The unit of observation is hearing-legislator. Standard errors clustered by legislator. Full results with controls are presented in Appendix Table B3. ** $p < 0.01$; * $p < 0.05$; † $p < 0.1$

7 CODEL Participation and Hearing Discourse

To estimate the effect of CODEL participation on what legislators say within hearings, I repeat the analyses from the previous section using the text measures of hearing discourse, first-person travel references and foreign policy specificity, as the main outcome variables. Importantly, the analysis with foreign policy specificity as the dependent variable is conditional on whether the legislator spoke within a hearing, as a legislator who does not speak cannot be assigned a meaningful foreign policy specificity score. For these specifications, I include a control for the number times a legislator speaks in a given hearing.

These results are presented in Table 5. *Traveled in Prev. Month* and *Logged Trip Days* is positive and significant at the 95% level for all specifications. Columns 1 and 2 show the results for CODEL travel on firsthand references relevant to a foreign travel experience. I find a legislator who travels on a CODEL to a particular region is 4.8 percentage points more likely to mention a firsthand travel experience in a subsequent, relevant hearing compared to when she does not travel to that region. Turning to the continuous measure of experiential mentions, recent CODEL travel is similarly associated with roughly a 5.2% increase in experiential mentions. While these coefficients are small, relative to the sample mean, these effects are substantial. A recent CODEL

trip is associated with a 233% increase in the probability of an experiential mention (column 1) and a 305% increase in the (logged) count of experiential mentions (column 2).²⁶ This represents a 0.34 and 0.42 of the sample standard deviation of the dependent variables, respectively.²⁷ In sum, both specifications show that, for a legislator focused on a particular region, CODEL travel can roughly triple the baseline rate of experiential language used in a hearing.²⁸

Columns 3 and 4 show the results for CODEL travel on specific and relevant to foreign policy statements made in hearings.²⁹ The positive and significant coefficient for *Traveled in Prev. Month* suggests that conditional on speaking in a given hearing, CODEL travel is associated with a 3.16 point increase in a legislator's average foreign policy specificity score. This represents an effect size that is a 5.7% increase relative to the mean and 0.28 of the within-unit standard deviation of the average specificity score.³⁰ Similarly, a one unit increase in logged trip days is associated with a 1.43 increase in average specificity score for a legislator focused on a given region. Given the difficulty of altering an individual legislator's use of foreign policy specific knowledge in the context of an HFAC hearing, this represents a meaningful effect. Compared to other effect sizes on similarly scaled dependent variables within the context of committee hearings, these effects are relatively high, suggesting that CODEL travel is doing meaningful work.³¹

Taken together, these results suggest that CODEL travel makes legislators more informative in the hearings that follow. Such information is an important input into the foreign policy oversight process, with implications for the quality of oversight more broadly. For example, specific, firsthand-grounded language can constrain how witnesses respond. A question grounded in a named program, a concrete figure, or a firsthand observation prompts a witness to either engage substantively or avoid the question altogether, whereas a vague or general question can invite an equally vague talking point in response. Specificity and firsthand information can also build a leg-

26. $0.0484/0.0208 \approx 233\%$ and $0.0515/0.0169 \approx 305\%$.

27. $0.0484/0.1426 \approx 0.34$ and $0.0515/0.1218 \approx 0.42$

28. $(0.0208 + 0.0484)/0.0208 \approx 3.33$

29. For these models, I also control for the number of speaking instances a legislator took within a hearing.

30. $3.157/55.37 \approx 5.7\%$ and $3.157/11.30 \approx 0.28$

31. For example, Park (2021, 233) finds that minority status under divided government is associated with a 0.608-point increase in grandstanding score, equivalent to approximately 0.14 within-individual standard deviations of that measure. Using an analogous comparison, this estimate for recent CODEL travel corresponds to 0.303 of the within-unit standard deviation (10.461 is the within-legislator standard deviation. $3.157/10.461 \approx 0.307$). This comparison should be interpreted with caution, as Park (2021) aggregates her analysis to the member-Congress level rather than the member-hearing level. Nonetheless, this provides a useful benchmark to interpret this effect size.

Table 5: CODEL Travel, Information, and Hearing Engagement

Dependent Variables:	Experiential Mention (0/1)	Logged Experiential Mentions	Average Specificity Score	
Traveled in Prev. Month	0.0484** (0.0157)	0.0515** (0.0164)	3.157** (1.012)	
Logged Trip Days				1.425* (0.5596)
Subcommittee	0.0044 (0.0045)	0.0039 (0.0041)	0.9937 (0.7469)	1.009 (0.7524)
Committee Seniority	-0.0235*** (0.0043)	-0.0158*** (0.0036)	5.905*** (0.8365)	6.069*** (0.8460)
Dist. from Median	0.0498 (0.0867)	0.0617 (0.0775)	-22.99 [†] (13.39)	-22.32 (13.39)
LES	-0.0015 (0.0025)	-0.0018 (0.0027)	0.7370 [†] (0.3826)	0.7259 [†] (0.3828)
Prev. Vote Share	-0.0005 (0.0004)	-0.0005 (0.0004)	-0.0539 (0.0591)	-0.0543 (0.0592)
Speaking Instances (n)			-0.4562*** (0.0487)	-0.4582*** (0.0483)
Legislator-region FE	Yes	Yes	Yes	Yes
Congress FE	Yes	Yes	Yes	Yes
Observations	6,889	6,889	1,421	1,421
Dependent variable mean	0.02076	0.01685	55.367	55.367
Within Adjusted R ²	0.00455	0.00731	0.12403	0.12177
Std. Dev. (dep. var.)	0.14258	0.12177	11.297	11.297

The unit of observation is hearing-legislator. Standard errors clustered by legislator. Columns 1 and 2 include the full sample. Columns 3 and 4 only include legislator-hearing observations in which a legislator spoke in that hearing.

** $p < 0.01$; * $p < 0.05$; [†] $p < 0.1$

islator’s credibility within Congress, signaling to colleagues that a foreign policy issue is serious enough to merit increased oversight (Carter and Scott 2004). This kind of signal may extend beyond Congress as well: specific, well-informed oversight can also signal to the executive branch that Congress has made a genuine institutional investment in a given issue. In other words, this has implications for how willing and effective Congress is at engaging in downstream oversight.

One potential concern is that the variation in the *Average Specificity Score* could be largely driven by procedural statements. While I omit short statements under 20 words to partially address for this problem, some procedural statements remain longer and are explicitly coded with having low levels of foreign policy specificity. To address this more rigorously, I rerun the regressions presented in Table 5 and omit hearing chairs, as these members will be the most likely to use procedural statements. The results are substantively similar to the main findings (see Appendix B5). Another potential concern is that the *Average Specificity Score* might be correlated with statement length and the effect uncovered could simply be reflecting the fact that legislators who travel will speak for longer. I assess the relationship between *Average Specificity Score* and *Total Words (Logged)* I find that total statement length and the foreign policy specificity score is not significantly correlated ($r = 0.04$; $p > 0.1$) (see Appendix Figure B2).

8 COVID-19, CODEL Cancellations, and Hearing Content

In this section, I provide a case study to trace the mechanism behind my proposed theory and illustrate substantively how CODEL participation can shape what a legislator says relevant hearings. I leverage the COVID-19 pandemic as a plausible exogenous shock to CODEL travel and evaluate the hearing behavior of HFAC member and Africa Subcommittee Chair, Rep. Karen Bass, before the COVID cancellation of CODELs and during the pandemic, where no CODELs occurred (2018-2021). I find that Rep. Bass’ engagement in similar hearings before and during pandemic were substantively different. Rep. Bass’ participation in hearings after a relevant CODEL was more detailed and explicitly grounded in firsthand observations from those CODELs. Her participation in hearings during the pandemic contained of references to broader foreign policy positions and explicit invitations to witnesses to supply firsthand observations instead.

This case offers a couple of methodological advantages that complement the quantitative anal-

ysis already presented. First, I minimize potential confounders by holding constant the congressional member, the region and country of CODEL destination and hearing focus, and the institutional setting. Second, the cancellation of all CODELs for public health reasons allows me to better make the argument that the change in CODEL participation and, subsequently, hearing related behavior during this time period was due to Rep. Bass not being able to participate on the CODEL, not because her interest in the region or topic had changed. This variation across time within one committee member allows me to trace the mechanisms in a more robust way.

8.1 Pandemic Disruptions

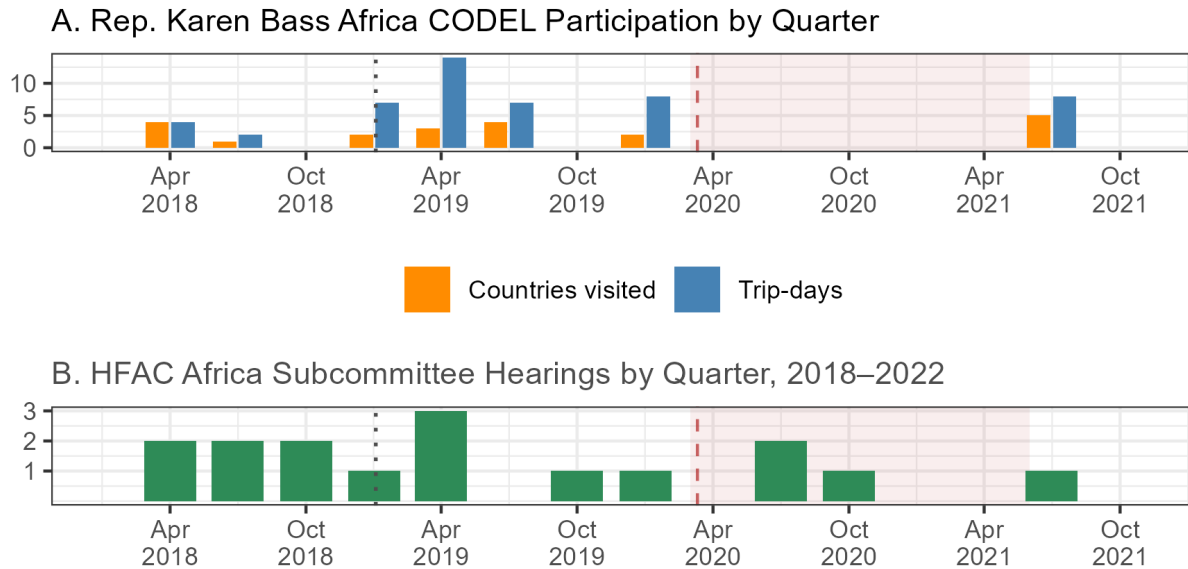
Beginning in 2018, Rep. Bass consistently led and participated in CODELs to Sub-Saharan Africa, typically once or more per year (See Panel A of Figure 4). During this study period, Rep. Bass travelled to Ethiopia (7), Angola (3), Djibouti (2), Niger (2), Botswana, Mozambique, Sierra Leone, Cameroon, Burkina Faso, and Namibia. This pattern was noticeably interrupted for roughly 18 months starting from March 2020 which coincided with the ban on CODELs due to the COVID-19 pandemic.³² While there is no explicit cancellation of these CODELs, I infer the disruption from the pattern itself: Rep. Bass took two Africa trips in the six weeks immediately preceding the pandemic's onset, consistent with her established cadence, and then took no further CODELs until August 2021. Given this consistency immediately before the gap, I treat the interruption as plausibly attributable to COVID-era travel restrictions rather than an independent change in Bass's travel preferences.

During this time period, the HFAC Africa Subcommittee consistently held hearings related to Sub-sahara Africa (see Panel B of Table 4). These 18 hearings, spanning the House Foreign Affairs Committee's Africa Subcommittee covers a mix of country-specific developments (Zimbabwe's post-Mugabe transition, Cameroon's crisis, Ethiopia's political developments and later conflict, Sudan's political crisis, Nigeria's elections) alongside broader thematic and regional hearings (China's role in Africa, civil society and religious freedom, public health/tuberculosis, peace-keeping, the Sahel, and youth demographics).

I compare how Rep. Bass engages in Africa subcommittee hearings before the pandemic period

32. <https://www.foxnews.com/politics/lawmakers-missing-votes-cancellations-coronavirus>. Last accessed: August 4, 2026.

Figure 4: Rep. Bass CODEL Participation & HFAC Africa Hearings



Panel A shows the number of CODELs took to Africa during 2018-2021. Panel B shows the number of HFAC Africa hearings during this time period. The red dotted line indicates the onset of COVID-19 and the red shaded region the pandemic period. The black dotted line denotes when Rep. Bass became chair of the Africa subcommittee.

when she consistently traveled on relevant CODELs to her hearing behavior during the pandemic period when she lost access to CODELs. More precisely, I compare the opening statements Rep. Bass makes across hearings. This has the additional benefit of holding the purpose and structure of the speaking instance constant across hearings, as opening statements are always given at the beginning of a hearing to set the agenda. Rep. Bass gave an opening statement in all of the hearings during this study period except for one during the pandemic (Hearing Title: “Updates on COVID-19 Pandemic Response in Africa; Hearing Date: July 30, 2020.).

It is also important to recognize some limitations about this case. First, COVID disrupted congressional activity broadly, including a shift to virtual hearing formats, which may independently affect how members engage regardless of travel history. Second, some hearings in the pre-COVID period before 2019 were held when Rep. Bass was the ranking member and not the chair. Thus, I caution against using this case to make a causal claim. Rather, I use this case to illustrate *how* CODELs allow legislators to provide information that is specific and relevant to first hand observations.

8.2 Changes in Hearing Discourse

In the months immediately preceding the pandemic, upon participating on a recent CODEL, Rep. Bass's opening statements consistently drew her travel experience to frame her remarks. Opening the November 2019 hearing on US policy in the Sahel, she stated that "several weeks back, I led a CODEL focused on US military engagement across Africa, and we traveled to parts of the Sahel to gain a better understanding of our military assistance in the region. I spoke with many heads of State, foreign ministers, our own Ambassadors, and military personnel, and security was never far from their minds." She went on to note that "members of the CODEL also heard that there are in some countries numerous positions that are unfilled within the embassies," using the trip directly to motivate her questions to the two bureaucratic witnesses from the Department of State and USAID. Furthermore, using specific, foreign policy relevant statements, Rep. Bass highlights developments within the region ("AQIM and Boko Haram have since capitalized on the region's political instability and reportedly merged with locally based and rural extremist groups in Mali and Burkina Faso.") and advocates for preferred policy positions ("the international community should support the efforts of the G5 Sahel force, which consists of troops from Burkina Faso, Chad, Mali, Mauritania, and Niger by helping to train soldiers and assisting in funding operations").³³

A similar pattern appears in the February 2020 hearing on Africa's youth demographics, where she recounted observations from a recent CODEL, stating: "I was recently in Darfur, meeting with a group of young activists, who played a significant role in Sudan's transition, and it was interesting talking to them because they have strong opinions about the leadership but, yet at the same time, they did not see themselves participating in the leadership. They saw themselves as standing on the outside." Rep. Bass uses this experience to make a case for deeper engagement in the region and points to existing policies to back up her claims ("young people have always been at the forefront of positive social change ... there are examples [of this], including Nigeria's Not Too Young to Run campaign, [the] LUCHA in DRC and Burkina Faso and Senegal, among many, many others.."). In both cases, the CODEL functioned as the explicit evidentiary basis for her framing of Rep. Bass's opening remarks.³⁴

This pattern seemingly shifts in each of the hearings held during the COVID gap where Rep.

33. <https://www.congress.gov/116/chrg/CHRG-116hhrg38457/CHRG-116hhrg38457.htm>. Accessed July 17, 2026.

34. <https://www.congress.gov/116/chrg/CHRG-116hhrg39653/CHRG-116hhrg39653.htm>. Accessed July 17, 2026.

Bass lost access to CODELs. Rep. Bass was less explicit in her references to Africa-specific first-hand information. Furthermore, her presence in these hearings was itself less consistent than in prior periods. She was absent from the first Africa Subcommittee hearing held during the pandemic (July 2020), leaving Rep. Susan Wild to preside in her place. In the third subcommittee hearing of the COVID era (“Democratic Backsliding in Sub-Saharan Africa”, Sept 2020), although she was the subcommittee chair, Rep. Bass delivered her opening remarks after the ranking member, Rep. Chris Smith. She spoke broadly about the harms of democratic backsliding, drawing analogies to other regions of the world rather than citing specific recent encounters on the continent: “There are many sub-Saharan countries of concern and I look forward to hearing which countries we should monitor closely and the witnesses’ recommendations for how we can support them... We see similar patterns in Côte d’Ivoire as the executive branch legalizes the deviation and democratic institutions to codify nondemocratic actions... In Europe, we see the egregious behavior of Belarusian President Alexander Lukashenko who claimed success in a disputed August 9th election.”³⁵

This evidence points to a general change in hearing discourse. One explanation, however, could be that the pandemic era reflects a change in hearing topics and priorities. To address this, I compare two hearings held on the same country and same general subject matter, conflict within Ethiopia, where one was held before COVID and other during the CODEL gap.

During the pre-pandemic period, Rep. Bass traveled to Ethiopia in August 2018³⁶, and her opening statement at the subsequent September 2018 hearing referenced the trip directly, stating: “we did face a little pushback when we were [in Ethiopia] over our resolution 128, and some people voiced why we did it at that particular time, what was our timing, what were we trying to say. And the chairman and I had an opportunity to explain that... the resolution was not trying to... slap Ethiopia right when change was occurring. We had an opportunity to really discuss that with many people.” She uses insights gained through her trip to highlight challenges the incoming Ethiopian prime minister might face (“we do know that he faces daunting challenges ahead, including regional security issues, the country’s past human rights record, ethnic tensions across the country, and hardliners within the EPRDF that hope to stall his reform agenda.”).³⁷

35. <https://www.congress.gov/116/chrg/CHRG-116hhrg41970/CHRG-116hhrg41970.pdf>. Accessed July 17, 2026.

36. During this hearing, given her minority party status, I evaluate Rep. Bass’s opening statement as a ranking member and not chair.

37. <https://docs.house.gov/meetings/FA/FA16/20180912/108670/HHRG-115-FA16-Transcript-20180912.pdf>. Accessed July 17, 2026.

By contrast, her opening statement at the December 2020 hearing on Ethiopia’s unfolding conflict contains no comparable firsthand claim. She described the general situation: “The current State of Ethiopia, not only has a ripple effect on its citizens, but neighboring countries such as Somalia, Sudan, and Eritrea are negatively affected I appreciate and I am looking forward to hearing from Tsedale Lemma, *who can give a firsthand report on what is happening in the country* [emphasis added].”³⁸ There was a stark shift in from Rep. Bass positioning herself as a direct source of on-the-ground information to explicitly deferring that role to a witness (Tsedale Lemma). Holding topic fixed in this way isolates the mechanism of interest: the shift is not simply that Bass discussed different subjects during the pandemic, but that on the same subject, she moved from direct, specific references and from positioning herself as a source of direct information to relying on witnesses to supply it. This pattern is consistent with the loss of independent information-gathering capacity that CODEL access is theorized to provide.

9 Conclusion

To conduct oversight, Congress relies on relevant information about policy implementation and outcomes. Scholars and practitioners recognize the importance of informational tools, especially in detecting potential deviations from intended policy goals. Legislators, however, often struggle to access this information, as reported information, through institutional and informal channels, can often be unavailable or unreliable. In this paper, I highlight the importance of mechanisms of direct observation. These tools give legislators the opportunity to gather firsthand information about policy implementation without solely relying on executive or third party accounts in order to make such assessments. In particular, I focus on CODELs as an example of such a tool in the context of foreign policy oversight. CODELs are particularly valuable for allowing legislators to observe foreign operations direct, question executive branch and foreign contacts, and understand the parameters that define foreign policy success.

Using newly collected hearing transcript data from the HFAC and data on CODEL participation, I analyze how CODELs affect the likelihood of hearing participation and the type of discourse

38. <http://www.govinfo.gov/content/pkg/CHRG-116hhrg42439/pdf/CHRG-116hhrg42439.pdf>. Accessed July 17, 2026.

used within relevant hearings. I show that, on average, CODEL participation does not change whether a legislator is more likely to participate in hearings. The effects, however, differ by committee setting: legislators are more likely to speak in a relevant subcommittee hearing, typically used for information provision, after traveling on a CODEL, but not in full committee hearings, which are more often used for political staging. These findings suggest that the relevant question concerns how CODELs shape the information legislators convey once they choose to speak. I find that the effects of CODEL participation are indeed more apparent in how legislators speak within hearings: legislators who travel are more likely to reference firsthand observations and use rhetoric that is richer in specific references to entities directly relevant to foreign policy. These findings have implications for how effectively legislators, and Congress as a whole, can conduct oversight.

This theoretical framework and analysis offers an updated perspective on how Congress conducts oversight. Legislators often rely on a range of different strategies - within Congress, outside of Congress, and even abroad - that inform each other in order to conduct oversight. Tools of direct observation, such as CODELs, are particularly valuable among these strategies because they allow Congress to gather information relatively independent of the executive branch, rather than solely relying on the executive to supply it. Consequently, these tools used to monitor and gather information can shape downstream venues of congressional oversight. Broadening our understanding of what tools Congress has and uses to monitor policy implementation is important for assessing the nature, effectiveness, and future of congressional oversight, as well as the policy produced afterward.

With respect to studying US foreign policy, scholars have highlighted that participation in formal oversight through hearings appears to have declined substantially. Fowler (2015)'s careful evaluation of oversight hearings over foreign policy issues traces a decline in Senate Armed Services and Foreign Relations hearings and concludes that congressional oversight over foreign affairs has weakened considerably. Moreover, Drezner and Fowler (2025) find that the decline in hearings is also paired with a decline in the number of witnesses and expert witnesses. Goldgeier and Saunders (2018, 145) argues more forcefully that "in Congress, the combination of declining foreign policy expertise among members and increasing political polarization has reduced the ability of legislators to supervise the executive branch even if they had the appetite to do so". This paper reveals that studying hearings in isolation, however, may give an incomplete picture of for-

foreign policy oversight, as hearings only capture the formal, public stage of the broader oversight process.

This paper introduces several avenues for future research across legislative studies and foreign policy. One avenue is to empirically test the downstream consequences of these informational tools for oversight itself. For instance, scholars can assess whether more specific, firsthand-grounded questioning as a result of participating in CODELs actually constrains witness testimony, or whether it translates into demonstrable changes in agency behavior or policy outcomes. In addition, Congress uses other observational tools of oversight across different policy domains, such as site visits and facility investigations (Garvey 2025; Levin and Bean 2018). This paper provides a framework upon which to study these mechanisms in domestic policy as well. Furthermore, by positioning CODELs as a strategy uniquely suited for foreign policy oversight, this study raises broader questions about domain-specific oversight tools. Future work should explore whether other policy domains, particularly those that are highly technical or require specialized expertise, benefit from similar strategies.

One limitation of this study is that it does not address the potential effects of polarization. Scholars have pointed to the encroachment of polarization into foreign policy, a policy domain that is typically portrayed as unified “at the water’s edge”. Others have found that the rise of ideological extremism and partisan warfare has led to increasingly partisan CODELs and a weakening of collaborative norms surrounding CODELs (McGee and Theriault 2022). With growing partisan division over key foreign policy issues (e.g., aid for Ukraine, the wars in the Middle East) and the politicization surrounding foreign policy relevant agencies (e.g., dissolution of USAID), the incentives to leverage informational tools such as travel to conduct oversight over foreign policy might also change. Certain forms of direct observation as forms of oversight, such as site visits to ICE detention centers, have already been subject to partisan debate³⁹. It is unclear whether CODELs and other tools of direct observation can persist as an informative tool of oversight in an era of rising polarization.

39. <https://apnews.com/article/ice-detention-centers-congress-oversight-f94f1226ee3d8c0c57302370485cad0a>. Last accessed: August 7, 2026.

References

- Aberbach, Joel. 1990. *Keeping a Watchful Eye: The Politics of Congressional Oversight*. Brookings Institution Press.
- Alduncin, Alex, David C. W. Parker, and Sean M. Theriault. 2017. "Leaving on a Jet Plane: Polarization, Foreign Travel, and Comity in Congress." *Congress & the Presidency* 44, no. 2 (May): 179–200. <https://doi.org/10.1080/07343469.2016.1270370>.
- Alduncin, Alexander, Sean Q Kelly, David C.W. Parker, and Sean M. Theriault. 2014. "Foreign Junkets or Learning to Legislate? Generational Changes in the International Travel Patterns of House Members, 1977–2012." *The Forum* 12, no. 3 (January). <https://doi.org/10.1515/forum-2014-5021>.
- Ban, Pamela, and Jaclyn Kaslovsky. 2024. "Local Orientation in the U.S. House of Representatives." *American Journal of Political Science* (November): ajps.12929. <https://doi.org/10.1111/ajps.12929>.
- Ban, Pamela, Ju Yeon Park, and Hye Young You. 2022. "How Are Politicians Informed? Witnesses and Information Provision in Congress." *American Political Science Review* (May): 1–18. <https://doi.org/10.1017/S0003055422000405>.
- . 2025. "Bureaucrats in Congress: The Politics of Interbranch Information Sharing." *The Journal of Politics* (December): 000–000. <https://doi.org/10.1086/735440>.
- Bellodi, Luca. 2023. "A Dynamic Measure of Bureaucratic Reputation: New Data for New Theory." *American Journal of Political Science* 67, no. 4 (October): 880–897. <https://doi.org/10.1111/ajps.12695>.
- Box-Steffensmeier, Janet M., Dino P. Christenson, and Alison W. Craig. 2019. "Cue-Taking in Congress: Interest Group Signals from Dear Colleague Letters." *American Journal of Political Science* 63 (1): 163–180.
- Canes-Wrone, Brandice, William G. Howell, and David E. Lewis. 2008. "Toward a Broader Understanding of Presidential Power: A Reevaluation of the Two Presidencies Thesis." *The Journal of Politics* 70, no. 1 (January): 1–16. <https://doi.org/10.1017/S0022381607080061>.
- 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, no. 4 (November): 835–843. <https://doi.org/10.1017/S0003055417000302>.
- Carter, Ralph G., and James M. Scott. 2004. "Taking the Lead: Congressional Foreign Policy Entrepreneurs in U.S. Foreign Policy." *Politics & Policy* 32, no. 1 (March): 34–70. <https://doi.org/10.1111/j.1747-1346.2004.tb00175.x>.
- Curry, James M. 2019. "Knowledge, Expertise, and Committee Power in the Contemporary Congress." *Legislative Studies Quarterly* 44, no. 2 (May): 203–237. <https://doi.org/10.1111/lsq.12219>.
- Curry, James M., and Jason M. Roberts. 2022. "Interpersonal Relationships and Legislative Collaboration in Congress." *Legislative Studies Quarterly* (April): lsq.12381. <https://doi.org/10.1111/lsq.12381>.

- DiGiuseppe, Matthew, and Michael E Flynn. 2025. *Scaling Open-ended Survey Responses Using LLM-Paired Comparisons*, January. <https://doi.org/10.31235/osf.io/39ajg>.
- Drezner, Daniel W, and Linda L Fowler. 2025. "Where Have All the Experts Gone? The Shifting Marketplace for Foreign Policy Ideas on Capitol Hill." *International Studies Quarterly* 69, no. 2 (March): sqaf035. <https://doi.org/10.1093/isq/sqaf035>.
- Eldes, Ayse, Christian Fong, and Kenneth Lowande. 2023. "Information and Confrontation in Legislative Oversight." *Legislative Studies Quarterly* (December): lsq.12440. <https://doi.org/10.111/lsq.12440>.
- Esterling, Kevin M., and Ju Yeon Park. 2024. "Flexible Estimation of Policy Preferences for Witnesses in Committee Hearings." *Political Analysis* 32, no. 4 (October): 476–492. <https://doi.org/10.1017/pan.2024.6>.
- Fowler, Linda L. 2015. *Watchdogs on the Hill: The Decline of Congressional Oversight of U.S. Foreign Relations*. Princeton University Press.
- Garvey, Todd. 2025. *Site Visits, Information Gathering, and Oversight by Members of Congress*. Technical report LSB11376. Congressional Research Service, November.
- Goldgeier, James, and Elizabeth N. Saunders. 2018. "The Unconstrained Presidency: Checks and Balances Eroded Long Before Trump." *Foreign Affairs* 97 (5): 144–148, 150–156.
- Grimmer, Justin, and Eleanor Neff Powell. 2013. "Congressmen in Exile: The Politics and Consequences of Involuntary Committee Removal." *The Journal of Politics* 75, no. 4 (October): 907–920. <https://doi.org/10.1017/S0022381613000704>.
- Hall, Richard L. 1987. "Participation and Purpose in Committee Decision Making." *American Political Science Review* 81, no. 1 (March): 105–127. <https://doi.org/10.2307/1960781>.
- Hall, Richard L., and Alan V. Deardorff. 2006. "Lobbying as Legislative Subsidy." *American Political Science Review* 100, no. 1 (February): 69–84. <https://doi.org/10.1017/S0003055406062010>.
- Halterman, Andrew, and Katherine A. Keith. 2026. "Codebook LLMs: Evaluating LLMs as Measurement Tools for Political Science Concepts." *Political Analysis* 34, no. 2 (April): 188–204. <https://doi.org/10.1017/pan.2025.10017>.
- Hathaway, Oona A, Tobias Kuehne, Randi Michel, and Nicole Ng. 2021. "Congressional Oversight of Modern Warfare: History, Pathologies, and Proposals for Reform." *William & Mary Law Review* 63 (1): 137–218.
- Jones, Bryan D., Frank R. Baumgartner, Sean M. Theriault, Derek A. Epp, Cheyenne Lee, and Miranda E. Sullivan. 2023. *Policy Agendas Project: Congressional Hearings*. Technical report.
- Joseph, Michael F., and Michael Poznansky. 2024. "Secret Innovation." *International Organization* 78 (4): 766–799. <https://doi.org/10.1017/S0020818324000250>.
- Krehbiel, Keith. 1991. *Information and Legislative Organization*. University of Michigan Press.

- Kriner, Douglas, and Liam Schwartz. 2008. "Divided Government and Congressional Investigations." *Legislative Studies Quarterly* 33, no. 2 (May): 295–321. <https://doi.org/10.3162/036298008784310993>.
- Kriner, Douglas L., and Eric Schickler. 2016. *Investigating the President : Congressional Checks on Presidential Power*. Princeton University Press.
- Levin, Carl, and Elise J. Bean. 2018. "Defining Congressional Oversight and Measuring Its Effectiveness." *Wayne Law Review* 64 (1).
- Lowande, Kenneth. 2018. "Who Polices the Administrative State?" *American Political Science Review* 112, no. 4 (November): 874–890. <https://doi.org/10.1017/S0003055418000497>.
- Lowande, Kenneth, and Rachel Augustine Potter. 2021. "Congressional Oversight Revisited: Politics and Procedure in Agency Rulemaking." *The Journal of Politics* 83, no. 1 (January): 401–408. <https://doi.org/10.1086/709436>.
- Lowande, Kenneth, and Justin Peck. 2016. "Congressional Investigations and the Electoral Connection." *Journal of Law, Economics, and Organization* (December): 1–27. <https://doi.org/10.1093/jleo/eww014>.
- MacDonald, Jason A., and Robert J. McGrath. 2016. "Retrospective Congressional Oversight and the Dynamics of Legislative Influence over the Bureaucracy." *Legislative Studies Quarterly* 41, no. 4 (November): 899–934. <https://doi.org/10.1111/lsq.12138>.
- McCubbins, Mathew D., and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American Journal of Political Science* 28, no. 1 (February): 165. <https://doi.org/10.2307/2110792>. JSTOR: 2110792.
- McGee, Zachary A, and Sean M. Theriault. 2022. "Partisanship in Congressional Travels Abroad." *International Politics*.
- McGrath, Robert J. 2013. "Congressional Oversight Hearings and Policy Control." *Legislative Studies Quarterly* 38, no. 3 (August): 349–376. <https://doi.org/10.1111/lsq.12018>.
- Mills, Russell W., Nicole Kalaf-Hughes, and Jason A. MacDonald. 2016. "Agency Policy Preferences, Congressional Letter-Marking and the Allocation of Distributive Policy Benefits." *Journal of Public Policy* 36, no. 4 (December): 547–571. <https://doi.org/10.1017/S0143814X15000252>.
- Montgomery, Jacob M., Florian M. Hollenbach, and Michael D. Ward. 2012. "Improving Predictions Using Ensemble Bayesian Model Averaging." *Political Analysis* 20 (3): 271–291. <https://doi.org/10.1093/pan/mps002>.
- Park, Ju Yeon. 2017. "A Lab Experiment on Committee Hearings: Preferences, Power, and a Quest for Information." *Legislative Studies Quarterly* 42, no. 1 (February): 3–31. <https://doi.org/10.1111/lsq.12139>.
- . 2021. "When Do Politicians Grandstand? Measuring Message Politics in Committee Hearings." *The Journal of Politics* 83, no. 1 (January): 214–228. <https://doi.org/10.1086/709147>.

- Parker, David C.W., and Matthew Dull. 2009. "Divided We Quarrel: The Politics of Congressional Investigations, 1947–2004." *Legislative Studies Quarterly* 34, no. 3 (August): 319–345. <https://doi.org/10.3162/036298009788897790>.
- Quirk, Paul J. 2005. "Deliberation and Decision Making." In *The Legislative Branch*, edited by Paul J. Quirk and Sarah A. Binder. Oxford University Press.
- Ritchie, Melinda N. 2023. *Backdoor Lawmaking : Evading Obstacles in the US Congress*. Oxford University Press.
- Ritchie, Melinda N., and Hye Young You. 2019. "Legislators as Lobbyists." *Legislative Studies Quarterly* 44, no. 1 (February): 65–95. <https://doi.org/10.1111/lsq.12221>.
- Rosenson, Beth A. 2009. "Congressional Frequent Flyers: Demand- and Supply-Side Explanations for Privately Sponsored Travel." *Legislative Studies Quarterly* 34 (2): 245–271.
- Schlesinger, Arthur. 1973. *The Imperial Presidency*. Houghton Mifflin.
- Scoville, Ryan M. 2013. "Legislative Diplomacy." *SSRN Electronic Journal*, <https://doi.org/10.2139/ssrn.2211651>.
- Selin, Jennifer L., and Grace Moore. 2023. "Keeping Tabs on the Executive." *Presidential Studies Quarterly* 53, no. 2 (June): 186–208. <https://doi.org/10.1111/psq.12829>.
- U.S. House Subcommittee on National Security. 2008. *U.S. Coalition Support Funds to Pakistan*. House Majority Staff Report. United States House of Representatives, Committee on Oversight and Government Reform, Subcommittee on National Security and Foreign Affairs, Majority Subcommittee Staff, September.
- Volden, Craig, and Alan E. Wiseman. 2014. *Legislative Effectiveness in the United States Congress*. Cambridge University Press.

Appendix

Contents of the Appendix

A	Descriptive Patterns	38
A1	Patterns of Travel	38
A2	HFAC Hearings	40
A3	CODELs Precede Hearings	41
B	Additional Tables and Figures	42
C	Measuring Foreign Policy Specificity	49
C1	Instructions for Human and LLM Coders	49
C2	Labeling and Validating Coded Statements	51
C3	Predicting Foreign Policy Specificity	53
C4	Validating Foreign Policy Specificity	53
D	Interview Procedure	59

A Descriptive Patterns

A1 Patterns of Travel

To provide some predictive validity for CODELs as a measure of foreign policy relevance, I assess the predictors of CODEL destinations. If used for oversight, CODELs should primarily target countries where executive branch activity warrants monitoring. One salient way to identify countries that receive significant executive branch attention and activity is by identifying countries mentioned in the Presidential State of the Union (SOTU) address (Jones et al. 2023). Lee (2008, 917) argues that “raising a policy issue in the State of the Union Address indicates that the president is asserting leadership on that issue”. The SOTU serves as a high-profile platform where presidents publicly signal their policy priorities, making it a useful indicator of the executive foreign policy and country-specific priorities. Furthermore, scholars have employed SOTU mentions to measure presidential priorities in both domestic policy (Russell and Eissler 2022) and foreign policy (Peake 2001). When a president highlights a particular country in the SOTU address, it signals executive interest and suggests that country may warrant congressional oversight attention. Alternatively, I consider countries that are also presidential and Secretary of State travel destinations (Malis 2021) and the amount of US security assistance⁴⁰ and foreign aid⁴¹ received by a country.

Using all CODEL data from 1995-2015 and a two-way fixed-effect linear model, I regress the number of CODELs to a given country in a given year on these measures of executive attention. I include country-level controls that might affect both travel probability and executive attention, such as regime type (Coppedge et al. 2024) and ideological proximity to the US as measured by UN General Assembly votes (Bailey, Strezhnev, and Voeten 2017).

The results, shown in Table A1, show a positive and significant correlation between executive attention and CODEL destinations. Each additional SOTU mention of a country is associated with approximately 0.55 more CODELs and 0.37 additional travel days to that country, suggesting that congressional travel patterns align with countries receiving presidential attention. Notably, other variables indicating foreign policy relevance such as executive branch trips and logged weapons sales, are also positive and significant. Together, these patterns suggest that CODELs target countries where executive branch interest is likely greatest.

40. Center for International Policy. 2025. Security Sector Assistance Database. Source: <https://web.archive.org/web/20250312035415/https://internationalpolicy.org/programs/sam/security-sector-assistance-database/>. Last Accessed: February 25, 2025.

41. U.S. Department of State and U.S. Agency for International Development. 2025. U.S. Foreign Assistance Data. Source: <https://foreignassistance.gov/data>. Last Accessed: February 25, 2025.

Table A1: Relationship between SOTU Mentions of a Country Destination on Number of CODELs to that Country (1995-2015)

	CODELs	CODELs	CODELs	CODELs	Days on CODELs
SOTU Mentions	0.65** (0.11)	0.60** (0.08)	0.57** (0.11)	0.55** (0.12)	0.37** (0.08)
V-Dem Polyarchy		0.09 (0.27)	1.28 [†] (0.76)	1.59* (0.70)	3.97** (1.36)
Logged US Weapons Sales		0.02 [†] (0.01)	0.01* (0.01)	0.03** (0.01)	0.06** (0.02)
UNGA Vote Prox. to US		1.70** (0.55)	−4.39** (0.94)	−0.87 (0.87)	−2.85 (1.80)
Ln(Aid + 1)		−0.04* (0.02)	−0.01 (0.02)	0.00 (0.02)	−0.07 (0.05)
Executive Branch Trips		0.83** (0.10)	0.47** (0.07)	0.38** (0.07)	0.75** (0.14)
Observations	4846	4190	4190	4190	4190
Country FEs	✓		✓	✓	✓
Year FEs	✓	✓		✓	✓
Adj. R ²	0.52	0.39	0.51	0.55	0.51

The unit of analysis is the country-year. Table entries are linear regression coefficients with standard errors (clustered by country) shown in parentheses. The dependent variable in columns 1-4 is the number of CODELs. Column 5 captures the number of days spent on CODELs. ** $p < 0.01$; * $p < 0.05$; [†] $p < 0.1$

A2 HFAC Hearings

Figure A1 presents the number of HFAC hearings (left hand panel) and number of hearing remarks (right hand panel) over time. Notably, the number of hearings and remarks generally increases over time. This reflects underlying foreign policy trends including the US involvement in the Middle East following 9/11.

Figure A1: Number of HFAC Hearings and Remarks Over Time (1999-2006)

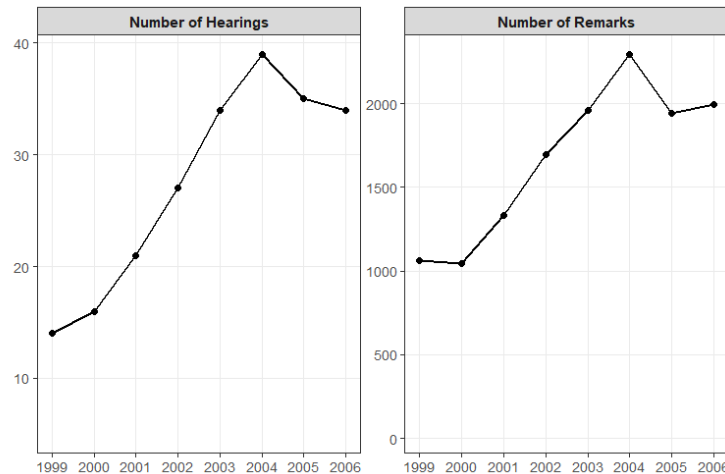
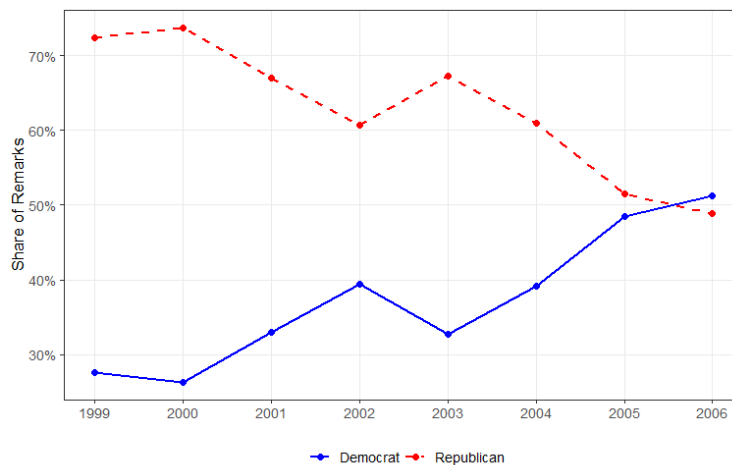


Figure A2 shows the share of HFAC hearing remarks by party over time. Notably, Republicans generally dominate HFAC hearing discussions, as they control the majority during this time period. Democrats become more active in HFAC hearings in the latter years of the time period and the Bush administration.

Figure A2: Share of all HFAC Remarks Made by Party (1999-2006)



A3 CODELs Precede Hearings

To estimate the within-region effect of CODELs on hearing onset, I estimate the following negative binomial regression regression:

$$\text{HFAC Hearings}_{t,r} = \beta \text{CODELs}_{t-1,r} + \alpha_r + \gamma_t + \varepsilon_{t,r} \quad (3)$$

where the subscripts indicate month (t) and CODEL destination region (r). The unit of observation is region-month, allowing me to estimate the within-region effect CODELs on oversight hearings. The outcome variable $\text{HFAC Hearings}_{t,r}$ is the number of HFAC hearings pertaining to region r in a given month t . The main independent variable are $\text{CODELs}_{t-1,r}$, which captures the number of CODELs directed at region r in the previous month $t - 1$. I lag this variable by one month to account for the delay between when CODELs inform members and when the hearings actually take place. I include region fixed effects (α_r) to account for characteristics of the CODEL destination region that might drive probability of travel and oversight hearings. I also include month-year fixed effects (γ_t) to account for time varying confounders.

Table A2 presents the results of estimating equation 3. In all specifications, CODELs are positively correlated with the number of subsequent, relevant HFAC hearings. This relationship is statistically significant in specifications with region fixed effects. While statistical significance is not archived with the inclusion of time fixed effects, the relationship is still positive.

Table A2: CODELs and Hearing Onset

Dependent Variable:	Number of Hearings			
Model:	(1)	(2)	(3)	(4)
CODELs (t-1)	0.06** (0.01)		0.03 (0.02)	
Logged Travel Days (t-1)		0.17* (0.06)		0.06 (0.03)
Region FE	Yes	Yes	Yes	Yes
Month-year FE			Yes	Yes
Observations	575	575	575	575
Dependent variable mean	0.83	0.83	0.83	0.83
Within Adjusted R ²	0.04	0.02	0.01	0.002

The unit of analysis is month-year-region. Standard errors, shown in parentheses, are clustered by region.

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

B Additional Tables and Figures

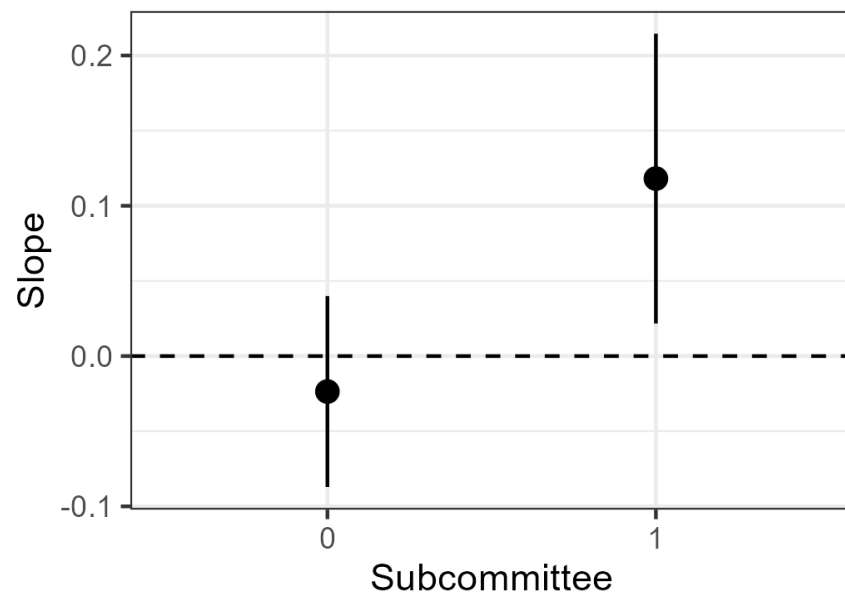
Table B1: Summary Statistics: All Hearings

	N	Mean	SD	Min	Max
Subcommittee	7306	0.36	0.48	0.00	1.00
Committee Seniority	7306	4.47	3.44	1.00	15.00
Distance from Median	7196	0.42	0.23	0.00	0.91
LES	7196	0.97	1.34	0.00	11.51
Prev. Vote Share	7113	70.84	13.30	51.00	100.00
Traveled in Prev. Month	7306	0.04	0.19	0.00	1.00
Logged Trip Days in Prev. Month	7306	0.06	0.31	0.00	2.94
Spoke in Hearing	7306	0.22	0.41	0.00	1.00
Speaking Instances (n)	7306	1.82	5.60	0.00	78.00
Logged Speaking Instances	7306	0.40	0.86	0.00	4.37
Total Words	7306	225.53	678.81	0.00	11321.00
Total Questions	7306	1.42	4.46	0.00	75.00
Total Bureaucrat Address	7306	0.73	2.59	0.00	59.00
Total Experiential Mentions	7306	0.03	0.23	0.00	6.00
Any Experiential Mentions	7306	0.02	0.14	0.00	1.00
Average Specificity Score	1552	55.03	11.43	9.63	90.47

Table B2: Summary Statistics: Committee and Subcommittee Hearings

	N	Mean	SD	Min	Max
Panel A: Committee Hearings					
Committee Seniority	4690	4.54	3.47	1.00	15.00
Distance from Median	4594	0.42	0.23	0.00	0.91
LES	4594	1.00	1.37	0.00	9.99
Prev. Vote Share	4541	71.03	13.41	51.00	100.00
Traveled in Prev. Month	4690	0.03	0.18	0.00	1.00
Logged Trip Days in Prev. Month	4690	0.05	0.27	0.00	2.94
Spoke in Hearing	4690	0.19	0.39	0.00	1.00
Speaking Instances (n)	4690	1.26	4.35	0.00	68.00
Logged Speaking Instances	4690	0.31	0.74	0.00	4.23
Total Words	4690	151.56	471.57	0.00	7982.00
Total Questions	4690	0.95	3.19	0.00	53.00
Total Bureaucrat Address	4690	0.52	2.23	0.00	52.00
Total Experiential Mentions	4690	0.02	0.19	0.00	6.00
Any Experiential Mentions	4690	0.02	0.12	0.00	1.00
Average Specificity Score	862	55.95	12.04	12.59	90.47
Panel B: Subcommittee Hearings					
Committee Seniority	2616	4.33	3.37	1.00	15.00
Distance from Median	2602	0.43	0.24	0.00	0.91
LES	2602	0.94	1.29	0.00	11.51
Prev. Vote Share	2572	70.51	13.09	51.00	100.00
Traveled in Prev. Month	2616	0.05	0.22	0.00	1.00
Logged Trip Days in Prev. Month	2616	0.08	0.37	0.00	2.77
Spoke in Hearing	2616	0.27	0.44	0.00	1.00
Speaking Instances (n)	2616	2.80	7.22	0.00	78.00
Logged Speaking Instances	2616	0.56	1.02	0.00	4.37
Total Words	2616	358.15	927.92	0.00	11321.00
Total Questions	2616	2.28	6.01	0.00	75.00
Total Bureaucrat Address	2616	1.12	3.09	0.00	59.00
Total Experiential Mentions	2616	0.04	0.27	0.00	5.00
Any Experiential Mentions	2616	0.03	0.17	0.00	1.00
Average Specificity Score	690	53.89	10.51	9.63	89.13

Figure B1: CODEL Participation and Hearing Participation by Hearing Setting



This figure plots the interaction effect of *Traveled in Prev. Month* $\times$ *Subcommittee*. Controls and legislator-region and congress fixed effects are included. Standard errors are clustered by legislator.

Table B3: CODEL Travel and Intensity of Hearing Participation (full table)

Dependent Variables:	Speaking Instances	Total Words	Total Questions	Total Bureaucrat Addresses
Traveled in Prev. Month	0.0728 (0.0709)	0.3928* (0.1941)	0.0612 (0.0689)	0.0366 (0.0540)
Subcommittee	0.1152* (0.0451)	0.1638 (0.1353)	0.1069** (0.0368)	0.0967*** (0.0242)
Committee Seniority	0.3064*** (0.0560)	0.9087*** (0.1478)	0.2268*** (0.0493)	0.2071*** (0.0354)
Dist. from Median	-0.0971 (0.6750)	0.2208 (2.074)	0.1777 (0.5738)	0.1915 (0.3504)
LES	0.0144 (0.0313)	0.0453 (0.0777)	0.0154 (0.0283)	0.0130 (0.0251)
Prev. Vote Share	-0.0048** (0.0016)	-0.0159** (0.0051)	-0.0045** (0.0016)	-0.0026* (0.0011)
Legislator-region FE	Yes	Yes	Yes	Yes
Congress FE	Yes	Yes	Yes	Yes
Observations	6,889	6,889	6,889	6,889
Dependent variable mean	0.41108	1.4335	0.35263	0.23933
Within Adjusted R ²	0.00702	0.00445	0.00626	0.00711
Std. Dev. (dep. var.)	0.87028	2.7322	0.80575	0.61126

Note: All dependent variables are logged. The unit of observation is hearing-legislator. Standard errors clustered by legislator. ** $p < 0.01$; * $p < 0.05$; $^{\dagger} p < 0.1$

Table B4: CODEL Travel, Information, and Hearing Engagement w/ two-way clustered standard errors

Dependent Variables:	Experiential Mentions (0/1)	Logged Experiential Mentions	Average Specificity Score	
Traveled in Prev. Month	0.0484** (0.0172)	0.0520** (0.0181)	3.109*** (0.9295)	
Logged Trip Days				1.411** (0.5010)
Subcommittee	0.0042 (0.0048)	0.0042 (0.0041)	1.374 [†] (0.8243)	1.384 [†] (0.8312)
Committee Seniority	-0.0249 (0.0254)	-0.0167 (0.0175)	5.761*** (0.6538)	5.926*** (0.6566)
Dist. from Median	0.0262 (0.0786)	0.0418 (0.0658)	-24.05* (10.17)	-23.36* (10.14)
LES	-0.0020 (0.0022)	-0.0022 (0.0021)	0.6991* (0.2959)	0.6881* (0.2953)
Prev. Vote Share	-0.0005 (0.0004)	-0.0004 (0.0003)	-0.0676 (0.0573)	-0.0678 (0.0571)
Speaking Instances (n)			-0.4664*** (0.0446)	-0.4684*** (0.0445)
Legislator-region FE	Yes	Yes	Yes	Yes
Congress FE	Yes	Yes	Yes	Yes
Observations	7,070	7,070	1,442	1,442
Dependent variable mean	0.02122	0.01722	55.087	55.087
Within Adjusted R ²	0.00415	0.00680	0.12862	0.12648
Std. Dev. (dep. var.)	0.14412	0.12292	11.312	11.312

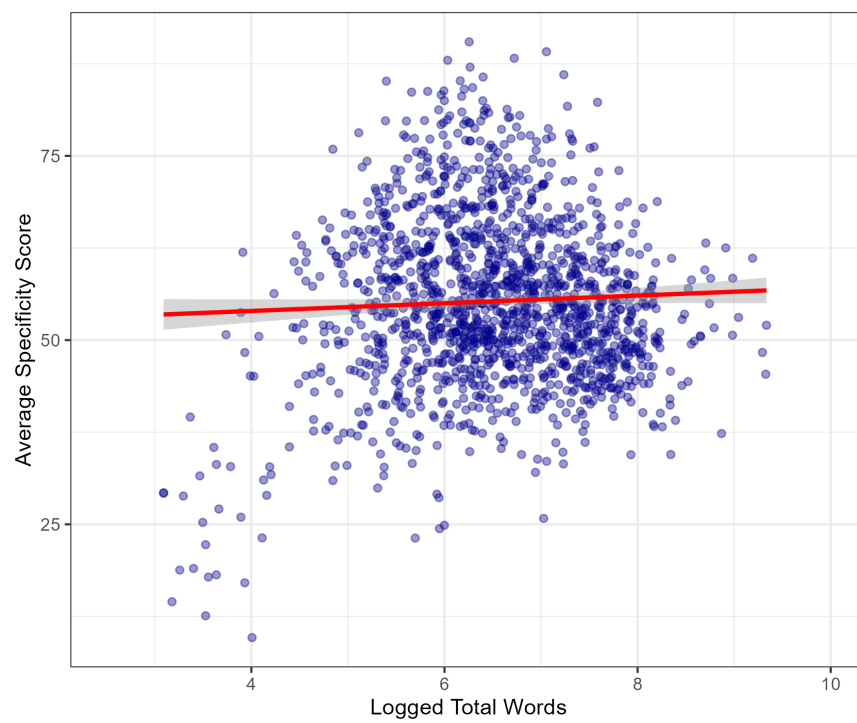
Note: The unit of observation is hearing-legislator. Standard errors clustered by legislator-region. Columns 1 and 2 include the full sample. Columns 3 and 4 only include legislator-hearing observations in which a legislator spoke in that hearing. ** $p < 0.01$; * $p < 0.05$; [†] $p < 0.1$

Table B5: CODEL Travel and Average Specificity Scores w/o Chairs

Dependent Variable:	Average Specificity Score	
Traveled in Prev. Month	3.226** (1.134)	
Logged Travel Days		1.421* (0.6643)
Subcommittee	2.016* (0.8511)	2.048* (0.8602)
Committee Seniority	5.743*** (1.027)	5.914*** (1.035)
Dist. from Median	-19.53 (14.10)	-19.07 (14.13)
LES	0.5465 (0.4750)	0.5348 (0.4787)
Prev. Vote Share	-0.0720 (0.0691)	-0.0729 (0.0693)
Speaking Instances (n)	-0.6443*** (0.0873)	-0.6435*** (0.0879)
Legislator-region FE	Yes	Yes
Congress FE	Yes	Yes
Observations	1,233	1,233
Dependent variable mean	56.124	56.124
Within Adjusted R ²	0.12135	0.11895
Std. Dev. (dep. var.)	11.486	11.486

The unit of observation is hearing-legislator. Standard errors clustered by legislator. Analysis include legislator-hearing observations in which a legislator spoke in that hearing and omits the chair of the hearing. ** $p < 0.01$; * $p < 0.05$; † $p < 0.1$

Figure B2: Foreign Policy Specificity and Statement Length



This figure plots the correlation between a legislator's average foreign policy specificity score and their logged total number of words spoken.

C Measuring Foreign Policy Specificity

C1 Instructions for Human and LLM Coders

You will be presented with pairs of statements taken from US congressional hearing transcripts. For each pair, your job is to choose which statement contains more specific, concrete information about the foreign policy situation being discussed - such as references to named people, places, policies, quantities, or on-the-ground conditions.

There are no right or wrong answers. We are interested in your judgment. Read each statement carefully and make your best choice, even when the difference is subtle.

Background: What Is a Congressional Hearing?

Congressional committees hold hearings for various purposes: to oversee the executive branch, to gather information for legislation, to confirm government nominees, or to review budgeting plans.

A typical hearing proceeds as follows:

- The committee chair opens with a statement, followed by opening remarks from other committee members and invited witnesses
- The chair then opens a Q&A session in which committee members ask questions of the witnesses
- Witnesses may include government officials, experts, advocates, or other individuals with relevant knowledge

Statements in your task come primarily from the Q&A session of hearings. Opening statements are omitted.

What Makes a Statement More Foreign Policy Specific?

A statement is more foreign policy specific when it contains concrete, detailed references to people, places, quantities, policies, events, or institutions that are *directly relevant to US foreign policy*.

Indicators of a more specific statement:

- Named individuals (e.g., a foreign leader, a U.S. official, a witness)
- Named regions, cities, or specific geographic locations
- Specific numbers, percentages, dollar amounts, or dates
- Named policies, treaties, programs, or legislation
- Concrete actions, events, or policy recommendations

Example of a MORE specific statement:

“Finally, Dr. Holmberg, you mentioned 14 million blood units are missing or short in Africa. Obviously, reducing unnecessary transfusions would be one way of mitigating that. Even though you want to build up supply, stopping those that are not necessary would be an important goal. How much progress are the African nations making on trying to stop those unnecessary transfusions? Finally, if one of you might answer this: WHO says 5 to 10 percent of the HIV infections are attributable to bad blood.

The number you gave was 3.2 percent. Is that because of better calculations or is it that the programs are beginning to work?"

Why more specific: references a named expert, a specific quantity (14 million blood units), a named region (Africa), a named organization (WHO), and specific statistics (5–10% vs. 3.2%).

What Makes a Statement Less Foreign Policy Specific?

A statement is less foreign policy specific when it is vague, abstract, procedural, or unclear in its connection to U.S. foreign policy.

Indicators of a less specific statement:

1. **Vague or general foreign policy posturing;** references to policy or countries without concrete detail:

"Thank you, Mr. Chairman. It seems clear the ball is in Hamas's court. There are three things they have to do — it is just that simple: Renounce violence, the absolute right and recognition of Israel to exist, and accept and honor all prior agreements, as any sovereign nation would be. We, in the United States, no matter what President was here, we have got to honor agreements. But the ball is in their court. We should not, and cannot, fund Hamas or send money over there until we know that they accept the terms of the agreement. It is just that plain and simple."

Why less specific: mentions Hamas and Israel but makes no reference to specific policies, amounts, agreements, events, or people beyond vague general claims.

2. **No clear foreign policy focus:**

"Thank you very much, Mr. Chairman. It pleases me to see that some of my colleagues now are understanding why some of us on this side of the aisle have been so hesitant about putting America's security in the hands of an organization like the United Nations. This difficult organization, as my colleague mentioned, is not something that I would trust with American security. I would now yield 1 minute of my time to Mr. Burton."

Why less specific: no discernible foreign policy content.

3. **Procedural statements** — logistics, scheduling, recognizing speakers:

"I'd like to recognize Mr. Menendez for his line of questioning before we break and take a short recess for votes. Mr. Menendez."

Why less specific: purely procedural, no substantive content.

Decision Rules

Hard Cases

- **One statement is specific but off-topic; the other is vague but about foreign policy:** Choose the foreign policy statement even if it is less specific in general terms.

Table C1: Decision Rules for Coding Statement Pairs

Statement A	Statement B	Choose
Contains specific foreign policy references	Contains no specific references	A
Contains no specific references	Contains specific foreign policy references	B
Both contain specific references	Both contain specific references	Whichever contains more explicit, more numerous, or more precise references
Neither contains specific references	Neither contains specific references	Whichever makes more substantive claims, even if vague. Prefer a general foreign policy claim over a procedural or purely rhetorical statement

- **Both statements are about foreign policy but one is only slightly more specific:** Make your best judgment and choose the more specific one. It is fine if the difference is small — we expect some pairs to be close calls.
- **A statement mixes foreign policy content with procedural filler:** Judge the statement by its substantive content, not its filler. If the core of the statement contains specific foreign policy references, treat it as specific.
- **You are unsure whether a topic is foreign policy relevant:** When in doubt, ask: Does this statement concern U.S. relationships, actions, or interests with respect to other countries, international organizations, or global issues? If yes, treat it as foreign policy relevant.

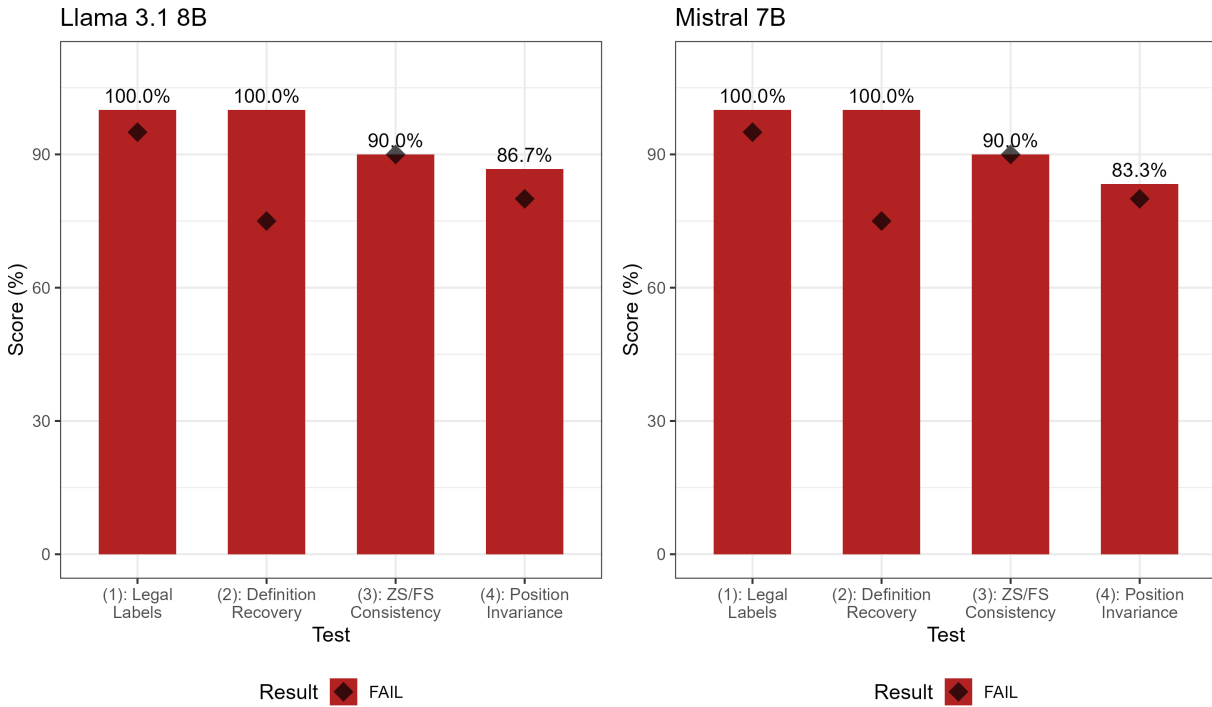
Important Reminders

- Do not use your own outside knowledge of a person, policy, or event to judge specificity. Judge only by what is explicitly stated in the text.
- There is no correct answer. We are measuring your perception of specificity, not factual accuracy.
- Make your best guess even when uncertain. Flag pairs you're unsure about. Close calls are expected and normal.
- Read both statements fully before making your choice.

C2 Labeling and Validating Coded Statements

To code the pairwise comparisons, I use a large language model (LLM) and validate the usage of an LLM with a human coded sample. To get the human coded comparisons, I randomly sampled 600 legislative statements made in HFAC hearings from 1999-2006. I omit all statements made by witnesses. First, I had 3 RAs code 450 unique pairs of statements following the procedures outlined in the codebook. To ensure that human coders would not get bogged down by lengthy legislative statements, I limit these comparisons to statements 120 characters or less. Each RA coded 50 of the same pairs to ensure reasonably consistent interpretations of the codebook and

Figure C1: Behavioral Test Results for LLMs



The bars show the achieved rates on each of the behavioral tests. Diamond markers denote pass thresholds for acceptable performance.

inter-coder reliability among human coders. The Cohen's Kappa between RAs range from 0.63-0.96 (82%-98% agreement), indicating a high level of inter-coder reliability among RAs. After correcting for coding mistakes (e.g., misapplications of the codebook), the RAs proceeded to code the remaining 400 unique pairs.

I do a couple of things to validate the use of an LLM to code the full sample of comparisons. First, I conduct a few behavioral tests to ensure that the LLM can follow instructions and make consistent predictions using a framework for applying LLMs to codebook instructions outlined by Halterman and Keith (2026). These tests assess: (1) Can the LLM return the correct labels and outputs? (2) Can the LLM correctly categorize examples from the codebook? (3) Can the LLM correctly classify pairs when given the codebook examples and non-codebook pairs? (4) Does swapping which statement is first and which is second consistently flip the LLM's answer? I run these behavioral tests on two open-source LLMs: Llama 3.1 8B and Mistral 7B.⁴² Both LLMs performed at acceptable rates (Table C1). I chose to use Llama 3.1 8B given its slightly better performance in these tests.

Second, I validate the LLM's outputs against the human coders. I conducted the same 450 pairwise comparisons using the LLM. The Cohen's Kappa between LLM and any given RA ranges

42. I choose open source models over larger, proprietary models (e.g., Claude, GPT) to maximize reproducibility. These were the two models that performed best in Halterman and Keith (2026).

between 0.45-0.63 (74%-82% agreement). On the 50 overlap pairs coded by all three coders, individual agreement with the majority vote ranged from 86% (LLM) to 100% among human coders (mean RA agreement = 96.7%). These results are consistent with Halterman and Keith (2026) finding that LLMs introduce modest additional variance relative to trained human coders.

To more rigorously assess whether the the addition of LLM significantly alters the comparison outcomes, I bootstrap ($n = 1000$) confidence intervals for the Krippendorff's alphas, used to assess inter-coder reliability among more than one coder, for the three human coders ($\alpha = 0.751$, 95% CI = [0.581, 0.886]) and the human coders plus the LLM ($\alpha = 0.651$, 95% CI = [0.487, 0.780]). The bootstrapped alphas overlapped substantially, suggesting the LLM's addition did not significantly alter coding reliability compared to human coders.

Next, I randomly sample 600 statements from my full corpus of 18,395 legislator statements within HFAC hearings for the training set. I omit all statements less than 20 characters to avoid as many procedural statements as possible. Each statement was compared ten times, generating 3000 comparison tasks for the LLM to code. Then, I fit a Bradley-Terry model on these codings.

C3 Predicting Foreign Policy Specificity

Of the 600 sampled legislative statements, I randomly select 500 statements for training and set aside 100 for validation. I pre-process the text, removing punctuation and lower-casing. I constructed document matrices using two methods: Term Frequency-Inverse Document Frequency (TF-IDF) and doc2vec. Because the doc2vec matrix struggled with statements longer than approximately 950 words these were split into sub-segments for embedding and their vectors averaged

I fit four machine learners that are used in Ban, Park, and You (2025), which also utilizes House hearing transcripts, on both document matrices. Using EBMAforecast Montgomery, Hollenbach, and Ward (2012), I aggregate and optimize across these eight models. All but one model was excluded from the ensemble. Using this final model, I predicted a foreign policy specificity score for each statement within the corpus and scaled this measure from 0-100.

C4 Validating Foreign Policy Specificity

To qualitatively validate this measure, I provide a sample of five of the most foreign policy specific statements, five of the least foreign policy specific statements, and five statements that are average on foreign policy specificity.

High Specificity

Score: 100

Equally distressing, humanitarian workers and journalists continue to be at risk, as witnessed by the brutal killing of Dutch journalist, Sander Thoenes on September 21st, and the more recent murder of the Indonesian journalist I just described. If the situation in East Timor remains dire, the conditions of those East Timorese stranded in West Timor is, in Secretary Albright's words, simply "appalling." As I speak, perhaps as many as 230,000 East Timorese, possibly over a quarter of the East Timorese

population, have fled or been forced to flee to West Timor. Despite concerted pressure from the international community, we still have no evidence that the Indonesian Government has disarmed or disbanded any of the militias. Assistant Secretary Taft will describe, in more detail, what she personally witnessed in the camps, in particular the heavy militia presence in the camps. While the Indonesian Government has indicated that the East Timorese in the camps will be allowed to return home, we have no evidence that they have been permitted to do so. To the contrary, reports tell us that the Indonesian Government is deporting some of these East Timorese to other parts of Indonesia. . . .

Score: 98.26

Mr. Chairman, Mr. Lantos, Members of the Committee, I am here today to tell you about a remarkable proliferation policy success. As a verifier in the world of mass destruction I don't often come bearing good tidings. As former congressional staff, I know that oftentimes the question that Members have to ask is what went wrong? What we need to do today is to focus on what went right. Our challenge is to ensure not only that Libya's WMD programs are verifiably eliminated completely but also to ensure that it marks the beginning of a trend. I would like to offer a brief chronology, highlight some of the elimination accomplishments achieved thus far, identify what needs to be done still, and highlight the elements of the President's counter-proliferation policy that have made it possible. . . .

Score: 97.80

Thank you, Mr. Chairman. And thank you for inviting us to speak today and testify on a situation in Angola and our policy toward that war-torn country. As you mentioned, unfortunately, 5 years ago you held a hearing on Angola. I would dare say that there have been dozens of hearings on Angola over the past 20 years, and we hope that we have now come to a place in Angola's history where we can have hearings not so much on the war or a potential peace process, but, in fact, on the reconstruction and redevelopment of that country. As you mentioned, on February 22, 2002, Jonas Savimbi was killed. His death set in motion a series of events that led to the March 30th ceasefire and demobilization agreement reached in Luanda. The accord reached between UNITA and the Angolan government was a very ambitious one. It established sites throughout the country for UNITA forces and neighboring areas for family members to congregate. . . .

Score: 97.67

I will try and keep it to 5 minutes or less. Thank you, Mr. Chairman, for this opportunity to discuss with you and your Committee U.S. policy toward Liberia and

the Department of Defense's role. The Defense Department's activities in Liberia build upon our security cooperation programs which are designed and developed to develop the capabilities of allied and friendly militaries for self-defense and coalition operations. In early June the Department of Defense deployed approximately 1,800 personnel offshore Liberia to assist if needed in securing the U.S. Embassy and evacuating Americans and foreign nationals due to the threat posed by advancing rebels and undisciplined government forces. Our forces withdrew on 17 June when the rebels halted their offensive and the security situation in the capital appeared to stabilize. ...

Score: 97.15

Madam Chairman, thank you. It is a pleasure to be here today. This is a very timely hearing, among other reasons that I will address, but just within the past few hours, we have been informed that Iran has announced a substantial resumption of its uranium enrichment program, reneging on the commitment that it made to the United Kingdom, Germany and France by informing them and the IAEA that it will begin next week the production of uranium centrifuge parts and equipment assembly and testing. This is yet another example of Iran thumbing its nose at the international community given that just last week the IAEA Board of Directors unanimously called on Iran to affirm its commitment to the three European countries not to do precisely that and to maintain the suspension of its uranium enrichment program that it previously committed to the Europeans. ...

Medium Specificity

Score: 55.05

Thank you, Mr. Chairman. I want to thank all of you for your testimony and for staying through a rather long hearing but an important one. I would disagree with you on one thing, Mr. Calzan. State secrets were revealed here today. The state secret of how despotic and how systematic are the violations of human rights inside of Cuba. That is a state secret, at least one that the state would like to keep. Not a very good secret, but it is nonetheless a state secret. I appreciate my colleague from New Jersey, who has said that he would express concern if any one of those three individuals suffered a consequence simply because of their testimony here today. You know, I am reminded by some who say that events such as these are a significant change. The question is, do they portend significant change? You know, Nelson Mandela spoke out forcefully about apartheid in South Africa. ...

Score: 55.05

Mr. Tancredo, just to answer that, if the KMT leader becomes the President of his country, he will be in the same category. It is actually the exact opposite. In other words, when people aren't government officials, we can talk and we work with them. But when they are the President of the country, under the One-China Policy, it is in a different set of category. So we try to have contacts with a lot of other lower-level

officials in Taiwan. And as I said, we can have them on the economic side. So I guess what I hope is, I hope we wouldn't discourage the contacts we have with the Taiwanese when they come.

Score: 55.05

It fits, and it was part of the ultimate peace agreement in El Salvador. So it is very interesting to go to El Salvador today. When I was USAID Director, there were three former FMLN guerillas who welcomed me with USAID hats on. They had a coffee co-op and they were selling coffee to some distributor in New York, I think. Mr. Chairman, people are not naturally guerillas, but they are naturally entrepreneurs. If you just give them the opportunity, they will become entrepreneurs.

Score: 55.05

Mr. Chairman, you mentioned the Intel example in Costa Rica, and you recall that was a multi-year strategy on the part of the Costa Rican government. It involved a carefully negotiated trade regime that made Costa Rica competitive for an Intel kind of investment. Looking at the Caribbean outside of the unique case of Haiti, a country that you are quite familiar with, the Caribbean has, historically had high standards of education. In some ways, that is not really the challenge. The challenge conceptually is education versus what has been a problem in the Caribbean, which is an insular vision or mentality. You can have highly educated people, but if they only look around their immediate neighborhood, it is not going to be fully articulated. But you probably have a consensus among the three statements that you have heard here, that in the Caribbean, the combination of high-value human resources and new technologies, will allow the region to enter the 21st century with a high degree of competitiveness.

Score: 55.05

I am trying to look at the time, Mr. Chairman. I just wanted to follow up with one last thing. I am not trying to monopolize your time, but I am still focused on that gap, and I guess my concern is that the social foundations, that net needs to be there whether it is health care or education, all those kinds of things. But there have to be jobs created, because they have done a lot of the homework. Some of this may only be on paper in terms of democratic institutions, but they really have done a lot of the hard work. They are not experienced in the financial remuneration. So my concern is the 80 percent who don't have- of the have-nots, how are we going to engage them so that they have the benefit of some of these economic rewards? That is the gap that I am really interested in, because I think a lot of the social progress will follow if the funding is there. The remuneration is there.

Score: 55.04

Mr. Pitts, the entire issue of Gujarat is one that has been of great concern to us; it is one that we have been following very closely. Immediately after it happened, our officers from Mumbai went down to observe the situation themselves. And we were very quick to express our concern, our real sympathy for the people who underwent this horrible event, and continue to do this.

Low Specificity

Score: 5.41

We are going to have a second round, so if I could indicate your time has expired, Dan, and then we will come back.

Score: 5.70

Thank you so much. We have been joined by Congressman Schiff. I do not know, Adam, if you would like to make some opening statements before we get to the questions.

Score: 5.77

Thank you, Dr. Salem. Next, Dr. Marius Deeb, a professor at Johns Hopkins University School of Advanced International Studies. Dr. Deeb.

Score: 5.89

Thank you so much, Congressman Engel. We appreciate your testimony. I know that we will be working together on some follow-up pieces of legislation. Thank you very much for your testimony today.

Score: 6.12

I have one that I submitted for the record, Mr. Chairman, but I would like to say thank you very much for holding this hearing.

To statistically validate this measure, I assess how the predicted scores compare against the Bradley-Terry scores from statements that were not used in the training set. I find that the ensemble model achieves a Pearson's correlation coefficient of 0.76 (see Table C2). This is also plotted in Figure C2.

Table C2: Model Performance Comparison

Model	Pearson r	RMSE	NRMSE (Range)	NRMSE (Mean)
svm_tfidf	0.618	2.623	0.236	39.71
ksvm_tfidf	0.686	1.936	0.174	29.32
lasso_tfidf	0.573	2.215	0.199	33.53
gbm_tfidf	0.710	1.845	0.166	27.94
svm_d2v	0.712	1.839	0.166	27.84
ksvm_d2v	0.715	1.831	0.165	27.72
lasso_d2v	0.722	1.814	0.163	27.47
gbm_d2v	0.603	2.107	0.190	31.90
ebma_ensemble	0.761	1.791	0.161	27.12

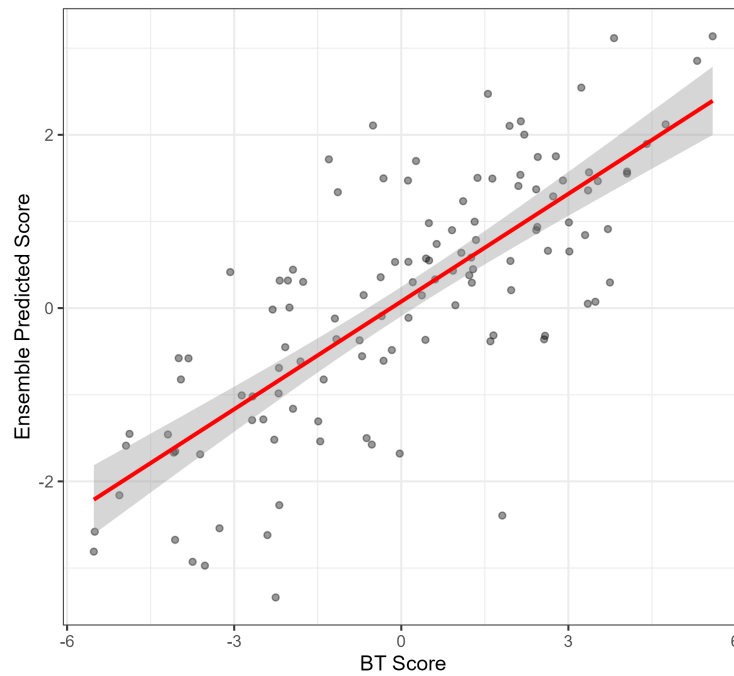
This table compares each model performance with the Bradley-Terry scores from statements that were not used in the training set.

In addition, I assess whether foreign policy specificity is associated with statement length in Table B2.

Figure C2: Ensemble Model Performance

Validation: Ensemble Predictions vs BT Scores

Pearson $r = 0.761$ | RMSE = 1.791 | $n = 120$



This figure compares the ensemble model scores with the Bradley-Terry scores from statements that were not used in the training set. Note that $n = 120$ because several statements were broken up due to length.

D Interview Procedure

I use several insights from interviews with former and current congressional staff members and State Department officials. These interviews were used to explore the nature and role CODELs played in foreign policymaking. Interviews were conducted in 2025 and 2026. All interviews conducted were approved by UCSD's IRB (IRB # 812103).

The target participants of this study are current or former congressional staff members or State Department officials who have observed, been involved with and/or have organized congressional delegations abroad. Research participants were recruited using a snowball sampling procedure. I contacted potential participants via email, LinkedIn messages, or phone calls using publicly available information or through recommendations from the research coordinator's professional network. Those who voluntarily consent to being interviewed will also be asked if they can provide additional contacts from their personal or professional networks.

Upon initial meeting, participants were first informed about the purpose of the study and asked if they would be willing to participate in the interview and/or recorded. Participants were also given to the

Respondents were given the following four options regarding anonymity. Most interviewers agreed to option # 2, "On Background". One participant chose to participate off the record; these insights were not included in the study. All names, specific occupational titles, and references to a staffer's former or current Congressional members were omitted to protect the respondent's privacy.

1. On Background: Anonymize the interviewee's name but can quote/paraphrase what the interviewee said.
2. Off the Record: Don't quote or cite any information from the interviews.

Interview participants were narrowed down based on whether they had experience with or could provide on CODELs specifically. Table D1 shows a list of interviewees along with their general status as it related to CODELs.

Table D1: Interview Subjects

Interview Number	CODEL Connection
1	Current State Department Official
2	Current State Department Official
3	Former Personal Staff Member
4	Former Personal Staff Member
5	Former Personal Staff Member
6	Current Personal Staff Member
7	Current Personal Staff Member

References

- Bailey, Michael A., Anton Strezhnev, and Erik Voeten. 2017. "Estimating Dynamic State Preferences from United Nations Voting Data." *Journal of Conflict Resolution* 61, no. 2 (February): 430–456. <https://doi.org/10.1177/0022002715595700>.
- Ban, Pamela, Ju Yeon Park, and Hye Young You. 2025. "Bureaucrats in Congress: The Politics of Interbranch Information Sharing." *The Journal of Politics* (December): 000–000. <https://doi.org/10.1086/735440>.
- Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Fabio Angiolillo, et al. 2024. *V-Dem Country-Year Dataset V14*. <https://doi.org/10.23696/vdemds24>.
- Halterman, Andrew, and Katherine A. Keith. 2026. "Codebook LLMs: Evaluating LLMs as Measurement Tools for Political Science Concepts." *Political Analysis* 34, no. 2 (April): 188–204. <https://doi.org/10.1017/pan.2025.10017>.
- Jones, Bryan D., Frank R. Baumgartner, Sean M. Theriault, Derek A. Epp, Rebecca Eissler, Cheyenne Lee, and Miranda E. Sullivan. 2023. *Policy Agendas Project: State of the Union Speeches*.
- Lee, Frances E. 2008. "Dividers, Not Uniters: Presidential Leadership and Senate Partisanship, 1981-2004." *The Journal of Politics* 70, no. 4 (October): 914–928. <https://doi.org/10.1017/S0022381608080961>.
- Malis, Matt. 2021. "Conflict, Cooperation, and Delegated Diplomacy." *International Organization* 75 (4): 1018–1057. <https://doi.org/10.1017/S0020818321000102>.
- Montgomery, Jacob M., Florian M. Hollenbach, and Michael D. Ward. 2012. "Improving Predictions Using Ensemble Bayesian Model Averaging." *Political Analysis* 20 (3): 271–291. <https://doi.org/10.1093/pan/mps002>.
- Peake, Jeffrey S. 2001. "Presidential Agenda Setting in Foreign Policy." *Political Research Quarterly* 54 (1).
- Russell, Annelise, and Rebecca Eissler. 2022. "Conditional Presidential Priorities: Audience-Driven Agenda Setting." *American Politics Research* 50, no. 4 (July): 545–549. <https://doi.org/10.1177/1532673X221074359>.