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

Leslie Huang† Patrick O. Perry‡ Arthur Spirling§

Summary
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.

Key words: Political Science, House of Commons, Stylometry, Text as Data, Regression

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

11,345 words (excl appendix)

---

*We thank Jack Blumenau for assistance with data. Jacob Eisenstein, Jennifer Holmes, Elena Labzina, Will Lowe and Juraj Medzhorsky provided helpful feedback on an earlier draft.

†Graduate Student, Center for Data Science, New York University, 60 Fifth Avenue, New York, USA, lesliehuang@nyu.edu

‡Senior Data Scientist, Oscar Health; Visiting Scholar, Center for Data Science, 60 Fifth Avenue, New York University, New York, USA. pperry@stern.nyu.edu

§Associate Professor of Politics and Data Science, New York University, 19 W4th St, New York, USA. arthur.spirling@nyu.edu
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 an idiosyncratic MP 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). Related, 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, floccinaucinahilipilification 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 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 so might be innately less distinctive. On the other, 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. 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

---

1These 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.
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 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 \( \eta_s \) 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 \( \eta_s \) from \( \eta_t \) 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 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 heterogenous 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
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 10 speeches of at least 3 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 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

\(^2\)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 2013–14 session, our data includes speeches up until March 2014. For the 2017–18 session, our data includes speeches up until March 2018.
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 \( i \).

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 \( \eta_s \).

Suppose that we know \( \eta_s \) 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 \( v \) 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

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

We define this quantity as the distinctiveness of target speaker \( t \) relative to reference set \( \{s, t\} \) in context \( c \).

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

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

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

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

In this case,

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

With this final expression we can compute $D_t$ by taking the difference between $\eta_t$ and the average $\bar{\eta}_s$ 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 $D_t$ that becomes our ‘distinctiveness’ dependent variable in what follows.

### 4.1 Standard Errors

The distinctiveness measure $D_t = D_t(S_c; I_t)$ 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 affect $D_t$:

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 $\eta_{sv}$ 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 $D_t$ 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 $\eta_{sv}$. 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 $\eta$ 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 $D_t$ 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 $D_t$ due to the variability in $I_t$ and $S_c$ first define for each $i \in \mathbb{I}_t$.

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

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

\[ E\{D_i(S)\} = \mu_i, \quad \text{var}\{D_i(S)\} = \frac{\sigma_i^2}{m}. \]

where $\mu_i$ and $\sigma_i^2$ are the mean and variance of $D_{is}$ as $s$ ranges over population $S_c$.

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, $\text{var}(D) = E\{\text{var}(D | I)\} + \text{var}\{E(D | I)\}$, where the outer expectation and variance on the right hand side are over the random set $I$. Using the independence of the speeches yields

\[ E(D | I) = \frac{1}{n} \sum_{i \in I} \mu_i, \quad \text{var}(D | I) = \frac{1}{n^2} \sum_{i \in I} \sigma_i^2. \]

Hence,

\[ \text{var}(D) = \frac{1}{n} \text{var}(\mu_i) + \frac{1}{nm} E(\sigma_i^2), \]

the variance and expectation being computed over a random $i$ drawn from population $\mathbb{I}_t$. Define estimate $\hat{\mu}_t = D_i(S_c)$ and set $\hat{\sigma}_i^2$ to be the empirical variance of $D_{is}$ as $s$ ranges over $S_c$. We estimate $\text{var}(\mu_i)$ by the empirical variance of $\hat{\mu}_t$ 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. 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 $\eta$ probabilities to be straightforwardly summed. But there are also differences from the usual set up. 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.

---

5We are grateful to an anonymous referee for pointing this out and encouraging us to make the links explicit.
Table 1: Summaries for variables in our data

<table>
<thead>
<tr>
<th>Description</th>
<th>Minimum</th>
<th>Median</th>
<th>Mean</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>2.40</td>
<td>9.92</td>
<td>9.88</td>
<td>15.68</td>
</tr>
<tr>
<td>Proportion Demoted</td>
<td>0.00</td>
<td>0.13</td>
<td>0.13</td>
<td>0.31</td>
</tr>
<tr>
<td>Distinctiveness</td>
<td>−0.76</td>
<td>0.45</td>
<td>0.44</td>
<td>1.83</td>
</tr>
<tr>
<td>Avg. num. of speeches per MP (per career)</td>
<td>10.00</td>
<td>38.74</td>
<td>52.08</td>
<td>482.94</td>
</tr>
<tr>
<td>Avg. num. of speeches per MP (per session)</td>
<td>26.14</td>
<td>57.87</td>
<td>56.68</td>
<td>135.87</td>
</tr>
<tr>
<td>Total speeches per session</td>
<td>3424</td>
<td>10979</td>
<td>11124</td>
<td>29212</td>
</tr>
<tr>
<td>Speaking MPs per session</td>
<td>102.0</td>
<td>187.0</td>
<td>193.0</td>
<td>288.0</td>
</tr>
<tr>
<td>Mean Speech Length</td>
<td>115.8</td>
<td>188.9</td>
<td>183.6</td>
<td>229.6</td>
</tr>
</tbody>
</table>

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. 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 take-away is that the (average) means and medians are generally very close, implying little skew in these variables.

---

6We augment the data with information on when—what session—the MPs speaking in the first period of the data entered Parliament.
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 as 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, 5 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 4 folds of data, repeating this procedure 5 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 (e.g.)
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 10
speeches of 3 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 (2015–16) 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 $D_i$ point estimate is by far the largest in the data, and appears to 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).

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 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

---

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).
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 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 $\bar{D}$, 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 (‘PCP’ in the tables below). We use 5-fold 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 20 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’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 ($p < 0.05$): 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 under-secretary 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 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 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 Euroskeptic, appears in this list too—in contrast to his rebellious colleagues who generally appear to be more distinctive overall.
Table 2: Most distinctive MPs November 1, 1995–October, 31, 1996 in parliament, by proportion of speeches correctly predicted.

<table>
<thead>
<tr>
<th>PCP</th>
<th>Name</th>
<th>Times Mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.41</td>
<td>Piers Merchant</td>
<td>1</td>
</tr>
<tr>
<td>0.41</td>
<td>Phil Gallie</td>
<td>3</td>
</tr>
<tr>
<td>0.43</td>
<td>Norman Fowler</td>
<td>3</td>
</tr>
<tr>
<td>0.43</td>
<td>Anthony Steen</td>
<td>2</td>
</tr>
<tr>
<td>0.43</td>
<td>Terence Higgins</td>
<td>6</td>
</tr>
<tr>
<td>0.43</td>
<td>Eric Pickles</td>
<td>0</td>
</tr>
<tr>
<td>0.44</td>
<td>James Pawsey</td>
<td>3</td>
</tr>
<tr>
<td>0.45</td>
<td>Andrew Hunter</td>
<td>4</td>
</tr>
<tr>
<td>0.46</td>
<td>Peter Brooke</td>
<td>5</td>
</tr>
<tr>
<td>0.47</td>
<td>Peter Bottomley</td>
<td>2</td>
</tr>
<tr>
<td>0.48</td>
<td>Edward Leigh</td>
<td>6</td>
</tr>
<tr>
<td>0.48</td>
<td>Robert Hughes</td>
<td>2</td>
</tr>
<tr>
<td>0.48</td>
<td>Patrick Nicholls</td>
<td>5</td>
</tr>
<tr>
<td>0.48</td>
<td>John Wilkinson</td>
<td>4</td>
</tr>
<tr>
<td>0.51</td>
<td>David Howell</td>
<td>4</td>
</tr>
<tr>
<td>0.52</td>
<td>David Mellor</td>
<td>3</td>
</tr>
<tr>
<td>0.64</td>
<td>Michael Colvin</td>
<td>0</td>
</tr>
<tr>
<td>0.66</td>
<td>Barry Field</td>
<td>4</td>
</tr>
<tr>
<td>0.72</td>
<td>Edward Heath</td>
<td>19</td>
</tr>
<tr>
<td>0.75</td>
<td>David Wilshire</td>
<td>4</td>
</tr>
</tbody>
</table>
Table 3: Least distinctive MPs November 1, 1995–October, 31, 1996 in parliament, by proportion of speeches correctly predicted.

<table>
<thead>
<tr>
<th>PCP</th>
<th>Name</th>
<th>Times Mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>Alan Haslehurst</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Michael Lord</td>
<td>2</td>
</tr>
<tr>
<td>0.00</td>
<td>Peter Tapsell</td>
<td>1</td>
</tr>
<tr>
<td>0.00</td>
<td>Tim Yeo</td>
<td>2</td>
</tr>
<tr>
<td>0.00</td>
<td>Charles Hendry</td>
<td>1</td>
</tr>
<tr>
<td>0.00</td>
<td>Nicholas Budgen</td>
<td>7</td>
</tr>
<tr>
<td>0.00</td>
<td>Matthew Carrington</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Michael Jopling</td>
<td>2</td>
</tr>
<tr>
<td>0.00</td>
<td>Gary Waller</td>
<td>2</td>
</tr>
<tr>
<td>0.00</td>
<td>Anthony Durant</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Hartley Booth</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>John Butcher</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Donald Thompson</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Edwina Currie</td>
<td>14</td>
</tr>
<tr>
<td>0.00</td>
<td>Jerry Hayes</td>
<td>2</td>
</tr>
<tr>
<td>0.00</td>
<td>Peter Butler</td>
<td>3</td>
</tr>
<tr>
<td>0.01</td>
<td>John Whittingdale</td>
<td>1</td>
</tr>
<tr>
<td>0.02</td>
<td>Michael Neubert</td>
<td>3</td>
</tr>
<tr>
<td>0.03</td>
<td>Alan Duncan</td>
<td>3</td>
</tr>
<tr>
<td>0.03</td>
<td>Nick Hawkins</td>
<td>0</td>
</tr>
</tbody>
</table>
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 ($p < 0.05$): 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. 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 NHS. Finally, Barry Jones was the chair of the Intelligence and Security Committee of Parliament.

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.

---

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

9One 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 signalling 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 4: Most distinctive MPs November 1, 1998–October, 31, 1999 in parliament, by proportion of speeches correctly predicted.

<table>
<thead>
<tr>
<th>PCP</th>
<th>Name</th>
<th>Times Mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.48</td>
<td>Hazel Blears</td>
<td>1</td>
</tr>
<tr>
<td>0.48</td>
<td>Denzil Davies</td>
<td>0</td>
</tr>
<tr>
<td>0.49</td>
<td>Geraint Davies</td>
<td>0</td>
</tr>
<tr>
<td>0.50</td>
<td>Gareth Thomas</td>
<td>1</td>
</tr>
<tr>
<td>0.51</td>
<td>Diane Abbott</td>
<td>3</td>
</tr>
<tr>
<td>0.51</td>
<td>Peter Temple-Morris</td>
<td>0</td>
</tr>
<tr>
<td>0.53</td>
<td>Tony Benn</td>
<td>19</td>
</tr>
<tr>
<td>0.53</td>
<td>Martin Linton</td>
<td>0</td>
</tr>
<tr>
<td>0.53</td>
<td>David Winnick</td>
<td>5</td>
</tr>
<tr>
<td>0.54</td>
<td>Harry Barnes</td>
<td>1</td>
</tr>
<tr>
<td>0.57</td>
<td>Norman Godman</td>
<td>1</td>
</tr>
<tr>
<td>0.58</td>
<td>Tam Dalyell</td>
<td>3</td>
</tr>
<tr>
<td>0.59</td>
<td>Bill Rammell</td>
<td>0</td>
</tr>
<tr>
<td>0.60</td>
<td>Joan Walley</td>
<td>0</td>
</tr>
<tr>
<td>0.62</td>
<td>John McDonnell</td>
<td>1</td>
</tr>
<tr>
<td>0.66</td>
<td>Roger Berry</td>
<td>1</td>
</tr>
<tr>
<td>0.70</td>
<td>David Hinchliffe</td>
<td>1</td>
</tr>
<tr>
<td>0.72</td>
<td>Donald Anderson</td>
<td>3</td>
</tr>
<tr>
<td>0.73</td>
<td>Stuart Bell</td>
<td>3</td>
</tr>
<tr>
<td>0.75</td>
<td>Barry Jones</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 5: Least distinctive MPs November 1, 1998–October, 31, 1999 in parliament, by our measure.

<table>
<thead>
<tr>
<th>PCP</th>
<th>Name</th>
<th>Times Mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>Charlotte Atkins</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Nigel Beard</td>
<td>1</td>
</tr>
<tr>
<td>0.00</td>
<td>Ben Bradshaw</td>
<td>1</td>
</tr>
<tr>
<td>0.00</td>
<td>Ronnie Campbell</td>
<td>3</td>
</tr>
<tr>
<td>0.00</td>
<td>Dennis Canavan</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Ivor Caplin</td>
<td>1</td>
</tr>
<tr>
<td>0.00</td>
<td>Helen Brinton</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Barbara Follett</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Linda Gilroy</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Win Griffiths</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>John Heppell</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Helen Jackson</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Jenny Jones</td>
<td>1</td>
</tr>
<tr>
<td>0.00</td>
<td>Jackie Lawrence</td>
<td>1</td>
</tr>
<tr>
<td>0.00</td>
<td>Ivan Lewis</td>
<td>1</td>
</tr>
<tr>
<td>0.00</td>
<td>Stephen McCabe</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>Shona McIsaac</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>John Maxton</td>
<td>2</td>
</tr>
<tr>
<td>0.00</td>
<td>Doug Naysmith</td>
<td>0</td>
</tr>
<tr>
<td>0.00</td>
<td>James Plaskitt</td>
<td>0</td>
</tr>
</tbody>
</table>
8 Validation III: Phrasing v 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 70 function words (i.e. stop words) from the original Mosteller and Wallace (1963) 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 5-fold cross validation) from the stop word model. The top line is the same quantity but our the model as in the rest of this paper, with no restrictions. It performs about three to five times as well as the pure phrasing model, on average. This implies 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.
Figure 2: The model picks up more than ‘style’: restricting the model to the Mosteller & Wallace function words only results in much poorer accuracy (bottom [green] line) relative to the full model that uses all words (top [orange] line).

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.

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 doesn’t 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.
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 at 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.
Plots can be misleading, but formal statistical examinations suggest initial impressions are correct. Standard (monotonic) tests—Cox-Stuart (Cox and Stuart, 1955) and Mann-Kendall (Mann, 1945; Kendall, 1975) (both $p > 0.05$)—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 80 years.

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 the 90th, 95th and 97.5th percentiles are trending down, Cox-Stuart tests cannot reject the null. Interestingly, by both tests ($p < 0.01$), 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’s 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 OLS 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 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 distinctive-
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.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable: distinctiveness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>experience</td>
<td>$-0.0003^{***}$</td>
<td>$-0.001^{***}$</td>
<td>$-0.002^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.0001)$</td>
<td>$(0.0001)$</td>
<td>$(0.0002)$</td>
</tr>
<tr>
<td>demoted</td>
<td>0.004</td>
<td>0.006</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>$(0.003)$</td>
<td>$(0.004)$</td>
<td>$(0.004)$</td>
</tr>
<tr>
<td>constant</td>
<td>0.443$^{***}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$(0.001)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MP-fixed effects</td>
<td>$\times$</td>
<td>$\checkmark$</td>
<td>$\checkmark$</td>
</tr>
<tr>
<td>Session-fixed effects</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\checkmark$</td>
</tr>
<tr>
<td>Observations</td>
<td>16,022</td>
<td>16,022</td>
<td>16,022</td>
</tr>
<tr>
<td>R$^2$</td>
<td>0.001</td>
<td>0.006</td>
<td>0.069</td>
</tr>
<tr>
<td>F Statistic</td>
<td>$4.995^{***}$</td>
<td>$39.602^{***}$</td>
<td>$11.889^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(df = 2; 16019)$</td>
<td>$(df = 2; 13396)$</td>
<td>$(df = 83; 13315)$</td>
</tr>
</tbody>
</table>

Note: $^{***}p<0.01$
ness ($p < 0.01$). 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 9 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 19th 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 Premiership that began in 1979 and ended in 1997 is marked by a coefficient in 1979. For each regression, we

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).
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 secondly, 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 dis-
tinctiveness (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
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.
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 set up. 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 20th 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 side-lining 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.

A Relationship Between our Approach and multinominal naive Bayes

In the conventional multinomial naive Bayes arrangement used for text, interest focuses on placing a document (for us, a speech, \( i \)) into one of two classes (for us, having been spoken by a member of parliament, \( s \) or \( t \)) using information about the terms used (for us, the counts \( x_{iv} \) of the tokens in speech \( i \)). We would express this as

\[
Pr(t|i) \propto Pr(t) \prod_{v \in V_c} Pr(v|t)^{x_{iv}},
\]

(2)

where, on the right hand side, \( Pr(t) \) is our prior that speech comes from member \( t \), and the likelihood is the product—over all terms making up speech \( i \)—of the probability of observing that word from member \( t \). Recall that the exponent \( x_{iv} \) is the number of times a particular \( v \) occurs in speech \( i \), and note that if that frequency is zero (i.e. that element of the vocabulary does not occur in the speech), simply multiplies the product by one. The “naive” independence assumption allows us to compose the product this way, while each \( Pr(v|t) \) is assumed to be a multinomial distribution. To reiterate then, this is the “multinomial naive Bayes” approach.

Keeping within the standard set up, consider the ratio of the probability that \( t \) spoke speech \( i \) relative to \( s \) (our only other speaker, initially). That is then the ratio

\[
\frac{Pr(t|i)}{Pr(s|i)} \propto \frac{Pr(t) \prod_{v \in V_c} Pr(v|t)^{x_{iv}}}{Pr(s) \prod_{v \in V_c} Pr(v|s)^{x_{iv}}},
\]

(3)

and appears in, for example, [Hand and Yu (2001)]. If one assumes equal priors on the members (which we do), this expression is

\[
\frac{Pr(t|i)}{Pr(s|i)} \propto \prod_{v \in V_c} Pr(v|t)^{x_{iv}}.
\]

(4)

Taking logs, we obtain

\[
\log \left( \frac{Pr(t|i)}{Pr(s|i)} \right) \propto \sum_{v \in V_c} x_{iv} \log(Pr(v|t)) - \sum_{v \in V_c} x_{iv} \log(Pr(v|s)).
\]

(5)

Collecting terms, and writing \( \log(Pr(v|s)) \) as \( \eta_{sv} \), we have

\[
\log \left( \frac{Pr(t|i)}{Pr(s|i)} \right) \propto \sum_{v \in V_c} x_{iv}(\eta_{tv} - \eta_{sv}).
\]

(6)
Recall that our core distinction measure as given in Equation (1) was

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

The resemblance between this and Equation (6) is obvious. Indeed, the only difference is that we divide by the length of the speech, but that does not affect the maximum \textit{a posteriori} class (i.e. speaker) designation for a given speech.

In the traditional arrangement, prediction into multiple classes can be done in various ways. This includes “one-vs-rest” techniques, which make the problem binary, in the sense that one is placing a document into one of two classes: the class of interest (for us, an MP) or an “other” category, composed of all other possible classes (for us, all other MPs). We follow a similar strategy in principle, but in practice, we take average pairwise distances from a given speaker to all others to obtain our estimates (which is not usually done in the “one-vs-rest” arrangement).

### B A Reexamination of the Federalist Papers Authorship Problem

Recall that in the original [Mosteller and Wallace (1963)](https://www.jstor.org/stable/20023437), the goal was to determine whether Hamilton or Madison wrote the twelve mystery papers. Now we re-examine that question, except that we also consider John Jay as a possible author. We can do this because unlike in [Mosteller and Wallace (1963)](https://www.jstor.org/stable/20023437), our method is not restricted to a single pairwise comparison of candidate producers of the text. To keep the problem otherwise similar, we use the same vocabulary as in the earlier presentation (the seventy function words).

Our results are pictured in Figure 5. There each bar sums to one as a total probability of authorship for each disputed Federalist Paper: the light gray part corresponds the probability that Madison was the author, dark gray corresponds to Hamilton and the mid-gray part to Jay. Reassuringly from a validation perspective, we get results essentially identical to [Mosteller and Wallace (1963)](https://www.jstor.org/stable/20023437). We mean this in the sense that Madison is by a large margin the most likely author in all cases, with the exception of \textit{Federalist} 55 where—as in the original presentation—the evidence is more equivocal. But again, note that our model is more flexible: we can include Jay (or indeed any other candidates) as a potential author “automatically” in our comparison. It so happens that the estimated probability for Jay is near zero, but that is as we would expect given our priors in this case. In short, our more general model works when taken to this canonical application.

### C Influential Tokens

A nice feature of our approach is that we can inspect the ‘most influential’ terms, both within certain sessions and over the data as a whole. Term influence is estimated separately for each speaker–term pair in each session, and is calculated for term \(v\) and speaker \(s\) by multiplying the mean (over speaker \(s\)’s speeches) term frequency count \(\frac{1}{|I_s|} \sum_{i \in I_s} x_{iv}\) with the mean-centered term frequency rate \(\eta_{sv}\) (the latter term is \(\eta_{sv} - \frac{1}{|S|} \sum_{s \in S} \eta_{sv}\)). Thus, the difference between the speaker’s rate of use of word \(v\) and the mean rate of use of \(v\) is weighted by how many times the speaker
Figure 5: Reexamining the “mystery” papers using a more flexible, three-way model. We find results similar to those in the original Mosteller and Wallace (1963) paper, with Madison almost certainly the author in all cases, bar one. In addition, we have very small but non-zero estimates for Jay in each case.
Table 7: Most influential terms, 1995–1996.

<table>
<thead>
<tr>
<th>the</th>
<th>of</th>
<th>to</th>
<th>and</th>
<th>that</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>we</td>
<td>in</td>
<td>is</td>
<td>a</td>
</tr>
<tr>
<td>it</td>
<td>hon</td>
<td>for</td>
<td>not</td>
<td>my</td>
</tr>
<tr>
<td>be</td>
<td>have</td>
<td>bill</td>
<td>are</td>
<td>will</td>
</tr>
<tr>
<td>they</td>
<td>was</td>
<td>he</td>
<td>as</td>
<td>labour</td>
</tr>
<tr>
<td>people</td>
<td>our</td>
<td>friend</td>
<td>on</td>
<td>local</td>
</tr>
<tr>
<td>would</td>
<td>european</td>
<td>has</td>
<td>education</td>
<td>which</td>
</tr>
<tr>
<td>committee</td>
<td>council</td>
<td>right</td>
<td>service</td>
<td>government</td>
</tr>
<tr>
<td>by</td>
<td>do</td>
<td>defence</td>
<td>schools</td>
<td>or</td>
</tr>
<tr>
<td>who</td>
<td>there</td>
<td>about</td>
<td>been</td>
<td>member</td>
</tr>
</tbody>
</table>

Table 8: Most influential terms, 1998–1999.

<table>
<thead>
<tr>
<th>the</th>
<th>of</th>
<th>to</th>
<th>that</th>
<th>and</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>in</td>
<td>a</td>
<td>we</td>
<td>is</td>
</tr>
<tr>
<td>it</td>
<td>not</td>
<td>for</td>
<td>london</td>
<td>have</td>
</tr>
<tr>
<td>be</td>
<td>people</td>
<td>was</td>
<td>health</td>
<td>are</td>
</tr>
<tr>
<td>they</td>
<td>hon</td>
<td>will</td>
<td>women</td>
<td>my</td>
</tr>
<tr>
<td>on</td>
<td>children</td>
<td>bill</td>
<td>local</td>
<td>government</td>
</tr>
<tr>
<td>he</td>
<td>as</td>
<td>house</td>
<td>industry</td>
<td>who</td>
</tr>
<tr>
<td>by</td>
<td>scottish</td>
<td>our</td>
<td>police</td>
<td>with</td>
</tr>
<tr>
<td>their</td>
<td>wales</td>
<td>friend</td>
<td>would</td>
<td>right</td>
</tr>
<tr>
<td>but</td>
<td>about</td>
<td>there</td>
<td>has</td>
<td>committee</td>
</tr>
</tbody>
</table>

actually used the word \( v \). One consequence of this arrangement is that stop words tend to be ‘influential’ because they are so commonly used, boosting the mean frequency rate across all speakers.

In Table 7 and Table 8 we list the ‘top 50’ most influential words, in terms of mean term influence, from the pre- and post-Blair election landslide (the time period we used to validate above). For each term, we compute the mean of the absolute value of the term’s influence over all speakers in each session. (We use absolute values because it is possible for an influence value to be negative for a term \( v \) since it is comprised in part of the difference between the speaker’s rate of use of \( v \) and the mean rate of use of \( v \).)

As is clear and expected, we indeed see a large number of stop words in both tables: ‘the’, ‘of’, ‘to’ and so on. More compellingly though, we also see words that make sense as important or distinctive during this time. For example, Tory backbenchers made ‘european’ influential, which accords to the protracted and fractious discussion of the various EU treaties at this time. Meanwhile, ‘defence’—a common right-wing concern—turns up as an influential word in 1995, but not in 1998 (among Labour MPs). After Blair’s election, it is words like ‘london’, ‘scottish’ and ‘wales’ that are most influential. Given that the Labour government was discussing and passing legislation to create new governing institutions in the capital, Scotland and Wales, this makes some sense. We also typical left-wing concerns pertaining to ‘women’, ‘children’ and ‘health’ at this time.

Another nice feature of our model is that we can examine “influential” tokens to get some
sense of how their use changes over time, and how that use affects distinctiveness. In Figure 6 we do exactly that, with a lowess imposed on the data for visualization purposes. In particular, we plot the ‘average influence’—defined as the mean of a term’s influence over all speakers within a session—of the various tokens for every session of parliament under study. Starting at the top left, we see that ‘defence’ surged in influence during the Cold War, and then fell away after it ended, around 1990. Moving right, we see ‘european’ coming to a peak during the Tory rebel years of the early 1990s, during the Maastricht negotiations. Moving to the bottom left, ‘Scottish’ peaks in the late 1970s and late 1990s—the periods in which some devolved rule was offered to Scotland (and referendums took place on the same). Finally, ‘health’ peaks in the late 1990s/early 2000s as the Blair government abolished the internal market, added large real term budget increases, but allowed independent foundation status for hospitals. This latter move was very unpopular with Labour rebels, and in 2003 they almost defeated the government’s bill in the Commons that dealt with this policy change.
Figure 6: Average influence of some key terms over time. The non-stop word tokens from the sessions of 1995 and 1998 accord with our priors, in terms of their behavior over the entirety of the data. For example, ‘defence’ ceases to be (as) influential once the Cold War ends.

References


