Measuring Agenda Setting in Interactive Political Communications*

Erin Rossiter
Department of Political Science
Washington University in St. Louis

ABSTRACT

While strategies exist to measure actors' efforts to set policy, media, and lawmaking agendas, political scientists lack a method for identifying and accurately measuring another form of agenda setting that lies under the surface anytime two people talk. Within interactions, such as debates, deliberations, and discussions, actors can set the agenda by shifting others' attention to their preferred topics. In this article, I use a topic model that locates where topic shifts occur within an interaction in order to measure the relative agenda-setting power of actors (Nguyen et al. 2014). Validation exercises show that the model accurately identifies topic shifts and infers coherent topics. Three empirical applications also validate the agenda-setting measure within different political settings: US presidential debates, in-person deliberations, and online discussions. These applications show that successfully setting the agenda can shape an interaction's outcomes, demonstrating the importance of continued research on this form of agenda setting.
Abstract word count: 149
Manuscript word count: 9969

*Valuable feedback for this project was provided by Political Data Science Lab members at Washington University in St. Louis. I thank participants at VIM 2019, New Faces in Political Methodology XI, Text as Data 2018, APSA 2018, PolMeth 2018, and SPSA 2018 for helpful comments. I am grateful to Christopher Karpowitz, Hans Hassell, Taylor Carlson, Jaime Settle, and Amber Boydstun for generously sharing data. Funding for elements of this project was provided by the National Science Foundation (SES-1558907, SES-1423788, SES-1938811) and the Weidenbaum Center on the Economy, Government, and Public Policy and Washington University in St. Louis.

1 INTRODUCTION

Who sets the agenda during political debates, deliberations, and discussions is fundamental to the study of political discourse but remains an elusive concept to quantify. While strategies exist to quantify how issues make their way to broader policy and media agendas, we lack a systematic way to measure the agenda setting that occurs when actors *interact* with each other. Yet, measuring agenda setting in political interactions is important because their purpose is often to confine the set of issues relevant for downstream stages of politics. Thus, interactive communications are an ideal venue for actors to advance issues on their personal agendas. In this article, I seek to fill this gap in the study of agenda setting by drawing on research from computer science to measure agenda setting within political interactions. Throughout my applications, I demonstrate the importance of studying agenda setting in debates, deliberations, and discussion by showing that it can shape the outcomes of these political interactions.

Studying agenda setting in interpersonal, interactive communications is important for both formal and informal political settings. Presidential debates, Congressional committee hearings, and Supreme Court oral arguments are all examples of formalized interactions that are embedded in the framework of American government to facilitate decision-making. Indeed, research shows these interactions have consequences on political outcomes, such as the ability of a lawyer's oral argument to sway Supreme Court votes (Johnson, Wahlbeck and Spriggs 2006). Beyond these formal settings, informal interactions are also ubiquitous in politics, for example, as lawmakers and lobbyists interact during the process of crafting and voting on legislation. Citizens also experience politics via informal interactions with others, and research shows that engaging in political conversation can influence subsequent behaviors like vote choice (Beck et al. 2002).

Because political interactions are an opportunity to set the stage for subsequent decisions, outcomes, and behaviors, I expect actors will seek to exercise power during them. It has long been understood that agenda setting is fundamental to understanding the political process (e.g., Schattschneider 1960; Cobb and Elder 1972; Kingdon 1995), and an important source of political power (e.g., Bachrach and Baratz 1962). And in this article, I consider how agenda setting is one form of power actors can exercise with political interactions, as well.

Specifically, actors can set the agenda during an interaction through attempts to shift the discussion to preferred topics. Like familiar forms of agenda setting in the literature, power is derived from shaping what issues do (and do not) receive attention by others. But unlike previously studied venues for agenda setting, interactions are uniquely a social experience. When engaging in a debate, deliberation, or discussion, a complex social exercise is underway as actors negotiate who is speaking and what is being discussed in real-time. Actors, therefore, have the opportunity to influence the agenda in these settings.

Take, for example, an interactive setting in which the fight over the agenda is particularly evident—United States presidential debates. Candidates seek to set the agenda, or to shift the debate toward topics they "own" (Petrocik 1996; Boydstun, Glazier and Pietryka 2013). The following lines from the first 2016 general election presidential debate between Donald Trump and Hillary Clinton, with moderator Lester Holt, demonstrate Clinton shifting the agenda to a preferred topic.

> *Holt:* We are at—we are at the final question.
>
> *Clinton:* Well, one thing. One thing, Lester.
>
> *Holt:* Very quickly, because we're at the final question now.
>
> *Clinton:* You know, he tried to switch from looks to stamina. But this is a man who has called women pigs, slobs and dogs, and someone who has said pregnancy is an inconvenience to employers, who has said...

Holt, as the moderator, tried to introduce his final debate question. Yet Clinton, in real-time, overpowered his efforts and successfully steered discussion during the final minutes of the debate toward an issue on *her* agenda—Trump's history of degrading women. Clinton's skill at setting the agenda not only affected what was discussed during the debate, but it also influenced subsequent media converge. As discussed below, news outlets reported that this exchange was a memorable moment from the first debate (Ross 2016; Mason 2016).

Although it is easy to see why measuring agenda setting in political interactions is important, standard approaches in political science offer no way to quantify Clinton's role in setting the debate's agenda. And more broadly, political scientists lack a systematic way to measure the relative agenda-setting power of actors in interactive settings. To be sure, political scientists have quantified other features of interactions. A common approach is to count easily observable quantities, such as the number of words spoken by a participant. Scholars have applied these measures to citizen deliberations, (e.g., Karpowitz, Mendelberg and Shaker 2012) legislative committee hearings (e.g., Kathlene 1994), and Supreme Court oral arguments (e.g., Epstein, Landes and Posner 2010). While these count-based measures are useful for studying participation patterns, they fall short when the goal is to measure what issues make it on the agenda and which actors succeeded in getting them there.

In this article, I build upon these previous efforts in political science to quantify how political actors interact with each other, and I focus my efforts on agenda setting. Specifically, I leverage the text of interactions as data, and I use the parametric Speaker Identity for Topic Segmentation (SITS) model from the computer science topic segmentation literature (Nguyen et al. 2014; Nguyen, Boyd-Graber and Resnik 2012). SITS extends Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan 2003), a topic model used widely in political science, to simultaneously estimate three sets of latent quantities of interest: what topics are on the agenda, where shifts in the agenda occur, and each actor's agenda-setting power (Nguyen 2015).

I proceed by first comparing agenda setting within political interactions to previously studied forms of agenda setting by the media or policymakers. I then outline SITS and present three validation exercises to show that the model can accurately identify where shifts in the agenda occur and can infer coherent topics. Then I use SITS to investigate agenda setting in three contexts. First, I assess agenda setting in electoral debates. I use SITS to replicate and extend findings in the literature about which candidates include the economy on their agendas during presidential debates (Boydstun, Glazier and Pietryka 2013; Vavreck 2009). Then, I assess how agenda setting relates to several common measures of participation using deliberation texts. I find that deliberators who set the agenda are more likely to shape the deliberation's outcome, but that this relationship does not hold with the participation measures. Lastly, I perform a more direct test of my central claim that agenda setting can shape the outcomes of political interactions, and I show that agenda setting correlates with achieving one's desired outcome in an online discussion. Taken together, these applications provide evidence for the validity of the agenda-setting measure across debates, in-person deliberations, and online discussions. I conclude with a discussion of the usefulness of studying agenda setting in interactions as an important form political power, and I provide suggestions for further areas of inquiry enabled by this method.

2 CONCEPTUALIZING AGENDA SETTING IN INTERACTIONS

Before introducing the SITS model of agenda setting in Section 3, I will first discuss the broader study of agenda setting in political science, how agenda setting manifests in interactive communications, and a framework to motivate measurement of this important concept.

As mentioned, the concept of agenda setting is important to several political science literatures. For example, scholars investigate how the media's agenda setting influences what issues the mass public perceives as important (e.g., McCombs and Shaw 1972). Moreover, there is a vast literature identifying the power of the media, Congress, and the President to set the policymaking agenda, and how issues rise to and fall from this agenda (e.g., Baumgartner and Jones 1993; Kingdon 1995). Scholars also investigate formal agenda-setting processes of specific institutions, such as how legislative parties seek control over which bills are considered on the floor by gaining agenda-setting powers through Congressional offices (Cox and McCubbins 2005) or how Supreme Court justices vote strategically on which cases are granted review (Black and Owens 2009). Regardless of the specific literature, "agenda setting" can be thought of as an influence over the set of issues that are (and are not) receiving attention, which in turn can influence the set of issues relevant for downstream outcomes and decision-making. As such, agenda setting has been viewed as an important source of political power (e.g., Bachrach and Baratz 1962).

Building on these various traditions, I argue agenda setting is one form of power actors seek to exercise during one of the most basic, ubiquitous political activities—talking with others. Like other forms of agenda setting, agenda setting within interactions is an influence over the set of issues that are receiving attention (i.e., being discussed) by others. However, unlike other forms of agenda setting which pertain to broad agendas that change over long periods of time, agenda setting in this article pertains to an actor's *immediate* efforts to shift the *specific* topics of discussion.

Also consistent with several agenda-setting traditions, agenda setting is a source of power within interactions because of its two-fold impact. First, agenda setting leads to an immediate control over what is being discussed. And because of this, agenda setting can shape subsequent outcomes and decision-making (e.g., Schattschneider 1960; Riker 1986). For example, successfully setting the agenda in a deliberation might result in keeping certain issues off the table, which then obviates the risk of the issue rising to a vote. Or, conversely, effective agenda setting during a debate might raise the status of an otherwise overlooked issue, which then shapes what the media reports the next day.

However, an analogy to previously studied forms of agenda setting is limited when it comes to how, exactly, actors set the agenda during an interaction. That is because interactions are uniquely a social game. As actors navigate the waters of a social setting, who is speaking and what is being discussed are negotiated in real-time. This informal negotiation of the agenda presents an opportunity for an actor to set the agenda—to gain the floor, introduce their preferred topic, and maintain others' attention on it.

To be sure, what agendas and agenda setting looks like will vary depending on whether the interaction is an adversarial debate or a thoughtful policy deliberation, which have different goals and norms. For example, in US presidential debates, two candidates may seek to shift attention to two distinct sets of issues (e.g., Petrocik 1996; Vavreck 2009). However, in a focused policy deliberation, actors may seek to advance competing framings of only a few issues. Either strategy—shifting attention to a new issue, or shifting attention to new attributes of an issue—can constituent agenda setting depending on the purpose of the interaction and breadth of discussion.[1]

Relatedly, it is important to note that agenda setting is just one form of power available to actors within political interactions. Actors may begin an interaction with pre-existing power in the form

[1]Considering both issues and frames as a part of agenda setting parallels theory on the media's "first" and "second" level agenda setting, where the media's emphasis on both issues (first level) and the attributes (e.g., frames) of those issues (second level) as agenda setting processes (Weaver 2007, e.g.,).

of a status differential (e.g., Karpowitz, Mendelberg and Shaker 2012). Or, actors may seek power beyond agenda setting during the interaction through persuasive arguments (e.g., Wang et al. 2017) or through other heresthetical stategies (Riker 1986). Agenda setting may not be the form of power most useful to every actor in every situation. Measuring other sources of power within interactions, and how they may or may not relate to agenda setting, are left to future research.

2.1. *Prior Approaches*

While agenda setting is an important concept for understanding political interactive communications, approaches currently used in political science to analyze interactions are not well-suited for systematically studying this concept. First, political scientists have used hand-coding methods to analyze interactions, including hand coding efforts to measure agenda setting (Boydstun, Glazier and Phillips 2013; Boydstun, Glazier and Pietryka 2013). While hand-coding is often considered a gold standard approach, it can have significant weaknesses. First, recruiting, adequately training, and compensating the work of research assistants can be prohibitively time-consuming and costly, especially with a large corpus. Second, research shows that even high quality coders can provide estimates that are unreliable (Mikhaylov, Laver and Benoit 2012).

Quantitative analysis of the text of interactions is a more popular approach. Scholars often count directly observable and quantifiable behaviors such as the number of turns taken or words spoken by participants (e.g., Kathlene 1994; Epstein, Landes and Posner 2010; Karpowitz, Mendelberg and Shaker 2012). While this approach measures participation, itself an important concept in the study of representation in group deliberations, count-based measures are limited when the goal is to assess an interaction's agenda and who is influencing it.

To be sure, automated text analysis has been applied to measure the concept of agenda setting (Quinn et al. 2010; Eggers and Spirling 2016). However, existing methods are not suited to studying agenda setting behavior of actors in *interactions*. Quinn et al. (2010) conceptualize the agenda as what issues are broadly gaining attention in the political arena and which are not. As a macro-level measure of the agenda, it is not equipped to measure the micro-level agenda setting within a political interaction. Moreover, Eggers and Spirling (2016) conceptualize the agenda as the relative importance placed on issues over over months and years, and propose a measure of an actor's influence on this long-term agenda. Thus, this measure is not a good fit for the task at hand, as interactions require an immediate negotiation of the agenda as the communication unfolds.

2.2. *Conceptual framework*

Despite the importance of agenda setting within interactive communications, political scientists currently lack an accurate and systematic means to measure it. In this section, I consider each speaker's agenda-setting ability, where shifts in the agenda occur, and the topical agenda itself as latent quantities of interest. And in Section 3, I build on this conceptual framework to outline a measurement strategy for these concepts using the text of the interactions as data.

Ideally, to assess agenda setting in interactions, we would be able to compare each actor's *preferred* distribution over topics to the *actual* distribution over topics that arises from interaction. If an actor's preferred topics completely aligned with the discussed topics, we would know the discussion's agenda unfolded in their favor. However, this underlying concept of interest is not directly observable. Even if the set of discussed topics were clear, it is impossible to know each actor's preferences over the attention devoted to each topic in the universe of topics. Therefore, I necessarily step away from this ideal understanding of agenda setting, and in this section, I discuss how I operationalize this concept by examining *who* is setting the agenda.

While we can not know if a discussion aligns with an actor's preferred discussion topics, we can gain an understanding of each actors' ability to shift topic. This approaches the ideal concept of agenda setting discussed above because presumably when an actors shifts to a new topic, they shift to a preferred topic. Therefore, in order to measure agenda setting, I treat whether an actor can shift the set of topics receiving attention by others as an indicator of their agenda-setting ability.

As an important aside—I consider agenda setting as a latent ability of an actor that when used in practice is a form of power, as discussed in Section 2. But, introducing a topic shift is not in and of itself power. Rather, agenda-setting power is also evidenced by others' attention on one's shifted-to topics (i.e., others *not* shifting topic themselves). Because actors are in competition with each other, seeking to advance their preferred distribution of discussion topics, agenda-setting power is inherently a relative measure. Therefore, an actor's agenda-setting power ought to be considered relative to the power of others who are fighting over the same agenda, within the same limited time.

To gain an understanding of each actors' agenda-setting ability, I discussed the need to locate where shifts in the agenda occur as an interaction unfolds. Because an interaction can be thought of as a sequence of different actors taking turns speaking, I focus on each of these separate turns. Since each speaking turn could shift the set of topics receiving attention (or could stay on topic), I treat this a binary latent variable. To be clear, I consider topics shifts as something we can not directly observe, but rather, something that needs to be inferred from the textual data. Shifts in topic are not directly observable because, in part, we simultaneously need to have a clear idea of the set of topics being discussed in the corpus.

Therefore, I consider an interaction's agenda—the set of issues and/or frames that arise during the interaction—as an additional latent quantity of interest. To infer the set of "topics" that make up the agenda, I adopt a similar strategy as prior research by using unsupervised topic model of the text of the interactions to explore the issues (e.g., Grimmer 2009) and/or frames (e.g., Aslett et al. 2020) within a text corpus.

In sum, we cannot ascertain an ideal sense of who has power over a discussion's agenda by comparing an actor's preferred discussion topics to how a discussion unfolds. However, identifying who can shift topic and maintain attention on their introduced topics is one way to operationalize agenda setting. But, understanding who has changed the topic requires we know the set of topics being discussed and where they shift. Therefore, these three interrelated concepts all need to be inferred simultaneously from the text.

3 A MODEL OF AGENDA SETTING IN INTERACTIONS

I measure agenda setting within interactions using the parametric Speaker Identity for Topic Segmentation (SITS) model (Nguyen, Boyd-Graber and Resnik 2012; Nguyen et al. 2014). Specifically, SITS builds upon a familiar topic model in political science, Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan 2003), to account for and measure the latent concepts motivated in Section 2.2: the agenda, where shifts in the agenda occur, and the agenda-setting power of actors

Figure 1: SITS data generating process

- **For each speaker $m \in [1, M]$, draw a speaker topic shift probability $\pi_m \sim$ Beta(γ).**
- For each topic $k \in [1, K]$, draw a topic-word distribution $\phi_k \sim$ Dir(β).
- For each turn $t \in [1, T_d]$, in each discussion $d \in [1, D]$ **(with speaker $a_{d,t}$)**:
  - **If $t = 1$, set the topic shift $l_{d,t} = 1$, otherwise draw $l_{d,t} \sim$ Bernoulli($\pi_{a_{d,t}}$).**
  - **If $l_{d,t} = 0$, set the topic distribution $\theta_{d,t} \equiv \theta_{d,t-1}$, otherwise** draw $\theta_{d,t} \sim$ Dir(α).
  - For each word index $n \in [1, N_{d,t}]$:
    - Draw a topic $z_{d,t,n} \sim$ Categorical($\theta_{d,t}$).
    - Draw a word $w_{d,t,n} \sim$ Categorical($\phi_{z_{d,t,n}}$).

Note: Figure adapted from Nguyen et al. (2014). The underlying data generating process of the parametric Speaker Identity for Topic Segmentation model. Bold text indicates extensions from Latent Dirichlet Allocation.

within an corpus of interactions. For ease of exposition, as I outline the model, I will refer to any single interaction as a "discussion," each actor participating in a discussion as a "speaker," and each uninterrupted utterance by a speaker as a "speaking turn."

3.1. *SITS data generating process*

The SITS data generating process is outlined in Figure 1. Since SITS follows in a line of research that extends topic models to estimate additional latent quantities of interest to political scientists (e.g., Grimmer 2009), I use bold text to denote extensions to LDA.[2]

First, for each speaker $m \in [1, M]$, their agenda-setting ability within the corpus (π_m) is drawn from a symmetric Beta distribution with parameter γ. Then, as with LDA, topics (ϕ_k), or probability distributions over the corpus vocabulary, are drawn for each of $k \in [1, K]$ from a symmetric Dirichlet distribution with parameter β. Then, a distribution over topics ($\theta_{d,t}$) needs to be drawn for each speaking turn $t \in [1, T_d]$ for each discussion $d \in [1, D]$. However, this part of the SITS generative process unfolds differently than LDA. That is because SITS seeks to find sequences of speaking turns on the same set of topics, or segments. Since the first turn of a discussion inherently changes the topic, this is noted by setting a turn-level topic shift binary variable equal to one ($l_{d,t=1} = 1$). For all other turns, whether or not a shift in topic occurs is drawn from a Bernoulli

[2]To make the data generating process of LDA comparable to a text arising from an interaction, consider what is usually referred to as a "document" as single speaking turn in an interaction. Alternatively, this could be defined at the discussion-level, however this would disregard the interactive nature of discussions by concatenating all speaking turns into one instance of text.

distribution parameterized by the speaker's agenda-setting measure ($\pi_{a_{d,t}}$, where $a_{d,t}$ is the observed speaker of the speaking turn). Therefore, whether or not a speaking turn changes the topic is influenced by its speaker's latent agenda-setting ability. If a topic change is indicated, a *new* topic distribution is drawn from a symmetric Dirichlet distribution with parameter α. Otherwise the topic distribution from the previous turn *carries over* to the current turn ($\theta_{d,t} \equiv \theta_{d,t-1}$) indicating those speaking turns belong to the same segment. Then, identical to LDA, for each word index $n \in [1, N_{d,t}]$ in the speaking turn, a topic assignment ($z_{d,t,n}$) is drawn given the speaking turn's distribution over topics and a word ($w_{d,t,n}$) is drawn given its assigned topic.

Note that the agenda-setting measure π_m is a speaker-level quantity. It describes the propensity of a speaker to shift topic when speaking. As such, this measure captures the theoretical quantity of interest as it is not simply a count of how often a speaker shifts topic (think of this as an "uptick" in the numerator). It also accounts for how often a speaker is willing to maintain attention on others' topics (think of this as an "uptick" in the denominator). Therefore, this measure is most meaningful when comparing the relative agenda setting abilities of actors who are seeking to influence the same agenda in the same limited amount of time, as motivated in Section 2.

To estimate SITS models in all simulations and applications that follow, I use a Gibbs sampler written in Java by Viet An Nguyen that is available to the public (Nguyen 2014). Appendix A provides additional details regarding the sampler, Appendix B details my pre-processing decisions for each corpus guided by metrics and tools in the text analysis literature (Denny and Spirling 2018), and Appendix C details my approach to choosing hyperparameter values by relying on extent advice in the unsupervised topic modeling literature.

4 VALIDATION EXERCISES

Before exploring applications of agenda setting across different political interactions, I present results from three validation exercises. The first exercise provides evidence that SITS can accurately identify where *latent shifts* in the agenda occur. Second, using crowd-sourced human judgments, I validate that SITS can infer semantically meaningful *latent topics*. Finally, I assess the interrelated nature of where shifts in topic occur and the topics themselves by examining the resulting *segments* of an interaction. Using crowd-sourced human judgments, I find that SITS segments are viewed as more coherent than segments derived from a hand-coding approach.

4.1. *Latent topic shifts*

Texts were generously shared by Jaime Settle and Taylor Carlson from a study conducted in the fall of 2015 examining disagreeable political discussion. Participants had an in-person discussion on several topics with a partner for approximately 10 minutes.[3] Participants read a prompt on a screen and discussed the prompted topic for a pre-specified length of time. At that point, the screen prompted the participants to stop discussing and wait for the next topic, producing sharp shifts between topics with known locations.[4]

This study contains 70 discussions among 140 participants. The conversations were an average of 38 turns long. I preprocessed the text by removing numbers, stemming, and removing infrequent terms. I also transformed all features to lower case and removed all punctuation. I estimated three SITS chains from the data with randomly drawn starting values. I averaged the posterior mean for each turn-level variable across the three chains.[5]

To assess if SITS can accurately identify where latent shifts in topic occur, I classify a speaking turn as shifting topic if the posterior probability of a shift is greater than or equal to 0.50. I then compare where shifts were inferred by SITS to where topic changes were prompted by the researchers. SITS identifies 81.4% of the locations, within two speaking turns, that begin a new prompted topic segment by the researcher.[6] I check for a SITS-inferred topic shift within two speaking turns after a researcher-prompted shift, because often after reading the prompt, the first few speaking turns would simply answer the prompt's question by saying "yes," "no," or "do you want to go first?" Instead of classifying this as a topic shift, SITS would classify the subsequent speaking turn as a shift that actually began discussing the topic at hand.

[3] Researcher provided topics are outlined in Appendix D.

[4] I watched video recordings of each discussion, and the participants complied by discussing the prompted topics.

[5] I estimate SITS with $K = 13$, $\alpha = 1/K$, $\beta = .1$, and $\gamma = 1$, as outlined in Appendix C.

[6] In terms of validation, I have no way to know whether participants shifted topic *within* a researcher prompted topic segment. Participants could, and likely did, change topic when not prompted to. Therefore, for this exercise, I only validate SITS against locations where topic changes were known to occur.

This exercise demonstrates that SITS can accurately identify the speaking turns that should be attributed with shifts in the agenda.[7] Moreover, SITS provides us a more nuanced view of how topics ebbed and flowed in these discussions than if we considered the locations of where the researchers prompted topic changes as ground truth.

4.2. *Latent topics*

Next I assess whether SITS can infer semantically coherent topics. Influential work in computer science proposes that crowd-sourced tasks are more useful than traditional metrics to assess if a topic model returns semantically meaningful and distinct topics (Chang et al. 2009). Therefore, I use the "topic intrusion" task proposed by Chang et al. (2009) to validate the topics from a SITS model I will return to in Section 5. This model is estimated on 20 U.S. general election presidential debates held between 1992-2016.[8]

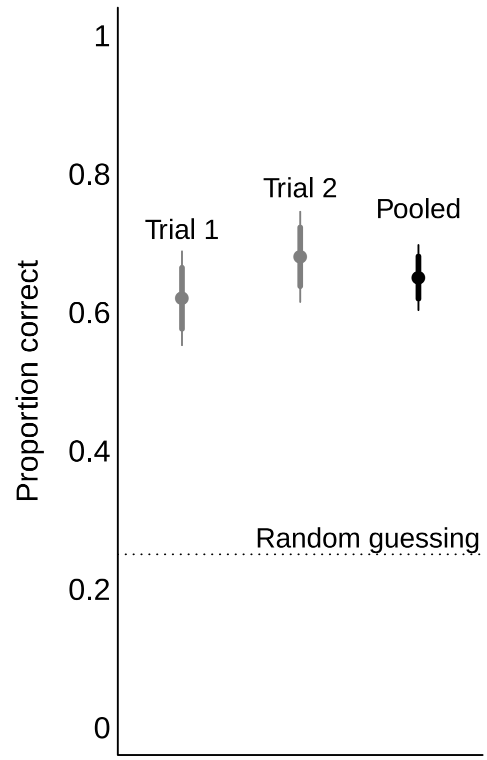
The topic intrusion task presents the human judge with a document, in this case a segment inferred by SITS. The judge is also presented with four word sets. Three of these word sets represent the three highest probability topics for the segment. The fourth word set is the intruder topic, drawn randomly from the segment's low probability topics. Each word set contains the top eight FREX topwords for the topic (Roberts et al. 2014). I set up the topic intrusion task for 200 randomly drawn segments from the debates. Then, Amazon Mechanical Turk workers were asked to choose which word set was most unrelated to the passage. In line with recent work on validation procedures for topic models by Ying, Montgomery and Stewart (2019), I ran two identical trials of the 200 tasks. Figure 2 plots the results for each trial separately as well as the pooled result.

Workers competed 62% and 68% of the tasks correctly in Trial 1 and Trial 2, respectively. A difference of proportions test indicates that Trial 1 and Trial 2 are not significantly different ($p = 0.249$). This result is comparable to one, and better than three, of four models assessed by Ying, Montgomery and Stewart (2019) using the topic intrusion task. In all, human coders and

[7] An additional validation study in Appendix D compares the performance of SITS to the use of standard automated text analysis methods in political science, and shows that these commonly used methods do not perform well when adapted to the task of identifying where shifts in topic occur within an interaction's text.

[8] I estimate SITS with $K = 44$, $\alpha = 1/K$, $\beta = .1$, and $\gamma = 1$, as outlined in Appendix C.

Figure 2: Test of semantically meaningful topics



Note: Proportion of correct answers to the topic intrusion task. Thick line shows 80% confidence interval, and thin line shows 95% confidence interval. Grey bars show the identical repeated trials of 200 tasks each. Black bar represents the result when pooling both trials. A difference of proportions test indicates that Trial 1 and Trial 2 are not significantly different ($p = 0.249$).

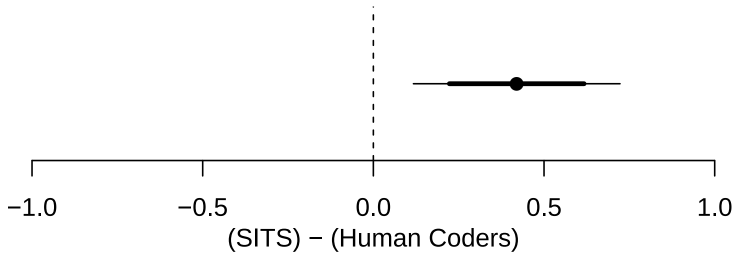
SITS largely agree about which topics are and are not associated with the inferred segments of the debates.

4.3. *Latent segments*

Next, I validate that SITS can identify coherent segments of an interaction. To do so, I follow the Grimmer and King (2011) procedure for evaluating the similarity of segments estimated to have similar topic distributions. Because SITS also infers the boundaries of these segments when estimating where shifts in topic occur, this exercise speaks to the validity of both the latent topics and how well SITS partitions sequences of speaking turns into coherent segments.

Importantly, I evaluate the coherence of topics and segments estimated by SITS against topics and segments assigned using a hand-coding approach. I use hand-coded data from Boydstun,

Figure 3: Coherence of topics and segments inferred by SITS vs hand-coding approach



Note: Difference in the Grimmer and King cluster quality measure between the SITS approach and a hand-coded approach to segmentation and topic assignment. Dot shows point estimate, thick line shows 80% confidence interval, and thin line shows 95% confidence interval.

Glazier and Pietryka (2013) from the 1992, 2004, and 2008 United States general election presidential debates. Boydstun, Glazier and Pietryka hand-code several variables from the debate transcripts, including the topic of each question posed to the candidates and the topic of each phrase in the candidates' responses. Then, they deem a candidate as going "off-topic" and thus, engaging in agenda setting behavior, if the phrase's topic does not correspond to the question's topic.

Comparing the topics and segments inferred by SITS to those derived from hand-coding required five steps. First, I determined where topic changes occurred (and thus, formed segments of the debates) according to each method. Second, I determined the similarity of these segments according to each method's topic assignments.[9] Third, I conducted the crowd-sourced exercise outlined by Grimmer and King. Separately with the segments from the SITS and the hand-coding approaches, I drew 25 random pairs of segments from the same topic and 25 random pairs of segments from different topics.[10] Fourth, four unique Amazon Mechanical Turk workers rated the similarity of the segments within each pair on a three point scale of (1) unrelated, (2) loosely related , or (3) closely related.[11] The measure of each method's "cluster quality" is then "the average similarity of pairs of documents from the same cluster minus the average similarity of pairs of documents from different clusters, as judged by human coders one pair at a time." (pg 5 Grimmer and King 2011).

[9] Appendix D details the specific steps taken to (1) estimate shifts in topic and (2) cluster the resulting segments by topic.

[10] Appendix D shows that the results are not likely to be due to differences in the length of the segments derived by the two approaches.

[11] Appendix D provides details of the task's instructions and an example of a pair of segments.

The fifth and final step is to use difference in means to compare the two methods. Figure 3 plots this point estimate along with the 80% (thick line) and 95% (thin line) confidence interval. The estimated difference between approaches is positive and significant. Therefore, the Grimmer and King (2011) evaluation suggests SITS can infer segments that are even more coherent than those derived from a hand-coding approach.

5 APPLICATIONS

I next explore agenda setting within three different interactive political communications. First, with U.S. presidential debates, I show that SITS can be used to test theories about the topics actors should favor in electoral debates. Second, using in-person deliberations, I shift emphasis from the topical agenda to who can set the agenda. I validate that participants who SITS identifies as setting the agenda indeed shift attention to their ideas, and do not do so by simply interrupting or out-talking others. And third, using a novel online discussion study, I show that agenda setting can shape the outcomes of political interactions.

5.1. *Electoral debates*

I first assess the contribution of SITS to the measurement of agendas and agenda setting in electoral debates. Boydstun, Glazier and Pietryka (2013) examined how theories of agenda setting throughout presidential campaigns may translate to agenda setting within a debate (see also Boydstun, Glazier and Phillips 2013). This research relied on a rigorous hand-coded content analysis, and therefore reasonably analyzed only three elections. I use SITS to replicate and extend their findings regarding which candidates' agendas feature the economy by analyzing all general election presidential debates from 1992-2016.

Vavreck (2009) crafts a typology to describe when presidential candidates should focus their campaigns on the economy. Clarifying candidates are those that benefit from doing so (outparty candidates in a bad economy and inparty candidates in a good economy), while insurgent candidates do not (outparty candidates in a good economy and inparty candidates in a bad economy).

Boydstun, Glazier and Pietryka assess how these patterns manifest in the context of presidential debates. As discussed in Section 4.3, Boydstun, Glazier and Pietryka hand code the 1992, 2004, and 2008 debates to assess two general hypotheses. In terms of Vavreck's typology, the two hypotheses are that the clarifying candidate should (1) talk more about the economy than the insurgent candidate and (2) focus more of their agenda-setting efforts on the economy than insurgent candidate.

I preprocessed the text by removing capitalization, punctuations and numbers. I also remove a set of stopwords that included common English stopwords, annotations to the transcripts (such as "applause"), and the name and titles of all speakers. I also perform stemming and remove infrequent terms.[12] The resulting corpus contained 994 unique terms used across 3,818 speaking turns in the 20 debates. I estimated three SITS chains from the data with randomly drawn starting values.[13] I use iterations from all three chains to estimate posterior means of turn-level topic shifts, and I use FREX topwords to describe topics from the best performing model.

Figure 4 describes patterns of how much clarifying and insurgent candidates discussed the economy during the debates. The barplots show topic proportions for the "economy" topic for all of a candidate's speaking turns in (a) and all of the turns in which they shifted topic in (b). [14] The dashed line represents the amount the economy would be discussed if all estimated topics were discussed equally. I use Vavreck's classification of candidates as clarifying or insurgent for the 1992-2008 elections (p. 38 Vavreck 2009). Across those elections, we see support for the hypotheses that clarifying candidates talk more about the economy in (a), and that they engage in agenda setting in order to shift the course of the debate to the economy in (b).

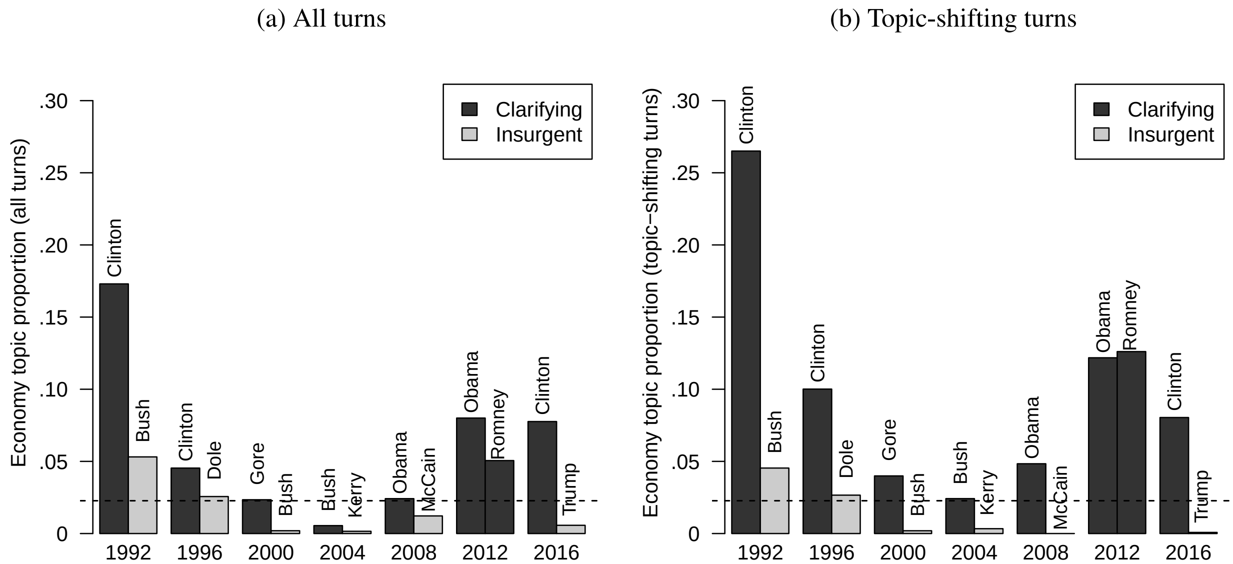
In particular, the results for 1992, 2004, and 2008 are in line with results from the hand-coded

[12]Appendix B discusses the preprocessing in detail. Because preprocessing decisions may have consequences for downstream, substantive results, I use the `preText` R package to assess this (Denny and Spirling 2018). I find that results may be sensitive to removing a common set of stopwords, so I replicate all results when retaining stopwords in Appendix E.

[13]I estimate SITS with $K = 44$, $\alpha = 1/K$, $\beta = .1$, and $\gamma = 1$, as outlined in Appendix C. Appendix E provides details on convergence and model selection.

[14]From the FREX topwords, I judged which topic(s) were on the economy. In this case, I found one topic distinctly on the economy—invest, econom, economi, growth, grow, creat, unemploy, trickl. Aggregating across all debates, this was the topic most discussed when candidates set the agenda by shifting topic. Appendix E presents FREX topwords and aggregate topic-shifting patterns for all 44 topics.

Figure 4: Clarifying candidates talk more about the economy



Note: Topic proportions for all of a candidate's speaking turns in (a) and all of the turns in which they shifted topic in (b). Dashed line represents if they talked about all topics equally. Clarifying candidates discuss the economy more than insurgent candidates, and engage in agenda setting to shift the course of the debate to the economy. Difference of proportions tests indicates that all within election differences are distinct ($p < .05$), except for topic-shifting turns in 2012.

content analysis. Boydstun, Glazier and Pietryka (2013) note that in 1992 and 2008, the economy was a salient issue, but in 2004, defense was uniquely more important to the public than the economy. These patterns hold in Figure 4, as we see the 2004 election indeed discussed the economy less than any other election.

Finally, the 2012 and 2016 elections were both outside the preview of the Boydstun, Glazier and Pietryka (2013) and Vavreck (2009) analyses. In 2012, both candidates were stressing the economy and running what looked like a clarifying campaign (Sides and Vavreck 2014). We see that pattern holds in Figure 4 as the candidates shifted attention to the economy in nearly equal amounts. Finally, in 2016, Clinton would be considered the clarifying candidate (the inparty candidate in a growing economy), even if the economic concerns were not the main campaign issue or motivation for voters (Sides, Tesler and Vavreck 2019). We see she also spoke about the economy much more than Trump—both in total and when she was setting the agenda.

5.2. *In-person deliberations*

To date, political scientists have lacked a systematic way to measure the agenda-setting features of interactions. Instead, participation is often measured in these contexts using count-based measures, such as the number of words spoken by a participant. However, participating in an interaction is a necessary but not sufficient component of setting its agenda. Setting the agenda requires participation that also shifts the set of topics currently receiving attention. Therefore, I assess the relationship between status-quo, count-based measures of an actor's participation and the proposed measure of agenda setting. Then, I examine how a participant's participation and agenda setting correlate with successfully shaping the outcome of the deliberation. I find that agenda setting positively correlates with shaping the deliberation's outcome, whereas I fail to find any correlation with the between the outcome and participation measures.

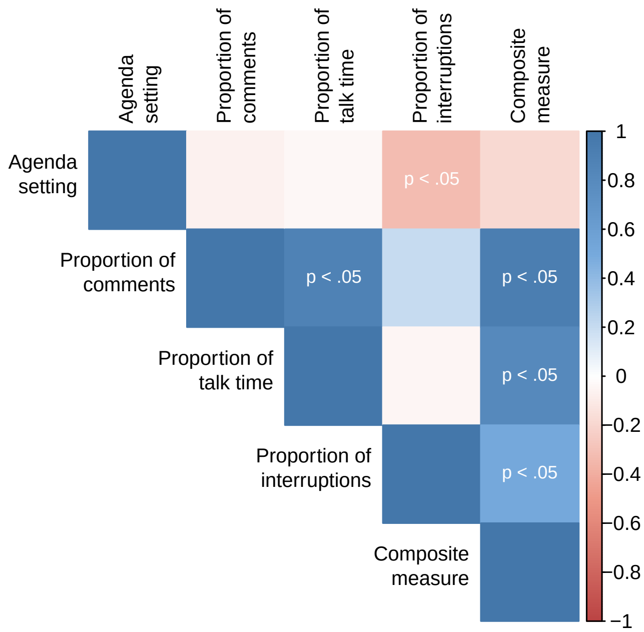
Deliberation texts were generously shared by Christopher Karpowitz and Hans Hassell from a pilot study conducted in June of 2016 examining the effect of stress on deliberation participation.[15] For this study, Brigham Young University (BYU) students were recruited to discuss the BYU Dress and Grooming Standards (DGS), a specific set of rules governing the appearance of all students and staff at the university.

The study included ten discussion groups, each composed of four members. Participants first completed a pre-discussion survey on their attitudes regarding the DGS. Participants then engaged in a discussion where they had 25 minutes to agree upon any recommended changes to the standards. Participations voted on their group's recommendations after the discussion. Recommendations with a majority of the post-discussion votes would be sent to the Honor Code office, with no guarantee that the changes would be implemented.

While these deliberations did not involve a conventional policy issue, this application is a useful case study as it involves an important issue for participants with relatively strong and conflicting attitudes. Students have petitioned and recently protested the Honor Code and how it is enforced, including the DGS portion, with their efforts even making national news (Turkewitz 2014; Levin

[15] I find no evidence of a treatment effect; therefore, I do not include the treatment as a variable in subsequent analyses.

Figure 5: Agenda setting does not correlate with quantity of participation



Note: Visualization of correlation matrix. Significant correlations noted with $p < .05$.

2019). Moreover, not only do pre-treatment survey responses suggest participants held conflicting views, but participants often expressed these conflicting perspectives during the deliberation. While the implications of this application are unclear for deliberative settings which may involve participants with weaker attitudes and a lower-stakes outcome, I will turn to this setting in Section 5.3.

I preprocessed the text removing capitalization, punctuations and numbers. I also perform stemming and remove infrequent terms.[16] The resulting corpus contained 667 unique terms used across 899 speaking turns in ten deliberations. I estimated three SITS chains from the data with randomly drawn starting values.[17] I use iterations from all three chains to estimate posterior means of the agenda setting measures.

I first investigate how agenda setting may correlate with commonly used participation measures. Figure 5 visualizes a correlation matrix between agenda setting and three count-based par-

[16]See Footnote 12. I find that results may be sensitive to removing stopwords, so I replicate all results when retaining stopwords in Appendix F.

[17]I estimate SITS with $K = 48$, $\alpha = 1/K$, $\beta = .1$, and $\gamma = 1$, as outlined in Appendix C. Appendix F provides details on convergence.

ticipation measures: proportion of the group's (1) comments made, (2) speaking time used, and (3) interruptions made by a participant. I also include a composite measure created by averaging a standardized version of each count-based measure. Significant correlations are noted with $p < .05$.[18] First, there is no evidence of a correlation between the first two count-based measures based on amount of participation. Second, we see a negative correlation between how often a participant interrupts others and their agenda setting. Taken together, these results suggest agenda setting is not achieved by out-talking or interrupting others. Finally, we see that if aggregating the count-based measures does not provide additional leverage for measuring agenda setting.

To further explore agenda setting and its impact in these deliberations, I assess how agenda setting relates to the deliberation's outcome—the group's DGS policy recommendations. For each deliberation, I note what was eventually written down as a group recommendation, and I then trace the recommendation back to which participant introduced it during the deliberation.[19] I then examine if agenda setting and participation are correlated with shaping the deliberation's outcomes.

Table 1 presents coefficients from logistic regression models with clustered standard errors at the group level in parentheses. Additionally, I include a variable indicating group membership to estimate an intercept shift for each group. The outcome of each model is introducing one of up to two ideas included as a group policy proposal ($y = 1$) or not ($y = 0$). Each model assesses the correlation between the outcome and a count-based measure of participation or agenda setting, all scaled to range between 0 and 1. It may be the case that those with strongly held DGS attitudes prioritized the outcome of the deliberation more than others and for that reason were more interested in seeking their preferred outcome. Therefore, I also control for pre-treatment DGS attitude strength in the models by creating an indicator for strong DGS attitudes.[20]

First we see the coefficients on each count-based measure is not distinct from zero in Models 1-4. Therefore, I find no evidence of a correlation between the participation measures and success in

[18] Appendix F presents full correlation coefficients.

[19] I do this exercise blind to the agenda-setting measures.

[20] Before the discussion, participants were asked to rate their agreement with 23 questions regarding the purpose and fairness of the Dress and Grooming Standards on a seven point scale. I used these questions to create an additive index of pre-treatment DGS attitude strength, then consider a participant to have "strong" attitudes if they fell at or below (above) the first (third) quantile.

Table 1: Agenda setters more likely to shape deliberation outcome

	Dependent variable: Introduced group proposal				
	(1)	(2)	(3)	(4)	(5)
Proportion of comments	2.59				
	(2.18)				
Proportion of talk time		3.76			
		(2.24)			
Proportion of interruptions			−3.14		
			(2.67)		
Composite measure				1.97	
				(1.18)	
Agenda setting					6.00*
					(2.35)
Strong DGS attitudes	1.50	1.23	0.933	1.61	0.867
	(0.839)	(0.993)	(0.924)	(0.956)	(1.03)
Group indicators	✓	✓	✓	✓	✓
Constant	−2.77	−2.94	0.047	−2.60	−2.90*
	(1.51)	(1.35)	(1.37)	(1.82)	(1.41)
Observations	40	40	40	40	40
AIC	68.585	64.936	69.021	69.587	63.703

Note: *p<0.05. Coefficients from logistic regressions with clustered standard errors at the discussion-group level in parentheses. Dependent variable is introducing one of up to two ideas included as a group policy proposal ($y = 1$) or not ($y = 0$). Explanatory variables are those from Figure 5, all scaled to range from 0 to 1.

including one's ideas as policy recommendations. However, the positive coefficient on the agenda-setting measure in Model 5 suggests that participants who succeeded at setting the agenda were also more likely to have their ideas included as a group recommendation.[21]

The agenda setting measure has potential payoffs for empirical scholars of deliberations akin to this one. For example, it can provide new insight into a core ideal of deliberation—equal

[21]The large agenda setting coefficient makes sense given how I measured the dependent variable—introducing a proposal eventually adopted by the group (the dependent variable) is itself a measure of agenda-setting.

participation—which pertains to all points of view receiving attention (e.g., Fishkin and Luskin 2005). As a measure of attention at its core, the agenda-setting measure proposed here may aid in identifying the conditions under which this deliberative ideal is more or less achieved.

5.3. *Online discussions*

Lastly, I validate that agenda setting is a source of power available to actors when they interact with each other. To do so, I fielded a novel online discussion study. Importantly, I constructed the discussions to feature disagreement, and I incentivize participants with additional compensation to achieve their preferred outcome. Specifically, my goal is to build upon Application 2 and more directly test whether setting a discussion's agenda correlates with shaping its outcome.

The study involved four stages. First, participants took an online pre-discussion survey. During the survey, participants learned about five prominent charities and indicated which charity they would prefer to receive a $1 donation from the researchers.[22] At the conclusion of the pre-discussion survey, participants were asked if they would be willing to return for an optional follow-up task at a specific time within the hour. Second, participants who indicated they were willing to return were randomly assigned a partner who disagreed about which charity should receive the donation.[23] Third, participants returned to the online platform at the pre-specified time and engaged in a ten minute online, written discussion with their assigned partner. Fourth, participants answered a short post-discussion survey.

This study was fielded between April 2016 and June 2020 on Amazon Mechanical Turk (MTurk). Participants were paid $1 for the pre-discussion survey and $3 for returning promptly and completing the follow-up discussion task. Participants also received a $1 bonus payment if their charity was chosen to receive the researcher's donation in the post-discussion survey by *both* participants. The bonus payment was intended to incentivize participants to pursue their preferred outcome. In addition to the bonus payment, the participants were further incentivized to agree because the donation would not be made unless participants indicated agreement in the post-discussion survey.

[22] Appendix G shows charity information given to participants.

[23] Participants answered an open-ended question asking why they chose their preferred charity. Participants were not considered for the discussion stage of the study if this answer was of poor quality.

MTurk is an online labor market increasingly used in social science research to quickly and inexpensively recruit samples (see Berinsky, Huber and Lenz 2012). It is important to consider the use of the MTurk subject pool for this application. First and foremost, ethical concerns have been raised regarding the compensation of MTurk participants. This work's compensation followed Williamson's (2016) guidelines with payment above federal minimum wage at approximately $14 per hour. Additionally, because the task given to participants—deciding on a charitable donation—it is important to note that research has failed to find evidence that monitoring cues influence MTurk participants' donations to charity (Saunders, Taylor and Atkinson 2016). Relatedly, research shows MTurk participants are not receptive to experimenter demand effects in survey experiments (Mummolo and Peterson 2019). Also, prior research shows MTurk participants are more attentive and cooperative than participants from other online samples, such as from Qualtrics or Facebook (Boas, Christenson and Glick 2020), making MTurk a useful subject pool for this discussion task requiring a high level of engagement. A limitation of the MTurk subject pool is their high digital literacy, which may moderate what I observe about how the participants interact with each other online (Munger et al. 2018). However, other online subject pools are not flexible enough for the coordination required of this study—both participants in a pre-specified partnership needed to return at the same pre-specified time for the discussion.

This procedure yielded 91 online deliberations; however, 10 of the deliberations did not reach an agreement. I preprocessed the text by transforming all word to lowercase, stemming, and removing infrequent terms. I also selectively removed punctuation, keeping exclamation marks, question marks, and punctuation meant to mimic emojis. I kept this punctuation because it was an important part of the communication in an online environment.[24] I estimated three SITS chains from the data with randomly drawn starting values.[25] I use iterations from all three chains to estimate posterior means of the speaker-level agenda setting.

The agenda-setting measure in this application had a mean of .382 and standard deviation of

[24]See Footnote 12. Appendix B shows that results are not likely to be sensitive the choice to use these common preprocessing steps.

[25]I estimate SITS with $K = 17$, $\alpha = 1/K$, $\beta = .1$, and $\gamma = 1$ as outlined in Appendix C. Appendix G provides details on convergence.

.158. Recall the agenda setting measure is interpreted as the probability a participant will shift topic when speaking. That these participants, on average, shift topic every third speaking turn makes sense in this context. The ten minute discussions were fairly quick, so participants shifted topic fairly rapidly to makes sure they came to a decision.

Unlike Application 2 where I assessed how agenda setting correlated with shaping the deliberation's outcome by backing out which respondents introduced the proposals, here I am able to more directly assess the discussion's outcome—who's preferred charity was chosen? To test if agenda setting during the deliberation correlates with achieving one's preferred outcome, I conduct a two-sided, paired Wilcoxon signed-rank test with the 81 deliberations that reached agreement. This non-parametric test is useful for the paired data (each deliberation has some who does and does not achieve their preferred outcome). Results suggest that the partner who was more successful at setting the discussion's agenda was significantly more likely to achieve their preferred outcome ($p = .0168$). As a robustness check, I find results of a two-sided, paired t-test are consistent.[26]

To assess the robustness of this finding, I test if other discussion tactics explain who was able to achieve their preferred outcome. Specifically, I assess if the partner who received the bonus and donation to their preferred charity simply spoke first or spoke more. I fail to find that speaking first is associated with achieving one's preferred outcome (two-sided t-test, $p = .274$). Moreover, I fail to find evidence that suggests the number of words used in the conversation was meaningfully different for partners that achieved their preferred outcome and those that did not (two-sided, paired Wilcoxon signed-rank test $p = .190$).

This application provides evidence that agenda setting is a form of power—participants who set the agenda get what they want out of the discussion. While these discussions were largely civil and deliberative, with participants engaging with each others' views, political discussions and comments on social media can often be uncivil (Coe, Kenski and Rains 2014), polarizing (Settle 2018), and even spread misinformation (Anspach and Carlson 2018). SITS could help extend this research by identifying when comments are derailed and what kinds of people are better at shifting

[26]The mean difference between the partner who achieved their preferred outcome and the parter who did not was .065 ($p = 0.0153$), which is .41 standard deviations.

others' attention in a negative way.

6 CONCLUSION

In this article, by considering the role of agenda setting in one of the most basic political actions—talking with others—I've introduced a measure of agenda setting applicable to the countless interactions that occur across the political sphere. Importantly, I validated the agenda setting measure across a diverse set of discursive settings. Debates are oppositional, strategic, and have the goal of identifying a "winner" and "loser." Conversely, ideal deliberations are characterized by collaboration and thoughtful consideration of all perspectives. Online discussions are often informal, even anonymous, and lack body language or other interpersonal cues to guide the communication. I validated SITS in each of these environments.

There are of course limitations to the measure of agenda setting proposed here. First, this measure does not specify if it was *what* the person said or *how* they said it that changed the course of the agenda. Perhaps someone's serious or humorous tone shifted the room's attention to their point, rather than how interesting or persuasive it was—this measure can not distinguish how the shift in topic was achieved.

SITS could be extended to measure additional latent structures within political interactions. For example, the SITS data generating process could be altered to account for a speaker's tendency to bring up similar topics over time. One might suspect Clinton is likely to bring up Russian interference in the election because she shifted to this topic earlier in the debate. If SITS is extended to account for the same person shifting to similar topic distributions, then one could further imagine detecting a latent *coalition* of speakers that shift to similar topic distributions.

The results presented in this article are intended to validate the agenda setting measure and stress its importance to the study of political interactions. Teasing out the theoretical role of agenda setting in interactions is left for future work. One open question is, under what conditions does agenda setting power equate to perceived power? Perhaps others tend to discount a woman's power over the communication when she sets the agenda. Or, depending on the social norms of the setting, setting the agenda could be perceived as rude, controlling, or irritating to others, so exercising agenda setting power may harm an individual's perceived power. Understanding the relationships between agenda setting as a form of power, perceived power, influence, persuasion, and more is left for future work.

Finally, SITS is just one of many automated topic segmentation methods. SITS is particularly useful in an interactive setting because it uses speaker identity to inform the segmentation task. However, other topic segmentation methods have promise for researchers needing to find coherent topic segments in their copora (see Purver 2011). For example, topic segmentation might be useful for researchers trying to segment a large document into more manageable sizes or theoretically useful quantities to then be coded by crowd sourced workers. Topic segmentation methods provide a principled way to approach tasks like these where the text is not already structured in the most theoretically or practically useful way.

This article set out to quantify the agenda setting dynamics that lie under the surface anytime two people talk. Agenda setting in debates, discussions, and deliberations is often overlooked by empirical researchers because we lack of methodological tools suited for these settings. This article helps shift the quantitative study of agenda setting to interactions by offering a technique for measuring this important concept across a variety of settings. The hope is that the measure provides opportunity for additional theoretical development and principled analysis regarding the agenda setting dynamics in political interactions.

6 References

Anspach, Nicolas M and Taylor N Carlson. 2018. "What to believe? Social media commentary and belief in misinformation." *Political Behavior* pp. 1–22.

Aslett, Kevin, Nora Webb Williams, Andreu Casas, Wesley Zuidema and John Wilkerson. 2020. "What Was the Problem in Parkland? Using Social Media to Measure the Effectiveness of Issue Frames." *Using Social Media to Measure the Effectiveness of Issue Frames (April 29, 2020)* .

Bachrach, Peter and Morton S Baratz. 1962. "Two Faces of Power." *American Political Science Review* 56:947–952.

Baumgartner, Frank R and Bryan D Jones. 1993. *Agendas and instability in American politics*. University of Chicago Press.

Beck, Paul Allen, Russell J Dalton, Steven Greene and Robert Huckfeldt. 2002. "The Social Calculus of Voting: Interpersonal, Media, and Organizational Influences on Presidential Choices." *American Political Science Review* 96(1):57–73.

Berinsky, Adam J, Gregory A Huber and Gabriel S Lenz. 2012. "Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk." *Political analysis* 20(3):351–368.

Black, Ryan C and Ryan J Owens. 2009. "Agenda setting in the Supreme Court: The collision of policy and jurisprudence." *The Journal of Politics* 71(3):1062–1075.

Blei, David M, Andrew Y Ng and Michael I Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3(Jan):993–1022.

Boas, Taylor C, Dino P Christenson and David M Glick. 2020. "Recruiting large online samples in the United States and India: Facebook, mechanical turk, and qualtrics." *Political Science Research and Methods* 8(2):232–250.

Boydstun, Amber E, Rebecca A Glazier and Claire Phillips. 2013. “Agenda Control in the 2008 Presidential Debates.” *American Politics Research* 41(5):863–899.

Boydstun, Amber E, Rebecca A Glazier and Matthew T Pietryka. 2013. “Playing to the Crowd: Agenda Control in Presidential Debates.” *Political Communication* 30(2):254–277.

Chang, Jonathan, Sean Gerrish, Chong Wang, Jordan L Boyd-Graber and David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*. pp. 288–296.

Cobb, Charles D and Roger W Elder. 1972. *Participation in American Politics: The Dynamics of Agenda Building*. Allyn and Bacon.

Coe, Kevin, Kate Kenski and Stephen A Rains. 2014. “Online and uncivil? Patterns and determinants of incivility in newspaper website comments.” *Journal of Communication* 64(4):658–679.

Cox, Gary W and Mathew D McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the US House of Representatives*. Cambridge University Press.

Denny, Matthew J and Arthur Spirling. 2018. “Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It.” *Political Analysis* 26(2):168–189.

Eggers, Andrew C and Arthur Spirling. 2016. “The Shadow Cabinet in Westminster Systems: Modeling Opposition Agenda Setting in the House of Commons, 1832–1915.” *British Journal of Political Science* pp. 1–25.

Epstein, Lee, William M Landes and Richard A Posner. 2010. “Inferring the Winning Party in the Supreme Court from the Pattern of Questioning at Oral Argument.” *The Journal of Legal Studies* 39(2):433–467.

Fishkin, James S and Robert C Luskin. 2005. “Experimenting with a democratic ideal: Deliberative polling and public opinion.” *Acta politica* 40(3):284–298.

Grimmer, Justin. 2009. "A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases." *Political Analysis* 18(1):1–35.

Grimmer, Justin and Gary King. 2011. "General purpose computer-assisted clustering and conceptualization." *Proceedings of the National Academy of Sciences* 108(7):2643–2650.

Johnson, Timothy R, Paul J Wahlbeck and James F Spriggs. 2006. "The Influence of Oral Arguments on the US Supreme Court." *American Political Science Review* 100(1):99–113.

Karpowitz, Christopher F, Tali Mendelberg and Lee Shaker. 2012. "Gender Inequality in Deliberative Participation." *American Political Science Review* 106(3):533–547.

Kathlene, Lyn. 1994. "Power and Influence in State Legislative Policymaking: The Interaction of Gender and Position in Committee Hearing Debates." *American Political Science Review* 88(3):560–576.

Kingdon, John W. 1995. "Agendas, alternatives, and public policies. 2nd." *NY: HaperCollins College Publisher* .

Levin, Dan. 2019. "Brigham Young Students Value Their Strict Honor Code. But Not the Harsh Punishments." *New York Times*. April 12. https://www.nytimes.com/2019/04/12/us/byu-honor-code.html. Last accessed on May 10, 2020.

Mason, Melanie. 2016. "The Most Memorable Moments from the First Presidential Debate." *Los Angeles Times*. September 26. https://www.latimes.com/politics/la-na-pol-debate-moments-20160926-snap-htmlstory.html. Last accessed on September 25, 2019.

McCombs, Maxwell E and Donald L Shaw. 1972. "The agenda-setting function of mass media." *Public opinion quarterly* 36(2):176–187.

Mikhaylov, Slava, Michael Laver and Kenneth R Benoit. 2012. "Coder Reliability and Misclassification in the Human Coding of Party Manifestos." *Political Analysis* 20(1):78–91.

Mummolo, Jonathan and Erik Peterson. 2019. "Demand effects in survey experiments: An empirical assessment." *American Political Science Review* 113(2):517–529.

Munger, Kevin, Mario Luca, Jonathan Nagler and Joshua Tucker. 2018. Everyone On Mechanical Turk is Above a Threshold of Digital Literacy: Sampling Strategies for Studying Digital Media Effects. Technical report Working Paper. https://csdp. princeton. edu/sites/csdp/files/media/munger .

Nguyen, Viet-An. 2014. "Speaker Identity for Topic Segmentation (SITS)." GitHub repository. https://github.com/vietansegan/sits. Last accessed September 25, 2019.

Nguyen, Viet-An. 2015. "Guided Probabilistic Topic Models for Agenda-Setting and Framing." PhD dissertation. University of Maryland, College Park.
https://drum.lib.umd.edu/handle/1903/16600.

Nguyen, Viet-An, Jordan Boyd-Graber and Philip Resnik. 2012. SITS: A Hierarchical Nonparametric Model Using Speaker Identity for Topic Segmentation in Multiparty Conversations. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics pp. 78–87.

Nguyen, Viet-An, Jordan Boyd-Graber, Philip Resnik, Deborah A Cai, Jennifer E Midberry and Yuanxin Wang. 2014. "Modeling Topic Control to Detect Influence in Conversations Using Nonparametric Topic Models." *Machine Learning* 95(3):381–421.

Petrocik, John R. 1996. "Issue Ownership in Presidential Elections, with a 1980 Case Study." *American Journal of Political Science* pp. 825–850.

Purver, Matthew. 2011. "Topic Segmentation." *Spoken Language Understanding: Systems for Extracting Semantic Information From Speech* pp. 291–317.

Quinn, Kevin M, Burt L Monroe, Michael Colaresi, Michael H Crespin and Dragomir R Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54(1):209–228.

Riker, William H. 1986. *The art of political manipulation*. Vol. 587 Yale University Press.

Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson and David G Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–1082.

Ross, Janell. 2016. "Trump on 'fat slobs,' housekeepers and women who dont have that 'presidential look'." *The Washington Post*. September 27. https://www.washingtonpost.com/news/the-fix/wp/2016/09/27/trump-on-fat-slobs-housekeepers-and-women-who-dont-have-that-presidential-look/. Last accessed on September 25, 2019.

Saunders, Timothy J, Alex H Taylor and Quentin D Atkinson. 2016. "No evidence that a range of artificial monitoring cues influence online donations to charity in an MTurk sample." *Royal Society open science* 3(10):150710.

Schattschneider, Elmer E. 1960. *The Semi-Sovereign People: A Realist's View of Democracy in America*. Wadsworth Publishing.

Settle, Jaime E. 2018. *Frenemies: How social media polarizes America*. Cambridge University Press.

Sides, John and Lynn Vavreck. 2014. *The Gamble: Choice and Chance in the 2012 Presidential Election-Updated Edition*. Princeton University Press.

Sides, John, Michael Tesler and Lynn Vavreck. 2019. *Identity crisis: The 2016 presidential cam-*

paign and the battle for the meaning of America. Princeton University Press.

Turkewitz, Julie. 2014. "At Brigham Young, Students Push to Lift Ban on Bears." *New York Times*. November 17.
https://www.nytimes.com/2014/11/18/us/campaigning-to-change-the-cleanshaven-look-at-brigham-young-university.html. Last accessed on May 10, 2020.

Vavreck, Lynn. 2009. *The message matters: The economy and presidential campaigns*. Princeton University Press.

Wang, Lu, Nick Beauchamp, Sarah Shugars and Kechen Qin. 2017. "Winning on the merits: The joint effects of content and style on debate outcomes." *Transactions of the Association for Computational Linguistics* 5:219–232.

Weaver, David H. 2007. "Thoughts on agenda setting, framing, and priming." *Journal of communication* 57(1):142–147.

Williamson, Vanessa. 2016. "On the ethics of crowdsourced research." *PS: Political Science & Politics* 49(1):77–81.

Ying, Luwei, Jacob M Montgomery and Brandon M Stewart. 2019. "Inferring Concepts from Topics: Towards Procedures for Validating Topics as Measures." *PolMeth XXXVI, Cambridge, MA. Society for Political Methodology* . Last accessed on June 5, 2020.
https://polmeth.mit.edu/sites/default/files/documents/YingMontgomeryStewart