

The U.S. House Committee Votes Dataset*

Josh M. Ryan[†]

Department of Political Science
Utah State University

Ryan Bell[‡]

Department of Politics
Princeton University

Abstract

The U.S. House Committee Votes Dataset is a new dataset of roll call votes taken within House standing committees. The data begin with the 104th Congress and extend through the 114th Congress, though future updates are planned. We provide a brief history of committee votes since 1970, discuss the data collection process, the structure of the data, and some limitations to the data. Using the dataset, we conduct a brief examination of majority party agenda control within committees. We find majority and minority party roll rates are similar to those observed on the floor, but with significant heterogeneity across committees and across time. Our analyses examine the effect of majority party size and preference distribution within committees on roll rates, and show that roll rates may not necessarily be due to majority party control of the committee's agenda.

*Working Paper, comments welcome. Please cite this paper when using the dataset. This project was funded by the Social Science Research Council's "Negotiating Agreement in Congress Research Grants."

[†]0725 Old Main Hill, Logan, UT 84322. josh.ryan@usu.edu.

[‡]001 Fisher Hall, Princeton, NJ 08540. rwbell@princeton.edu.

Legislative scholars have long recognized standing committees as the center of legislative action in Congress (Wilson 1885). Powerful party leadership is an important component of the modern Congress, but committees and their members still take the lead on legislative development, oversight, and processing introduced bills prior to floor consideration (Curry 2019). Despite their importance, committee activity has been less amenable to empirical analysis because, among other reasons, record-keeping has been inconsistent and less accessible to congressional scholars. As a result, a wide variety of theoretical questions about the operation and effects of congressional committees, and about how individual members behave within them, remain unanswered.

The U.S. House Committee Votes Dataset illuminates one aspect of committee activity: roll call votes taken during the markup of legislation. These data will allow researchers to test a variety of theories that heretofore have not been subject to quantitative analysis. For example, scholars may be interested in the extent to which the majority party controls the committee's agenda (Bussing and Treul Forthcoming), the ideological extremity of individual committee members, how partisan the committee itself is (Ryan Forthcoming), or whether legislators are consistent in their voting behavior from the committee to the floor (Hamm 1982; Maltzman 1995; Unekis 1978).

This research note provides a brief overview of the dataset. We provide background on recorded committee votes since 1970, explain the data collection strategy and the structure of the data. We also offer best practices for using the data, along with some cautions on the limitations of the dataset. To demonstrate the utility of the new dataset in helping answer theoretical questions, we aggregate individual member votes within committees to the vote-level and explore how successful the majority party is in controlling committee agendas. We find that the roll rate of both the majority and minority party vary across committees, though they are usually similar to those on the House floor. These results indicate majority party committee agenda control is not constant across time or committees, and that observed roll rates may be due to the number and preferences of majority and minority party committee members, rather than to a majority party cartel.

Background on Recorded Committee Votes

The Legislative Reorganization Act (LRA) of 1970 introduced a number of reforms to Congress intended to make the body more transparent. Changes made by the LRA include the introduction of electronic voting (implemented in 1973), and requirements that committee meetings be open to the

public¹, and their votes be “made available”. Specifically, the LRA requires that committee votes be recorded, open for public inspection, that information on each vote include a description of the issue, and that members’ individual roll call vote choice be listed (see Appendix A for more details). The law also requires that a vote on the motion to report the bill be printed in the committee report. Even after enactment of the LRA, House Rules only required votes to be kept in committee offices, and it remained very difficult to systematically collect committee vote data in the subsequent years. Unekis (1978) reports that he was able to access some votes from 1971 to 1974 kept in offices by staff, and that *most* committees had separated the votes from committee minutes, but he was required to transcribe—by hand—the votes for the committees he examined (see footnote 8 in Unekis 1978). Using this approach, he was able to collect data on 128 total votes across 21 committees from 1971-1974).

The new Republican majority in the 104th House drastically changed the availability of committee votes. Rules adopted at the start of the term codified reporting votes in committee reports by calling for “accountability for committee votes.” This rule was likely adopted to give the party leadership better oversight of committee activities, and increase party discipline within committees; every subsequent Congress has adopted a similar rule (House Rule XIII, Clause 3(b), see Appendix A). Thus, the data collection strategy relies on the scraping and parsing of roll call votes within committee reports. While some committees were reporting roll call votes within committee reports prior to the 104th, no House rule required it, nor did the LRA—with the exception of motions to report—and even this requirement seems to have been unenforced. As a result, the existing dataset begins with the 104th Congress because prior committee vote availability and quality were deemed to be too poor and inconsistent (in committee reports) to extend the dataset further backward in time.

In recent congresses a report typically has a table of contents listing the page on which committee votes are recorded and a separate section listing all votes taken in committee. The committee votes section in the report typically has a brief introduction, a description of each vote, and the results. Formatting varies across committees, presenting some challenges to the data collection process as described in the following section. Figure 1 shows an example of how the votes are indicated in the report’s table of contents (left panel), and then in the report language itself (right panel).

¹Unless a majority of committee members vote to close the meeting.

Figure 1: Committee Report 114-660, Accompanying H.R. 4854, “Supporting America’s Innovators Act of 2016”

<p>Mr. UPTON, from the Committee on Energy and Commerce, submitted the following</p>																							
<p style="text-align: center;">R E P O R T</p>																							
<p style="text-align: center;">together with</p>																							
<p style="text-align: center;">DISSENTING VIEWS</p>																							
<p style="text-align: center;">[To accompany H.R. 1770]</p>																							
<p style="text-align: center;">[Including cost estimate of the Congressional Budget Office]</p>																							
<p>The Committee on Energy and Commerce, to whom was referred the bill (H.R. 1770) to require certain entities who collect and maintain personal information of individuals to secure such information and to provide notice to such individuals in the case of a breach of security involving such information, and for other purposes, having considered the same, report favorably thereon with an amendment and recommend that the bill as amended do pass.</p>																							
<p style="text-align: center;">CONTENTS</p>																							
<table> <tr> <td>Purpose and Summary</td> <td>9</td> </tr> <tr> <td>Background and Need for Legislation</td> <td>9</td> </tr> <tr> <td>Hearings</td> <td>10</td> </tr> <tr> <td>Committee Consideration</td> <td>10</td> </tr> <tr> <td>Committee Votes</td> <td>10</td> </tr> <tr> <td>Committee Oversight Findings</td> <td>17</td> </tr> <tr> <td>Statement of General Performance Goals and Objectives</td> <td>17</td> </tr> <tr> <td>New Budget Authority, Entitlement Authority, and Tax Expenditures</td> <td>17</td> </tr> <tr> <td>Earmark, Limited Tax Benefits, and Limited Tariff Benefits</td> <td>17</td> </tr> <tr> <td>Committee Cost Estimate</td> <td>17</td> </tr> <tr> <td>Congressional Budget Office Estimate</td> <td>17</td> </tr> </table>	Purpose and Summary	9	Background and Need for Legislation	9	Hearings	10	Committee Consideration	10	Committee Votes	10	Committee Oversight Findings	17	Statement of General Performance Goals and Objectives	17	New Budget Authority, Entitlement Authority, and Tax Expenditures	17	Earmark, Limited Tax Benefits, and Limited Tariff Benefits	17	Committee Cost Estimate	17	Congressional Budget Office Estimate	17	
Purpose and Summary	9																						
Background and Need for Legislation	9																						
Hearings	10																						
Committee Consideration	10																						
Committee Votes	10																						
Committee Oversight Findings	17																						
Statement of General Performance Goals and Objectives	17																						
New Budget Authority, Entitlement Authority, and Tax Expenditures	17																						
Earmark, Limited Tax Benefits, and Limited Tariff Benefits	17																						
Committee Cost Estimate	17																						
Congressional Budget Office Estimate	17																						
	<p style="text-align: center;">COMMITTEE VOTES</p> <p>Clause 3(b) of rule XIII of the Rules of the House of Representatives requires the Committee to list the record votes on the motion to report legislation and amendments thereto. The sole record vote in committee was a motion by Chairman Hensarling to report the bill favorably to the House as amended. That motion was agreed to by a recorded vote of 57 yeas to 2 nays (Record vote no. FC-110), a quorum being present.</p>																						

Note: Left panel shows how committee votes are usually displayed in table of contents, right panel shows how committee votes section is usually displayed in body of report. Highlighting in left panel by authors.

Data Collection Process

Python was used to scrape committee reports from [Congress.gov](https://www.congress.gov), then parse the text from the report itself into a usable format. Within each report, the Python script searches for the section in which the votes are contained, divides that section into individual votes, then determines whether the vote was a roll call or non-roll call vote (i.e., voice vote or unanimous consent). Within each vote, the script also captures the vote topic, as described in the report, and the yeas and nays on the vote itself.

The format of the vote itself varies. Some votes are embedded in the text and the yeas and nays are given along with individual member names as text. A substantial number of votes are contained in tables. In the first case, the parsing process is straightforward. In the second case, when the vote is given as a table, the process is more complicated. In most reports, the table can be read as text using an optical character recognition (OCR) process, and the script will then save the votes of individual members. In a substantial number of cases, however, the table itself is contained within the committee report as an image document. We were not able to develop an effective way of using the OCR process to parse these images. Instead, the Python script saved the individual image files and they were coded by hand by research assistants. There are approximately 510 votes contained in image files that were

hand coded out of more than 9,800 total votes. Some other votes were hand-coded for idiosyncratic reasons including poor text quality that prevented an accurate OCR process, or a table format that was not easily parsed by the Python script. Random sampling of these hand-coded votes was performed to check for accuracy. Figure 2 shows how votes are contained within committee reports, with the left panel displaying the table style, while the right panel shows the text style. Appendix B provides additional technical details on the scraping and parsing of roll call votes from committee reports.

Figure 2: Example Committee Roll Call Vote Formats

Committee on Natural Resources

U.S. House of Representatives

114th Congress

Date: 09.08.16

Recorded Vote: #1

FC Mark Up on 4 bills: **Grijalva_037 Amendment to H.R. 3764 (Rep. Rob Bishop)**, To provide that an Indian group may receive Federal acknowledgment as an Indian tribe only by an Act of Congress, and for other purposes. "Tribal Recognition Act of 2015"

MEMBERS	Yes	No	Pres	MEMBERS	Yes	No	Pres
Mr. Bishop, UT, Chairman		X		Mr. LaMalfa, CA		X	
Mr. Grijalva, AZ, Ranking Member	X			Mrs. Dingell, MI			
Mr. Young, AK				Mr. Denham, CA		X	
Mrs. Napolitano, CA	X			Mr. Gallego, AZ	X		
Mr. Gohmert, TX		X		Mr. Cook, CA		X	
Mrs. Bordallo, Guam	X			Mrs. Capps, CA	X		
Mr. Lamborn, CO		X		Mr. Westerman, AR			
Mr. Costa, CA	X			Mr. Polis, CO	X		
Mr. Wittman, VA		X		Mr. Graves, LA		X	
Mr. Sablan, CNMI	X			Mr. Clay, MO	X		
Mr. Fleming, LA		X		Mr. Newhouse, WA		X	
Mrs. Tsongas, MA	X			Mr. Zinke, MT			
Mr. McClintock, CA		X		Mr. Hice, GA		X	
Mr. Pierluisi, Puerto Rico				Mrs. Radewagen, AS		X	
Mr. Thompson, PA				Mr. MacArthur, NJ		X	
Mr. Huffman, CA				Mr. Mooney, WV		X	
Mrs. Lummis, WY				Mr. Hardy, NV		X	
Mr. Ruiz, CA	X			Mr. LaHood, IL		X	
Mr. Benishek, MI		X					
Mr. Lowenthal, CA	X						
Mr. Duncan, SC		X					
Mr. Cartwright, PA	X						
Mr. Gosar, AZ		X					
Mr. Beyer, VA							
Mr. Labrador, ID							
Mrs. Torres, CA	X			TOTALS	14	20	

dismissing the election contest against loretta sanchez

The Committee on House Oversight, having had under consideration the resolution H. Res. 355, dismissing the election contest against Loretta Sanchez, reports the same to the House with the recommendation that the resolution be agreed to.

committee action

On February 4, 1998, by a vote of 8-1, a quorum being present, the Committee agreed to a motion to report the resolution favorably to the House. Yeas: Mr. Thomas, Mr. Ney, Mr. Ehlers, Mr. Boehner, Ms. Granger, Mr. Gejdenson, Mr. Hoyer, Ms. Kilpatrick. Nay: Mr. Mica.

Note: Left panel shows table style committee vote from Committee Report 114-847, Accompanying H.R.3764, "Tribal Recognition Act of 2015". Right panel shows text style committee vote from Committee Report 105-416, Accompanying H.Res. 355, "Dismissing the election contest against Loretta Sanchez".

Committee reports are numbered sequentially within a Congress and not all reports are relevant to the dataset. Committees issue some reports to the chamber that do not accompany a bill, including end-of-the-year summaries on their activities (commonly called committee activity reports, annual reports, summary of committee activities, etc.) These reports are scraped by the script but do not contain votes and as a result are absent from the dataset. Even for committee reports related to a bill, the vast majority do not have a recorded roll call vote and may either have no votes listed, or may only contain voice votes. In either case, these committee reports will not appear in the dataset. About 30% of all committee reports (including conference committee reports and activity reports) issued within a congressional term contain votes, with more recent congresses having a higher percentage of reports with votes.

Voice votes are sometimes identified in the text of the report, but this appears to be the case for only some committees, some of the time. Further, they were not specifically sought out by the parser so they may not appear if they are not with associated text identifying them as committee votes. Some of these votes were collected, but are excluded from the roll call data because of the inconsistency with their reporting and collection. Any other type of vote that may occur in committees (e.g., division votes) are not contained in the data.

All types of legislation with an accompanying report are captured by the parser. There are 2,638 separate pieces of legislation in the dataset across the 11 congresses in the sample. The vast majority are House bills (H.R.), with only eight Senate bills for which a committee reported the bill with roll call votes. We speculate that important or controversial Senate legislation referred to committee is instead reported through a House legislative vehicle as either a companion H.R. bill or through the inclusion of its text in another bill (Kirkland and Kroeger 2018; Wilkerson, Smith and Stramp 2015). The majority of House Resolutions (H. Res.) in the dataset are the rules for consideration attached to another piece of legislation as reported by the Rules Committee. The dataset captures the bill to which the resolution is attached.

Structure of the Dataset

The completed, parsed dataset is composed of rows of legislator-vote, nested within a vote number, committee and committee report. Because the committee reports are the means by which the committee roll call votes data were collected, all committee votes are nested within a committee report. And,

because all committee reports are associated with a bill, votes are also nested within bills, though a bill may have more than one committee report if, for example, the bill was referred to multiple committees. Each vote is uniquely identified at the committee report-level, which itself is nested within congresses, and we recommend researchers view the report number, not the bill number, as the appropriate nesting unit. Each vote within a committee report-congress is sequentially numbered, though these numbers are not meaningful other than providing a unique identification number for each vote. The vote numbers will likely not match the vote number printed in the committee report. Some committees include voice votes in the vote number, number votes within the committee rather than within the report, or include vote results from from other committees. Most vote numbers within the dataset begin with zero and are numbered sequentially.

Votes are almost always taken within the committee which issues the report. Exceptions occur when multiple committees take votes on the same bill, but only one committee report for a given bill is issued by a particular committee. Most notably, the Budget Committee, which compiles the budget resolution, sometimes reports reconciliation bills and issues a report in which votes from other committees are included. This occurs because those other committees contribute text to the reconciliation bill. For example, the Sequester Replacement Act of 2012 (H.R. 5652) was a bill, “to provide for reconciliation pursuant to section 201 of the concurrent resolution on the budget for fiscal year 2013.” That section required six committees, “to achieve specified amounts of deficit reduction from programs within their jurisdictions (House Report 112-740, 3).” These reconciliation bills are, in essence, a combination of text developed by other standing committees. In these cases, the vote is assigned to the committee which took it, not the committee which produced the report. Users of the data would not be aware that the vote was included in a committee report issued by a different committee unless they examined the reporting committee for the bill (which is not included in the data). This also sometimes occurs in rules votes embedded in Rules Committee reports, and in a few other cases.

The current dataset, for the 104th through 114th Congresses, contains more than 295,000 individual roll call vote choices across approximately 2,700 total committee reports with about 9,800 committee roll call votes, or approximately 30 individual member votes per roll call. See Appendix D for the number of votes by committee-congress. Figure 3 shows a ten row sample of the dataset to demonstrate its structure, with individual legislator-votes identified by vote number within a report.

Figure 3: Structure of the Dataset

Congress	icpsr	Last Name	Report Number	Bill Type	Bill Number	Vote Number	Vote as Scraped	Vote
114	10713	CONYERS	114-235	HR	9	1	yea	1
114	10713	CONYERS	114-235	HR	9	2	nay	0
114	10713	CONYERS	114-235	HR	9	3	yea	1
114	10713	CONYERS	114-235	HR	9	4	nay	0
114	10713	CONYERS	114-8	HR	285	1	yea	1
114	10713	CONYERS	114-228	HR	348	1	yea	1
114	10713	CONYERS	114-228	HR	348	2	yea	1
114	10713	CONYERS	114-228	HR	348	3	yea	1
114	10713	CONYERS	114-228	HR	348	4	yea	1
114	10713	CONYERS	114-228	HR	348	5	yea	1

Note: Ten observations from the dataset, sorted by legislator, showing how votes are nested within vote numbers and reports. “Vote as Scraped” shows the text in the report, while the “Vote” column indicates the vote as coded, where one equals “yea” and 0 equals “nay”. See Appendix C for more details.

Integrating the Committee Votes Data with Committee Assignments

For the data to be useful to researchers, legislators need to be identified in a way that allows their votes to be merged with existing data sources. This proved to be somewhat challenging as committee reports display the names of members in different ways, and reporting standards within committee reports have changed across time. In most cases, last names within committee-congresses are unique, and in some cases, states are listed for members. The technique used to identify legislators was to subset known committee members using Stewart’s committee data (Stewart and Woon 2016), which contains a list of legislators with their ICPSR identification codes, along with other variables, by committee-congress. The last names of members in these data were matched to those taken from each committee-bill report vote. Each legislator whose parsed last name matches the last name of a legislator on the same committee-congress in the Stewart data is then identified with their unique ICPSR number, which is taken from the NOMINATE data (Lewis et al. 2022).

This process worked well except in situations in which there are two legislators with the same last name in the same committee-congress, and either there is no state identification by legislator, or the report does have a state identification but there are two legislators with the same last name from the same state. For example, both Ann Davis and Jim Davis served on the International Relations committee in the 107th Congress. There are approximately 200 vote choices cast by members with the same last name on the same committee. Attempts were made to identify these members by state, but in many cases that information was not listed and we were not able to match these legislators with an ICPSR number from the Stewart data.

In other instances, last names were sometimes difficult to merge due to differences between the Stewart data and the committee report. This occurs because of suffixes (e.g., Sr.), because of errors in the OCR (e.g., incorrect letters, parsing issues, for example “SMITHAND”), or because the names differed across the two datasets. For example, Enid Greene Waldholtz (R-UT) is listed as “Waldholtz” in the Stewart committee data, but her votes are recorded as “Greene” in the committee reports. After the initial matching process, the non-matched names were “fuzzy” matched in R. This was very effective at capturing problematic last names and fixed the vast majority of issues. After using this technique, any remaining issues were hand-corrected, which was usually a straightforward process (e.g., FROBES became FORBES). For example, in the 104th Congress, there were 637 total legislator-votes that did not match with the Stewart committee data. 570 were matched using fuzzy match, while 66 were corrected by hand. One scraped last name could not be matched to the Stewart committee data using these techniques.

Legislator-votes which could not be matched to a corresponding legislator in the Stewart committee data were coded as having a missing ICPSR number and vote choice within the dataset. These include members whose last name did not match a committee member in the Stewart data, legislators on the same committee-congress with the same last name and no other identifying information (i.e., state), or legislators whose vote was not scraped correctly and could not be coded. That is, a missing ICPSR number indicates a legislator-vote was cast for a given committee roll call vote, but we could not determine who cast the vote, or occasionally, what the vote choice was. There are approximately 700 legislator-vote choices missing an ICPSR number and in most cases, researchers should remove these legislator-votes from their analyses. We are not aware of any component of the data collection process that would produce non-random missingness for these legislator-vote choices so their exclusion from the dataset should not bias inferences drawn from the data.

The dataset also reports the total number of yeas and nays and the total number of votes cast for a given vote. These values are given in the vote text within the committee report, but are not scraped from there and instead are calculated from within the dataset. As a result, these values may not match what is given in the committee report due to missing votes. Sometimes, the vote totals listed in the committee report text itself does not match the listed yeas and nays. Legislators who cast a vote other than yea or nay (e.g., present) or who did not vote despite serving on the committee at the time are not included in the dataset. There were approximately ten committee reports that contain votes but were not printed with the report that are missing from the data.

Variables Included in the Dataset

The resulting dataset contains a legislator's last name, their ICPSR number, and the committee identification number (as coded within the Stewart data) in which the vote was taken. It also includes the committee's name, though as these change across time, committee names as of the 114th Congress are used. As noted, votes are nested within vote numbers, committee reports, bills, and congresses. The dataset includes the text of the vote as scraped from the committee report, which varies across committees and across congresses. Some reports use yeas and nays, while others report yes and no, aye and nay, Y and N, etc. A separate vote column classifies votes cast as a one for an affirmative vote (yea/aye/yes/y), and a zero for a negative vote (nay/no/n).

The dataset includes aggregated columns for total number of votes cast for a given vote, total number of affirmatives and negatives, and the percentage of the committee voting for and against on the vote. There are separate columns for total number of affirmative and negative votes for each party as well, along with the percentages of each party voting in favor or against. These values do not vary within votes and do not include votes with missing ICPSR numbers (and missing votes). Other variables included are taken directly from the Stewart committee data and include party, majority or minority party status, and the legislator's state, if it is available from the Stewart data. See Appendix C for a detailed description of all variables included in the dataset.

Application: Agenda Control within Committees

The ability of the majority party to prevent internally divisive proposals from reaching the floor and damaging the majority party's brand is the cornerstone of party cartel theory (Cox and McCubbins 2002, 1993, 2005). A substantial literature uses majority party "rolls", in a variety of settings, to determine the extent to which variation in institutional rules empowers the majority party to control the agenda. Majority roll rates on the House floor in a given term are less than 5% (and often less than 1%). Similar results have been shown in other institutional settings, and vary based on the extent to which institutional rules empower the majority party is able to control the agenda (Anzia and Jackman 2012; Cox, Kousser and McCubbins 2010). The low majority party roll rate, in which a majority of the majority party votes on the losing side of the issue, is taken as evidence that the majority party successfully prevents divisive issues from reaching the floor, and more broadly, solves the collective action problems associated with heterogeneous preferences within the party.

Standing committees pose an interesting test of cartel theory as votes that occur within committees directly affect the content of legislation which will be addressed on the floor. Further, while committees have norms about majority party control of the agenda (e.g., the chairman's mark, Curry 2015), there is ambiguity in the formal rules governing which legislative items are considered. For example, in the 117th Congress, the agenda rules for Energy and Commerce require only that, "items of business...shall be provided to each member of the Committee at least 36 hours in advance of such meeting" (Rule 2(d), see Appendix D.) The lack of formal rules may empower the majority party, through the chairperson or median committee member, to cartelize the agenda. Alternatively, the committee agenda may be more open to the minority than the floor, making the markup process more bipartisan than is commonly assumed. And, if this is the case, the minority's strategy may be to moderate policy by proposing items which roll the majority party more frequently than occurs on the floor.²

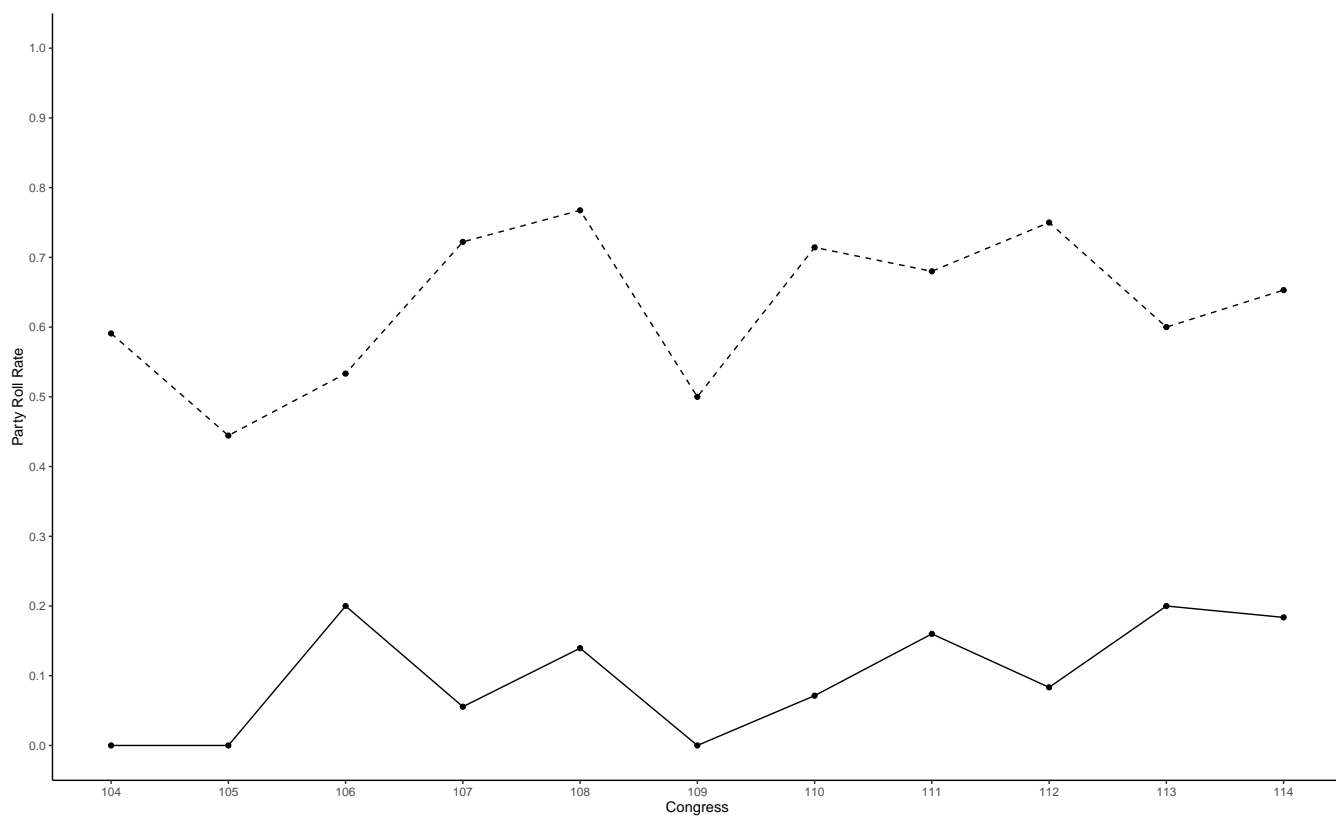
As a first, descriptive analysis, we examine majority and minority party roll rates within committees by aggregating individual committee votes and generating roll rates for both parties in each committee-congress. In some committees very few votes are taken, so we exclude committee-congresses in which the committee took fewer than 10 votes. This removes 54 committee-congress observations for a total of 146 from the 104th through 114th Congresses (see Appendix D).

The results seemingly support the claim that the majority party controls committee agendas. Across all committees and congresses, the majority roll rate is about 6% and the minority roll rate is about 76% for a ratio of nearly 13 to 1. There is significant variation across committees, and within committees across time, however. For example, the majority party roll rate in the Agriculture committee ranges from zero in a number of congresses to 38% in the 112th Congress. The majority party roll rate in Appropriations is about 8%, but in the 107th and 114th Congresses the rates were 13% and 14%, respectively. The data also show that the majority party in the Rules committee was *never* rolled, while the minority party roll rate was 97%, consistent with expectations about the role of the Rules Committee in the modern House (Finocchiaro and Rohde 2008). Figure 4 shows roll rates for the majority and minority parties in a sample committee, Armed Services. Appendix Figure F1 roll rates for all committees within the sample.

Krehbiel (2007) suggests that variation in majority and minority party rolls is expected in a legislature due exclusively to policy predispositions and variation in party sizes. That is, roll rates themselves are not dispositive evidence for the majority party exercising negative agenda control because there

²It is also important to note that committee rules allow for a recorded roll call vote on the request of 1/5 of members.

Figure 4: Majority and Minority Party Roll Rates by Congress in the Armed Services Committee



Note: Dashed line shows roll rate for minority party, solid line shows roll rate for majority party.

exists a baseline roll rate within a legislature based solely on majority party size and the distribution of preferences within and across the parties, irrespective of the strength of agenda control. Similarly Patterson Jr. and Schwartz (2020) show that low roll rates similar to those observed in the House could occur by chance alone, and that sometimes the majority party has an incentive to allow bills which roll the party.³

To take these criticisms seriously, and to investigate agenda power at the committee stage, we model roll rates based on preferences and majority party size. Previous work on roll rates uses the legislature itself (i.e., the House floor) as the unit of interest, so here we use committees, which we see as semi-autonomous “mini” legislatures in which policy is crafted and in which the majority party’s control over the agenda is ambiguous. The use of the committee votes data allows for an investigation of variation in roll rates across similar institutional settings, holding the policy agenda and other committee-level factors constant through the inclusion of committee fixed effects. We can also control for confounders that change across time, such as the party leadership or polarization, through the use of congress fixed effects.

Party theories suggests that as party heterogeneity increases, the number of party rolls should increase (for both the majority and minority). Conversely, in a non-partisan legislature, Krehbiel expects no relationship between intraparty heterogeneity and roll rate. Second, as the distance between the two parties increases, the number of minority party rolls should increase, while the number of majority party rolls should decrease “marginally.”⁴ Finally, a larger majority party will increase the roll rate for the minority and decrease the rate for the majority (also see Jackman 2013 for evidence of this claim).

Three independent variables are created from the committee votes data and follow the measures used by Krehbiel (2007). Two of the measures are an aggregation of individual legislator ideal points within a committee-congress. To find these ideal points, we use the “oc” package in R to create an optimal classification score for each legislator, which are similar to W-NOMINATE scores and more robust to small sample sizes (see Ryan Forthcoming for more details on creating OC scores from committee votes).

Intraparty heterogeneity is measured within a committee-congress through the sum of the proportion Republican committee members more liberal than the most conservative Democrat and the pro-

³These situations may occur as part of voting on an overall set of bills in which the majority benefits from allowing legislators “free” votes.

⁴According to Krehbiel, the expectation for the majority is marginal because, “the rate is bounded below at zero and, under most parameter settings, low to begin with [19].”

portion of Democratic committee members more conservative than the most liberal Republican. The variable ranges from zero to 1.9 with a mean of .46 and a median of zero. Interparty heterogeneity is the absolute ideal point distance of the committee's Republican and Democratic medians. The variable ranges from zero to 1.52 with a mean of .73 and a median of .78. Finally, the proportion of committee seats controlled by the majority party is taken from the Stewart committee data and is the total number of majority party members at the start of the term divided by the total number of majority and minority party members on the committee.⁵ The value ranges from .52 to .67, with a mean of .56 and a median of .56.

Table 1: Predicting Minority Party Roll Rates in Committees, 104th-114th Congresses

	Minority Party Roll Rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Intraparty Heterogeneity	−0.02 (0.02)		0.01 (0.02)	−0.01 (0.02)		0.02 (0.02)
Proportion Overlap		0.06 [#] (0.03)	0.06 [#] (0.04)		0.07* (0.03)	0.09* (0.04)
Proportion of Majority Party		1.26* (0.59)	1.28* (0.60)		0.47 (0.68)	0.48 (0.68)
Constant	0.71* (0.07)	−0.06 (0.34)	−0.07 (0.35)	0.62* (0.07)	0.30 (0.38)	0.26 (0.38)
Committee Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Congress Fixed Effects				Yes	Yes	Yes
N	135	135	135	135	135	135
R ²	0.52	0.55	0.55	0.64	0.66	0.66

Note: [#]p<0.1; **p<0.05. Dependent variable is minority party roll rate, only committees with more than ten votes in a congressional term are included. Members' OC votes used to generate intraparty heterogeneity and proportion overlap. A member will not have an OC score generated if they cast fewer than ten votes within a committee-congress.

The results for minority party roll rate are shown in Table 1. The first two columns include each of the preference measurements separately with committee fixed effects, while column three includes both preference measurements and the party size measure. Consistent with baseline expectations for a non-partisan legislature, the intraparty heterogeneity variable is not a statistically significant predictor of minority party rolls. In both models two and three, the proportion overlap between the two parties is positive and statistically significant at the .1 level, also consistent with the non-partisan legislature predictions. Finally, an increase in the size of the majority party also increases the number of minority party rolls.

⁵Bernie Sanders is coded as a Democrat

In models 4-6, fixed effects are included for congressional term. The results are substantively altered by controlling for confounding factors which are constant between committees but vary across time. The proportion overlap becomes statistically significant at the .05 level, while the proportion of the majority party on the committee becomes insignificant, though remains positive. The proportion of the majority party varies across committees but depends largely on the overall size of the majority within a congress. Thus, including fixed effects for congress controls for majority size within the chamber, and may mute the effects of the committee-level variable.

Table 2 shows the same models, but predicts majority party roll rates within committee-congresses. As with minority party roll rates, intraparty heterogeneity has no effect in any of the models on majority party roll rates, inconsistent with predictions from partisan theories. The proportion overlap, which is predicted to have a “marginal” negative effect is positive but no statistically significant. Finally, the proportion of the majority party on the committee is negative and statistically significant (at the .1 level in model 2 and at the .05 level in model 3), the opposite sign from the coefficient predicting minority party roll rates. This is evidence in support of the claim that the observed roll rates are not the result of partisan agenda control. However, as with the previous results, including congress fixed effects makes these results insignificant.

To summarize, the null hypothesis in a non-partisan legislature that intraparty heterogeneity should affect roll rates is borne out by the data. Hypotheses about the proportion of party overlap are also largely supported; with more overlap comes a greater minority party roll rate and there is no effect on the majority party roll rate. The hypotheses about the effect of the proportion of the majority party within the committee are weakly supported. The effect is positive and significant for minority party roll rates, and negative and significant for majority party roll rates if congress fixed effects are not included.

Evaluating the overall evidence suggests that roll rates in House standing committees do not provide strong evidence of majority party control of the agenda. Of course, the distribution of preferences and the size of the majority party on committees are themselves functions of majority party power. It also seems likely that the majority exercises more control over the agenda in some committees than others. The party leadership stacks certain committees with certain types of members to produce more favorable policy outcomes, and perhaps that is where party power manifests, rather than through overt control of the committee’s agenda. And, perhaps some committees are more important to the majority party’s brand, or parties seek to balance the inherent trade-off between strong control over some com-

mittees and weaker control over others. Indeed, the significant variation in majority and minority party roll rates across committees and congresses, despite the same institutional rules, suggests that majority party agenda control is conditional on other factors. Future research should investigate this variation in order to further inform theories of legislative organization.

Table 2: Predicting Majority Party Roll Rates in Committees, 104th-114th Congresses

	Majority Party Roll Rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Intraparty Heterogeneity	−0.01 (0.01)		−0.01 (0.02)	−0.02 (0.01)		−0.01 (0.02)
Proportion Overlap		0.02 (0.02)	0.01 (0.02)		0.02 (0.02)	0.01 (0.02)
Proportion of Majority Party		−0.75 [#] (0.38)	−0.78* (0.39)		−0.27 (0.46)	−0.28 (0.46)
Constant	0.18* (0.05)	0.59* (0.22)	0.62* (0.22)	0.20* (0.05)	0.32 (0.26)	0.34 (0.26)
Committee Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Congress Fixed Effects				Yes	Yes	Yes
N	135	135	135	135	135	135
R ²	0.52	0.55	0.55	0.64	0.66	0.66

Note: [#]p<0.1; **p<0.05. Dependent variable is majority party roll rate, only committees with more than ten votes in a congressional term are included. Members' OC votes used to generate intraparty heterogeneity and proportion overlap. A member will not have an OC score generated if they cast fewer than ten votes within a committee-congress.

Discussion

The House Committee Votes dataset offers new opportunities to better understand legislative action inside committees. That data will allow researchers to examine committee agendas, voting behavior, representation, and the effect of exogenous influence such as interest groups on committee action. The data currently extend from the 104th Congress to the 114th, but we hope to maintain and update the dataset, adding additional committee votes at the end of each congressional term.

The House recently created a Modernization Committee to study ways of improving the operation of Congress. One of its key suggestions was to make committee votes more easily available to the public, likely through posting vote results online, similar to how floor votes are currently posted. This would allow researchers to access voting data in future congresses, and would connect with this dataset to allow for a long time-series of committee votes. Unfortunately, the House has taken few steps to

implement the suggestions made by the Modernization Committees.

As an example of how useful these data will be, we offered an initial examination of agenda control within standing committees by aggregating the individual voting observations in the dataset. Our results suggest that roll rates in committees are similar to those on the floor, but there is significant heterogeneity across committees and across time. We also applied a simple test of roll rates to determine whether they are different than would be predicted in a non-partisan legislature by majority party size and the preference distributions of legislators within committees. On balance, the evidence suggests they are not, though the Committee Votes dataset will allow researchers to undertake more complex and detailed analyses of this question.

References

- Anzia, Sarah F. and Molly C. Jackman. 2012. "Legislative Organization and the Second Face of Power: Evidence from U.S. State Legislatures." *The Journal of Politics* 75(1):210–224.
- Bussing, Austin and Sarah A. Treul. Forthcoming. "Majority Party Agenda Setting: Picking Fights or Avoiding Them?" *Legislative Studies Quarterly*.
- Cox, Gary W. and Mathew D. McCubbins. 2002. Agenda Power in the U.S. House Of Representatives, 1877-1986. In *Party, Process, and Political Change in Congress*. Stanford, CA: Stanford University Press.
- Cox, Gary W. and Matthew D. McCubbins. 1993. *Legislative Leviathan: Party Government in the House*. Berkeley: University of California Press.
- Cox, Gary W. and Matthew D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the US House of Representatives*. Cambridge, UK: University of Cambridge Press.
- Cox, Gary W., Thad Kousser and Mathew D. McCubbins. 2010. "Party Power or Preferences? Quasi-Experimental Evidence from American State Legislatures." *Journal of Politics* 72(3):799–811.
- Curry, James M. 2015. *Legislating the Dark: Information and Power in the House of Representatives*. Chicago, IL: University of Chicago Press.
- Curry, James M. 2019. "Knowledge, Expertise, and Committee Power in the Contemporary Congress." *Legislative Studies Quarterly* 44(2):203–237.
- Finocchiaro, Charles J. and David W. Rohde. 2008. "War for the Floor: Partisan Theory and Agenda Control in the U.S. House of Representatives." *Legislative Studies Quarterly* 33(1):35–61.
- Hamm, Keith E. 1982. "Consistency Between Committee and Floor Voting in U.S. State Legislatures." *Legislative Studies Quarterly* 7(4):473–490.
- Jackman, Molly C. 2013. "Parties, Median Legislators, and Agenda Setting: How Legislative Institutions Matter." *The Journal of Politics* 76(1):259–272.
- Kirkland, Justin H. and Mary A. Kroeger. 2018. "Companion Bills and Cross-Chamber Collaboration in the U.S. Congress." *American Politics Research* 46(4):629–670.

- Krehbiel, Keith. 2007. "Partisan roll rates in a nonpartisan legislature." *The Journal of Law, Economics, & Organization* 23(1):1–23.
- Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin and Luke Sonnet. 2022. "Voteview: Congressional Roll-Call Votes Database." Voteview.com.
- Maltzman, Forrest. 1995. "Meeting Competing Demands: Committee Performance in the Postreform House." *American Journal of Political Science* 39(3):653–682.
- Patterson Jr., Shawn and Thomas Schwartz. 2020. "Parties, Agendas, and Roll Rates." *Journal of Theoretical Politics* 32(2):348–359.
- Ryan, Josh M. Forthcoming. "The Partisanship of House Committees and Member Self-Selection." *Legislative Studies Quarterly* . <https://doi.org/10.1111/lsq.12313>.
- Stewart, III, Charles and Jonathan Woon. 2016. "Congressional Committee Assignments, 103rd to 114th Congresses, 1993–2017: House of Representatives." http://web.mit.edu/17.251/www/data_page.html.
- Unekis, Joseph K. 1978. "From Committee to the Floor: Consistency in Congressional Voting." *The Journal of Politics* 40(3):761–769.
- Wilkerson, John, David Smith and Nicholas Stramp. 2015. "Tracing the flow of policy ideas in legislatures: A text reuse approach." *American Journal of Political Science* 59(4):943–956.
- Wilson, Woodrow. 1885. *Congressional Government*. Cleveland, OH: Meridian.

Appendix

Appendix A: Committee Vote Requirements in Law and House Rules

Legislative Reorganization Act of 1970 P.L. 91-510, Section 104(b), Requiring Committee Votes to Be Recorded and Made Available

(b) Clause 27(b) of Rule XI of the Rules of the House of Representatives is amended by adding at the end thereof the following: “The result of each rollcall vote in any meeting of any committee shall be made available by that committee for inspection by the public at reasonable times in the offices of that committee. Information so available for public inspection shall include a description of the amendment, motion, order, or other proposition and the name of each Member voting for and each Member voting against such amendment, motion, order, or proposition, and whether by proxy or in person, and the names of those Members present but not voting. With respect to each record vote by any committee on each motion to report any bill or resolution of a public character, the total number of votes cast for, and the total number of votes cast against, the reporting of such bill or resolution shall be included in the committee report.”

H.Res.6, 104th Congress, Requiring Committee Votes to be Reported in Committee Reports

Accountability for Committee Votes

Sec. 209. In clause 2(1)(2) of rule XI amend subdivision (B) to read as follows: “(B) With respect to each rollcall vote on a motion to report any measure or matter of a public character, and on any amendment offered to the measure or matter, the total number of votes cast for and against, and the names of those members voting for and against, shall be included in the committee report on the measure or matter.”.

House Rule XIII, Clause 3(b), Committee Vote Reporting Requirement as of the 117th Congress (2021-2023)

(b) With respect to each record vote on a motion to report a measure or matter of a public nature, and on any amendment offered to the measure or matter, the total number of votes cast for and against, and the names of members voting for and against, shall be included in the committee report. The preceding sentence does not apply to votes taken in executive session by the Committee on Ethics, and applies only to the maximum extent practicable to a report by the Committee on Rules on a rule, joint rule, or

the order of business.

(e)(1) Each committee shall keep a complete record of all committee action which shall include a record of the votes on any question on which a roll call vote is demanded. The result of each such roll call vote shall be made available by the committee for inspection by the public at reasonable times in the offices of the committee. Information so available for public inspection shall include a description of the amendment, motion, order, or other proposition and the name of each Member voting for and each Member voting against such amend. meant, motion, order, or proposition, and whether by proxy or in person, and the names of those Members present but not voting.

Appendix B: Scraping and Parsing Details

Scraping

Reports from the 104th through the 114th Congress are scraped from Congress.gov. If scraping by Congress, the url is:

```
'https://www.congress.gov/search?q="report-chamber":"House+Report","congress": "%s"'
```

where %s is replaced with the specified Congress. If scraping all congresses at once, the url is:

```
'https://www.congress.gov/search?q="report-chamber":"House+Report"'
```

The text from each report, along with other information gathered from the report's webpage, is saved and stored in JSON Lines format.

Parsing Scraped Reports

The *ReportData* class serves as a vehicle to store the data from each report. Using load, each entry in the JSON Lines files—one line per report for each Congress—is loaded into Python.

Each report loaded is then passed to the *Report* class to be parsed. The *Report* class serves as a vehicle to parse the report's information and locates the roll call votes within the report. The information from each JSON Lines entry are added as attributes to the *Report* class instance, and then additional information not scraped is parsed and added to the report. The text of the report is divide into its sections and the section containing the votes is identified.

The section containing the vote information is divided into individual votes, and each vote is passed to *Vote*, with each *Vote* instance stored post-parsing in a list attribute for the *Report* instance. The text of the vote is parsed to identify whether the vote is a roll call or non-roll call vote.

Roll Call Votes

Appended votes—votes which are not contained in the text of the report, but as a scanned image in the pdf—are identified for later processing. For in-text votes, information is collected from the text and the members' votes collected.

The pdfs for reports that contain appended vote images are downloaded using 'download_report_url'.

Scanned images from the reports are extracted using 'extract_images'. All images from the pdf are extracted, so images that do not contain a vote are removed. Using Adobe, OCR is then performed on the appended vote images. The text from the newly-created appended image pdfs is extracted using 'extract_image_text'. The text files are cleaned as necessary.

The appended votes are parsed similar to non-appended or in-text-votes, except that the text of the vote is loaded from the vote's stored .txt. file. After completing this process, the next time the data is loaded, the appended votes will be parsed as well.

Appendix C: Codebook

Variable Names and Descriptions

Column Name	Description
cong	Congress of vote.
billtype	one of: House Concurrent Resolution (HCONRES), House Joint Resolution (HJRES), House of Representatives Bill (HR), House Resolution (HRES), Senate Bill (S).
billnumber	Bill Number of report and vote.
Bill.Name	Bill title taken from committee report.
Report.Number	Unique within bills; report numbers with a number after a second hyphen indicates multiple parts, treated separately (e.g., 104-280-1 indicates part 1 of report 104-280).
Vote.Number	Unique vote identifier within a report; not necessarily in same order or match roll call vote number reported in report.
stewartcommid	Stewart committee data identification number of committee which took the vote; not necessarily the same committee which wrote the report or reported the bill.
icpsr	ICPSR number of voting member; merged in from Stewart committee assignment data.
lastname	formatted last name of voting member; used to merge with Stewart committee assignment data.
Member	In most cases, name as scraped from report; in some cases (hand-coded image file) taken directly from Stewart committee assignment data.
Vote	Member's vote as coded, 1=yes, 0=no. Other votes (e.g., present) not coded.
Vote.Scrape	Text of member's vote as scraped from committee report.
Party	Member's party, 100=Democrat, 200=Republican, other parties included; taken from Stewart committee assignment data.
Maj.Min	Party status code, collapsed (i.e., majority=1 or minority=2 party); taken from Stewart committee assignment data, see Stewart codebook for more details.
State.Name	State abbreviation for member; taken from Stewart committee assignment data; not available for all members.
Total.Yea	Total voting in favor on vote; vote-level variable.
Total.Nay	Total voting against on vote; vote-level variable.

Table 3

Column Name	Description
percentyea	percentage voting in favor on vote; vote-level variable.
percentnay	percentage voting against on vote; vote-level variable.
totalvotes	Sum of Total.Yea and Total.Nay.
demyesvotes	Number of Democrats voting in favor on vote; vote-level variable.
demnovotes	Number of Democrats voting against on vote; vote-level variable.
demttotalvotes	Sum of demyesvotes and demnovotes; vote-level variable.
dempereya	demyesvotes divided by demttotalvotes; vote-level variable.
gopyesvotes	Number of Republicans voting in favor on vote; vote-level variable.
gopnovotes	Number of Republicans voting against on vote; vote-level variable.
gopttotalvotes	Sum of gopyesvotes and gopnovotes; vote-level variable.
goppereya	gopyesvotes divided by gopttotalvotes; vote-level variable.
Consideration.of	Rules Committee vote on resolution considering bill; not all consideration votes coded as such.

Appendix D: Frequency Table of Votes Per Committee-Congress

Number of Committee Votes by Committee-Congress

Committee	Congress											Total
	104	105	106	107	108	109	110	111	112	113	114	
Agriculture	10	2	3	2	0	1	11	7	20	12	1	69
Appropriations	21	8	13	10	6	16	2	6	17	7	11	117
Armed Services	10	8	9	9	28	8	11	14	12	18	28	155
Budget	39	20	22	19	26	8	0	0	8	4	3	149
Education and the Workforce	23	8	43	12	12	18	14	21	16	5	20	192
Energy and Commerce	41	5	11	17	31	65	7	38	21	6	17	259
Financial Services	18	17	15	13	4	5	20	12	13	5	3	125
Homeland Security	NA	NA	NA	19	6	25	10	11	25	5	7	108
House Administration	6	8	4	5	2	5	12	7	2	0	0	51
Intelligence (Select)	1	1	1	0	11	10	13	16	4	6	1	64
International Relations	21	12	4	7	7	8	2	1	32	0	2	96
Judiciary	41	19	29	18	14	32	17	17	21	12	9	229
Natural Resources	20	7	10	19	3	19	23	12	11	6	7	137
Rules	30	9	17	26	22	16	16	65	87	12	7	307
Oversight and Government Reform	21	3	2	0	7	2	3	5	12	7	3	65
Science, Space, And Technology	24	2	5	1	11	2	4	4	5	4	32	94
Small Business	0	0	1	0	0	0	4	2	10	7	4	28
Transportation and Infrastructure	21	4	3	4	0	1	3	0	17	5	2	60
Veterans Affairs	2	0	1	1	1	0	0	2	1	0	4	12
Ways and Means	35	18	7	9	10	15	21	26	9	5	17	172

Note: Homeland Security was a select committee in the 107th and 108th Congresses, became a standing committee in the 109th Congress.

Appendix E: Rules of the House Energy and Commerce Committee—Agenda

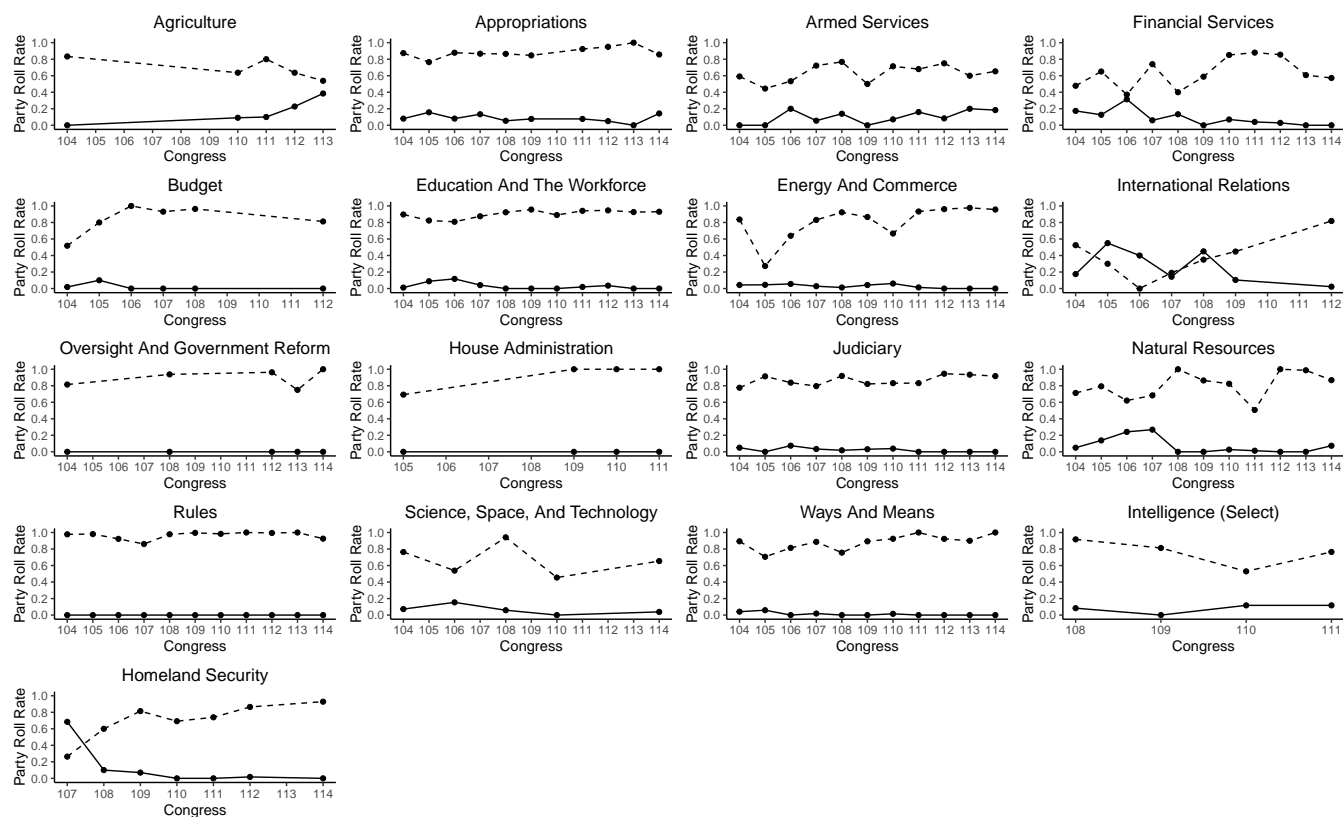
RULE 2. BUSINESS MEETINGS/MARKUPS

(d) Agenda. The agenda for each Committee meeting, setting out all items of business to be considered, shall be provided to each member of the Committee at least 36 hours in advance of such meeting.

Accessed at: <https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/117th%20Congress%20Committee%20on%20Energy%20and%20Commerce%20Resolution%20Rules%20Package.pdf>

Appendix F: Party Roll Rates Within Committees Across Congresses

Figure F1: Majority and Minority Party Roll Rates by Committee-Congress



Note: Dashed line shows roll rate for minority party, solid line shows roll rate for majority party. Committee-congresses in which there are fewer than ten committee votes are dropped from the sample. In the graph, those congresses without a scatter point indicate the observation is missing from the sample.