A General Model of Author "Style" with Application to the UK House of Commons, 1935–2018

Leslie Huang[1], Patrick O. Perry[2] and Arthur Spirling[3]

[1] Graduate Student, Center for Data Science, New York University, 60 Fifth Avenue, New York, NY 10011, USA. Email: lesliehuang@nyu.edu

[2] Senior Data Scientist, Oscar Health; Visiting Scholar, Center for Data Science, New York University, 60 Fifth Avenue, New York, NY 10011, USA. Email: pperry@stern.nyu.edu

[3] Professor of Politics and Data Science, New York University, 19 W4th St, New York, NY 10012, USA. Email: arthur.spirling@nyu.edu

Abstract

We consider evidence for the assertion that backbench members of parliament (MPs) in the UK have become less distinctive from one another in terms of their speech. Noting that this claim has considerable normative and substantive implications, we review theory and findings in the area, which are ultimately ambiguous on this question. We then provide a new statistical model of distinctiveness that extends traditional efforts to statistically characterize the "style" of authors and apply it to a corpus of *Hansard* speeches from 1935 to 2018. In the aggregate, we find no evidence for the claim of more homogeneity. But this hides intriguing covariate effects: at the MP-level, panel regression results demonstrate that on average, more senior backbenchers tend to be less "different" in speech terms. We also show, however, that this pattern is changing: in recent times, it is more experienced MPs who speak most distinctively.

Keywords: House of Commons, stylometry, text as data, regression

1 Introduction

Floccinaucinihilipilification is not a word regularly encountered in the House of Commons; between 1803 and the present, it was uttered just twice. The first time was 1947; subsequently, it was used in 2012 by Jacob Rees-Mogg. So apparently interesting was the second invocation that it earned Rees-Mogg an interview on the BBC's current affairs program *The Daily Politics*. This special treatment of idiosyncratic members of parliament (MPs) and their unusual word choice speaks to a broader concern with modern Westminster politics. Crudely, it is that contemporary members are simply too similar to one another and speak about the same substantive matters and in identical ways. This alleged linguistic homogeneity, which we will refer to as a decline in "distinctiveness," has obvious substantive and normative implications for representation of an increasingly diverse society. How should we assess such claims statistically, and what is their truth value in this particular case? The purpose of this paper is to answer both questions.

On the core substantive claim, there is much previous work on the differences between MPs. But those efforts yield orthogonal predictions for our particular question. For example, we see "rebellious" behavior (on roll calls) increasing—moving the UK away from the traditional Westminster archetype of pliant backbenchers. Furthermore, we observe that MPs increasingly seem to play to constituency preferences, rather than the party line, on at least some votes (Vivyan and Wagner 2012; Hanretty, Lauderdale, and Vivyan 2017). Relatedly, MPs try to cultivate a personal vote separate to their party (Jackson and Lilleker 2011), although the size of the effect is debatable (Eggers and Spirling 2017). But we *also* observe that MPs are more "professional" (Rush 2001) and career oriented (King 1981) than in the past (see also O'Grady 2018). To the extent that being distinctive in one's parliamentary behavior (of whatever type) is likely to be costly to ministerial ambitions (Kam 2009) and more MPs have such ambitions, we should expect to see decreased distinctiveness, on average, over time.

British legislators have unusual latitude to speak freely in debates (Proksch and Slapin 2012). Given the above, though, the optimal way that we expect those opportunities to be used—and how that has changed over time—is unclear. That is, there are good reasons to imagine legislators have become either *more* or *less* distinctive, in accordance with the relative weight one places on the forces discussed. What complicates the picture considerably is that there likely are *within career* effects, but these are also ambiguous. On the one hand, more senior MPs are more likely to engage in distinctive or rebellious behavior (Benedetto and Hix 2007); on the other, the longer they serve, the more members become socialized to party norms and expectations (Rush and Giddings 2011).

More fundamental than these ambiguous predictions, it is not clear how one should assess evidence for either side in the purported debate. Thus, the deeper contribution of this paper is to show how to statistically estimate the distinctiveness of backbench MPs in aggregate and individually. We do this using a new incarnation of a well-established technique from "stylometry." Namely, a general version—across all words, across all speakers—of an approach made famous by Mosteller and Wallace (1963) to determine the authorship of the (mystery) *Federalist Papers*. We expend considerable efforts validating this measure. Ultimately, we find no evidence that MPs are becoming less distinctive as a whole. To the extent that service length matters, it typically has a negative effect: that is, senior MPs tend to be less different, junior ones are more distinct. Yet, this relationship is changing: in recent times, more experienced members are emerging as the most distinctive legislators in the Commons.

2 Motivation: Competing Pressures in Westminster Systems

We will shortly define distinctiveness in a precise, mathematical way. For now, "distinctiveness" is literally how different, on average, an MP's words are to everyone else's. More concretely, an MP is distinct relative to others if she uses words at different *rates* to her colleagues. In the limit, and subject to some smoothing we will introduce below, this could mean using a unique word that no one else ever uses ("floccinaucinihilipilification"). But that is a special case, and the general idea is that the MP in question uses words with different frequencies than do other members. From our perspective, this is only an interesting quantity to the extent that speakers have choices about what to say. For this reason, we study (government party) backbench MPs only. We exclude those on government payroll since such ministers have a policy brief and are *de facto* required to speak on certain issues and topics regardless of their own personal investment in them.

In one sense, our understanding of the distinctiveness of political speech is narrow: we are aware that a given member's contribution may be notable for reasons not measured via their raw words themselves—including the pitch of the voice or the vocal stresses used. But within the world of "bag of words," there is another sense in which our understanding of distinctiveness is deliberately broad. It subsumes "mere" stylistic differences and more substantive ones. While this measurement decision avoids having to precisely decide what constitutes "style" as opposed to "substance," it is nonetheless helpful to convey our perceptions. In keeping with the earlier "stylometry" literature, we would argue that function words, and the within-speaker variance they exhibit, are a particularly narrow but internally consistent definition of "style." In that world, "substance" is everything else—the verbs, the nouns, floccinaucinihilipilification, etc. In practice, we conceive of "style" as being somewhat larger: intuitively, it is the *way* a given latent position or argument is delivered, separate to its content. Put otherwise, two MPs might have identical substantive interests, and views on those substantive interests, yet, they make the argument in rhetorically different ways. For example, they both support Brexit for the same reasons, yet, one wants to "take back control," while the other speaks of the importance of "restoring parliamentary sovereignty." We would argue that these differences may or may not be immediately recognizable to human observers. But as with our particular interpretation of style and substance, this fact does not affect the statistical derivation or the results below. And in any case, as we show in

Section 8, our approach allows users to make different assumptions about which understanding of distinctiveness is of interest and assess the relative performance of models that arise from those beliefs.

Legislators may speak differently to others based on conscious, strategic decisions or due to some latent, sincere preference (perhaps ideological) that is simply manifested in the speech act. And, of course, some combination of these factors is possible. Whatever the data generating process, MPs in Westminster systems face constraints in their optimal level of distinctiveness (Kam 2009; Lijphart 2012). For the ambitious (see Rush and Childs 2004), being distinctive by taking positions out of step with the party leadership has negative effects on ministerial promotion (Rush 2001). This is true of rebelling on divisions (Cowley 2002; Benedetto and Hix 2007), and one can imagine that it is also true of speech. Beyond career incentives, there is evidence of socialization in Commons behavior, which should also suppress potential distinctiveness (Crowe 1986; Rush and Giddings 2011). The *quid pro quo* for this acquiescence to a united message may be private arrangements in which party leaderships listen to backbench preferences on personnel choices (Kam *et al.* 2010).

Outside the Commons itself, there are other reasons to think that MPs have little to gain by being deliberately or "accidentally" distinctive. "Dyadic representation" (Weissberg 1978)—that is, the notion that MPs might seek to legislate in a way that reflects constituent preferences—is thought to be rare in the UK. Instead, voters are responsive to party identification (e.g. Heath *et al.* 1991) and perceptions of national leaders (e.g. Green and Hobolt 2008). Any incumbency bias is commensurately small (Gaines 1998; Eggers and Spirling 2017).

Given the above, why would MPs ever (wish to) appear distinctive from one another? Possibly because behaving differently may be a way to attract positive attention, including from the media. As noted, MPs are more rebellious (Cowley 2002) and more strategic in their rebellion (Slapin *et al.* 2017) than they used to be. This seems to be rewarded by their voters (Pattie, Fieldhouse, and Johnston 1994; Kam 2009; Vivyan and Wagner 2012)—in addition to other responses to constituent opinion (Hanretty, Lauderdale, and Vivyan 2017). Related, MPs seek to actively manage their personal "brand" via social media (Jackson and Lilleker 2011). If being distinctive in these ways is self-promotional, speech may be a helpful extra resource. Certainly, there is qualitative evidence that MPs understand the need to produce soundbites and interesting content for local media consumption (Negrine and Lilleker 2003).

Matching these behavioral changes, and possibly a cause of the same, are changes to the sociological make-up of the House of Commons. Since the mid-1850s, parliament has become more "professional" (Rush 2001). Members are increasingly "career" politicians (in the sense of King 1981) and view their positions as full-time jobs. Since the 1960s, MPs are increasingly drawn from the university educated middle classes, and whatever distinctions traditionally existed between Labour and Conservative backgrounds are now much weaker (Norris and Lovenduski 1995; Heath 2015). At least part of this trend is the increasing tendency to draw MPs from the ranks of "special advisors" and other professional party workers (Shaw 2001). These changes cut both ways. On the one hand, MPs are from similar social backgrounds, and so might be innately less distinctive. On the other hand, they have more "streetwise" knowledge of their political environment, which may facilitate more independent thought and action.

All told, we have reasons to think distinctiveness may be higher or lower than in the past. To know how these forces play out, we need a valid measure of the concept for MP speech.

3 Measuring Distinctiveness: Intuition

To measure distinctiveness, we rely on the basic principle of *stylometry*—that authors have idiosyncratic markers in the documents they produce. The typical goal in that literature is detecting the most likely author of a given text of uncertain origin, by way of the candidate authors'

known preferences in word use. In terms of applications to politics, the work of Mosteller and Wallace (1963; 1964) is well known. Their challenge was to identify the most likely author—Madison or Hamilton—of twelve "disputed" *Federalist Papers*. They used seventy function words as the features for this task.[1] Using essentially the same model, Airoldi, Fienberg, and Skinner (2007) look at radio addresses in the 1970s which may (or not) have been authored by Ronald Reagan.

Our strategy here is similar in that we care about the distinctiveness of one speaker/author (MP) relative to another. But in detail, it differs markedly. In particular, our interest is not in identifying origins of texts: we have labels for all our data (we know which MP gave each speech). Instead, authors/speakers are our focus, and the extent to which any linguistic features exist that mark them apart from one another.

To see the intuition, note that in the case of Mosteller and Wallace, what is of interest is the comparison of two probabilities for an unlabeled text. The model implies some probability the paper was written by Madison given the counts of the words (**w**) it contains, Pr(Madison|**w**). That is then compared with the probability the essay is, conditioned on its contents, from Hamilton, Pr(Hamilton|**w**). Subject to some mathematical housekeeping, the larger of these probabilities then yields the author prediction.

But in our application, we care about how different Madison is from Hamilton *in general* for all the (labeled) data. To fix ideas, suppose we have an essay we know Madison wrote. Given the words it contains, our model can still provide us with an *estimate* of the probability Madison penned it: while we know the true probability to be one, the model will give us an in-sample prediction (which is not one). We can do the same thing for Hamilton for that given Madison speech: obtain a model probability that Hamilton wrote it (though we know the true probability to be zero) by plugging in Hamilton's word-use tendencies from the essays which we know *he* wrote. Suppose that we do this for every one of Madison's known texts. If, in general, the predicted probability that Madison wrote them is much higher than the predicted probability that Hamilton wrote them, we have *prima facie* evidence that the authors differ in style terms. That is, the model is finding features that enable it to distinguish one author from the other for the Madison documents. But if the Madison essays have predicted probabilities that are always similar for Madison and Hamilton, then the opposite lesson applies: the model simply cannot distinguish between the two men as most plausible authors for the selection of texts. To summarize, and in keeping with our simple statement of the problem, we are interested in the *distinctiveness of Madison*, $\mathbb{D}_M$. We will shortly generalize this to an arbitrary number of essays (speeches) and authors (MPs).

Below, we give a full mathematical derivation of our statistical model, but, first, we give a non-technical overview to guide readers:

1. For a given session of parliament (approximately one year of law-making), each member's speeches are summarized as a vector with word counts for every one of the terms spoken by any MP in that session. These counts are converted into *probabilities* by dividing them by the sum of the vectors (which have length equal to the size of the vocabulary used in a session). For convenience, those probabilities are then logged (with appropriate smoothing) to produce a vector η_s, where s denotes a specific MP.

2. To compare MP s with MP t as a pair for a given word in a random speech i from one of them, we subtract the value of η_s from η_t for that word in the speech (multiplied by the number of times that word is used in the speech). If this number is "large," then t is *distinctive* relative

---

1 These words were: a, all, also, an, and, any, are, as, at, be, been, but, by, can, do, down, even, every, for, from, had, has, have, her, his, if, in, into, is, it, its, may, more, must, my, no, not, now, of, on, one, only, or, our, shall, should, so, some, such, than, that, the, their, then, there, things, this, to, up, upon, was, were, what, when, which, who, will, with, would, your.

to *s* in her use of this word. We can take the average of this quantity over all the words in the vocabulary to get a more general measure of the distinctiveness of *t*.

3. For the random speech *i*, we generalize this measure to all members—that is, we produce a number that compares *t* to everyone else pairwise—by taking the average of the quantity in (2) *over all speakers*. The second generalization is *over all speeches* given by *t* which essentially requires summing the (average) distinctiveness per word and dividing by the number of speeches *t* gave.

This quantity has a Bayesian interpretation but, in essence, represents the evidence that a given speaker (relative to all others) produced the words she spoke—averaged over all words, all speeches, and all possible pairwise comparisons to other members. Substantively, we are producing a one-dimensional summary statistic: a line from least distinctive to most distinctive on which each MP "sits" as a real-valued point estimate. Crudely, *ceteris paribus*, if an MP starts to use words in ways (frequencies per speech) that deviate from others, they will move "up" the scale in terms of distinctiveness; if they instead use words in a way more like others (e.g. reducing their use of idiosyncratic terms and substituting for ones more commonly used by others), they move down the scale.

Of course, as with all measurement strategies, we must make some assumptions. In our case, they are threefold. First, all distinctiveness is "relative" to some particular reference set of speakers: change the reference set, and a given speaker may not appear very different. Our reference set will be all other backbenchers in the governing party. Second, distinctiveness is heterogeneous *within* speakers. That is, within a given session of parliament and with respect to the same reference set of colleagues, a member's distinctiveness may vary depending on what they are speaking about and why. Presumably, if they are speaking about procedural issues, their opportunities to be distinctive are limited—at least relative to more free-form debate. This could cause problems for historical comparisons if the proportions of such speeches are changing systematically over time. In practice, we will use session-fixed effects to dampen down variation from this source. Finally, we will define distinctiveness such that, provided certain independence assumptions are in force, the measure converges in probability to a fixed value as the number of speeches from context *c* increases. To the extent that our assumptions are reasonable approximations of reality, our measure will accurately quantify distinctiveness. Even when our assumptions are unreasonable, it may still be the case that our distinctiveness measure approximates that which we hope to capture. We will rely on external validation to argue this case.

Before giving the full derivation, we clarify our approach relative to two popular alternatives. First, in line with the original Mosteller and Wallace (1963) article, some scholars have shown that specific information about social dynamics can be gleaned from studying "function" word usage rates (see, e.g, Gonzales, Hancock, and Pennebaker 2010). Our approach is broader, in that we deliberately study substantive differences between speakers. Second, then, why not fit a topic model (in the sense of Quinn *et al.* 2010) and see how different MPs contribute to different types of debates (in the sense of Roberts *et al.* 2014)? Certainly, other scholars have used related ideas for parliamentary systems (e.g. Lauderdale and Herzog 2016). Put simply, we do not care about topics *per se*. In a Westminster system, governments have profound power to organize what will or will not be debated (i.e. the topic of discussion). So while topics might tell us the types of issues MPs are interested in, they do not tell us how *distinctive* MPs are from one another. Crudely, we will not know whether an MP contributing to a particular debate has something different to say relative to another. And, more importantly, if government preferences for debate (qualitatively or quantitatively) change over time, it becomes difficult to make cross-time comparisons between MPs. Our model is not constrained in this way.

4 Measuring Distinctiveness: Mathematical Derivation

Consider the context c of a particular parliamentary session. Let S_c denote the reference set of speakers for this context: all speaking backbenchers from the governing party with at least ten speeches of at least three tokens. Let t denote the target speaker, any member of S_c. Our formal goal is to define the distinctiveness of speaker t relative to speaker set S_c in context c. For the purposes of defining the distinctiveness measure, we suppose that each speaker s belongs to a unique context $c(s)$; a member appearing in two parliamentary sessions is treated as two speakers.[2]

To start, we will need a probabilistic model relating a speaker to the text of a speech. For purposes that will become clear momentarily, recall that a word "type" is a distinct entity or concept in a text, while a word "token" is an instance of that entity. Thus, the phrase "Dog eat dog world" contains four tokens (the words), but only three types (the second dog token is a second incidence of the first type). Of course, it could be the case that the same token is used to mean different things in practice: in one speech, a "pound" might be a reference to a unit of currency (£), whereas in another, it is a unit of weight (*lb*). We are aware that there exist word-embedding techniques that can resolve these ambiguities (Rheault and Cochrane 2019; Rodman 2019), but we prefer to keep things simple here. Note, in addition, that we redefine our set of types per parliamentary session (lasting approximately one year). This somewhat mitigates the concern that we may be conflating the same word used in different ways—over time, at least.

We begin by reducing the text of each speech to a sequence of word tokens drawn from fixed vocabulary V_c specific to the context c. In our application, V_c is the total set of word types for session c. That is, the union of the types used by all the MPs who speak during that period.[3] Let I_s denote the set of all speeches from speaker s. For speech $i \in I_s$ and word type $v \in V_c$, let x_{iv} denote the number of word tokens in speech i equal to v; let $n_i = \sum_{v \in V_c} x_{iv}$ denote the length of speech i.[4]

Our first simplifying assumption is that each speaker $s \in S_c$ has a set of word type probabilities that determine how speeches from s are generated. It will be convenient to parameterize these probabilities in terms of their natural logarithms. Specifically, take a particular speech i given by speaker $s \in S_c$. Let w denote a randomly chosen word token from this speech, and suppose that the probability that this word is $v \in V_c$ in terms of its logarithm by

$$\log \Pr(w = v | s) = \eta_{sv},$$

the same for all speeches by s in context c. Denote the speaker-specific vector of such log probabilities by η_s.

Suppose that we know η_s for each speaker (or that we have estimated these vectors using speakers' empirical word frequencies). Our next task will be to use these quantities together with the speeches from context c to define the distinctiveness of each speaker.

To define the distinctiveness of target speaker t with respect to reference set S_c in context c, we start by taking the simple case where the reference set contains only two speakers, $S_c = \{s, t\}$, and the context contains only a single speech, i. Suppose that we randomly pick a word token from speech i. If, on the basis of this token, it is easy to identify whether t is the speaker, then we will say that t is distinctive. If, on the other hand, it is difficult to identify whether t is the speaker, then we will say that t is typical (not distinctive). In particular, suppose that we randomly pick word type v

---

2 This is simply for measurement purposes in the sense that we have to define the unit of observation: when making *inferences* from the data below, we will use fixed effects to look at within-MP variation.

3 For the 2013–2014 session, our data includes speeches up until March 2014. For the 2017–2018 session, our data includes speeches up until March 2018.

4 In practice, we smooth by adding $\frac{1}{2}$ to all counts to avoid zeros/undefined quantities in the calculations that follow.

from the speech. Using Bayes' rule and equal prior probabilities for whether s or t is the speaker, the log posterior odds ratio that t is the speaker are given by $(\eta_{tv} - \eta_{sv})$. The expected value of this quantity for a random word type drawn from speech i is

$$\frac{1}{n_i} \sum_{v \in V_c} x_{iv}(\eta_{tv} - \eta_{sv}). \tag{1}$$

We define this quantity as the distinctiveness of target speaker t relative to reference set $\{s, t\}$ in the context of speech i.

With the simple case covered, it is straightforward to generalize our distinctiveness measure to larger reference speaker sets and larger contexts. In the first direction, for an arbitrary reference set S_c containing t, we take the average pairwise distinctiveness for a randomly chosen alternative $s \in S_c$:

$$\frac{1}{|S_c|} \sum_{s \in S_c} \left\{ \frac{1}{n_i} \sum_{v \in V_c} x_{iv}(\eta_{tv} - \eta_{sv}) \right\},$$

where $|S_c|$ denotes the size of set S_c. The second generalization, to larger numbers of speeches, can be obtained by taking the expectation over a randomly chosen speech $i \in I_t$. This gives our final measure of distinctiveness, which, after re-arranging the sums, can be expressed as

$$\mathbb{D}_t = \mathbb{D}_t(S_c; I_t) = \frac{1}{|I_t|} \frac{1}{|S_c|} \sum_{i \in I_t} \sum_{s \in S_c} \sum_{v \in V_c} f_{iv}(\eta_{tv} - \eta_{sv}),$$

where $f_{iv} = \dfrac{x_{iv}}{n_i}$. We can further simplify this expression by defining

$$\bar{f}_{tv} = \frac{1}{|I_t|} \sum_{i \in I_t} f_{iv} \quad \bar{\eta}_{cv} = \frac{1}{|S_c|} \sum_{s \in S_c} \eta_{sv}.$$

In this case,

$$\mathbb{D}_t = \sum_{v \in V_c} \bar{f}_{tv}(\eta_{tv} - \bar{\eta}_{cv}).$$

With this final expression, we can compute $\mathbb{D}_t$ by taking the difference between η_t and the average η_s over all speakers $s \in S_c$, then taking the dot product with the empirical frequencies $\bar{f}_{tv}$ computed from target speaker's speeches I_t. It is $\mathbb{D}_t$ that becomes our "distinctiveness" dependent variable in what follows.

4.1 Standard Errors

The distinctiveness measure $\mathbb{D}_t = \mathbb{D}_t(S_c; I_t)$ is an empirical average over all observed speeches I_t by the target speaker t and all other speakers in the context S_c of the quantity $\sum_{v \in V_c} f_{iv}(\eta_{tv} - \eta_{sv})$. There are at least three sources of variability in that effect $\mathbb{D}_t$:

1. The observed speeches I_t can be considered as a sample of all speeches that could have been delivered by the target t in the same context.
2. The reference speakers S_c (in our context, the other backbenchers in the governing party) can be considered a sample of all potential reference speakers that could have been present.

3. The speaker-specific log word type frequencies η_{sv} are estimates based on empirical frequencies; these estimates depend on the actual observed speeches by s, which, again, can be considered as a sample of all potential speeches by s.

To assess the variability in our computed value in $\mathbb{D}_t$, we make the simplifying approximation that the largest sources of variability come from the random process that determined I_t and S_c. That is, we ignore variability in determining the estimates of η_{sv}. In practice, this means that we assume that the empirical word-use probabilities themselves do not require their own standard errors. The motivation for this is that we constrain MPs to a minimum threshold (described below) of speeches before we estimate an η for them. It is a maintained assumption that we therefore have enough data on everyone to avoid wildly misleading point estimates.

Set $D_{is} = \sum_{v \in V_c} f_{iv}(\eta_{tv} - \eta_{sv})$. The quantity $\mathbb{D}_t$ is an average of D_{is} over sets I_t and S_c. We will condition on the sizes of the sets I_t and S_c, but, otherwise, we will treat these sets as random. Specifically, set $n = |I_t|$ and $m = |S_c|$; take I_t to be a set of independent identical draws from some population $\mathbb{I}_t$ and take S_c to be a set of m independent draws from some population $\mathbb{S}_c$. Treat D_{is} as a deterministic function of the speech $i \in \mathbb{I}_t$ and the speaker $s \in \mathbb{S}_c$.

To assess the uncertainty in $\mathbb{D}_t$ due to the variability in I_t and S_c, first, define for each $i \in \mathbb{I}_t$.

$$D_i(S) = \frac{1}{m} \sum_{s \in S} D_{is}$$

If S is a set of size m drawn independently and identically from population $\mathbb{S}_t$, then define the expectation and variance over random S as

$$E\{D_i(S)\} = \mu_i, \quad \mathrm{var}\{D_i(S)\} = \frac{\sigma_i^2}{m},$$

where μ_i and σ_i^2 are the mean and variance of D_{is} as s ranges over population $\mathbb{S}_c$.

Express the distinctiveness over a random set speaker set S of size m drawn as before and a random speech set I of size n drawn independently and identically from population $\mathbb{I}_t$ as a random variable

$$D = D(I, S) = \frac{1}{n} \sum_{i \in I} D_i(S).$$

Note that $\mathbb{D}_t = D(I_t, S_c)$. Now, $\mathrm{var}(D) = E\{\mathrm{var}(D \mid I)\} + \mathrm{var}\{E(D \mid I)\}$, where the outer expectation and variance on the right-hand side are over the random set I. Using the independence of the speeches yields

$$E(D \mid I) = \frac{1}{n} \sum_{i \in I} \mu_i, \quad \mathrm{var}(D \mid I) = \frac{1}{n^2} \sum_{i \in I} \frac{\sigma_i^2}{m}.$$

Hence,

$$\mathrm{var}(D) = \frac{1}{n} \mathrm{var}(\mu_i) + \frac{1}{nm} E(\sigma_i^2),$$

the variance and expectation being computed over a random i drawn from population $\mathbb{I}_t$. Define estimate $\hat{\mu}_i = D_i(S_c)$ and set $\hat{\sigma}_i^2$ to be the empirical variance of D_{is} as s ranges over S_c. We estimate $\mathrm{var}(\mu_i)$ by the empirical variance of $\hat{\mu}_i$ and we estimate $E(\sigma_i^2)$ by the empirical mean of $\hat{\sigma}_i^2$. This gives us an estimate of the variance of $\mathbb{D}_t$; we use the square root of this quantity as a standard error for $\mathbb{D}_t$.

4.2 Relation to Other Approaches

Above, we were explicit that we were building on the traditional Mosteller and Wallace (1963) setup. Their core Poisson distribution model, for the case with more than two classes, is identical to the well-known multinomial naive Bayes model. Consequently, our approach may be seen as an extension of the multinomial naive Bayes approach.[5] We lay out the relationship in detail in Appendix A, and like the traditional model, we make the "naive" assumption of independence across terms, allowing the relevant η probabilities to be straightforwardly summed. But there are also differences from the usual setup. First, we propose using a version of the posterior log odds expressly as a measure of distinctiveness. Second, we average across all speeches, which is not typically done. And, third, we derive a standard error, which is important for social science uncertainty statements.

Of course, other approaches are available for measuring distinctiveness. On the more complex front, one could use *n*-grams, in addition to unigrams, to capture phrases and word order. And as we noted above, one could, in principle, use a word-embedding approach to represent the concepts of interest. We think these are interesting, but beyond the scope of the current effort, where the intention is to extend the statistical machinery of Mosteller and Wallace (1963), rather than the feature representations. A far simpler approach would be to use distance vectors, perhaps averaged in some way: for example, the mean cosine distance between MP *s* and all her colleagues. This would be fast to calculate though unlike our approach, would not as naturally allow for uncertainty statements, nor the identification of influential terms.

More broadly, one could move away from a naive Bayes-influenced model altogether. There are obviously more complex and flexible machine-learning tools on offer to do the core task required above. We are in no doubt that improvements in performance (however defined) are available, but we would argue that our contribution here is beyond "mere" technical innovation. That is, we are engaged in defining political distinctiveness itself—and demonstrating how it can be measured, with attendant uncertainty estimates, from text. By doing this with a simple model derived from a classic approach, we hope to provide a foundation for more advanced work.

5 Data

Our data supplements Rheault *et al.* (2016) with speeches to the present and consists of approximately three million speeches in the House of Commons (1935–2018). These are organized into "sessions" lasting approximately a year each. There is meta-data pertaining to party membership and ministerial position. We calculate the *experience* of a given MP as being the number of sessions they have served in parliament since their first speech.[6] We also introduce a variable that records whether an MP has ever been *demoted* from ministerial office.

For the purposes of this paper, we are only interested in MPs that actually speak (at least ten times, speeches of at least three words) in a given session, and, in particular, we are focused on government *backbenchers*. That is, our analysis of distinctiveness pertains to MPs that are in the governing party (the party of the Prime Minister) but who do not hold ministerial office. This allows us to compare like with like over time in terms of the incentives MPs have, even though power shifts across parties.

In Table 1, we give more details regarding the averages and ranges of (average) values of our data by session. A takeaway is that the (average) means and medians are generally very close, implying little skew in these variables.

---

5   We are grateful to an anonymous referee for pointing this out and encouraging us to make the links explicit.
6   We augment the data with information on when—what session—the MPs speaking in the first period of the data entered parliament.

Table 1. Summaries for variables in our data.

Description	Minimum	Median	Mean	Max
Experience	2.40	9.92	9.88	15.68
Proportion demoted	0.00	0.13	0.13	0.31
Distinctiveness	−0.76	0.45	0.44	1.83
Avg. no. of speeches per MP (per career)	10.00	38.74	52.08	482.94
Avg. no. of speeches per MP (per session)	26.14	57.87	56.68	135.87
Total speeches per session	3424	10979	11124	29212
Speaking MPs per session	102.0	187.0	193.0	288.0
Mean speech length	115.8	188.9	183.6	229.6

5.1 Vocabulary Standards: Obtaining V_c

Our derivation above requires that we define a fixed vocabulary V_c for a given context, c. To reiterate, the context for us is a session (and, further, it is the speeches of government backbenchers in that session). Above, when introducing the model, we implied that this would be all terms spoken in the relevant period. In practice, this can cause problems. In particular, very rare terms can lead to misleading results. For example, terms spoken by one individual once will imply that this MP is maximally distinctive relative to all other members. Yet, this seems unreasonable in terms of our substantive understanding of the quantity of interest. Similar problems may occur with typographic errors in *Hansard*. To avoid depending on outliers in this way, we prune our vocabulary with a straightforward and replicable mechanical procedure.

In particular, we obtain V_c via an n-fold cross-validation procedure: from the raw set of all (backbencher) words spoken during each session, we compute the frequency of each type. We select types above a given percentile p as a candidate vocabulary v_p for $p \in \{50, 55, \ldots, 90, 95, 99\}$, e.g. the vocabulary above the 50th percentile is comprised of terms that occur more frequently than the median.

For the cross-validation procedure, we partition the data from each session into n folds (in our case, five folds each comprised of 20 percent of speeches). For a given v_p, we hold out one of the folds and fit a model to the other four folds of data, repeating this procedure five times per v_p. We use the fitted model to predict the MPs who spoke speeches in the held-out fold of data and compute the mean prediction rate obtained. This procedure is repeated for each $v_{p \in P}$ for each session. We select as V_c the vocabulary with the maximum mean prediction rate for a given session.

To give intuition, our approach selected a subset of the vocabulary based on a frequency threshold (e.g. the 80th percentile) that maximized the model's ability to correctly predict speakers of speeches. This practice discards extremely "rare" words that may be the result of errors in transcription rather than true distinctiveness. However, our software is designed such that the end user can use the entire vocabulary from their corpus if they so choose.

In practice, our frequency-based cross-validation procedure eliminates extremely rare words from the vocabulary. This avoids situations where a speaker may appear to be distinctive because of typos in the transcriptions of his/her speeches. Speakers who introduce "new" terms, provided that those terms are not pruned from the vocabulary by the cross-validation, are likely to appear distinctive, assuming that other MPs do not immediately adopt the term during the same session.

6 Validation I: Are We Measuring What We Think We Are Measuring?

Our measure of distinctiveness is derived from a classical approach—but is it any good? We examine that question in several ways.

First, we will briefly discuss the range of distinctiveness scores and how they are impacted by various speaking strategies. The distinctiveness measure is essentially a weighted average log-odds ratio, comprised of both (a) the speaker's rate of usage of a term v, and (b) the difference in that speaker's (log) probability of using v versus other speakers' (log) probability of using v. The scope of a speaker's speech—does he/she utilize a small or large subset of the vocabulary for the session, and does he/she speak many tokens during the session?—is affected through both of these channels. A speaker who speaks infrequently will have a usage rate of zero for many words (which we smooth; minimizing this phenomenon is why we dropped speakers who spoke fewer than ten speeches of three tokens in a session). Additionally, when the speaker's probability of speaking v will also be small, their log probability for speaking v may be negative, which means that the log-odds ratio for v may be negative (but note that this also depends on the log probabilities of other speakers saying v). The distinctiveness of a speaker who uses a small subset of the total vocabulary will be similarly affected by low word usage rates and low (individual) log probabilities that contribute to the log-odds ratios.

6.1 Intrusion

One basic requirement is that our method ought to label "intrusive" texts, i.e. ones that were not clearly produced by the parliamentary data generating process, as distinctive. To put this to the test, we took the set of backbenchers from a modern session (2015–2016) added to them the State of the Union speeches given by U.S. President Barack Obama. We randomly sampled n "speeches" of m sentences each from the SotU speeches, where n is the mean speeches per MP and m is the mean speech length in sentences for the 2015 session. After inserting Obama in the corpus, we select the vocabulary using our standard cross-validation procedure. We used Obama because although his works are approximately contemporaneous with our data, his style is distinctive relative to our MPs: they come from an American rather than British political system, and they are long oratories consumed by the general public rather than speeches directed primarily at other politicians. With that in mind, our model should identify Obama as easily the most distinctive author. As we see from Figure 1, this is indeed the case: Obama's $\mathbb{D}_t$ point estimate is by far the largest in the data and appears at the far top right. Its confidence interval does not overlap with any other MP in the distribution (note that we do not include every MP in the graphic due to limited space, but we do include the full range in terms of distinctiveness estimates).

Of course, such a test might be cherry-picking, and there is no obvious baseline for performance (other than identifying the intruder).[7] So we now turn to a domain-specific assessment.

7 Validation II: Applying our Measure to MPs

For more substantive performance evaluation, we look at the "most distinctive" and "least distinctive" backbench MPs for the parliamentary sessions on either side of Blair's election landslide in 1997 (that is, 1995–1996 and 1998–1999). This has the advantage of being a period in which (a) control of the Commons switched (from Conservative to Labour), meaning we have variation in the party of the backbenchers, and (b) there are academic accounts which help ground our understanding of MPs at this time (Cowley 2002; Spirling and Quinn 2010; Kellermann 2012). In terms of measurement, we use a *convergent validity* approach: we compare our measure to another (computed independently) and show that they are related as expected.

To see how we proceed in practice, note that for each MP t, in each session, we have an estimate of their distinctiveness in log-odds terms: our $\mathbb{D}$, above. For current purposes, however, we focus on something related but more concrete and directly interpretable: the proportion of their speeches which are correctly predicted as being from them relative to all other MPs (proportion

---

7 With that in mind, we also used our method to examine the original Mosteller and Wallace (1963) data using their findings as a "gold standard" with which to assess ours: our results can be found in Appendix B.
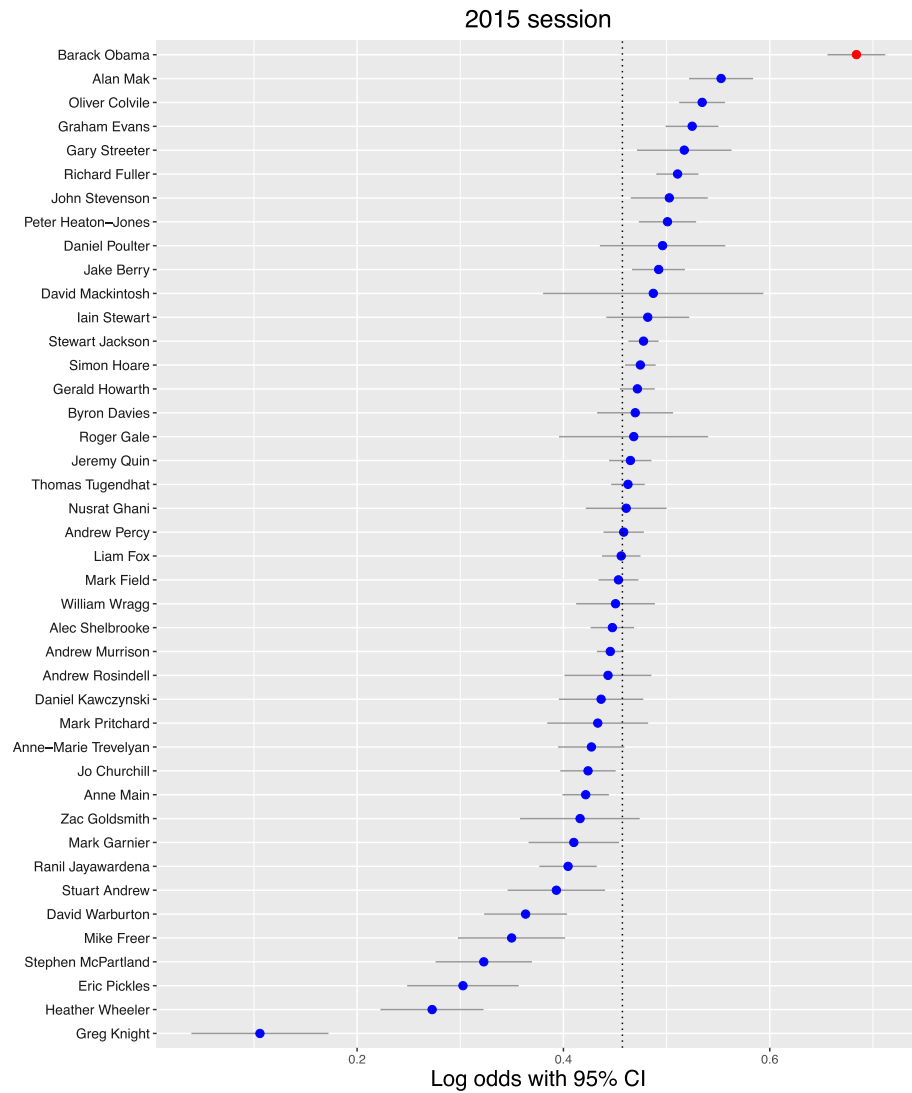
Figure 1. Basic validation: our model identifies U.S. President Barack Obama as an "intruder." Obama—far top right [red]—is the most distinctive author relative to all backbench MPs in the 2015 session (only some shown for visualization purposes).

of speeches correctly predicted, or "PCP," in the tables below). We use fivefold cross-validation to fit a model to texts from a given session, predict the speakers of held-out texts using this model, and calculate each speaker's rate of correct predictions; we report each speaker's mean PCP. To validate these estimates, we consider their extrema—their minimums and maximums. In the subsection tables below, we list the twenty names of the MPs who were most distinct and least distinct by this measure (subject to having made a minimum of twenty speeches). We do this for the two sessions in question: one in 1995–1996 and one in 1998–1999. We also list the number of mentions of each MP in the *Times* newspaper archives (via Gale Group Digital Archive) for the same period, specifically the "Politics and Parliament" subsection of the "News." Note that we searched for the person's (professional) first name and last name together (as a bigram). Our maintained assumption here is that more distinct MPs will tend to be discussed more: whether in the news, editorials, or in parliamentary sketches. Of course, there are various reasons why that assumption might not hold, and, indeed, there are technical issues with this measure. For example, a politician might be mentioned for something they did prior to the relevant search. And, on inspection, it is apparent that we sometimes miss mentions since the searching of an OCR

(Optical Character Recognition)'d newspaper has less than perfect recall. Nonetheless, we hope that with our qualitative comments, this helps validate our measure.

7.1 Tory Backbenchers, 1995–1996

In Table 2, we give the *Times* mention counts for our "most distinctive" Tory backbenchers, and in Table 3, we give the same information for what we claim are the least distinctive ones. A Wilcoxon signed-rank test returns a statistically significant result for the former having a higher mean: that is, that the MPs we claim are more distinctive are indeed mentioned more.

More substantively, we note the presence of several well-known Eurosceptics in Table 2. These include some of the so-called "Maastricht Rebels" like John Wilkinson and Andrew Hunter who abstained or voted against their Conservative government on the relevant treaty (note that Barry Legg, Teresa Gorman, Teddy Taylor, Bill Cash, and Rupert Allason were all just outside our top 20). Also present are other Eurosceptics like Patrick Nicholls, Eric Pickles, and Edward Leigh (who was fired by Major from an undersecretary role for his views). Michael Colvin, unusually for the time, was opposed to restricting gun ownership after the Dunblane Massacre in 1996. David Mellor voted similarly to Colvin. Barry Field had initially decided that he was sufficiently interesting to challenge John Major for the leadership of the Tory party itself, but ultimately withdrew following the emergence of John Redwood. Peter Bottomley was described (by the *Independent*, "The maverick with 'five ideas: four good, one mad'," 11 July, 1993) as being notable for "his delight in surprising colleagues with a range of apparently perverse causes." Our most distinctive MPs are Edward Heath and David Wilshire. Heath is a former Prime Minister, and someone who actively criticized the Conservative governments of Thatcher and Major for being too economically liberal at the expense of social cohesion. Meanwhile, David Wilshire was a right-wing MP responsible for the initial introduction of Section 28 (which meant that local authorities could not "promote homosexuality or... promote the teaching in any maintained school of the acceptability of homosexuality"). He also criticized initial plans (in 1995) to allow Hong Kong Chinese to settle in the UK after the hand-back of the territory to China.

Among the least distinctive MPs, Edwina Currie is perhaps the only one worthy of further comment. Currie had been a controversial cabinet minister (forced to resign in 1988), and by the mid-1990s, was a novelist with two popular tomes written. This perhaps inflates her mentions in the *Times*. We note candidly that Nicholas Budgen, a known Eurosceptic, appears in this list too—in contrast to his rebellious colleagues who generally appear to be more distinctive overall.

7.2 Labour Backbenchers, 1998–1999

In Table 4, we give the *Times* mention counts for our "most distinctive" Labour backbenchers, and in Table 5, we give the same information for what we claim are the least distinctive ones. A Wilcoxon signed-rank test returns a statistically significant result for the former having a higher mean: that is, that the MPs we claim are more distinctive are indeed mentioned more.

More substantively, we note the presence of several Labour "rebels" among the most distinct. These include Tony Benn, Diane Abbott, John McDonnell, Roger Berry, and Tam Dalyell, all of whom consistently voted against the Labour government's plan to reform the welfare state.[8] Peter Temple-Morris was a party switcher, and "interesting" for that reason—he was elected as a Tory MP in 1997, but then crossed the floor to Labour the same year. The most interesting MPs here include Stuart Bell, who was the Church Estates Commissioner, meaning that he was one of the managers of the Church of England's property. David Hinchliffe, chairman of the Select Committee on Health, was subsequently extremely critical of the Blair government's proposed reforms to the National

---

8   Jeremy Corbyn had a PCP of 0.216 for this session, placing him around the 60th percentile for distinctiveness by this metric.

Table 2. Most distinctive MPs, November 1, 1995–October, 31, 1996, in parliament by proportion of speeches correctly predicted.

PCP	Name	*Times* mentions
0.41	Piers Merchant	1
0.41	Phil Gallie	3
0.43	Norman Fowler	3
0.43	Anthony Steen	2
0.43	Terence Higgins	6
0.43	Eric Pickles	0
0.44	James Pawsey	3
0.45	Andrew Hunter	4
0.46	Peter Brooke	5
0.47	Peter Bottomley	2
0.48	Edward Leigh	6
0.48	Robert Hughes	2
0.48	Patrick Nicholls	5
0.48	John Wilkinson	4
0.51	David Howell	4
0.52	David Mellor	3
0.64	Michael Colvin	0
0.66	Barry Field	4
0.72	Edward Heath	19
0.75	David Wilshire	4

Table 3. Least distinctive MPs, November 1, 1995–October, 31, 1996, in parliament by proportion of speeches correctly predicted.

PCP	Name	*Times* mentions
0.00	Alan Haslehurst	0
0.00	Michael Lord	2
0.00	Peter Tapsell	1
0.00	Tim Yeo	2
0.00	Charles Hendry	1
0.00	Nicholas Budgen	7
0.00	Matthew Carrington	0
0.00	Michael Jopling	2
0.00	Gary Waller	2
0.00	Anthony Durant	0
0.00	Hartley Booth	0
0.00	John Butcher	0
0.00	Donald Thompson	0
0.00	Edwina Currie	14
0.00	Jerry Hayes	2
0.00	Peter Butler	3
0.01	John Whittingdale	1
0.02	Michael Neubert	3
0.03	Alan Duncan	3
0.03	Nick Hawkins	0

Table 4. Most distinctive MPs, November 1, 1998–October, 31, 1999, in parliament by proportion of speeches correctly predicted.

PCP	Name	*Times* mentions
0.48	Hazel Blears	1
0.48	Denzil Davies	0
0.49	Geraint Davies	0
0.50	Gareth Thomas	1
0.51	Diane Abbott	3
0.51	Peter Temple-Morris	0
0.53	Tony Benn	19
0.53	Martin Linton	0
0.53	David Winnick	5
0.54	Harry Barnes	1
0.57	Norman Godman	1
0.58	Tam Dalyell	3
0.59	Bill Rammell	0
0.60	Joan Walley	0
0.62	John McDonnell	1
0.66	Roger Berry	1
0.70	David Hinchliffe	1
0.72	Donald Anderson	3
0.73	Stuart Bell	3
0.75	Barry Jones	1

Health Service. Finally, Barry Jones was the chair of the Intelligence and Security Committee of Parliament.[9]

The set of least interesting MPs contains fewer obvious "stars" and consists mostly of loyalists like Doug Naysmith, Ivan Lewis, and Charlotte Atkins. Exceptions include Ronnie Campbell who would subsequently become rebellious regarding the Iraq War and Dennis Canavan who was dismissed (in 1999) from the Labour Party after a dispute over whether he could be an official party candidate for the Scottish Parliament (although he was not particularly notable as a government critic prior to this development).

While this completes the validation of the crucial output of our model for this paper, we now push further. In particular, we consider the validity of both the auxiliary quantities it produces and the nature of what we are actually estimating.

8   Validation III: Phrasing vs. Substance

Finally, with respect to a broader understanding of validity, we ask what exactly we are capturing as "distinctiveness" in our measure? As regards our comments at the opening of Section 2, is it mere "phrasing" (different ways of saying the same thing) or "substance" (saying something different)? Put more directly with respect to the extant literature, does our model "improve" (in fit terms at least) over the original Mosteller and Wallace (1963) approach and capture something more than function word usage?

To assess this, we conducted a simple experiment. We ran a special case of the estimation using only the seventy function words (i.e. stop words) from the original Mosteller and Wallace (1963)

---

9   One possibility is that those who are distinctive in speech are also distinctive in terms of their division voting profiles. And the motivation for this would be signaling valence (in the sense of Cowley *et al.* 2016) to voters. This does not seem to be the case. In particular, the correlation of distinctiveness (for 1998–1999) and rebellion rate for the 1997–2001 parliament (defined as the proportion of times an MP did not vote with the majority of their party) is only around 0.17.

Table 5. Least distinctive MPs, November 1, 1998–October, 31, 1999, in parliament by our measure.

PCP	Name	*Times* mentions
0.00	Charlotte Atkins	0
0.00	Nigel Beard	1
0.00	Ben Bradshaw	1
0.00	Ronnie Campbell	3
0.00	Dennis Canavan	0
0.00	Ivor Caplin	1
0.00	Helen Brinton	0
0.00	Barbara Follett	0
0.00	Linda Gilroy	0
0.00	Win Griffiths	0
0.00	John Heppell	0
0.00	Helen Jackson	0
0.00	Jenny Jones	1
0.00	Jackie Lawrence	1
0.00	Ivan Lewis	1
0.00	Stephen McCabe	0
0.00	Shona McIsaac	0
0.00	John Maxton	2
0.00	Doug Naysmith	0
0.00	James Plaskitt	0

study. Our contention is that if our model is simply capturing idiosyncratic stylistic differences (in the narrow sense meant in that earlier literature), the restricted version should perform approximately as well as the more general one that uses all words in the vocabulary. Studying Figure 2, we see this is clearly false: there, the bottom line with triangle points is the mean prediction rate (for each speaker, with fivefold cross-validation) from the stop word model. The top line is the mean prediction rate from our model, which has no restrictions on stop words (as in the rest of this paper). It performs about three to five times as well as the pure phrasing model, on average. This implies that there is certainly something more than expressive manner going on: we happily refer to that residual variation as "substance." This does not mean, of course, that the Mosteller and Wallace (1963) approach vocabulary is "wrong" (it is just a special case of ours), but it does suggest our model is doing something statistically useful in terms of capturing practical variation between contemporaneous MPs.

Why do we see this performance difference? From inspection, we note that the fit improvement comes mostly from the middle of the distribution (that is, both our approach and the more simple one perform similarly for the most and least distinctive MPs but not for the median and mean—at least for the sessions we looked at in detail). We suspect this is because while almost everyone will have non-zero use of all of the Mosteller and Wallace (1963) words, our richer vocabulary has much higher variance in use. At the top of the distribution—MPs who are distinctive whatever the vocabulary—this makes no difference. Conversely, at the bottom of the distribution—MPs who use neither vocabulary very much—this also makes no difference. But for MPs in the middle, our much larger vocabulary offers more opportunities to distinguish oneself (for a fixed amount of speaking), and, thus, our model does better for these people.

Before moving to the results, we note that readers may be qualitatively interested in the underlying *tokens* that affect distinctiveness of individuals in the model: in Appendix C, we discuss how these might be obtained and examined.
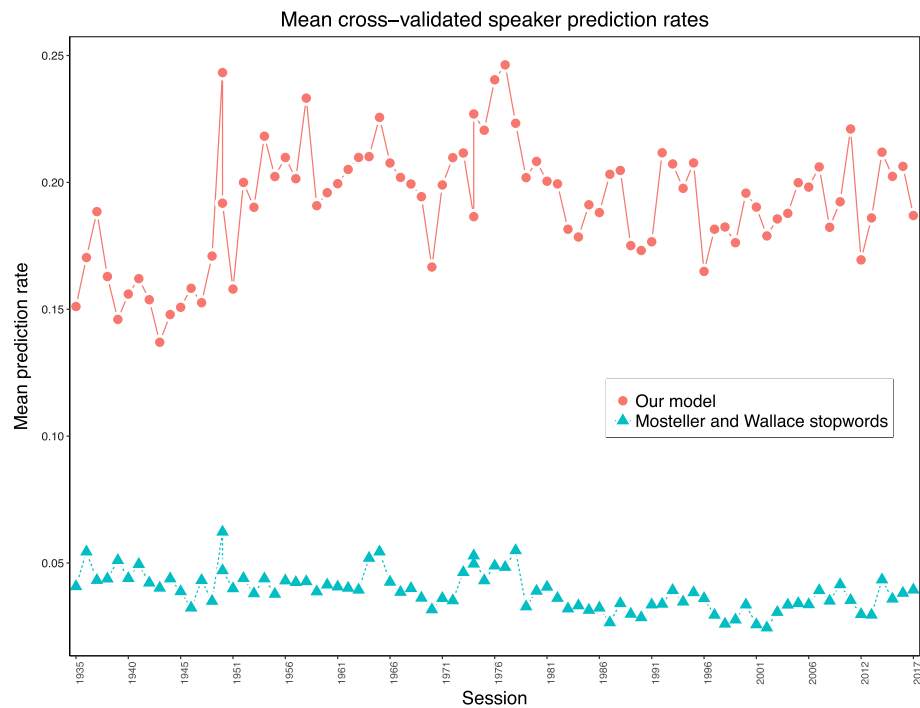
Figure 2. The model picks up more than "style": restricting the model to the Mosteller and Wallace function words only results in much poorer accuracy (bottom [green] line) relative to the full model that uses all words (top [orange] line).

9   Results: Aggregates, Inference and Effects of Service

Our model validated, we now turn to the results of secondary analysis using its output as a measure. Of course, we are limited to observational data in what follows, so any causal claims must be necessarily cautious. Nonetheless, we demonstrate how such measures might be used to draw conclusions here and elsewhere.

9.1   Time Series: Distinctiveness is Not Decreasing

For every session in the aggregate, it is trivial to produce an average (median or mean) distinctiveness score, along with a variance. In Figure 3, we do exactly that. Two observations are immediate. First, as demonstrated in the top panel, the median distinctiveness of MPs does not appear to decrease over time. At the very least, the points and the solid [red] lowess of the same seem fairly stable and certainly not moving downwards. This is also true of the mean (shown via the broken [blue] lowess curve). Indeed, if anything, these averages appear to be *increasing*. For completeness, we also plot the tails of the distribution of scores. These are the broken lines at the top and bottom of the upper plot. While they show some variance, there is no obvious trend in the extremes.

Nor does it seem to be the case that MPs are, on average, becoming more or less spread out in terms of their distinctiveness. This can be seen in the lower plot which reports the variance of the scores over time. The spread decreases until around 1975 (note [red] lowess) before rising again to reach a level approaching the beginning of the data.

Plots can be misleading, but formal statistical examinations suggest that initial impressions are correct. Standard (monotonic) tests—Cox-Stuart (Cox and Stuart 1955) and Mann–Kendall (Mann 1945; Kendall 1975)—reveal no presence of a "trend" in the medians. To summarize then, the average MP has not become more or less distinctive in terms of style over the past eighty years.
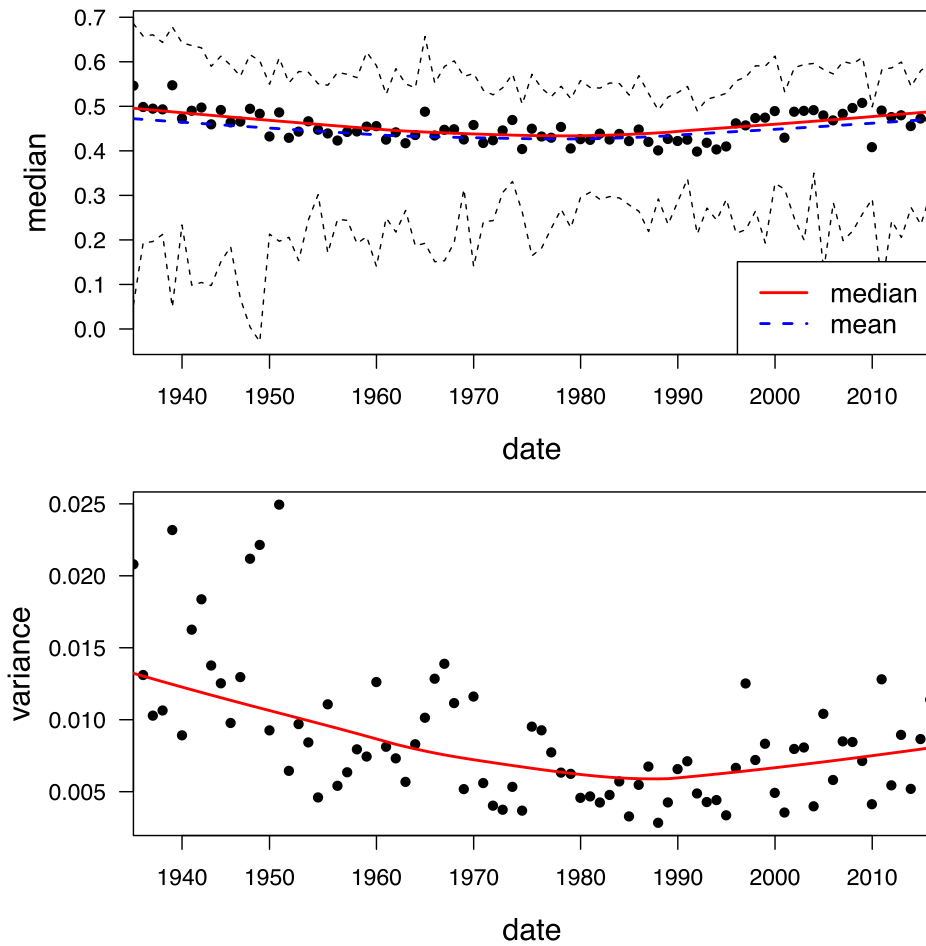
Figure 3. Time series of distinctiveness. In the top panel, the points represent the median, while the solid [red] line represents the lowess of the same. The [blue] broken line is the mean. The [black] broken lines at the top and bottom of the plot are the (empirical) 2.5th and 97.5th percentiles of the data. The bottom panel is the variance over time, plotted with a lowess.

Possibly, assessing "average" MPs is unfair in that the "real" action of decline is at the top end of the distribution: that is, it is the maverick outliers who have disappeared. But the evidence for this is equivocal: while a Mann–Kendall test suggests that the 90th, 95th, and 97.5th percentiles are trending down, Cox–Stuart tests cannot reject the null. Interestingly, by both tests, there is a trend in the lowest percentile we checked—the 2.5th—but it is upwards, not downwards.

9.2 Inference: Serving Longer, Becoming Less Distinctive

Of course, aggregates can oversimplify: it is possible that over time the relationship between being distinctive and other MP features has changed in such a way as to disguise something more profound. To look at this possibility, we begin with panel regressions. Here, the cross-section time series is MP-by-session and is unbalanced since members only serve for a limited number of years. As noted above, we have two covariates (in addition to the fixed effects) for predicting distinctiveness: experience (in session terms) and whether the MP has ever been demoted from government front-bench responsibilities.

In Table 6, we report the results from three specifications. In Column (1), the relevant regression is *pooled*: that is, we treat the entire sample of MPs as a cross-section and provide ordinary least square estimates for the coefficients on the same. In Column (2), we add MP-level (i.e. unit-level) fixed effects. In Column (3), we use MP-level and session-level (that is, time) fixed effects. The

Table 6. Effect of experience and demotion on distinctiveness of MPs. The first model pools all observations. The second adopts a panel structure with MP-fixed effects. The third uses MP-fixed and session-fixed effects.

	Dependent variable: distinctiveness		
	(1)	**(2)**	**(3)**
Experience	−0.0003	−0.001	−0.002
	(0.0001)	(0.0001)	(0.0002)
Demoted	0.004	0.006	0.003
	(0.003)	(0.004)	(0.004)
Constant	0.443		
	(0.001)		
MP-fixed effects	✗	✓	✓
Session-fixed effects	✗	✗	✓
Observations	16,022	16,022	16,022
R^2	0.001	0.006	0.069
F statistic	4.995	39.602	11.889
	(df = 2; 16,019)	(df = 2; 13,396)	(df = 83; 13,315)

results of an F-test and a Baltagi and Li (1990) Lagrange-Multiplier test suggest that time-fixed effects are indeed warranted in this case.

Regardless of the specification, there is a negative association of experience with distinctiveness.[10] That is, as MPs serve longer terms, they appear less and less different (relative to others) in terms of their speech. Being demoted at some point does not seem to change this dynamic. Substantively (for the best fitting model), a one standard deviation increase in experience (around nine sessions) decreases distinctiveness by around 0.02 (around one-fifth of that variable's standard deviation). So these effects are not huge.

What causal story could we tell here? One possibility is obvious: MPs have new or passionate beliefs when they enter parliament, and communicate those. Over time, to curry favor with party bosses or simply through some broader socialization (or indeed, human aging) process, they moderate the expression of those views. So it is not that MPs get "stuck in their ways"—it is more that they get stuck in the ways of the House. Certainly, this would be commensurate with findings from Eggers and Spirling (2016) for division voting in the nineteenth century. The point there is that institutions exert considerable pull on behavior, and it is almost impossible in the long term to be a maverick. Another possibility is that selection effects play a role. In particular, it could be that only the most "obedient" (from the perspective of distinctiveness) MPs survive multiple rounds of elections. One route for this might be that highly distinctive MPs (especially those out of step ideologically with the rest of the parliamentary party) find themselves less able to attract central party support in their constituencies and, thus, lose elections.

9.3 Model Inference: The Effect of Seniority is Changing

Figure 4 reports our final findings. Now, we divide the data into periods of uninterrupted party rule by one party. For each period, we record the coefficient on the experience of government party backbenchers in terms of its effects on distinctiveness. This point estimate is plotted via a letter symbolizing (L)abour or (C)onservative control at the time, and each estimate is plotted at the beginning of the period in question. Thus, for example, the Conservative control of the

---

10 As a robustness test, we also fit the panel regression using heteroscedastic "robust" standard errors (in the sense of White 1980). This makes no difference to our conclusions. Nor does, in addition, correcting for potential autocorrelation (in the sense of Arellano 1987).
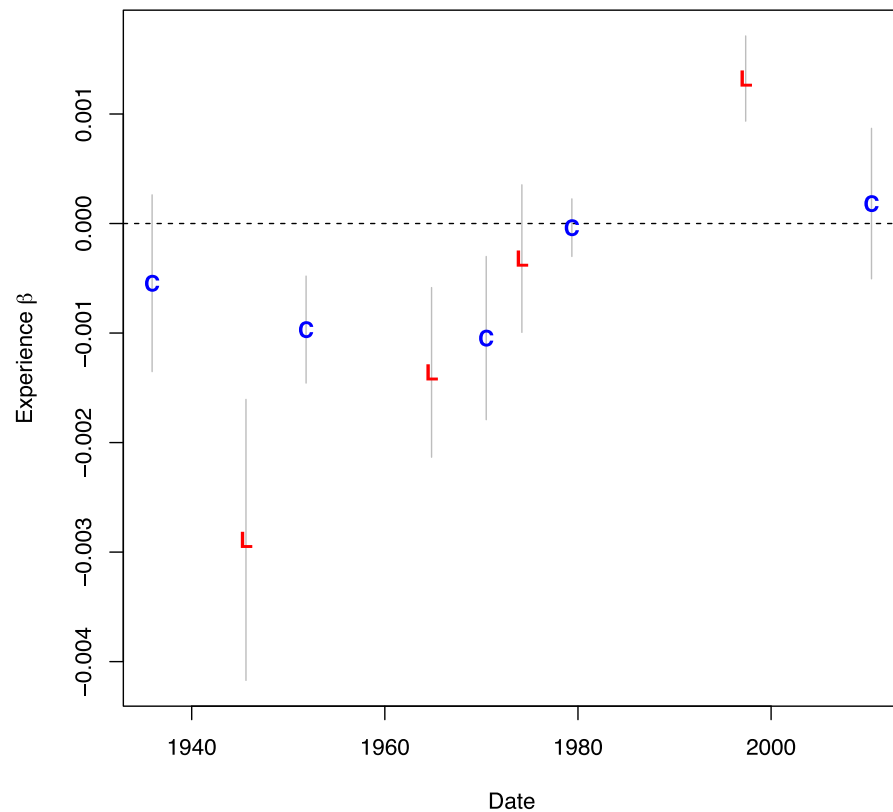

Figure 4. Effect of longer service on distinctiveness, over time. Data is broken up into periods of uninterrupted one party rule, with one coefficient (and 95% confidence interval) per period.

Premiership that began in 1979 and ended in 1997 is marked by a coefficient in 1979. For each regression, we report the 95% confidence interval on the figure. We see that for the vast majority of our data, the effect of seniority is negative: the coefficient is less than zero, implying that more senior members are less distinctive. This switches during the Blair government, with a positive coefficient for that period; it continues into the coalition government and beyond, though the coefficient is not distinguishable from zero.

We interpret this finding as being a consequence of two related events. Most immediate was the change to the internal ideological composition of the ruling Labour party at this time. In particular, Blair's backbenchers included a sizable number of senior "Old Labour" legislators whose views were fundamentally at odds with his "New Labour" governing agenda (see Spirling and Quinn 2010). But, second, owing to the large majority the Labour party had between 1997 and 2010, these "rebels" were more able than they might have been previously to speak their (distinctive) minds without fear of bringing the government down. It is unclear whether the age effect we note is now a permanent fixture of Westminster life.

10   Discussion

From the time of Burke to the present, ideal MP behavior has been debated. Today, voters say they like independent-minded MPs willing to express their positions (Vivyan and Wagner 2015); but they make voting decisions based on many other factors. Thus, it is unclear whether MP distinctiveness (from others) is likely to be electorally helpful or not. Regardless, a first step in the process of investigation is measuring the quantity, validating that measure, and using it to draw aggregate and individual-level conclusions about how MPs may have changed over time.

In this paper, we proposed and implemented a new statistical approach informed by earlier efforts in "stylometry." We began with the work of Mosteller and Wallace ([1963]) but provided a more general technique not limited to a pre-selected vocabulary; nor is it limited to comparing authors in a pairwise fashion. It is, nonetheless, based on simple naive Bayes principles. This model works well insofar as it provides substantively valid results along with useful auxiliary uncertainty estimates. While we undertook various validation tasks, it was beyond the scope of the paper to use human ratings to assess the merits of our measure. We would suggest that a next step for this research is to see whether coders (possibly crowdworkers) agree with what the model avers is a distinctive speaker or speech. We also think that the model itself can be made more flexible and more "realistic." For example, as we discussed above, there are a multitude of machine-learning tools that can outperform a naive Bayes setup. And one might want to relax, i.e. be more careful about the independence assumption in a legislative setting where speakers are likely responding to others.

In principle, it could be taken to many other political environments, including parliaments—especially if actors there have the opportunity for relatively unconstrained speech (as backbenchers do in the House of Commons). Of course, properly interpreting distinctiveness estimates in other contexts will require thinking carefully about the incentives and constraints that actors face. For us, measuring distinctiveness was of intrinsic interest, in the sense that we wanted to comment on how this quantity changed over time. But we can imagine the model being extrinsically helpful in cases where one wishes to monitor the input of different (but known) actors to an otherwise "anonymous" document—for example, determining who wrote which passages of a Congressional bill.

In the UK case we studied, there is very little evidence that MPs are becoming less distinctive. Possibly, there are fewer "big beasts," but most MPs are as different or as similar to their colleagues as they ever were. More interestingly, perhaps, the effect of seniority is changing. For most of the twentieth century, longer-serving backbenchers tended to be less distinctive. But at least since the Blair victory in 1997, perhaps due to the intake of young MPs more solely focused on career promotion or perhaps due to a structural change in the nature of leftist politics, it is older MPs who emerge as more distinctive speakers.

This finding is in line with work on the determinants of rebellion in recent times, from the likes of Cowley and Childs ([2003]) and Benedetto and Hix ([2007]). Perhaps, we are seeing the sidelining of socialization as the traditionally dominant force in MP careers (Eggers and Spirling [2016]). If so, it would be good to know more about *what* distinctive MPs talk about: is it simply a matter of linguistic choices, or something more related to topics of debate? And we would like to know more about *who* is speaking differently, in terms of their (re)election prospects, their career histories, and ideological positions. We leave such efforts for future work.

Acknowledgements

Data Availability Statement

Open-source R software that implements our model is available at:
https://cran.r-project.org/web/packages/stylest/index.html.
The replication materials for this paper can be found at Huang, Perry, and Spirling ([2019]).

Supplementary material

For supplementary material accompanying this paper, please visit
https://doi.org/10.1017/pan.2019.49.

References

Airoldi, E. M., S. E. Fienberg, and K. K. Skinner. 2007. "Whose Ideas? Whose Words? Authorship of Ronald Reagan's Radio Addresses." *PS: Political Science and Politics* 40(3):501–506.

Arellano, M. 1987. "PRACTITIONERS CORNER: Computing Robust Standard Errors for Within-groups Estimators." *Oxford Bulletin of Economics and Statistics* 49(4):431–434.

Baltagi, B. H., and Q. Li. 1990. "A Lagrange Multiplier Test for the Error Components Model with Incomplete Panels." *Econometric Reviews* 9(1):103–107.

Benedetto, G., and S. Hix. 2007. "The Rejected, the Ejected, and the Dejected: Explaining Government Rebels in the 2001–2005 British House of Commons." *Comparative Political Studies* 40(7):755–781.

Cowley, P., R. Campbell, N. Vivyan, and M. Wagner. 2016. "Legislator Dissent as a Valence Signal." *British Journal of Political Science*.

Cowley, P. 2002. *Revolts and Rebellions: Parliamentary Voting under Blair*. London: Politico's.

Cowley, P., and S. Childs. 2003. "Too Spineless to Rebel? New Labour's Women MPs." *British Journal of Political Science* 33(3):345–365.

Cox, D. R., and A. Stuart. 1955. "Some Quick Sign Tests for Trend in Location and Dispersion." *Biometrika* 42(1–2):80–95.

Crowe, E. 1986. "The web of authority: Party Loyalty and Social Control in the British House of Commons." *Legislative Studies Quarterly* 11(2):161–185.

Eggers, A. C., and A. Spirling. 2016. "Party Cohesion in Westminster Systems: Inducements, Replacement and Discipline in the House of Commons, 1836–1910." *British Journal of Political Science* 46(3):567–589.

Eggers, A. C., and A. Spirling. 2017. "Incumbency Effects and the Strength of Party Preferences: Evidence from Multiparty Elections in the United Kingdom." *The Journal of Politics* 79(3):903–920.

Gaines, B. J. 1998. "The Impersonal Vote? Constituency Service and Incumbency Advantage in British Elections, 1950–92." *Legislative Studies Quarterly* 23(2):167–195.

Gonzales, A. L., J. T. Hancock, and J. W. Pennebaker. 2010. "Language Style Matching as a Predictor of Social Dynamics in Small Groups." *Communication Research* 37(1):3–19.

Green, J., and S. B. Hobolt. 2008. "Owning the Issue Agenda: Party Strategies and Vote Choices in British Elections." *Electoral Studies* 27(3):460–476.

Hanretty, C., B. E. Lauderdale, and N. Vivyan. 2017. "Dyadic Representation in a Westminster System." *Legislative Studies Quarterly* 42(2):235–267.

Heath, A., J. Curtice, R. Jowell, G. Evans, J. Field, and S. Witherspoon. 1991. *Understanding Political Change: the British Voter, 1964–1987*. Oxford: Pergamon.

Heath, O. 2015. "Policy Representation, Social Representation and Class Voting in Britain." *British Journal of Political Science* 45(1):173–193.

Huang, L., P. O. Perry, and A. Spirling. 2019 "Replication Data for: A General Model of Author "Style" with Application to the UK House of Commons, 1935–2018." https://doi.org/10.7910/DVN/VJ9QDB, Harvard Dataverse, V1.

Jackson, N., and D. Lilleker. 2011. "Microblogging, Constituency Service and Impression Management: UK MPs and the Use of Twitter." *The Journal of Legislative Studies* 17(1):86–105.

Kam, C. J. 2009. *Party Discipline and Parliamentary Politics*. New York: Cambridge University Press.

Kam, C., W. T. Bianco, I. Sened, and R. Smyth. 2010. "Ministerial Selection and Intraparty Organization in the Contemporary British Parliament." *American Political Science Review* 104(2):289–306.

Kellermann, M. 2012. "Estimating Ideal Points in the British House of Commons Using Early Day Motions." *American Journal of Political Science* 56(3):757–771.

Kendall, M. 1975. *Rank Correlation Methods*. 4th edn. London: Charles Griffin.

King, A. 1981. "The Rise of the Career Politician in Britain and Its Consequences." *British Journal of Political Science* 11(3):249–285.

Lauderdale, B. E., and A. Herzog. 2016. "Measuring Political Positions from Legislative Speech." *Political Analysis* 24(3):374–394.

Lijphart, A. 2012. *Patterns of Democracy: Government Forms and Performance in Thirty-six Countries*. New Haven, CT: Yale University Press.

Mann, H. B. 1945. "Non-parametric Tests against Trend." *Econometrica* 13(3):163–171.

Mosteller, F., and D. Wallace. 1964. *Inference and Disputed Authorship: The Federalist*. Reading, MA: Addison-Wesley Publishing Company.

Mosteller, F., and D. L. Wallace. 1963. "Inference in an Authorship Problem." *Journal of the American Statistical Association* 58(302):275–309.

Negrine, R., and D. Lilleker. 2003. "The Rise of a Proactive Local Media Strategy in British Political Communication: Clear Continuities and Evolutionary Change 1966–2001." *Journalism Studies* 4(2):199–211.

Norris, P., and J. Lovenduski. 1995. *Political Recruitment: Gender, Race and Class in the British Parliament*. Cambridge: Cambridge University Press.

O'Grady, T. 2018. "Careerists Versus Coal-Miners: Welfare Reforms and the Substantive Representation of Social Groups in the British Labour Party." *Comparative Political Studies* 52(4):1–35.

Pattie, C., E. Fieldhouse, and R. J. Johnston. 1994. "The Price of Conscience: The Electoral Correlates and Consequences of Free Votes and Rebellions in the British House of Commons, 1987–92." *British Journal of Political Science* 24(3):359–380.

Proksch, S.-O., and J. B. Slapin. 2012. "Institutional Foundations of Legislative Speech." *American Journal of Political Science* 56(3):520–537.

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

Rheault, L., K. Beelen, C. Cochrane, and G. Hirst. 2016. "Measuring Emotion in Parliamentary Debates with Automated Textual Analysis." *PLOS ONE* 11(12):1–18.

Rheault, L., and C. Cochrane. 2019. "Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora." *Political Analysis* 28(1):112–133.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, B. Albertson, and D. G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–1082.

Rodman, E. 2019. "A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors." *Political Analysis* 28(1):87–111.

Rush, M. 2001. *The Role of the Member of Parliament Since 1868: From Gentlemen to Players*. Oxford: Oxford University Press.

Rush, M., and P. Giddings. 2011. *Parliamentary Socialisation*. London: Palgrave Macmillan.

Rush, M., and S. Childs. 2004. *A Changing Culture*. London: Palgrave Macmillan.

Shaw, E. 2001. "New Labour: New Pathways to Parliament." *Parliamentary Affairs* 54(1):35–53.

Slapin, J., J. Kirkland, J. Lazzaro, P. Leslie, and T. O'Grady. 2017. "Ideology, Grandstanding, and Strategic Party Disloyalty in the British Parliament." *American Political Science Review* 112(1):15–30.

Spirling, A., and K. Quinn. 2010. "Identifying Intraparty Voting Blocs in the UK House of Commons." *Journal of the American Statistical Association* 105(490):447–457.

Vivyan, N., and M. Wagner. 2012. "Do Voters Reward Rebellion? The Electoral Accountability of MPs in Britain." *European Journal of Political Research* 51(2):235–264.

Vivyan, N., and M. Wagner. 2015. "What Do Voters Want From Their Local MP?" *The Political Quarterly* 86(1):33–40.

Weissberg, R. 1978. "Collective vs. Dyadic Representation in Congress." *American Political Science Review* 72(2):535–547.

White, H. 1980. "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica: Journal of the Econometric Society* 48(4):817–838.