Legislative Hitchhikers: Re-envisioning Legislative Process and Effectiveness*

Andreu Casas † Matthew J. Denny‡ John Wilkerson§

Abstract

For more than half a century, scholars have been studying legislative effectiveness using a single metric - whether the bills a member sponsors progress through the legislative process. We investigate a less orthodox form of effectiveness - bill proposals that become law as provisions of other bills. Counting these "hitchhiker" bills as additional cases of bill sponsorship success reveals a more productive, less hierarchical and less partisan lawmaking process. We argue that agenda and procedural constraints are central to understanding why lawmakers pursue hitchhiker strategies. We also investigate the legislative vehicles that attract hitchhikers and find, among other things, that more Senate bills are enacted as hitchhikers on House laws than become law on their own.

Word Count: 7,004
(8,754 with Supporting Information)

*First version: March 20, 2017. This version: December 29, 2017. This research is based upon work supported by the National Science Foundation under Grant No. 1224173 and IGERT Grant DGE-1144860. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank Lexi Greenberg for her research assistance, Barry Pump for his procedural insights, and Amber Boydstun for suggesting the "hitchhiker" term.

†acasas2@uw.edu; Box 353530, University of Washington, Seattle, WA 98195

‡mdenny@psu.edu; 203 Pond Lab, Pennsylvania State University , University Park, PA 16802

§jwilker@uw.edu; Box 353530, University of Washington, Seattle, WA 98195

1 Introduction

In 2014, a Washington Post article described the legislative record of retiring Representative Robert Andrews (D-NJ) as the worst in Congress: "Andrews proposed 646 bills, passed 0: worst record of past 20 years." [1] In response, Andrews objected that journalists were using the wrong metric: " *I'm just a bill* is not the way it works."

Legislative scholars have also challenged this orthodox view of lawmaking: "The Schoolhouse Rock! cartoon version of the conventional legislative process is dead, if it was ever an accurate description in the first place" (Gluck et al., 2015). Increasingly, a process of considering bills on an individual basis has been replaced by a leader-centered process of constructing larger omnibus bills that combine multiple policy proposals into one (Sinclair, 2016; Krutz, 2005; Curry and Lee, 2016).

Andrews' advice to journalists was that they should also count policy proposals that "germinate in a larger bill." We develop an approach to identifying bills that are enacted into law as provisions of other bills. We then show that accounting for these hidden "hitchhiker" successes leads to different conclusions about legislative effectiveness. The next section of this paper reviews the longstanding legislative effectiveness literature and its limitations. We then propose and implement a new text-based methodology for accurately identifying hitchhiker bills. Examining two decades of lawmaking (1993-2014), we find that as many bills become law as hitchhikers as become law on their own.

We argue that agenda and procedural constraints are central to understanding why lawmakers pursue hitchhiker strategies. The most important differences between successful law and hitchhiker bill sponsors are that law sponsors are more likely to hold agenda setting positions that provide opportunities to sponsor the bills that advance for reasons that have little to do with who sponsors them. When these effects are discounted, the predictors of law

[1] Farenthold, David A. (2014, February 4), see new story in this following **link.**

success an hitchhiker bill success are very similar. We also find that the types of bills that are more likely to become law are also more likely to serve as vehicles for hitchhikers. Finally, we find that the procedural constraints lead the Senate to employ hitchhiker strategies more frequently than the House. As a result, more Senate bills become law as hitchhikers than become law on their own.

2 Effectiveness Research and its Limits

Studies of legislative effectiveness fit into a broader literature examining legislative influence (see, for example: Meyer, 1980; Hall, 1992; Thomas and Grofman, 1992; Kessler and Krehbiel, 1996; Arnold et al., 2000; Crisp et al., 2004; Fowler, 2006; Miquel and Snyder, 2006; Kirkland, 2011; Sulkin, 2011; Desmarais et al., 2015). They include some of the earliest quantitative analyses of legislative behavior. In *US Senators and their World*, Donald Matthews observed: "To the extent that the concept as used on Capitol hill has any distinct meaning, *effectiveness* seems to mean the ability to get one's bills passed" (Matthews, 1960). Matthews found bills sponsored by senators who adhered to chamber "folkways," such as specializing and spending less time giving floor speeches were more likely to pass the chamber. A decade later, Olson and Nonidez (1972) asked whether Representatives who adhered to similar norms were also more legislatively successful (they weren't). Subsequent research has continued to investigate what bill success reveals about norms and coalition building (see, for example: Matthews, 1960; Olson and Nonidez, 1972; Krutz, 2005; Baughman, 2006; Koger and Fowler, 2007; Hasecke and Mycoff, 2007). An equally important body of research seeks to discover (in the words of Anderson et al. (2003)) the "remarkable skills" of more effective lawmakers (Frantzich, 1979; Bratton and Haynie, 1999; Jeydel and Taylor, 2003; Anderson et al., 2003; Cox and Terry, 2008; Volden and Wiseman, 2009, 2014).

The methods of these effectiveness studies have become considerably more sophisticated

over time, but the measure has changed very little. Effectiveness continues to be defined in terms of whether a sponsor's bill progresses through the legislative process although different studies define progress differently. Some define it by whether a bill receives any committee consideration (Krutz, 2005) whereas others define it by whether a bill passes the chamber. Some focus on "hit rates" —the percentage of a legislator's bills that succeed (Anderson et al., 2003)— whereas others focus on whether individual bills progress. The most recent research also offers the most thoughtful measure. Volden and Wiseman (2014) compute summary "legislative effectiveness scores" by counting the number of bills introduced by members, and weighting them for how far they progress.

Bill success has also recently attracted interest from scholars in other fields and from entrepreneurs. Instead of trying to understand why some sponsors are more successful than others, the goal is to predict bill success as one might predict the winners of sporting events or elections (Yano et al., 2012; Nay, 2017). Several commercial ventures are or soon will be offering bill success prediction services [2]

An important limitation of these efforts is that bills are vehicles. The progress of a bill and the progress of a policy are sometimes one and the same. But there are systematic reasons to believe that this implicit assumption of the effectiveness literature is often violated. The bill that becomes law can also be very different from the one the sponsor originally proposed. The 900 page Affordable Care Act (HR 3590) was originally a 7 page bill providing a first time home buyer credit for service personnel (https://www.congress.gov/bill/111th-congress/house-bill/3590/text/ih?format=txt). Current effectiveness approaches give the original bill's sponsor (Rep. Charles Rangel (D-NY) full credit even though the substance of the ACA was added later by the Senate.

The lawmaking process has also fundamentally changed since Matthews first equated bill passage and effectiveness. A process that used to be driven by largely autonomous

[2]https://www.skoposlabs.com/, https://www.govtrack.us/, https://statehill.com

committees recommending bills on an individual basis has been replaced, to an increasing extent, by leadership-driven negotiations. These negotiations often produce large "omnibus" bills that combine policy proposals that often address diverse subjects (Krutz, 2001; Sinclair, 2016; Curry and Lee, 2016). Other research emphasizes the compulsory quality of lawmaking. Walker (1977, p. 424-6) argues that most of the items on the Senate's agenda are "recurring" issues that the chamber is "virtually forced" to take up, while Adler and Wilkerson (2012)) find that expiring laws and programs are a very important predictor of the congressional issue agenda.

We propose an approach to studying effectiveness that gets closer to what scholars (and citizens) ultimately care about - legislators' ability to get their *policy proposals* enacted into law. One implication of unorthodox approaches to lawmaking is that the legislative opportunity structure increasingly favors "hitchhiker" strategies. Our approach counts not only bills that become law on their own, but also the bills that are enacted law as provisions of other laws. This approach reveals, for example, that the Affordable Care Act includes almost 50 "hitchhikers."

Another implication of unorthodox lawmaking is focusing on bill progress probably exaggerates the perceived effectiveness of the small number of legislators (mostly committee leaders) who typically sponsor the vehicles frequently used in collective lawmaking efforts.

Accounting for hitchhikers represents an improvement rather than a complete solution to these limitations. We do not identify cases where only part of a bill is successfully enacted as a hitchhiker, or cases where a legislator successfully proposes language (e.g. an amendment) that does not originate with a bill. We also continue to give sponsors credit for bills that become law even when the bill progresses for reasons that have little to do with their efforts (as with the ACA). Nevertheless, we believe that hitchhikers provide new opportunities to explore how laws are made, and to better understand legislative effectiveness.

3 Why hitchhikers?

Why would a sponsor advance a bill as a hitchhiker when authoring a stand-alone law would seem to offer more visible credit claiming opportunities? The main reason is that opportunities to advance stand-alone bills are limited. In addition, hitchhiker strategies can be procedural efficient and, in some cases, procedurally necessary. In this section, we present several novel hypotheses about the bills that become law and bills enacted as hitchhikers that are tested later in the paper.

However, before proceeding, it is worth noting that legislators do take credit for their hitchhiker successes. For example, Rep. Carolyn Maloney's (D-NY) official website includes a "Laws Enacted" page.[3] The majority of the enactments listed (40 out of 74) are either bills that were "included" in other laws, or laws that were "versions" of her bills. Maloney also highlights hitchhikers in her direct communications with constituents. Her "Report to Manhattan" newsletter (Spring 2010) specifically points to provisions of the recently passed Affordable Care Act that are "based on" bills she sponsored.

We expect that many of the covariates that predict bill sponsorhip success in prior effectiveness studies will also predict hitchhiker bill successes. However, we propose several additional hypotheses that are related to several understudied aspects of lawmaking: bicameralism, agenda control, and procedural constraints. Although these subjects receive a lot of theoretical attention, very little of this research examines their implications for the progress of bills.

3.1 Bicameral support

Our first hypothesis proposes an additional explanation for why some bills are more likely to progress that reflects the bicameral nature of lawmaking. Members of the House and

[3] https://maloney.house.gov/my-work-in-congress/accomplishments/laws-enacted

Senate often work together to advance proposals. One of the indicators of such an effort is when a House and Senate sponsor introduce identical or nearly identical bills (Oleszek, 2017; Kirkland and Kroeger, 2017).

Hypothesis 1 – Companion legislation: *A bill is more likely to become law if there is a companion bill in the other chamber. We define a companion bill by whether the text of an introduced bill in the other chamber is at least 95% similar to the bill in question (after preprocessing).*

3.2 Agenda-setting

Congressional agenda space is a scarce commodity. It has always been the case that only a small percentage of bills make it beyond introduction but party polarization and lawmakers' willingness to engage in obstruction have made passing bills through the regular order increasingly difficult (Sinclair, 2016; Curry and Lee, 2016). These developments have contributed to a decline in the number of laws enacted by Congress (Taylor, 2013, p.145, Figure 7.1), and a lengthy incubation process for legislative proposals (Burstein et al., 2005).

The majority party monopolizes the limited opportunities to claim credit for bills that become law (Cox and McCubbins, 2005). In the 113th Congress (the last of the Congresses we analyze in this study), about 30% of all non-minor laws were sponsored by a small number of committee and subcommittee leaders.[4] Combined, majority party members (who make up 50-60% of the chamber) sponsor about 75% of laws. However, we believe that many of these successes have little to do with individual entrepreneurship. Agenda control provides these lawmakers with exceptional opportunities to sponsor the vehicles used in collective lawmaking efforts. If this is the case then a lawmaker's ability to advance a bill as a hitchhiker may be a better indicator of their true effectiveness. Leaders should also

[4]This large proportion of bills was sponsored by 63 members representing less than 12% of all policymakers in Congress.

be more willing to support hitchhikers sponsored by minority party members if doing so advances good public policy or increases support for other legislation (Fenno, 1973; Curry and Lee, 2016).

Hypothesis 2 – Agenda Control: *Agenda control (serving as a committee or subcommittee chair or member of the majority party) will be a less important predictor of hitchhiker success than law success.*

We also test four hypotheses that more directly control for the types of legislation typically sponsored by committee leaders but not authored by them. We expect to find that these types of bills are more likely to become law. And because they are more likely to become law, we expect them to be attractive vehicles for hitchhikers.

Leadership bills. The first ten House bill numbers are reserved for the majority leader, who allocates them to bills in order to highlight the party's legislative agenda (Oleszek, 2017, 5). In the 115th Congress, for example, HR 1 (Tax Cuts and Jobs Act) was introduced by Ways and Means Chair Kevin Brady (R-TX) 11 months after the beginning of the Congress.

Hypothesis 3 – Reserved (Leadership) Bills: *The first 10 House bills in every Congress are more likely to become law regardless of who sponsors them.*

Administration bills. The administration also initiates legislation, which is then introduced "by request" by a member designated by House and Senate party leaders. Clause 7 of House Rule XXII prohibits the requesting party from being named, but House rules identify a small number of subjects where legislation can only be initiated by request (i.e. by the administration). Most are trade or international agreements and by tradition, annual defense authorizations are also frequently introduced by request. We therefore designate any "by request" bill that is primarily about defense, trade or international affairs to be administration bills.[5]

[5]Using the 16, 18, 19 major topic codes of the Policy Agendas Project.

Hypothesis 4 – By request (Administration) Bills: *Bills proposed by the administration are more likely to become law regardless of who sponsors them.*

Reauthorizations. Many important programs include sunset provisions. We assume that the introduction of a reauthorization bill reflects an impending deadline which makes them more likely to advance (Adler and Wilkerson, 2012). As with administration bills and clean bills, there is no official tag for reauthorization bills. We define them as any bill that includes "reauth*" in its title. This approach is an underestimate because many reauthorization bills do not include this term.[6]

Hypothesis 5 – Reauthorization Bills: *Bills reauthorizing existing programs are more likely to become law, regardless of who sponsors them.*

Clean bills. Instead of marking up an existing bill, a committee will often draft its own proposal. The end product of this type of markup process is a "clean" bill *afterwards* that is typically sponsored by the subcommittee or committee chair (Heitshusen, 2017, p. 6). We designate a clean bill to be a bill that is introduced and reported from committee on the same or next day. The measure is also limited because it does not capture clean bills drafted in subcommittee (where much of the work of commmittees takes place).[7]

Hypothesis 6 – Clean Bills: *Bills that are products of committee markup sessions are more likely to become law, regardless of who sponsors them.*

3.3 Procedural constraints

Whereas the agenda setting hypotheses above propose new explanations for why some bills are more likely to become law on their own, the procedural constraints hypotheses presented here propose novel explanations for why some bills are more likely to advance as hitchhikers.

[6]Such as HR 1, the 'No Child Left Behind Act of 2001" sponsored by House Education and Workforce Chair John Boehner (R-OH).

[7]An example is the Clean Air Act Amendments of 1977 (HR 6161), which was introduced as a "clean bill" (his words) by Paul Rogers (D-FL) who was the Health Subcommittee chair of the House Energy and Commerce Committee (see HR 6161: An Act of Congress http://www.gpifilms.com/filmography.html).

Revenue bills. The clearest example of a procedural constraint that incentivizes hitchhiking is the "origination" clause of Article I of the Constitution. All all laws raising revenue must originate in the House (Rybicki, 2015). The practical implication is that Senate bills that include revenue-related provisions can only be enacted as hitchhikers on House laws.[8]

Hypothesis 7 – **Revenue Bills** : *Senate revenue-related bills are less likely to become law on their own but not less likely to become law as hitchhikers. We use referral to the Senate Finance Committee as our indicator of a Senate revenue-related bill (because all revenue-related bills must be referred to this committee).*

Amendments between chambers. Bills passed by the other chamber can be attractive vehicles for hitchhikers for other procedural reasons. In both chambers, bills from the other chamber are considered under different procedures than the chamber's own bills (Rybicki, 2015). For example, instead of being referred to a Senate committee, a House-passed bill is typically placed on the Senate's Calendar of Business. To bring up a Senate bill, the majority leader must negotiate a motion to proceed (which is subject to filibuster). A House bill, in contrast, is already on the calendar (Christopher, 2017). This is why Majority Leader Harry Reid used a House passed bill (H.R. 3590) as the vehicle for the Affordable Care Act (Cannan, 2013). Another reason to expect amendments between chambers to be an important entry point for hitchhikers is because the President can sign only one bill into law. A common practice as the chambers resolve their differences is for one chamber to take up the other chamber's bill, "strike all after the enacting clause" and insert its own proposal (Rybicki, 2015, p. 3). We should therefore expect that the process of resolving differences leads to many cross-chamber hitchhikers.

Hypothesis 8 – **Amendments Between Chambers**: *Cross-chamber hitchhikers will be common, and especially common in the Senate.*

[8]In practice, this requirement also extends to appropriations but we exclude appropriations from our analysis (Rybicki, 2015, p. 2).

4 Finding Hitchhikers: A Supervised, Active Learning Approach

In this section we describe how we identify hitchhiker bills. The general goal was to examine every bill-law pairing using supervised learning methods. To do this, we made use of a corpus of bill texts collected by Handler et al. (2016) which includes 104,005 unique versions of 81,801 bills that did not become law and 4,073 laws for the 103rd-113th Congresses (1993-2014). The standard approach to such a project is to first label a large, random sample of bill-law pairs for whether they include a hitchhiker or not, train a supervised learning classifier on part of this sample, and test its performance on a held out set of cases. When prediction accuracy for the test set is sufficiently high, the trained classifier is used to predict cases outside of the sample.

The first problem with this standard approach for the current project was that hitchhikers were assumed to be rare. If they are as rare as laws, we would need to visually examine and label about 10,000 bill-law pairs to obtain a reasonable sample of true hitchhiker cases (3-400). One alternative is to use "active learning" to iteratively assemble a training sample of sufficient size (Olsson, 2009). A small number of likely cases is first identified and manually labeled. This initial sample is then used to train a classifier to predict additional likely cases. These likely cases are then labeled, added to the corpus and the process is repeated. This method made it possible to identify substantial corpus of true hitchhikers from less than 1,000 manually labeled cases (see a detailed explanation of the active learning method and output in Supporting Information C).

A second challenge, discovered as we were labeling these likely cases, was the bill and law texts could be quite different for true hitchhikers. This was because the bill in the bill-law pair often contained non-substantive front matter (such as the title and date of introduction) and sections (e.g. Findings and Definitions) that were removed from what was incorporated

into the law. We addressed this by implementing a custom pre-processing protocol that removed non-substantive language from the bills and laws (see Supporting Information A for a full description of the pre-processing steps).

4.1 A new sequence-based algorithm for characterizing document similarity

Preprocessing dramatically improved predictions in many cases. However, our early experiments also revealed a new challenge. Even after pre-processing, the substantive language of the law and bill in true hitchhikers cases could still be still quite different as when the law reordered, deleted or added new words or sentences. We first tested several text reuse algorithms widely used in computational linguistics and information retrieval.[9] All of them predicted the clean cases, but none predicted the messier cases very well. However, our efforts to understand why (by inspecting the false positive and false negative cases) inspired us to develop an entirely new approach. We describe in this section the basic intuition behind the approach but a more detailed description can be found in Supporting Information B.

Hitchhikers are similar to plagiarism. They are characterized by lengthy sequences of matching text (between the bill and law), with occasional shorter sequences of mismatched text. "Bag of words" approaches (e.g. Cosine, Dice) do not consider the proximity of shared words. Alignment algorithms do (e.g. Smith-Waterman), but these algorithms ask the researcher to specify, in advance, how the differences between two sequences are scored.

Our approach accounts for word proximity without committing to a single parameterization (as Smith-Waterman requires). We propose a "sequence-based" algorithm that, as with alignment algorithms, only uses information about patterns of matching and non-matching text. It does not consider (for example) the frequency of particular words. However, it dif-

[9] `diff`, `wdiff`, Dice coefficients (Dice, 1945), Cosine similarity, and the Smith-Waterman algorithm (Waterman et al., 1976)

fers from other approaches that focus on shared word sequences in important ways. Below are two versions of the same section of the Dodd-Frank Wall Street Reform and Consumer Protection Act. The first is from the bill as introduced in the House:

```
SEC. 1008. OVERSIGHT BY GAO.

    (a) Authority to Audit.--The Comptroller General of the United
States may audit the activities and financial transactions of--
            (1) the Council; and
            (2) any person or entity acting on behalf of or under the
        authority of the Council, to the extent such activities and
        financial transactions relate to such person's or entity's work
        for the Council.
```

The second is from the version signed into law by President Obama:

```
SEC. 122. GAO AUDIT OF COUNCIL.

    (a) Authority To Audit.--The Comptroller General of the United
States may audit the activities of--
            (1) the Council; and
            (2) any person or entity acting on behalf of or under the
        authority of the Council, to the extent that such activities
        relate to work for the Council by such person or entity.
```

These two versions are similar but alo contain differences (such as different section titles). We first characterize each document as a set of overlapping "n-grams." A n-gram is a contiguous sequence of n words. Overlap means that adjacent n-grams share words. In this example we use 5-grams that overlap by n-1 words. Thus two 5 grams in the first version are "to work for the Council by" and "work for the Council by such." We then compare each n-gram in one version every n-gram in the second and represent whether there is a match or not as an entry in a vector. In Figure 1 blue rectangles indicate cases of matching 5 grams while orange rectangles indicate that a 5-gram does not have a match in the other document. Thus, a series of blue rectangles indicates a longer block of shared text whereas a series of orange rectangles indicates a longer mismatched block.

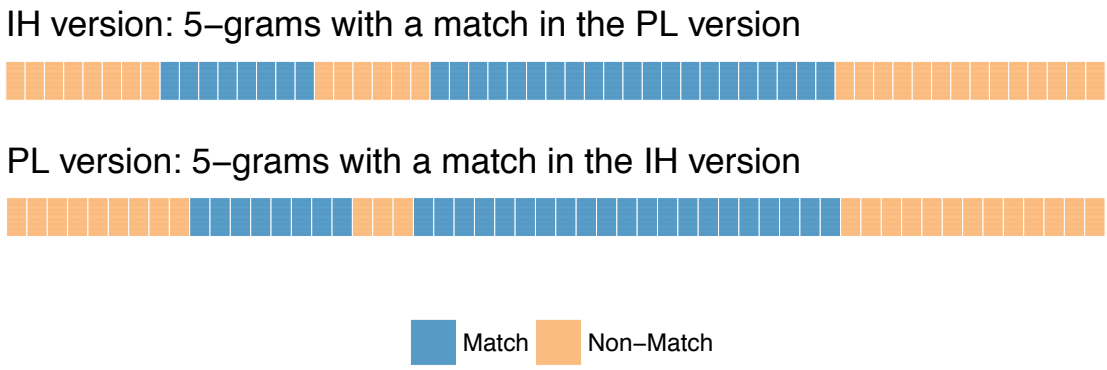

Figure 1: A comparison of two versions of the same section from the Dodd-Frank Wall Street Reform and Consumer Protection Act: Introduced in The House (IH), and the version that become Public Law (PL) using 5-grams on the un-preprocessed text. Blue rectangles indicate places where a 5-gram in the IH version exactly matches a 5-gram in PL version in the top plot, and vice-versa in the bottom plot.

The unique value of this approach is that whereas other alignment algorithms produce a single similarity score, we can construct many sequence-based similarity statistics (e.g. longest matching sequence, average matching sequence length, etc.). These statistics can then be used as features in models to discover which do the best job of predicting known hitchhiker cases. The best performing models can then be used to make final predictions.

For the active learning process described earlier, we tested over 1,500 models combining 21 different statistical features, identified the best performing, and used them to predict additional likely hitchhiker cases. We labeled the new cases, and repeated this process until no more hitchhikers were predicted for the 111th Congress. The majority vote of this final trained ensemble of 22 models was then used to predict hitchhikers across 20 years of lawmaking. Compared to our earlier experiments with other algorithm, this approach proved very accurate.[10]

[10]The majority vote of this ensemble had 95% precision (5% false positive rate) and 92% recall (8% false negative rate) based on 300-fold cross validation. The other o the shelf algorithms had higher recall on average (99%), but much lower precision (75%). In this respect our approach errs on the conservative side of underestimating the number of true hitchhikers..

5 Findings

Our findings begin with a brief overview of hitchhiker bills across 11 recent Congresses. We then test regression models predicting if bills become law on their own or as hitchhikers. In addition to comparing standard explanations of legislative effectiveness for laws versus hitchhikers, we also include covariates for the bicameral, agenda setting and procedural constraints hypotheses discussed earlier. The last section of the findings then explores hitchhiker strategies in more detail. What kinds of laws are particularly attractive vehicles? At what stage of the process are hitchhikers most likely to be added to these laws?

5.1 Hitchhiker bills in Congress, 1993-2014

Figure 2 confirms the importance of unorthodox lawmaking. The figure distinguishes between non-minor (left) and minor (right) laws and hitchhiker bills using the Important bill filter of the Congressional Bills Project. For the 1993-2014 time period, we identify 2,997 non-minor hitchhikers compared to 2,905 laws.[11]. Thus, hitchhikers double the number of legislative enactments and the number of successful bill sponsorships.

The pattern for minor bills (e.g. to name federal buildings or produce commemorative coins) is also noteworthy. These bills are much more likely to be enacted as stand alone laws than as hitchhikers. This is consistent with the agenda setting argument proposed earlier. Minor bills do not compete for limited agenda space. They do not consume the time of committees and typically pass under expedited procedures that limit debate (Suspension of the Rules in the House and Unanimous Consent in the Senate). So there is less need to resort to hitchhiker strategies.

[11] We exclude appropriations bills, private bills and duty suspension/ tariff bills as described in Supporting Information A.

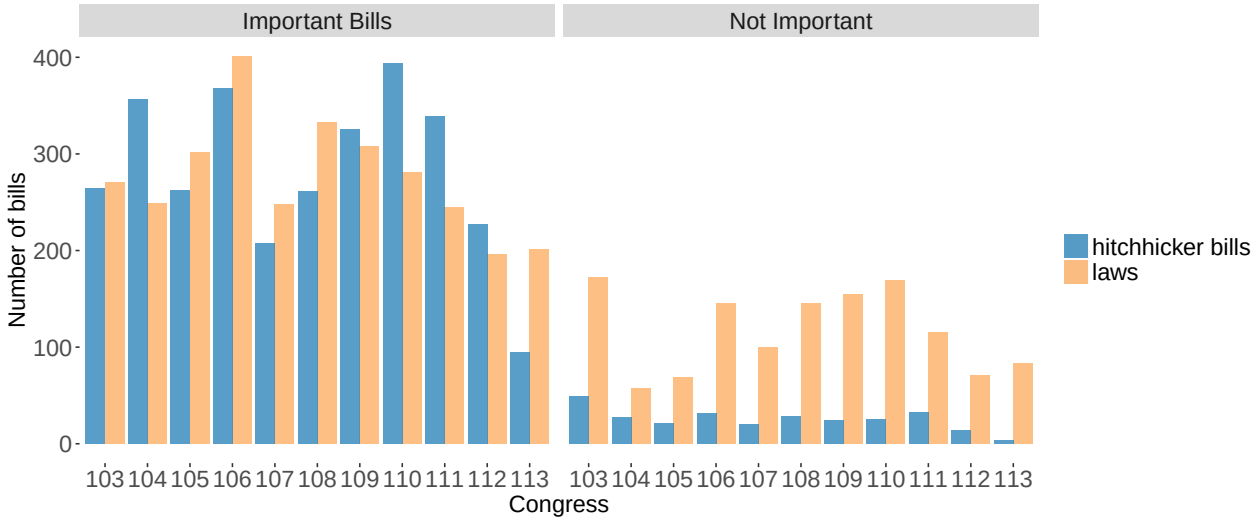

Figure 2: Counts of laws *versus* Hitchhiker bills (103rd-113th Congresses).

5.2 Sponsor and procedural predictors of bill success

Does accounting for the 50% of legislative enactments that are hitchhikers alter current understandings of who is effective in Congress? Current approaches measure effectiveness by either a single threshold of success (e.g. was the bill passed by chamber?), or by assigning more weight to bills that advance further in the process (e.g. the LES scores of Volden and Wiseman (2014)). If, for example, successful hitchhikers always pass the chamber on their own, then we would not expect much difference. However, Figure 3 indicates that this is not the case. Most non-minor hitchhiker bills do not progress beyond introduction on their own.

To compare the predictors of successful laws and hitchhikers, we estimate two logistic regressions. These regressions include the sponsor characteristics commonly found to be important in prior effectiveness research. However, where prior studies ask only whether the sponsor was a leader of *any* committee, we ask whether they led the committee with jurisdiction over the bill (or one of its subcommittees). This is a more precise test of the advantages of agenda control.[12] We also include a dummy for whether the sponsor is a

[12]For bills referred to multiple committees, this variable indicates whether the sponsor served on at least one of them.

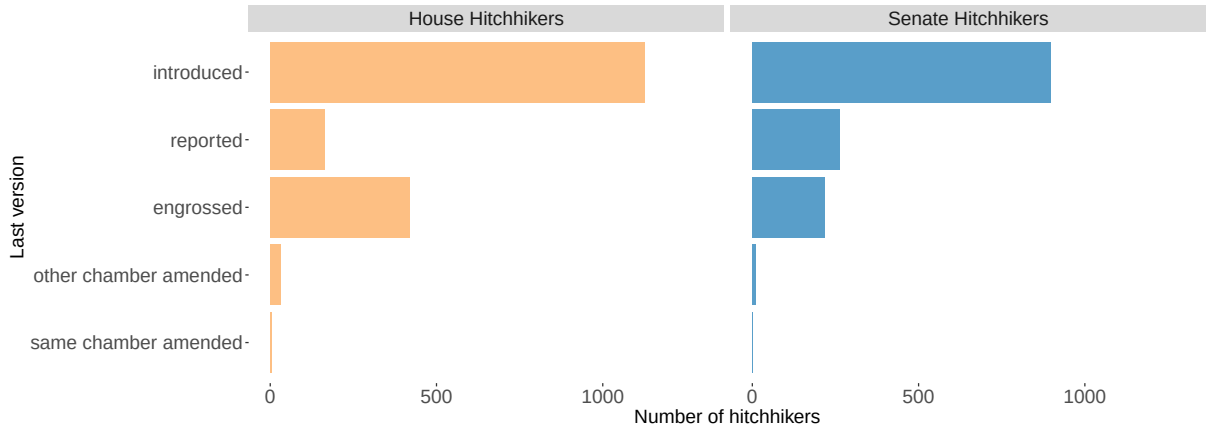

Figure 3: How far do hitchhiker bills get on their own?

senator. Finally, to test the bicameral support, agenda setting and procedural constraint hypotheses introduced in section 3), we include dummies for whether a bill has a companion in the other chamber, and for different types of bills that we expect will be more likely to progress regardless of their sponsors.

Figure 10 displays marginal effects and 95% confidence intervals for these two regression models.[13] Each figures has two scales to reflect the much larger effects of the bill type variables at the bottom (the triangles). Null effects for the upper variables (on the x0-4 scale) are represented by the black vertical line and the dashed line for the lower variables on the x0-16 scale.

The models indicate that sponsors of successful hitchhikers are similar to sponsors of successful laws with one main difference. Compared to other members, committee leaders and majority party members are much more likely to sponsor the bills that become law.[14]

[13]The full results are found in Table Supporting Information D in Supporting Information D. The estimates are based on min-max values because many of the independent variables are dummies where a one standard deviation change is meaningless.

[14]A version of the model on the left –LAW– without the bill level covariates (whether a proposal is a reauthorization, revenue, companion, clean, or administration bill) shows that, when not taking into consideration these special cases, Committee Chairs look about 15% more productive (from 3.2 to 3.7 more likely to sponsor bills that become law). Despite using crude measures for some of these bill level covariates (we only detect for example a small proportion of clean bills), this already indicates a tendency to overestimate the productivity of Committee Chairs since these are bills that become stand alone laws for reasons that have little to do with them.

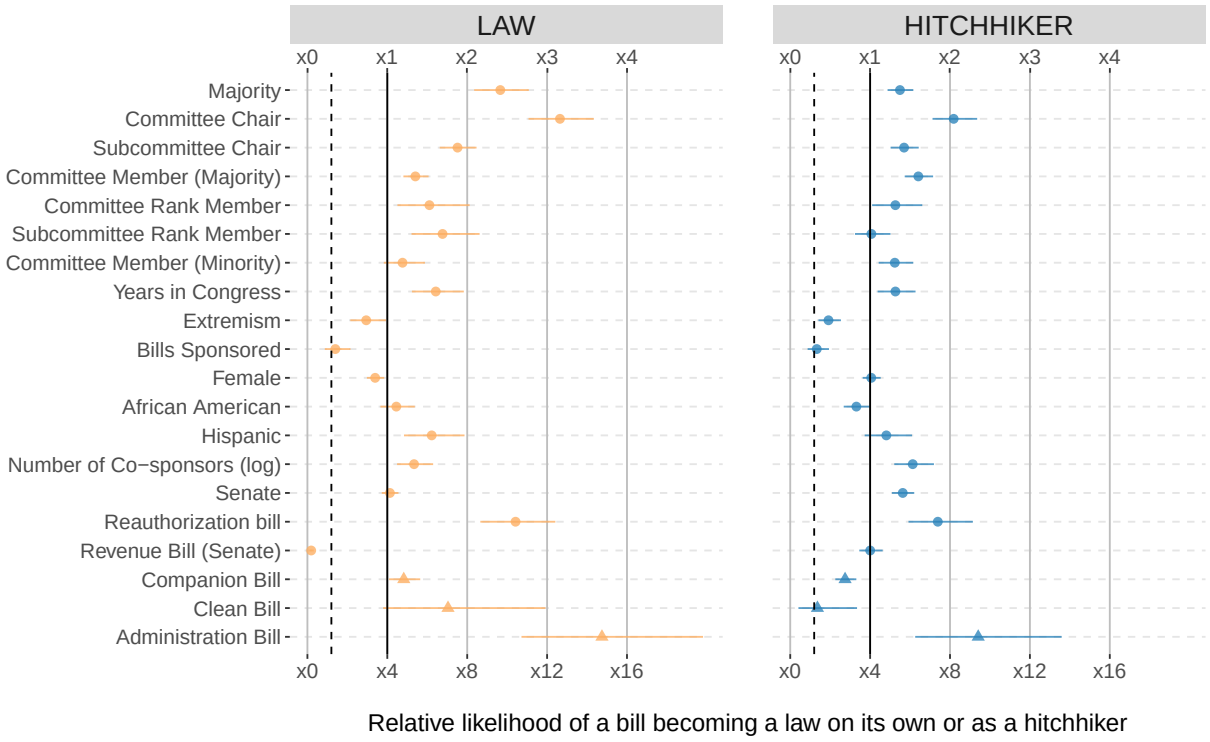

Figure 4: Marginal effects of sponsor and bill characteristics on law *versus* hitchhiker success.

Substantively, this means that rank and file and minority party members are relatively more successful advancing bills as hitchhikers than as stand alone laws. The model comparison also indicates that Senators are more likely than House members to advance bills as hitchhikers.

Figure 5 provides another perspective on how hitchhikers alter assessments of effectiveness. The blue line represents the percentage of legislators in each Congress that sponsor at least one non-minor law. The orange line represents those who sponsored at least one law or hitchhhiker. The largest increase is for minority party members. Twice as many are successful (from 16% to 32%) compared to laws alone. In every category, however, more lawmakers are successful when hitchhikers are considered.

Figure 6 reports correlations between sponsor LES scores and their enactments (laws plus hitchhikers).[15] LES scores are much more strongly correlated with the actual effectiveness of

[15]Each dot represents the number of enactments and the LES score for a House representative and

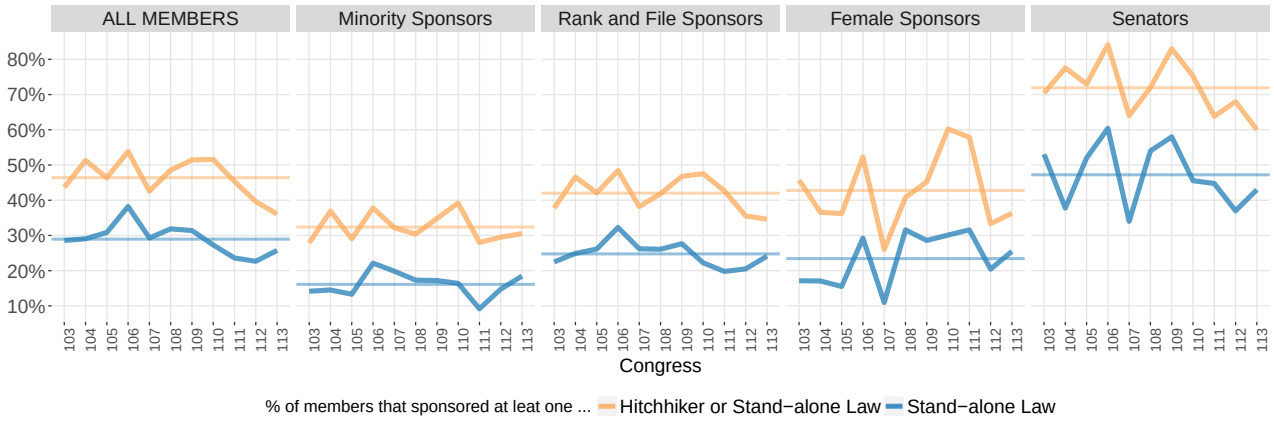

Figure 5: Successful sponsorship Legislators sponsoring successful lawswith at least one oScope of bill sponsorship success.

committee leaders (.86) than other members (.47-54).[16] This is probably because committee leaders sponsor more of the bills that make it out of committee, including those that become law on their own. While the LES methodology does a good job of capturing these patterns, it assigns little value to the many bills that do not make it beyond introduction on their own but are later enacted as hitchhikers.

The regressions in Figure 10 also demonstrate that certain types of legislation are much more likely to become law. Bills with companions are about 6 times more likely; by request bills are 14 times more likely to become law; reauthorizations are 10 times more likely; committee developed "clean" bills 7 times more likely; and leadership highlighted bills 4 times more likely. As expected, Senate revenue-related bills are much less likely to become law on their own (due to the origination clause).[17]

Congress.

[16]The *Rank and File* category includes members of the majority and minority party that are neither Committee Chairs nor Ranking Member

[17]We include in both models an interaction between revenue bills (whether a –House or Senate– bill has been referred to the House Ways and Means or Senate Finance committee) and the chamber the bill is introduced. In Figure 10 we plot only the marginal effect for the scenario of interest in Hypothesis 7: Senate revenue bills

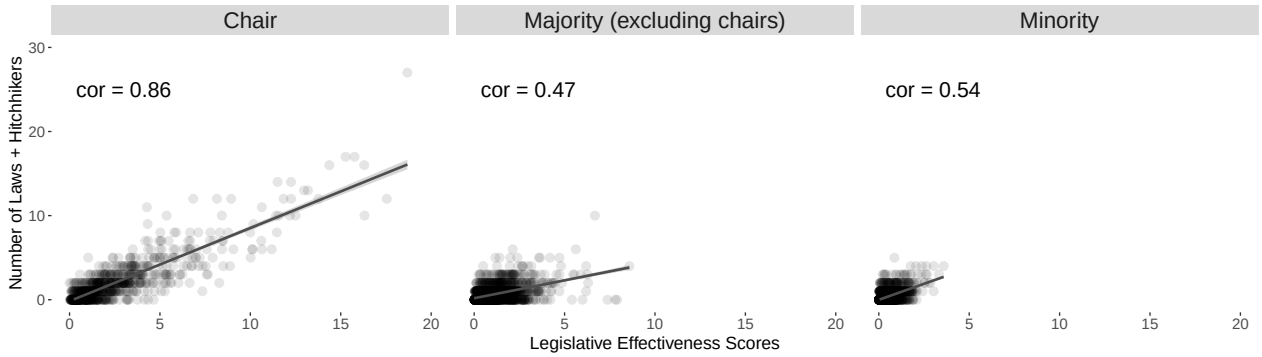

Figure 6: Correlation between the sponsor's enactments and the Legislative Effectiveness Scores of Volden and Wiseman (2014)

5.3 Hitchhiker Vehicles

Bills that are more likely to become law should be particularly attractive vehicles for hitchhikers. Figure 7 tests this by examining the hitchhikers attached to different types of laws. The upper left quadrant includes House hitchhikers attached to different types of House vehicles. The upper right includes Senate hitchhikers attached to House vehicles, etc. The most common scenario are 234 House hitchhikers sponsored by committee members that became law as provisions of non-leadership, non-administration, non-revenue, and non-reauthorization laws sponsored by a chair of the same committee or subcommittee. In terms of law type, the most common vehicle for hitchhikers is a revenue-related bill, which attracted a total of 681 House and Senate hitchhikers.[18] Another common vehicle are (House and Senate) administration-initiated laws, which attract a total of 285 hitchhikers.

Altogether these 5 types of vehicles represented just 31% (925) of all laws (2,904) but attracted abut 50% (1,335) of all non-minor hitchhikers (2,997). The figure also clearly demonstrates that House laws are much more common vehicles for both House and Senate hitchhikers than Senate laws. In fact, more Senate bills became law as hitchhikers on House laws (1,118) than were enacted on their own (1,037).

[18]Most of these hitchhikers add a substantive amount of content to the bill. On average, the length of these revenue bill hitchhikers represent about 15% of the length of the enacted law they get attached to.

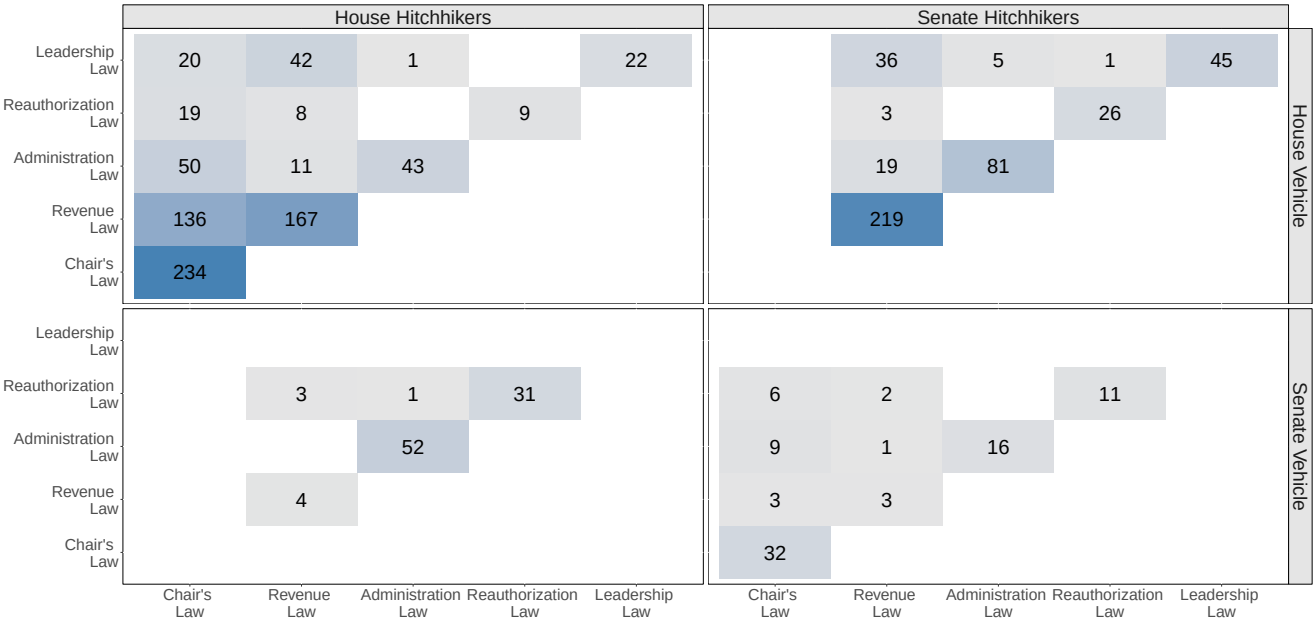

Figure 7: Exploring the vehicles of hitchhiker bills

A final hypothesis yet to be examined is whether hitchhikers are frequently added while the chambers are resolving their differences. As discussed, House-passed bills should be especially attractive vehicles for Senate hitchhikers because Senate leaders do not need to negotiate a motion to proceed to bring them to the floor. Experts also note that when both chambers pass different versions of the same proposal, the second chamber frequently substitutes its version before adopting the first chamber's bill. To investigate this hypothesis we count the number of hitchhikers added at the different stages of a law's consideration (Figure 8).[19] The top clusters are House laws while the bottom are Senate laws. The bars indicate the number of hitchhikers, which are color coded for whether they address the same primary topic as the law, of their primary topics when that is not the case. [20]

Figure 8 indicates that hitchhikers get added at every stage of the lawmaking process. As expected, many (the plurality) are added while the chambers are resolving their differences,

[19]To construct the figure we compare the hitchhiker to the different versions of the bill that became law. The last most similar alignment is when it is added.

[20]Using the major topic codes of the Policy Agendas Project

with more of these hitchhikers originating in the Senate. However it is important to note that the figure does not indicate which chamber added a hitchhiker. For example, in drafting the Affordable Care Act, the Senate drew on several other major House and Senate markup bills. The final law (HR 3590) did not align with any other single Senate bill, but with many other smaller bills (that were incorporated into the Senate's 900 page amendment). It is therefore likely that the Senate's amendment included many House hitchhikers. Just as it is a mistake to infer that the House has more policy influence because more laws originate in the House, it would be a mistake to infer that the Senate has more influence when it completely replaces the substance of a House-passed bill. The method presented here makes it possible to trace policy successes back to their sources.

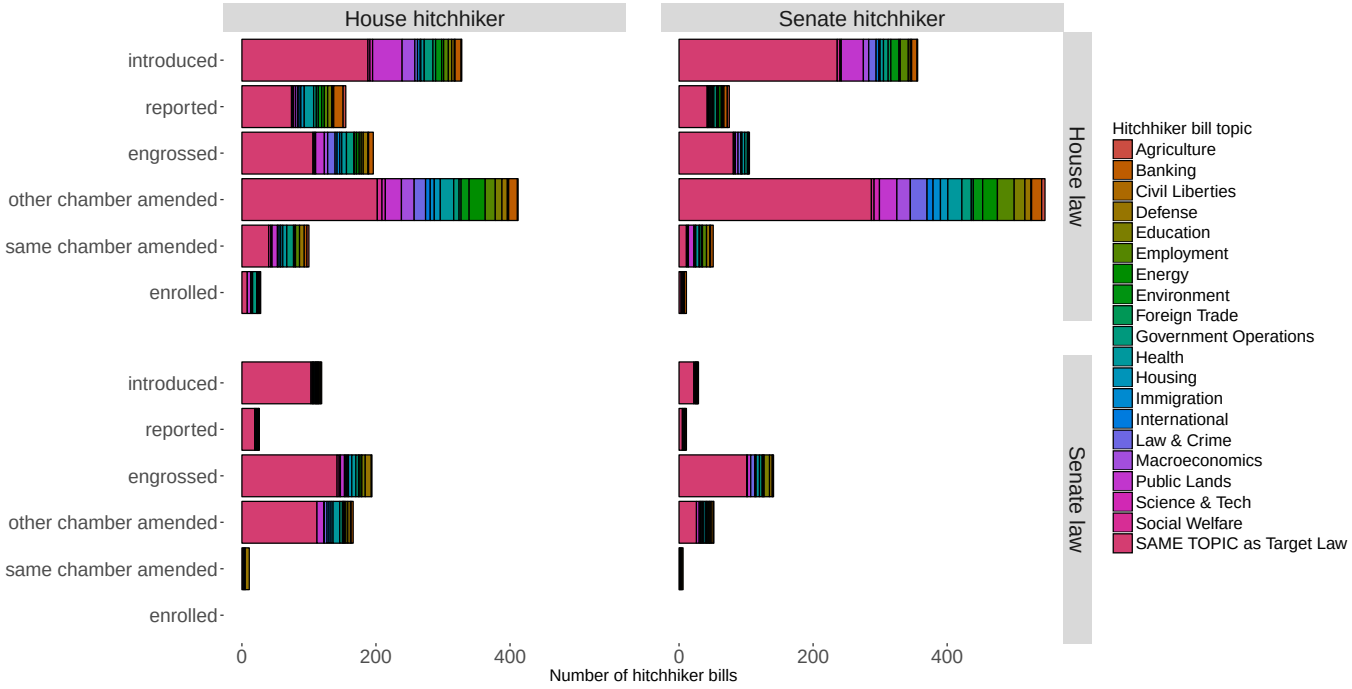

Figure 8: Where hitchhikers bills get picked up during the legislative process.

6 Discussion

In this paper we reexamine a longstanding subject of legislative studies. In 1960, Donald Matthews observed that "[t]o the extent that the concept as used on Capitol hill has any distinct meaning, *effectiveness* seems to mean the ability to get one's bills passed." For more than 50 years scholars have defined legislative effectiveness by whether the bills legislators sponsor progress through the legislative process. We redefine getting "one's bills passed" to include bills that are enacted into law as provisions of other bills.

Accounting for these "hitchhiker" successes offers new insights into effectiveness and into lawmaking more generally. The congressional opportunity structure appears less hierarchical and less partisan. Many more lawmakers can claim credit for a least one legislative enactment.

We have also tried to provide a more realistic and wholistic perspective on why bills progress and on lawmaking more generally. Effectiveness matters, but many bills also progress for reasons that have little to do with who sponsors them. These bills also tend to attract a large proportion of 50% of legislative enactments that are hitchhikers. Failing to account for these types of bills leads to misleading conclusions about effectiveness. Perhaps the best reason to be concerned about bill progress as a measure of effectiveness is the fact that more Senate bills become law as hitchhikers on House laws than become law on their own. And, as we saw, most of these hitchhikers never make it beyond introduction on their own.

There is much more about hitchhikers to be explored. Although we tracked who sponsored successful hitchhikers, and where in the process they were added, we have not examined many other questions about the venues through which hitchhikers advance. To what extent do House hitchhikers get added by the Senate and vice versa? Are some lawmakers more effective in getting their proposals added in the other chamber and if so why? Which policy

areas tend to attract the most hitchhikers, and why? Related to this bicameralism. Whereas current research on bicameral negotiations examines just one or a very small number of legislative enactments (see Monroe (2012) for a summary), the text based methods introduced provide opportunities to assess the relative influence of the House and Senate across many bills, issues, and partisan circumstances (e.g. unified versus divided government).

Another area of opportunity is legislative productivity research. There are currently two approaches to measuring legislative productivity - counts of laws and counts of "major laws." We find that counting hitchhikers as enactments has a dramatic impact on the first measure: Congress is about twice as productive. But hitchhikers also offer new opportunities for systematically categorizing the importance of laws by (among other things) the number and diversity of topics they include, as well as the amount of text attention these topics receive. This measure might then be compared with other approaches to measuring the importance of laws (beyond major law or not) (e.g. Jones and Baumgartner (2005)). [**Need more cites here!**]

More broadly, whereas current approaches to studying legislative history focus on a small number of historically significant laws, the similarly algorithm introduced in this paper can be used to investigate how thousands of individual bills evolve as they move through the lawmaking process. One basic yet to be examined question is — how much do the bills that become law change from one stage of the lawmaking process to the next? Statistical features derived from the algorithm can also be used to study more specific questions such as: Are bill edits mostly additions of new text or deletions? Do they tend to be granular (indicating focused word-smithing) or coarse (indicating the introduction or deletion of new provisions? Are new additions typically on-topic or off-topic? Do editing patterns differ depending on stage of the process (committee vs. floor), chamber, topic, or political context? Can editing patterns predict cosponsorship or whether a bill will progress?

Hitchhiker bills are just one way that lawmakers are able exercise policy influence. They

are closer to the "ground truth" of effectiveness than approaches that focus on how far bills progress in the legislative process on their own. But we have not examined partial bill hitchhikers or successful amendments. While both present additional challenges in terms of validation and data availability, they are amenable to investigation using the methods introduced here. We are hopeful that researchers will soon be studying them.

References

Adler, E. Scott and John D. Wilkerson. Agenda Setting in a Problem-Solving Legislature. In *Congress and the Politics of Problem Solving*, pages 116–140. Cambridge University Press, New York, 2012.

Anderson, William D., Janet M. Box-Steffensmeier, and Valeria Sinclair-Chapman. The Keys to Legislative Success in the U.S. House of Representatives. *Legislative Studies Quarterly*, 28(3):357–386, 2003.

Arnold, Laura W., Rebecca E. Deen, and Samuel C. Patterson. Friendship and Votes : The Impact of Interpersonal Ties on Legislative Decision Making. *State & Local Government Review*, 32(2):142–147, 2000.

Baughman, John. *Common Ground: Committee Politics in the U.S. House of Representatives*. Stanford University Press, Stanford, CA, 2006.

Bloomfield, Louis. WCopyFind. *Software. http://plagiarism.bloomfieldmedia.com/wordpress/software/wcop (accessed February 20, 2017)*, 2008.

Bratton, Kathleen A. and Kerry L. Haynie. Agenda setting and legislative success in state legislatures: The effects of gender and race. *The Journal of Politics*, 61(3):658–679, 1999.

Burstein, Paul, Shawn Bauldry, and Paul Froese. Bill Sponsorship and Congressional Support for Policy Proposals, from Introduction to Enactment or Disappearance. *Political Research Quarterly*, 58(2):295 –302, 2005.

Cannan, John. A Legislative History of the Affordable Care Act: How Legislative Procedure Shapes Legislative History. *Law Library Journal*, 105(2):131–173, 2013.

Christopher, Davis. How measures are brought to the senate floor: A brief introduction. Technical Report Congressional Research Service RS20668, April 2017.

Cox, Gary W and Mathew D McCubbins. *Setting the agenda: Responsible party government in the US House of Representatives*. Cambridge University Press, 2005.

Cox, Gary W. and William C. Terry. Legislative Productivity in the 93d105th Congresses. *Legislative Studies Quarterly*, 33(4):603–618, 2008.

Crisp, Brian F., Kristin Kanthak, and Jenny Leijonhufvud. The Reputations Legislators Build: With Whom Should Representatives Collaborate? *American Political Science Review*, 98(4):703–716, 2004.

Curry, James M. and Frances E. Lee. Congress at Work: Legislative Capacity and Entrepreneurship in the Contemporary Congress. In *Anxieties of Democracy Working Group on Institutions*. Princeton University, 2016.

Denny, Matthew J. and Arthur Spirling. Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Available at SSRN 2849145*, 2017.

Desmarais, Bruce a., Vincent G. Moscardelli, Brian F. Schaffner, and Michael S. Kowal. Measuring legislative collaboration: The Senate press events network. *Social Networks*, 40:43–54, jan 2015.

Dice, Lee R . Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26 (3):297–302, 1945.

Fenno, Richard R. *Congressmen in Committees*. Little, Brown, Boston, 1973.

Fowler, James H. Connecting the Congress: A Study of Cosponsorship Networks. *Political Analysis*, 14(4):456–487, mar 2006.

Frantzich, Stephen. Who Makes Our Laws? The Legislative Effectiveness of Members of the U.S. Congress. *Legislative Studies Quarterly*, 4(3):409–428, 1979.

Gluck, Abbe, Anne O'Connell, and Rosa Po. Unorthodox Lawmaking, Unorthodox Rulemaking. *Columbia Law Review*, 115(7):1789–1865, 2015.

Hall, Richard L. Measuring Legislative Influence. *Legislative Studies Quarterly*, 17(2):205, may 1992.

Handler, Abram, Matthew J. Denny, Hanna Wallach, and Brendan O'Connor. Bag of What? Simple Noun Phrase Extraction for Text Analysis. In *Proceedings of the Workshop on Natural Language Processing and Computational Social Science at the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016.

Hasecke, EB and JD Mycoff. Party Loyalty and Legislative Success: Are Loyal Majority Party Members More Successful in the U.S. House of Representatives? *Political Research Quarterly*, 60(4):607–617, 2007.

Heitshusen, Valerie. Introduction to the Legislative Process in the U.S. Congress. Technical report, Congressional Research Service, 2017.

Jeydel, Alana and Andrew J. Taylor. Are women legislators less effective? evidence from the u.s. house in the 103rd-105th congress. *Political Research Quarterly*, 56(1):19–27, 2003.

Jones, Bryan and Frank Baumgartner. *The Politics of Attention: How Government Prioritizes Problems*. The University of Chicago Press, Chicago, 2005.

Kessler, Daniel and Keith Krehbiel. Dynamics of Cosponsorship. *American Political Science Review*, 90(3):555–566, 1996.

Kirkland, Justin H. The Relational Determinants of Legislative Outcomes: Strong and Weak Ties Between Legislators. *The Journal of Politics*, 73(03):887–898, aug 2011.

Kirkland, Justin H. and Mary A. Kroeger. Companion bills and cross-chamber collaboration in the u.s. congress. *American Politics Research*, 0(0):1532673X17727094, 2017.

Koger, Gregory and James H. Fowler. Parties and agenda-setting in the senate, 1973-1998. *Available at SSRN: https://ssrn.com/abstract=1017901*, 2007.

Krutz, Glen S. Tactical Maneuvering on Omnibus Bills in Congress. *American Journal of Political Science*, 45(1):210–223, 2001.

Krutz, Glen S. Issues and Institutions: "Winnowing" in the U.S. Congress. *American Journal of Political Science*, 49(2):313–326, 2005.

Matthews, Donald R. *United States Senators and Their World*. University of North Carolina Press, 1960.

Meyer, Katherine. Legislative influence: Toward theory development through causal analysis. *Legislative Studies Quarterly*, 5(4):563–585, 1980.

Miquel, Gerard Padro I. and James M. Snyder. Legislative Effectiveness and Legislative Careers. *Legislative Studies Quarterly*, 31(3):347–381, 2006.

Monroe, Nathan. Bicameral Resolution: The Politics and Policy Implications of Creating Identical Bills. In Carson, Jamie L., editor, *New Directions in Congressional Politics*. 2012.

Nay, John. Predicting and understanding law-making with word vectors and an ensemble model. *Plos One*, 12(5):e0176999, 2017.

Oleszek, Mark J. Introducing a senate bill or resolution. Technical Report Congressional Research Service R44195, January 2017.

Olson, David M. and Cynthia T. Nonidez. Measures of legislative performance in the u.s. house of representatives. *Midwest Journal of Political Science*, 16:269–77, 1972.

Olsson, Frederik. *A literature survey of active machine learning in the context of natural language processing*. SICS Technical Report T2009:06, Kista, Sweden, 2009.

Rybicki, Elizabeth. Amendments Between the Houses: Procedural Options and Effects. *CRS Report for Congress*, 2015.

Sinclair, Barbara. *Unorthodox Lawmaking: New Legislative Processes in the US Congress (5th Edition)*. CQ Press, 2016.

Sulkin, Tracy. *The Legislative Legacy of Congressional Campaigns*. Cambridge University Press, Cambridge, 2011.

Taylor, Andrew J. *Congress: A Performance Appraisal*. Westview Press, 2013.

Thomas, Scott J. and Bernard Grofman. Determinants of Legislative Success in House Committees. *Public Choice*, 74(2):233–243, 1992.

Volden, Craig and AE Wiseman. Legislative effectiveness in Congress. In *Annual Meeting of the Midwest Political Science Association.*, number March in Annual Meeting of the Midwest Political Science Association., 2009.

Volden, Craig and Alan E Wiseman. *Legislative Effectiveness in the United States Congress: The Lawmakers*. Cambridge University Press, 2014.

Walker, Jack L. Setting the agenda in the u.s. senate: A theory of problem selection. *British Journal of Political Science*, 7(4):423–445, 1977.

Waterman, M. S., T. F. Smith, and W. A. Beyer. Some biological sequence metrics. *Advances in Mathematics*, 20(3):367–387, 1976.

Yano, Tae, Noa A. Smith, and John Wilkerson. Textual predictors of bill survival in congressional committees. *2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 793–802, 2012.

Supporting Information A Pre-processing

This appendix describes our approach to identifying hitchhiker bills. We propose an original active, supervised learning methodology that is tailored to studying legislative editing processes. As noted in the discussion, this new method offers research opportunities beyond the identification of hitchhiker bills. Its distinguishing attribute is the ability to create a wide variety of statistical features from a single, comparatively fast, algorithm. The software is available via the `document_similarities()` function of the SpeedReader R package.[21]

We also decided to exclude certain types of bills from our analysis. Private bills, duty suspension/tariff bills, and continuing appropriations bills. The problem in each case is that bills are very similar in content (often differing by just a word or two). We also exclude larger appropriations legislation because successful appropriations bills are always sponsored by Appropriations Committee leaders.

Research demonstrates that pre-processing decisions can have important consequences for prediction (Denny and Spirling, 2017). Our pre-processing steps are tailored to the task at hand. Early on we discovered that stand-alone bills often contain language that is not retained when its policy provisions are incorporated into a law. To improve the fidelity of our bill-law comparisons, we systematically remove certain non-substantive content from each text:

- Exclude *Private*, *Duty Suspension/Tariff*, and *Appropriations* bills from the analysis.
- Remove the procedural head and tail of the bill (head = bill number, date, sponsors, etc. & tail = date, place of signature, etc.)
- Remove *Table of Contents*
- Remove *Findings, Definitions*, and *Authorization of Appropriations* sections.
- Remove a very frequent sentence: "Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled" from the text.
- Remove common procedural words (the top 100 words across all of the bills) (Above this threshold, the word-distribution was essentially flat).
- Transform all text to lowercase.
- Remove all punctuation and numbers.
- Remove standard "stop words" ("the", "and", "it", "we", etc.).

[21] The implementation is parallelized and scales to tens of billions of comparisons per day, depending on the number of cores and memory available.

Figures 1 and 2 illustrate the value of pre-processing. In Figure 1, the left side contains the complete text of a bill, The Southern Nevada Limited Transition Area Act (sponsored by Dan Heller (R-NV)) while the right includes a portion of a much larger law, the Omnibus Public Land Management Act of 2009 (sponsored by Rush Holt (D-NJ)). The red text highlights the parts of each bill that match language in the other.[22] There is a lot of common text, but there is also a lot of non-matching text. In addition, some of the matching text (such as the very first part of the bill) does not seem particularly relevant.

HR-408-IH	HR-146-ENR
[Congressional Bills 111th Congress] [From the U.S. Government Printing Office] [H.R. 408 Introduced in House (IH)] 111th CONGRESS 1st Session H. R. 408 To direct the Secretary of the Interior to convey to the City of Henderson, Nevada, certain Federal land located in the City, and for other purposes. IN THE HOUSE OF REPRESENTATIVES January 9, 2009 Mr. Heller introduced the following bill; which was referred to the Committee on Natural Resources A BILL To direct the Secretary of the Interior to convey to the City of Henderson, Nevada, certain Federal land located in the City, and for other purposes. Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled, SECTION 1. SHORT TITLE. This Act may be cited as the "Southern Nevada Limited Transition Area Act". SEC. 2. DEFINITIONS. In this Act: (1) City.–The term "City" means the City of Henderson, Nevada. (2) Secretary.–The term "Secretary" means the Secretary of the Interior. (3) State.–The term "State" means the State of Nevada. (4) Transition area.–The term "Transition Area" means the approximately 502 acres of Federal land located in Henderson, Nevada, and identified as "Limited Transition Area" on the map entitled "Southern Nevada Limited Transition Area Act" and dated March 20, 2006. SEC. 3. SOUTHERN NEVADA LIMITED TRANSITION AREA. (a) Conveyance.–Notwithstanding the Federal Land Policy and Management Act of 1976 (43 U.S.C. 1701 et seq.), on request of the City, the Secretary shall, without consideration and subject to all valid existing rights, convey to the City all right, title, and interest of the United States in and to the Transition Area. (b) Use of Land for Nonresidential Development.– (1) In general.–After the conveyance to the City under subsection (a), the City may sell, lease, or otherwise convey any portion or portions of the Transition Area for purposes of nonresidential development. (2) Method of sale.– (A) In general.–The sale, lease, or conveyance of land under paragraph (1) shall be through a competitive bidding process. (B) Fair market value.–Any land sold, leased, or otherwise conveyed under paragraph (1) shall be for not less than fair market value.	SEC. 2602. SOUTHERN NEVADA LIMITED TRANSITION AREA CONVEYANCE. (a) Definitions.–In this section: (1) City.–The term "City" means the City of Henderson, Nevada. (2) Secretary.–The term "Secretary" means the Secretary of the Interior. (3) State.–The term "State" means the State of Nevada. (4) Transition area.–The term "Transition Area" means the approximately 502 acres of Federal land located in Henderson, Nevada, and identified as "Limited Transition Area" on the map entitled "Southern Nevada Limited Transition Area Act" and dated March 20, 2006. (b) Southern Nevada Limited Transition Area.– (1) Conveyance.–Notwithstanding the Federal Land Policy and Management Act of 1976 (43 U.S.C. 1701 et seq.), on request of the City, the Secretary shall, without consideration and subject to all valid existing rights, convey to the City all right, title, and interest of the United States in and to the Transition Area. (2) Use of land for nonresidential development.– (A) In general.–After the conveyance to the City under paragraph (1), the City may sell, lease, or otherwise convey any portion or portions of the Transition Area for purposes of nonresidential development. (B) Method of sale.– (i) In general.–The sale, lease, or conveyance of land under subparagraph (A) shall be through a competitive bidding process. (ii) Fair market value.–Any land sold, leased, or otherwise conveyed under subparagraph (A) shall be for not less than fair market value. (C) Compliance with charter.–Except as provided in subparagraphs (B) and (D), the City may sell, lease, or otherwise convey parcels within the Transition Area only in accordance with the procedures for conveyances established in the City Charter. (D) Disposition of proceeds.–The gross proceeds from the sale of land under subparagraph (A) shall be distributed in accordance with section 4(e) of the Southern Nevada Public Land Management Act of 1998 (112 Stat. 2345). (3) Use of land for recreation or other public purposes.– The City may elect to retain parcels in the Transition Area for public recreation or other public purposes consistent with the Act of June 14, 1926 (commonly known as the "Recreation and Public Purposes Act") (43 U.S.C. 869 et seq.) by providing to the Secretary written notice of the election.

Table 1: HR-146 bill insertion example. Matches highlighted in red.

Figure 2 presents the same comparison after the texts have been subjected to our pre-processing protocol. The texts are now nearly identical.

[22]Here we use a repeated n-gram algorithm WCopyFind (Bloomfield, 2008) to define matching text

HR-408-IH	HR-146-ENR
cited southern nevada limited transition area conveyance notwithstanding federal land policy management et seq request city without consideration subject valid existing rights convey city right interest united transition area use land nonresidential development general conveyance city city sell lease otherwise convey portion portions transition area purposes nonresidential development method sale general sale lease conveyance land competitive bidding process fair market value land sold leased otherwise conveyed less fair market value compliance charter except paragraphs city sell lease otherwise convey parcels within transition area accordance procedures conveyances established city charter disposition proceeds gross proceeds sale land distributed accordance southern nevada public land management stat use land recreation public purposes city elect retain parcels transition area public recreation public purposes consistent june commonly known recreation public purposes et seq providing written notice election noise compatibility requirements city plan manage transition area accordance united code relating airport noise compatibility planning regulations promulgated accordance agree land transition area sold leased otherwise conveyed city sale lease conveyance contain limitation require uses compatible airport noise compatibility planning reversion general parcel land transition area conveyed nonresidential development reserved recreation public purposes years enactment parcel land discretion revert united inconsistent use city uses parcel land within transition area manner inconsistent uses specified discretion parcel revert united make election city	march southern nevada limited transition area conveyance notwithstanding federal land policy management et seq request city without consideration subject valid existing rights convey city right interest united transition area use land nonresidential development general conveyance city city sell lease otherwise convey portion portions transition area purposes nonresidential development method sale general sale lease conveyance land competitive bidding process fair market value land sold leased otherwise conveyed less fair market value compliance charter except subparagraphs city sell lease otherwise convey parcels within transition area accordance procedures conveyances established city charter disposition proceeds gross proceeds sale land distributed accordance southern nevada public land management stat use land recreation public purposes city elect retain parcels transition area public recreation public purposes consistent june commonly known recreation public purposes et seq providing written notice election noise compatibility requirements city plan manage transition area accordance united code relating airport noise compatibility planning regulations promulgated accordance agree land transition area sold leased otherwise conveyed city sale lease conveyance contain limitation require uses compatible airport noise compatibility planning reversion general parcel land transition area conveyed nonresidential development reserved recreation public purposes years enactment parcel land discretion revert united inconsistent use city uses parcel land within transition area manner inconsistent uses specified discretion parcel revert united make election clause

Table 2: HR-146 bill insertion example after pre-processing. Matches highlighted in red.

Supporting Information B Constructing Statistical Features

Unfortunately, not all cases of true hitchhikers are as clean as the example above. Laws incorporating language from other bills often delete, add or rearrange the original language. Thus an approach for distinguishing these messier hitchhiker cases from other cases of shared language was needed. We initially experimented with off the shelf similarity algorithms before developing the new approach that is described here.

We first tokenized the pre-processed text of each document in a way that preserved information about word ordering. We then represent each document as a set of overlapping n-grams. Here we opt for five grams (e.g. "any land sold under this") and a one word overlap. The tradeoff that must be made in terms of n-gram length is that longer n-grams (e.g. 50-grams) provide a tougher standard for shared text but open the door to more false negative predictions. Imagine two long documents that are identical except for every 50th word. A 50-gram approach will find no matches. Shorter n-grams will find the same two documents to be highly similar, but they open the door to false positive predictions. Imagine two documents that include the exact same words, but completely reversed. A unigram approach will conclude that the two documents are identical. Our decision to use 5 grams represents a middle ground approach.

We next record whether each 5-gram in a document has a match in the other document as a vector to retain information about each n-gram's location in the document. One limitation of simply asking if each n-gram has a match is that two matches are recorded when (for example) "increase funding for this program" occurs 2 times in the first document but only once in the second. On the other hand, an approach that excludes matched n-grams would (in the same example) arbitrarily conclude that the second occurrence does not have a match.

The resulting vectors capture a lot of information about each document's similarity to the other. Instead of simply comparing the proportion of n-grams that are shared, we can also compute statistics that also consider the locations of the shared n-grams. For example, we expect the matched n-grams of a hitchhiker to be located in a compact area of the law. The statistics computed for the current study are listed below (many more are possible). $\textbf{bill}_1$ refers to the bill that did not become law, and $\textbf{bill}_2$ refers to the law.[23]

- **Shared n-grams**: For each bill-law pair, we compute the simple proportion of shared n-grams in $\textbf{bill}_1$ that have a match in $\textbf{bill}_2$ and vice versa. We do this for bi-grams, trigrams, 4-grams, 10-grams, and 20-grams (10 metrics in all).
- **Addition Scope**: This is calculated as the simple proportion of n-grams in $\textbf{bill}_2$ that do not have a match in $\textbf{bill}_1$.
- **Deletion Scope**: This is calculated as the simple proportion of n-grams in $\textbf{bill}_1$ that do not have a match in $\textbf{bill}_2$.
- **Scope**: This is calculated as mean of **Deletion Scope** and **Addition Scope** and gives a general characterization of the degree of difference between the two bills.

  The remaining statistics do leverage information about matching n-gram location.

- **Maximum Match Length** ($\textbf{bill}_1$): The longest contiguous overlapping n-gram match in $\textbf{bill}_1$. This captures the size of the "biggest chunk" of shared text in $\textbf{bill}_2$ from $\textbf{bill}_1$.
- **Mean Match Length** ($\textbf{bill}_1$): The mean length of contiguous overlapping n-gram matches in $\textbf{bill}_1$.
- **Mean Match Length** ($\textbf{bill}_2$): The mean length of contiguous overlapping n-gram matches in $\textbf{bill}_2$.
- **Number of Matching Blocks** ($\textbf{bill}_1$): The number of separate matching n-gram sequences in $\textbf{bill}_1$.
- **Number of Non-Matching Blocks** ($\textbf{bill}_1$): The number of separate non-matching n-gram sequences in $\textbf{bill}_1$.
- **Number of Matching Blocks** ($\textbf{bill}_2$): The number of separate matching n-gram sequences in $\textbf{bill}_2$.

[23] Only bill version published prior to the law's enrollment date are considered.

- **Number of Non-Matching Blocks (bill_2)**: The number of separate non-matching n-gram sequences in bill_2.
- **Average Deletion Size**: The average length of non-matching sequences (the purple sequences in Figure 1) of overlapping n-grams in bill_1.
- **Proportion of Possible Deletions**: The proportion of separate non-matching n-gram sequences in bill_1 relative to the possible separate non-matching sequences (if one token were different every n-gram size + 1 tokens).
- **Deletion granularity**: We start by dividing the average length of non-matching sequences (the purple sequences in Figure 1) by the total number of overlapping n-grams in bill_1. When this proportion is equal to one, none of the text of bill_1 is present in bill_2. When it is zero, bill_1 is identical to bill_2. To calculate the deletion granularity (from bill_1 to bill_2), we subtract this proportion from 1.

Supporting Information C Active Learning with a Massive Ensemble

These statistics are then combined as features/variables in logistic regression models predicting whether a bill-law pair includes a hitchhiker. As discussed in the main text, the initial challenge was that there is no corpus of hitchhikers to train on so we needed to develop our own. The first step in this process was to use a simple bigram algorithm (Dice) to find all bill-law pairs where at least 80% of the bill's unique words (after pre-processing) matched words in the law. This filter reduced the candidate pairs by 99% (from about 400 million to about 5 million). We then identified a single law that matched 164 bills at this 80% threshold level (HR-146, the Omnibus Public Lands Management Act of 2009).

One of the authors examined and labeled these cases (using WCopyFind) and found 89 of the 164 to be true hitchhikers. The next step was to use these 164 examples to train regression models to predict additional likely cases that could also be labeled and added to the corpus. We constructed over 1,500 different models using the statistics described above.[24] We then trained these models on the initial corpus and used the best of them to predict additional likely hitchhiker cases.

In this first iteration, the 99 models that had precision and recall above 90% predicted 480 additional hitchhikers in the 111th Congress.[25] Twelve graduate students, one undergraduate, and one faculty member then labeled these cases (once again using WCopyFind to visually compare how the texts overlapped). This was easier than expected as we observed perfect agreement for the 10% of cases labeled by two or more individuals.

[24] All possible 1-to-3 variable combinations for a total of: $\sum_{n=1}^{3} \frac{21!}{n!(21-n)!} = 1,561$ models.

[25] Precision and recall are calculated using an n-fold approach that averages results across 300 partitions of the corpus into 80% train and 20% test sets.

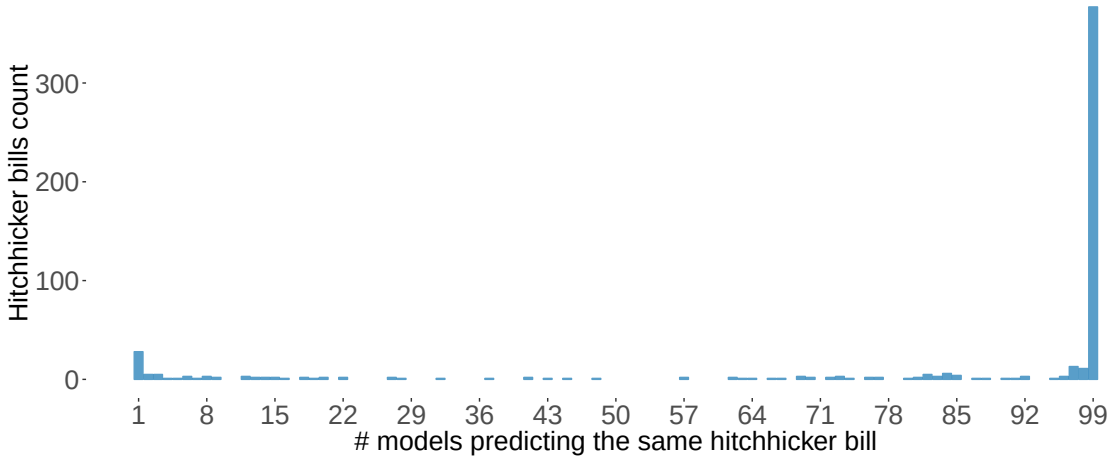

Figure 9: Bill insertion predictions for an ensemble of 99 models

We then retrained all 1,561 models using this larger corpus of 640 examples. In the second iteration, 39 models that exceeded the high performing threshold predicted just 5 additional cases. We labeled these cases and iterated the process two more times. The final ensemble of 22 high performing models - subsequently used to predict hitchhikers across all ten Congresses - had 92% precision and 95% recall. Closer inspection revealed that most of the false positive predictions (8%) were cases where a substantial portion (but not all) of the bill was in the law. The rest were very short bills that contained very similar language (such as duty suspension bills or continuing appropriations resolutions). The false negative cases (5%) tended to be cases where the annotator still judged it to be a hitchhiker case even though there was a fair amount of language difference between the overlapping text of the bill and law.

	Iteration 1	Iteration 2	Iteration 3	Iteration 4
Training Corpus Size	164	644	649	651
True Positives & Negatives	(P=89,N=75)	(477,167)	(481,168)	(483,168)
# High Performing Models	99	39	24	22
New Hitchhiker Predicted	480	5	2	1
Precision		91%	93%	92%
Recall		95%	94%	95%

Table 3: Summary of the Active Learning Process.

As a final step we used the same corpus of 650 cases to compare the performance of several off the shelf algorithms.%).[26] Their average recall was higher (99%) but their precision was

[26]Cosine similarity, Dice coefficient, WDiff, Smith-Waterman, Needleman-Wunsch

much lower (75%). This indicates that compared to the other methods, our approach is conservative. It is much less likely to make false positive predictions (92% versus 75%) at the expense of making a few more false negative predictions (95% versus 99%).

Supporting Information D Logistic Regression Models

In this appendix we first present a descriptive statistics table for all the variables included in the logistic regression models presented in the paper.

	Minimum	Maximum	Mean	Standard Deviation	Mode
Majority	0.00	1.00	0.593	0.491	1.00
Committee Chair	0.00	1.00	0.06	0.237	0.00
Subcommittee Chair	0.00	1.00	0.087	0.281	0.00
Committee Rank Member	0.00	1.00	0.029	0.167	0.00
Subcommittee Rank Member	0.00	1.00	0.05	0.217	0.00
Committee Member	0.00	1.00	0.45	0.498	0.00
Years in Congress	0.00	51.00	13.899	10.567	4.00
Extremism	0.00	1.26	0.442	0.175	0.00
Bills Sponsored	1.00	232.00	36.105	27.828	16.00
Female	0.00	1.00	0.16	0.366	0.00
African American	0.00	1.00	0.058	0.233	0.00
Hispanic	0.00	1.00	0.021	0.144	0.00
Number of Co-sponsors (log)	0.00	6.07	1.398	1.462	0.00
Senate	0.00	1.00	0.353	0.478	0.00
Reauthorization bill	0.00	1.00	0.017	0.127	0.00
Revenue Bill	0.00	1.00	0.248	0.432	0.00
Leadership Bill	0.00	1.00	0.001	0.033	0.00
Companion Bill	0.00	1.00	0.025	0.156	0.00
Clean Bill	0.00	1.00	0.001	0.026	0.00
Administration Bill	0.00	1.00	0.003	0.056	0.00
Congress	103.00	113.00			110.00
Major Policy Agendas Topic code	1.00	21.00			3.00

Table 4: Model Data: Descriptive Statistics Table

Then in the following Figure we show the coefficients (and standard errors in parenthesis) for the two logistic regression models for which we plotted marginal effects in Figure 10.

	LAW	HITCHHIKER
Majority	0.9224* (0.0751)	0.3405* (0.0628)
Committee Chair	1.1633* (0.0672)	0.72* (0.0698)
Subcommittee Chair	0.6462* (0.064)	0.3509* (0.0642)
Committee Rank Member	0.4453* (0.1538)	0.2972* (0.1233)
Subcommittee Rank Member	0.5069* (0.1347)	0.025 (0.1135)
Committee Member	0.1752 (0.1138)	0.2981* (0.0867)
Committee Member x Majority	0.1065 (0.1279)	0.2071* (0.1034)
Years in Congress	0.0092* (0.002)	0.0056* (0.0019)
Extremism	-0.2426 (0.1251)	-0.6482* (0.1218)
Bills Sponsored	-0.0042* (0.001)	-0.0047* (9e-04)
Female	-0.1738* (0.0667)	-0.0123 (0.0581)
African American	0.104 (0.1002)	-0.159 (0.1053)
Hispanic	0.418* (0.1281)	0.1781 (0.1303)
Number of Co-sponsors (log)	0.0522* (0.0142)	0.0784* (0.0138)
Senate	0.0235 (0.053)	0.3464* (0.0518)
Reauthorization bill	0.971* (0.0917)	0.6066* (0.1101)
Revenue Bill	-0.5499* (0.0743)	-0.1682* (0.0663)
Revenue Bill x Senate	-2.5336* (0.2931)	0.1744 (0.0931)
Leadership Bill	1.2757* (0.279)	0.6237 (0.3923)
Companion Bill	1.5826* (0.0856)	1.0302* (0.0998)
Clean Bill	1.6005* (0.3177)	0.1566 (0.5402)
Administration Bill	2.6487* (0.1555)	2.2447* (0.1985)
Constant	-5.5093* (0.1935)	-3.986* (0.1432)
N	84,913	82,009
AIC	21,509	23,763

Note: *p<0.05

Table 5: Results for two logistic regression models predicting whether a bill becomes a stand alone law (LAW) and, conditional on that not happening, whether it becomes law as a hitchhiker (HITCHHIKER). We include Congress (103 to 113th) and topic (Policy Agendas major topic) fixed effects, although for simplicity we do not include the fix effect coefficients in the table.

Finally, itn the following figure we report marginal effects for the same LAW and HITCHHIKER models when splitting the bills by the chamber they are introduced.

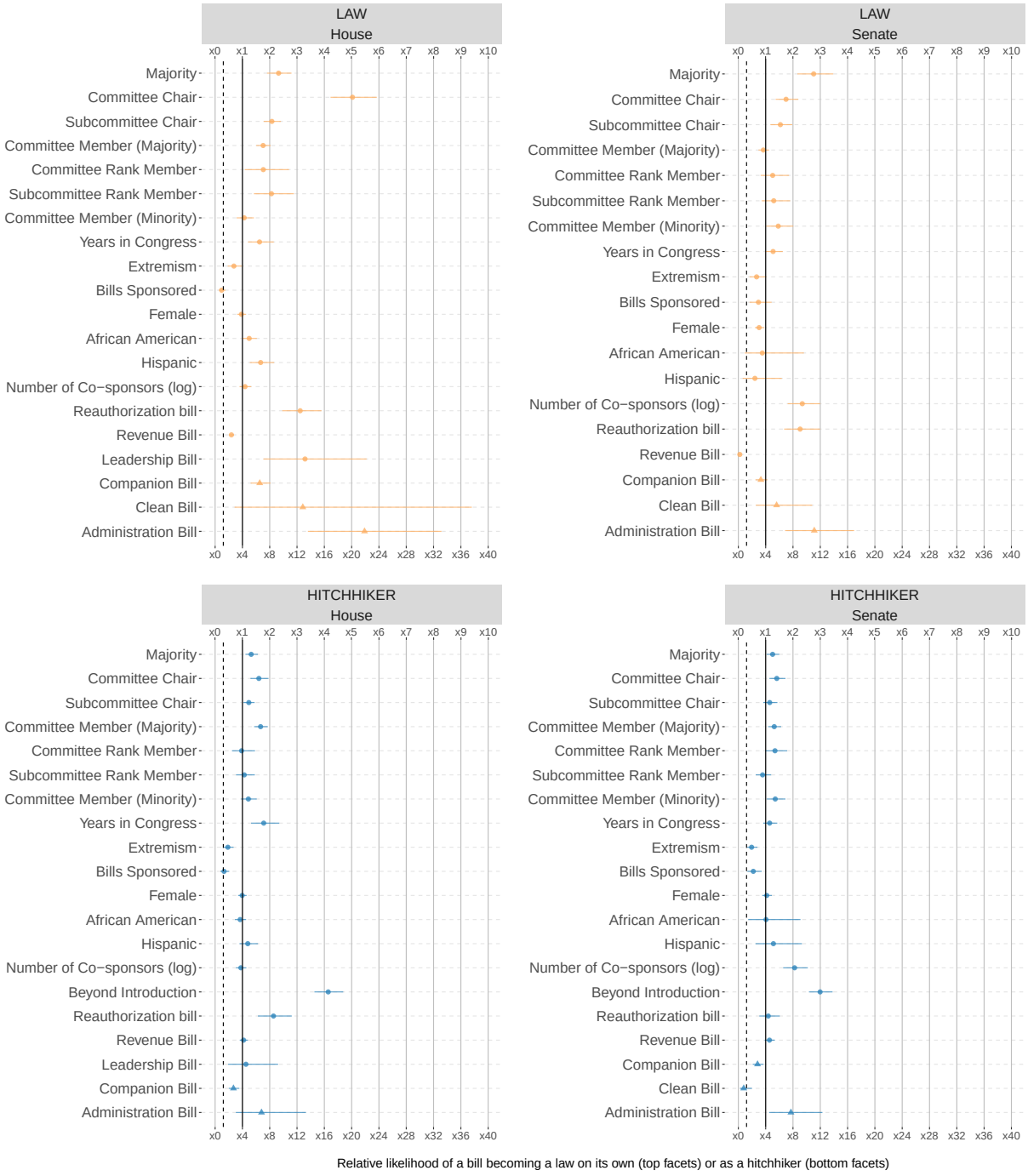

Figure 10: Marginal effects of sponsor and bill characteristics on law *versus* hitchhiker success.