Leaders' Prominence and Cultural Perspectives

Jeffrey B. Arnold and H. E. Goemans

University of Rochester

April 11, 2012

Abstract

What makes leaders prominent? And is this culture-specific? Our paper
provides the first steps to answer these questions, which could shed new
light on why and which leaders establish their policy priorities.

THIS IS VERY MUCH A DRAFT, PLEASE DO NOT CITE WITHOUT PERMISSION.

1 Introduction

In international relations, comparative politics and economics, scholars are turning their focus onto leaders to examine a remarkably broad range of puzzles. In this research, scholars incorporate an impressive variety of factors that could influence leaders' decision including political constraints (Bueno de Mesquita et al. 2003), expectations about the leaders' post-tenure fate (Goemans 2008), and leaders' personal life-experiences and attributes (Horowitz and Stam 2012). We seek to further extend the range of relevant factors by examining how leaders are evaluated by the public-at-large in the broadest possible sense. This will allows to accomplish at least three things. First, it can be argued that at least some leaders care about their prominence in history and that this figures directly in their utility function and thus influence their policy decisions. Second, we evaluate if different 'cultures' – here operationalized as different languages – privilege some policies and policy outcomes in how they rank political leaders' prominence in history. It could be the case, that in some cultures it is war that determines a leader's prominence, while in others it is economic growth We should expect that leaders concerned about their place in history in the former and latter cases to systematically prefer different policy options over others based on the culturally mediated effects of their policy on their prominence. Third, our way of proceeding will allow us to make frequent updates to measure changes in a leader's prominence on a montly basis; this will allow us to gain some traction on a current leader's 'standing' and 'popularity' and how this changes over time, a particularly valuable tool since it is much cheaper to implement than surveys and can be applied to even the most secretive and closed countries – as long as its inhabitant have access to the internet.

There is significant variation in the historical prominence of international leaders. For example in a 2007 poll of American adults, 96 percent of respondents had an opinion George Washington and Abraham Lincoln, but only 24 percent had any opinion of John Tyler and only 42 percent had any opinion of Millard Fillmore. Given this variation, we ask, what separates the George Washingtons and Abraham Lincolns from Millard Filmores and John Tylers?[1]. Similarly, when ranking the British Prime Ministers in 2010, Ben McIntyre wrote

> The premiership may be the highest elected office in the land but some of its occupants remain all but invisible. There are 6,364 books about Winston Churchill on Amazon.com. About Andrew Bonar Law, "The unknown prime minister" (1922-23) there is only one book in print, with a sales rank of 495,490 (Macintyre 2010)

We use Wikipedia to obtain admittedly noisy indicators of the prominence of leaders. This means that we assume that several observable attributes of an individual's Wikipedia article — page views, size, number of edits, number of

---

[1]See Wikipedia:Historical Rankings of Presidents of the United States and Rasmussen Reports

editors, and number of inlinks — combine to reflect the historical prominence of that individual. Thus, we use these indicators and a Bayesian factor model to estimate the prominence of individuals. Since there are multiple Wikipedias in different languages, we can estimate the latent prominence of individuals in different languages, with the assumption that these languages systematically represent different (political) cultures; an assumption that we empirically evaluate. This way we estimate the latent prominence for all leaders of international states between 1875 and 2004, in each of the ten languages with the most-visited Wikipedias.

In the remainder of this paper, we first explain the potential promise and piftfalls of our approach. In the next section we describe our methodology and then turn to the statistical model. In the fourth section we present some <u>initial</u> findings. The final section describes where we plan to go from here.

2 Fame, Infamy and Prominence

It is useful to think of the *fame* of leaders as consisting of two dimensions: Prominence and sentiment. *Prominence* (or recognition) is the extent to which a leader is widely known. *Sentiment* (or valence) is the extent to which people have a favorable or unfavorable view of the leader. Thus, to take one example, Adolf Hitler is a very prominent leader who is viewed unfavorably by most: he is *infamous*. Prominence and sentiment correspond to concepts of name-recognition and favorability questions that appear in polls and marketing.[2]

To start our project, this paper will focus first on prominence for two reasons. The first is that measures of prominence and recognition are easier to derive than measures of sentiment. Secondly, for at least some, being remembered by history may be more important than how one is remembered. If you do not take our words for it, consider the words of Jack Sparrow in *Pirates of the Caribbean*,[3]

> Norrington: . . . You are without doubt the worst pirate I've ever heard of.
> Jack Sparrow: But you have heard of me. [4]

Of course we would like to also use the Wikipedia's pages to also assess sentiment and favorability; the previous literature suggests some approaches, but we foresee – hopefully not insurmountable – obstacles, since we would need valence terms for each of the ten different languages. The problem is further complicated by Wikipedia's policy that articles should be written from a neutral point of view (NPOV).[5]

---

[2] For a marketing example, see Q-scores.

[3] Hat tip to Jean-Baptiste Michel who used this example when presenting Michel et al. (2011a).

[4] http://www.imdb.com/title/tt0325980/quotes (March 13, 2012).

[5] Wikipedia's NPOV and other policies make it harder to find the subjective phrases that typical sentiment analysis algorithms rely upon. In fact, we have been unable to find any paper that has been no paper that has done sentiment analysis on Wikipedia articles. If the reader is aware of one, please let us know.

The development of a quantitative measure of the prominence of leaders holds several promises. First, (anticipated) prominence is likely a component of leaders' utility functions. Prominence can be a consumption good, valued for its own sake or prominence can be an intermediate good, valued because more (anticipated) prominence increases the ability of a leader to implement his desired policies. Second, even if prominence is not part of leaders' utility functions then measures of leaders' prominence can still provide information about the perceived competence of leaders. Measures of leaders' prominence can then supplement statistics such as GDP growth and war victories. Third, quantifying the prominence of leaders potentially provides insights about how leaders form beliefs and learn from history. Regardless of how leaders process historical events, they can only learn from those events that are recorded by history. The historical record does not apply equal weights to lives of all leaders and thus understanding how the sample of leaders recorded by history is constructed may help us to understand the biases in the information available to leaders.

As an argument for why older leaders may be willing to make risky decisions, Horowitz, McDermott, and Stam (2005, p. 668) claim that older leader are concerned with the "creation of a legacy" in their limited time, both alive and in office. They claim that "A leader attempting to build a long-term legacy and needing consolidate immortaility quickly may implicitly view victory in militarized contests as a way to gain lasting reputation." (Horowitz, McDermott, and Stam 2005, p. 668) There is evidence from historians' views of American presidents that serving during times of political crisis increases greatness, but we do not know if a simlar dynamic extends to other countries and other cultures. This leads us to the second main thrust of this paper.

If their own (anticipated) or their predecessors' prominence affects the decisions of leaders, we might assume that *all* leaders are held to the same global standard, or, perhaps more reasonably, that their nationality or culture imposes certain standards of behavior on their own. Then, we would expect that different cultures – and nationalities – uses different yardsticks to measure and attribute the prominence of leaders. Some evidence we report below strongly suggests that cultural background does indeed have a strong effect on the prominence of leaders, as in all languages (associated with different nations) the most prominent leader(s) are from their own culture or linguistic background. We therefore examine whether and how a leader's *global* prominence differs from his or her *cultural* prominence and what factors explain this divergence. If, for example, the economic growth achieved while the leader was in office helps explain the difference between a leader's cultural and global prominence, this could tell us something about the leader's priorities in office, or how subsequent leaders establish their policy priorities.

Finally, our approach will hopefully allow us to see how the prominence of leaders changes over time, as we download subsequent Wikipedia pages. This could provide a close to real-time status report on the performance of *current* leaders, akin to a noisy opinion poll. The promise, here, of course is that this is a relatively cheap procedure which can provide feedback on the great majority

of current leaders.

3 Literature Review

The recent international relations literature on leaders is focused on two motivations and assessment of leaders: duration of tenure (Bueno de Mesquita et al. 2003) and post-tenure outcomes (Goemans 2008). We consider another measure of leader's tenures and potential motivation of leaders: their 'prominence.'

There is an extensive literature that quantifies and explains the greatness of American presidents (Simonton 1986; Simonton 1987; Simonton 1991; Simonton 1992; Simonton 1993; McCann 1995; Simonton 2001; Cohen 2003; Simonton 2006; Kenney and Rice 1988; Curry and Morris 2010). In 1948, Arthur M. Schlesinger conducted the first survey asking historians to rank presidents (Schlesinger 1948). Since then, there have been numerous surveys asking primarily experts, but also the general public, to rank the American presidents. The central question in this literature is whether presidential greatness is determined by attributes of the time period or attributes of the president. The canonical model in literature, often called the "Simonton model" Simonton (1987), explains presidential greatness rankings with six variables: tenure in office, scandal, assissination, war hero, war years, and intellectual brilliance. Curry and Morris (2010) critique the literature for ignoring measures of presidential performance, and find that economic growth, but not military victories, are also associated with presidential greatness.

Literature on the greatness of leaders has only been extended to a handfull of countries other than the United States: prime ministers of the United Kingdom Theakston and Gill (2006), Canada Ballard and Suedfeld (1988), and surveys by MacLean's in 1997 and 2011, New Zealand Sheppard (1998), and Australia (see Wikipedia page).[6]

While the work on American presidents focuses on greatness, earlier work by Simonton (1984) focuses on the "eminence" of leaders and historical figures. Like the literature on presidential greatness, the central question in this literature is whether eminence is the result of attributes of the situation (the eponymity hypothesis) or the attributes of the individual (the genius hypothesis). Simonton (1984) compares European hereditary monarchs and finds their eminence is explained by the number of historical events in their epoch; positive and negative events contribute to eminence, although the positive events are weighted slightly higher; intelligence is positively related to eminence; morality has a U-shaped relationship to eminence, meaning that the both the "saintly" and "decidedly evil" are remembered. Even earlier work by Sorokin (1925) found that the "most esteemed" monarchs were those with the longest lives and reigns.

The literature on greatness and eminence has required on costly to generate data. The greatness literature primarily relies on expert surveys, which are infrequently conducted, exist for only a small number of countries, and would

---

[6]There may be others and we'd be grateful for references.

be costly to expand. The research on eminence has been able to take advantage of existing secondary secondary sources to quantify eminence. Simonton (1984) generates an eminence index from citations in various sources, including histories, encyclopedias, and biographical dictionaries. This paper follows a similar approach. However, the rapid and massive digitalization of information has made gathering this information easier than ever. Michel et al. (2011a) use a sample of books digitalized by Google Books, consisting of 4 percent of all books ever published and spanning 1800 to 2000, to investigate cultural trends. They use the presence of the names of individuals in this corpus to measure the priminence of individuals and how the rise and fall of prominence has changed over time, and differs between occupations.

4 Measuring Prominence

Our methodology assumes that several observable attributes of an individual's Wikipedia article — page views, size, number of edits, number of editors, and number of inlinks — are noisy indicators of the prominence of that individual. Thus, we use these indicators and Bayesian factor model to estimate the prominence of individuals. Moreover, since there are multiple Wikipedias in different languages, we can estimate the latent prominence of individuals in different languages. We use this method to estimate the latent prominence for all leaders of international states between 1875 and 2004, in each of the ten languages with the most-visited Wikipedias (listed in 1).

In this section, we will discuss (1) the set of leaders used, (2) the set of Wikipedias used, (3) the observed indicators of prominence, (4) the statistical model used to estimate prominence, and (5) the results of that model.

4.1 Choice Of Leaders

For the set of leaders for whom we estimate their prominence, we use the 2476 people identified as leaders of international states between 1875 and 2004 in Archigos 2.9 (Goemans, Gleditsch, and Chiozza 2009).[7] Archigos defines the leader of international states as the "effective leader", i.e. "the person that de facto exercised power in a country." (Goemans, Gleditsch, and Chiozza 2009, p. 271) The effective leader is usually the prime minister in parliamentary systems, the president in presidential systems, and the chairman of the party in communist states. Exceptions to that rule include more complicated political systems, such as mixed parliamentary and presidential systems, such as Finland, France and Portugal, cases in which a foreign power imposes a leader, and cases in which the *de facto* ruler did not have a formal title, e.g. Rafael Trujillo in the Dominican Republic between August 1938 and May 1942.[8]

---

[7]Archigos defines international states according to the Gleditsch-Ward state system (Gleditsch and Ward 1999). There are 184 unique states in time period covered by Archigos.

[8]For more detail on the coding rules see Goemans, Gleditsch, and Chiozza (2009).

4.2 Choice of Wikipedias

For this analysis, we will use data from the ten Wikipedias with the highest monthly page views. These are listed in Table 1.

	Language	Views (per day)	Articles	Users
en	English	20,927,397	3,884,116	16,341,507
ja	Japanese	3,497,489	794,372	598,843
es	Spanish	2,875,342	871,245	2,131,214
de	German	2,787,900	1,368,841	1,382,923
ru	Russian	1,887,215	826,231	805,230
fr	French	1,864,612	1,219,020	1,232,356
it	Italian	1,281,963	895,288	723,110
pl	Polish	1,070,320	882,616	476,717
pt	Portuguese	972,146	715,191	968,747
zh	Chinese	480,091	404,145	1,149,239

Table 1: Top ten Wikipedias by page views with the number of page views, articles, and users. Users are non-anonymous editors of Wikipedia, individuals with user pages.

As of March 2012, there are 271 active Wikipedias, each with a different language.[9] The largest and most popular of the Wikipedias is the English Wikipedia with 3.8 million articles, 520 edits, 16 million users, and 15 billion page views per month. One of the smallest is the Kikuyu Wikipedia. with 122 articles, 999 edits, 2,037 users and 35 thousand views per month.[10']

For this paper, we restrict ourselves to the top 10 Wikipedias by page views.[11] These are listed in Table 1. While these ten Wikipedias are a small fraction of the total number of Wikipedias, they account for the majority of the total content, users, and views for wikipedia.org. These Wikipedias account for 57 percent of the articles, 79 percent of the users, and 91 percent of the page views for all of Wikipedia. [12]

Figure 1 plots, for each language Wikipedia, the share of views and edits from each country. In the German, French, Italian, Portuguese, and Russian languages, one country accounts for over 60 percent of the views and edits; although for the Portuguese Wikipedia, it is Brazil, not Portugal, which is the primary editor/viewer. In the English Wikipedia, the US accounts for about 45 percent of the views and edits, with the United Kingdom and Canada combine for another 16 percent, and miscellaneous other countries account for the

---

[9]http://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias&oldid=3524180. There are 284 language Wikipedias; 13 of these are closed or deprecated.

[10]See http://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias&oldid=3524180 and http://stats.wikimedia.org/EN/TablesPageViewsMonthly.htm.

[11]The page views in February 2012. These are the ten Wikipedias which appear around the logo on the Wikipedia portal.

[12]February 2012, http://stats.wikimedia.org/EN/TablesPageViewsMonthly.htm (retrieved on March 13, 2012).

remaining 39 percent. The Spanish Wikipedia is unusual for the number of different countries which view and edit it in significant amounts; seven countries have shares greater than 5 percent.
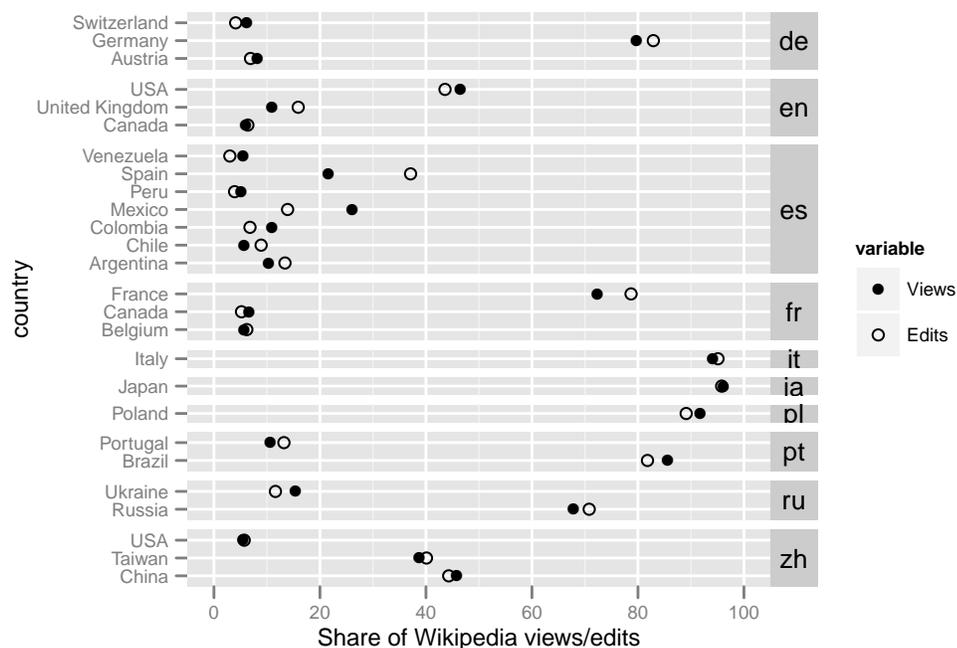

Figure 1: For each language Wikipedia, the share of views and edits from each country, for all countries with a share greater than 5 percent. For example, 46 percent of views and 43 percent of edits in the English language Wikipedia are from the United States.

4.3 Choice of Observed Indicators

Since international leaders easily exceed the notability requirement for Wikipedia, almost all leaders have an article in Wikipedia.[13] Only 38 leaders do not have an article in any of the ten Wikipedias we use. However, the number of leaders with an article differs by language, plotted in Figure 2. Wikipedias with fewer

---

[13]One advantage of our data source compared is that our measure of prominence unambiguously measures the leader. If we were to use the number of search results in Google or the Google Books n-gram data, we would not be able to unambiguously match our data to the leaders as leaders names could refer to multiple people. E.g. using only names, it would be hard to distinguish Michael Moore, the New Zealand prime minister, from Michael Moore, the filmmaker. However, since these distinct individuals have distinct Wikipedia pages, we can estimate the prominence of Michael Moore, the New Zealand prime minister, without worrying that our results are contaminated by the other Michael Moore.

articles tend to contain fewer articles of leaders. The English Wikipedia has articles for 98 percent of the leaders, while the Japanese and Chinese Wikipedias have articles for only about 42 percent of the leaders.
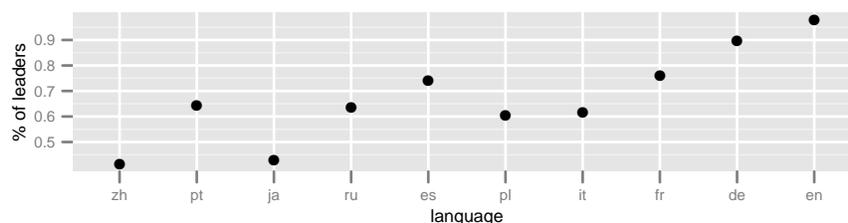

Figure 2: By Wikipedia, percent of leaders with an article. Wikipepdias are ordered from left to right by the total number of articles.

We use five observable attributes of a leader's Wikipedia article in a given language to measure the size and popularity of that leader's article. These attributes are listed below,

`num_links`	Number of other articles in the Wikipedia which link to the article.
`pageviews`	Number of page views between January 2011 and February 2012.
`page_size`	Size (in bytes) of the latest revision.
`revisions`	Number of revisions.
`users`	Number of unique users who have edited the article.

If a leader did not have a page in wikipedia, we assigned that leader a value of 0 for all the attributes. The sources for the values of `page_size`, `revisions`, `users`, and `num_links` are the Wikipedia API.[14] The source of the values for `pageviews` is stats.grok.se. The data were retrieved in early March 2012.

Figure 3 plots the distributions of each variable for each Wikipedia. For all of these statistics the distributions are highly skewed, meaning a select few leaders are orders of magnitude more prominent than all the others – even after a square root transformation. `pageviews` appears to be the most skewed, while `page_size` appears to be the least skewed. The lesser skew of `page_size` is almost certainly due to a norm on the upper bound of the size of an article before it is split into multiple articles.[15]

Our choice of response variables is supported by the decisions made in Michel et al. (2011b). Although that work measured prominence using n-grams in the

---

[14] See http://www.mediawiki.org/wiki/API:Main_page.

[15] If an article becomes too long, it will generally be split into multiple articles. For example, for George W. Bush, in addition to his main page, there are pages for his professional life, governorship, presidency, first term as president, second term as president, domestic policy, economic policy, and foreign policy. Thus page size may be work well for distinguishing unknown from moderately known, but works poorly for distinguishing the well known from the super prominent. This is one reason for our use of multiple indicators; the number of inward links will be higher in the presence of the aforementioned situation.
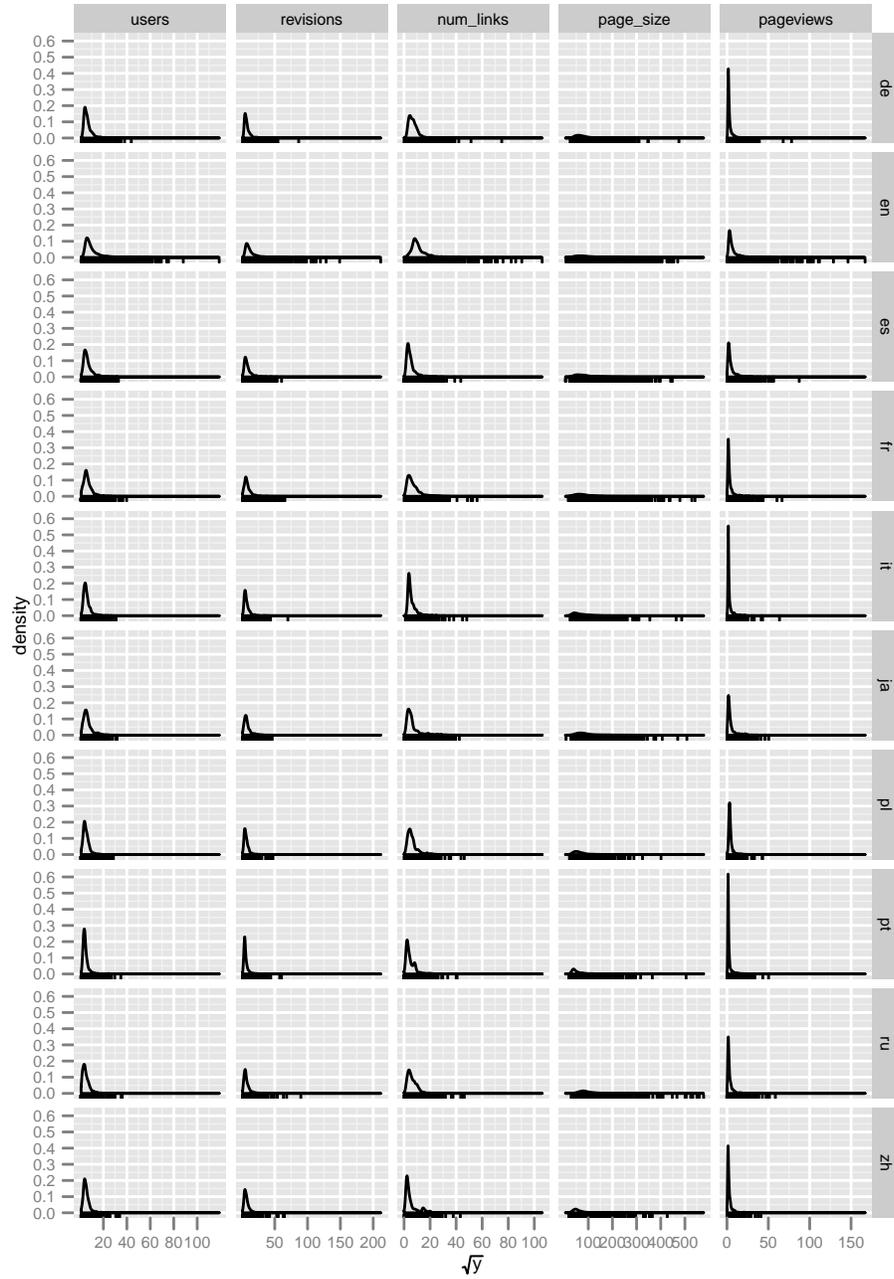
Figure 3: Distribution statistics of response variables conditional on the existence of a Wikipedia article in that wiki. See the text for descriptions of the variables.

Google Books Corpus, in order to disambiguate n-grams when there were multiple people to which an n-gram could refer, they used the article length and number of page views to determine the "more famous" person.

5 Statistical Model

Using the observed data from the Wikipedias considered, we estimate a Bayesian factor model in order to estimate the leaders' latent prominence. We estimate both the prominence of the leader common to all the languages, their *global prominence score*, and language-specific effects to their latent prominence. Using a factor model on citations and presence in multiple sources is the standard methodology for constructing a measure of prominence, see (Simonton 1984). Our method only differs in the use of Wikipedia as the source of the indicators.

Let $W = \{\texttt{de}, \texttt{en}, \dots, \texttt{zh}\}$ be the set of Wikipedias, discussed in Section 4.2. Let I be the set of leaders, discussed in Section 4.1. And let J be the set of variables, discussed in Section 4.3. Let $y_{i,j,w}$ be the value of the response variable $j \in J$ for leader $i \in I$ for the Wikipedia language $w \in W$. As is standard in factor analysis, the data are standardized by subtracting the mean and dividing by the standard deviation. Additionally, since all the data are highly skewed, we first transform all the variables with the square root function.

$$y_{i,j,w}^{\frac{1}{2}} \sim N(\lambda_{j,w}^G \phi_i^G + \lambda_{j,w}^w \phi_i^w, \sigma_{j,w}^2) \tag{1}$$

$$\phi_i^d \sim N(0,1) \qquad\qquad \text{for } d \in D \tag{2}$$

$$\tag{3}$$

There are eleven factors in this model: one global (or common) factor and ten language-specific factors. The factor score of the leader on the first dimension (ϕ^G), represents each leader's global prominence, i.e. the prominence of the leader common to all languages. The factor scores on the other ten dimensions (ϕ^w for $w \in \{\texttt{de}, \dots, \texttt{zh}\}$) are leaders' prominence in the ten languages considered, relative to their global prominence. We index the dimensions with the set $D = \{G, \texttt{de}, \texttt{en}, \dots, \texttt{zh}\}$.

All the indicators $y_{i,j,w}$ load only on the global dimension, and the appropriate language-specific dimension. For example, $\texttt{pageviews}_{\texttt{en}}$ loads on the global factor ($\lambda_{\texttt{pageviews},\texttt{en}}^G$) and the English language factor ($\lambda_{\texttt{pageviews},\texttt{en}}^{\texttt{en}}$), but is zero for all other language specific factors, e.g. $\lambda_{\texttt{pageviews},\texttt{en}}^{\texttt{de}} = 0$. Thus an observed indicator from a leader's article is a function of that leader's global prominence and the language-specific prominence of that leader in the language of the article.

The parameters $\sigma_{j,w}^2$, called uniqueness in factor analysis, measurement errors. These represent the variance in $y_{i,j,w}$ which cannot be explained by the global and language-specific factors.

We estimated the model in equation (1) using Markov Chain Monte Carlo with the R package **MCMCpack** (Martin, Quinn, and Park 2011). For each
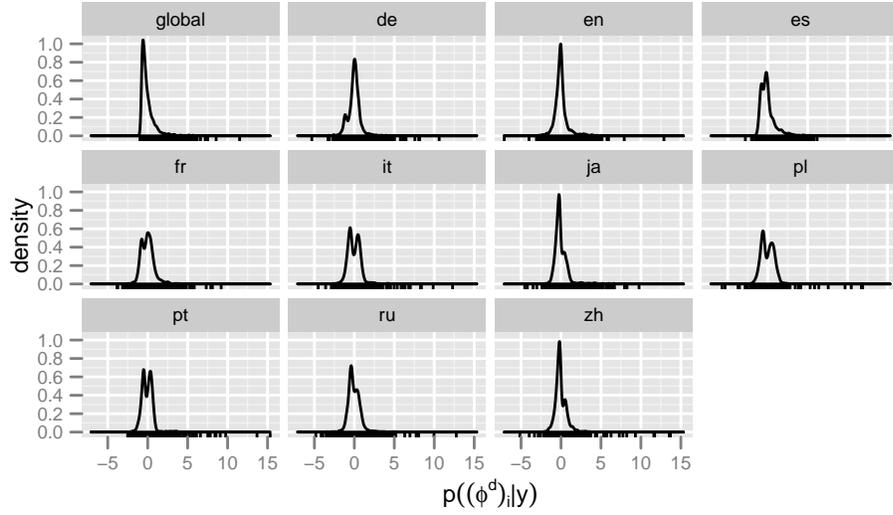
Figure 4: Distribution of the means of the posterior distributions of prominence ϕ_i^d for each dimension.

Wikipedia, we collected a sample of 1,999 from a single chain with a burnin of 5,100 and a thinning interval of 100. We tested that the chains had converged using Geweke's convergence diagnostic.

5.1 Factor Scores

- Global distribution is highly skewed.

- Language-specific factor scores are less skewed, but still have a larger right tail.

- Spanish is the most skewed of the language specific distributions. Many leaders which are much more prominent in the Spanish Wikipedia than elsewhere.

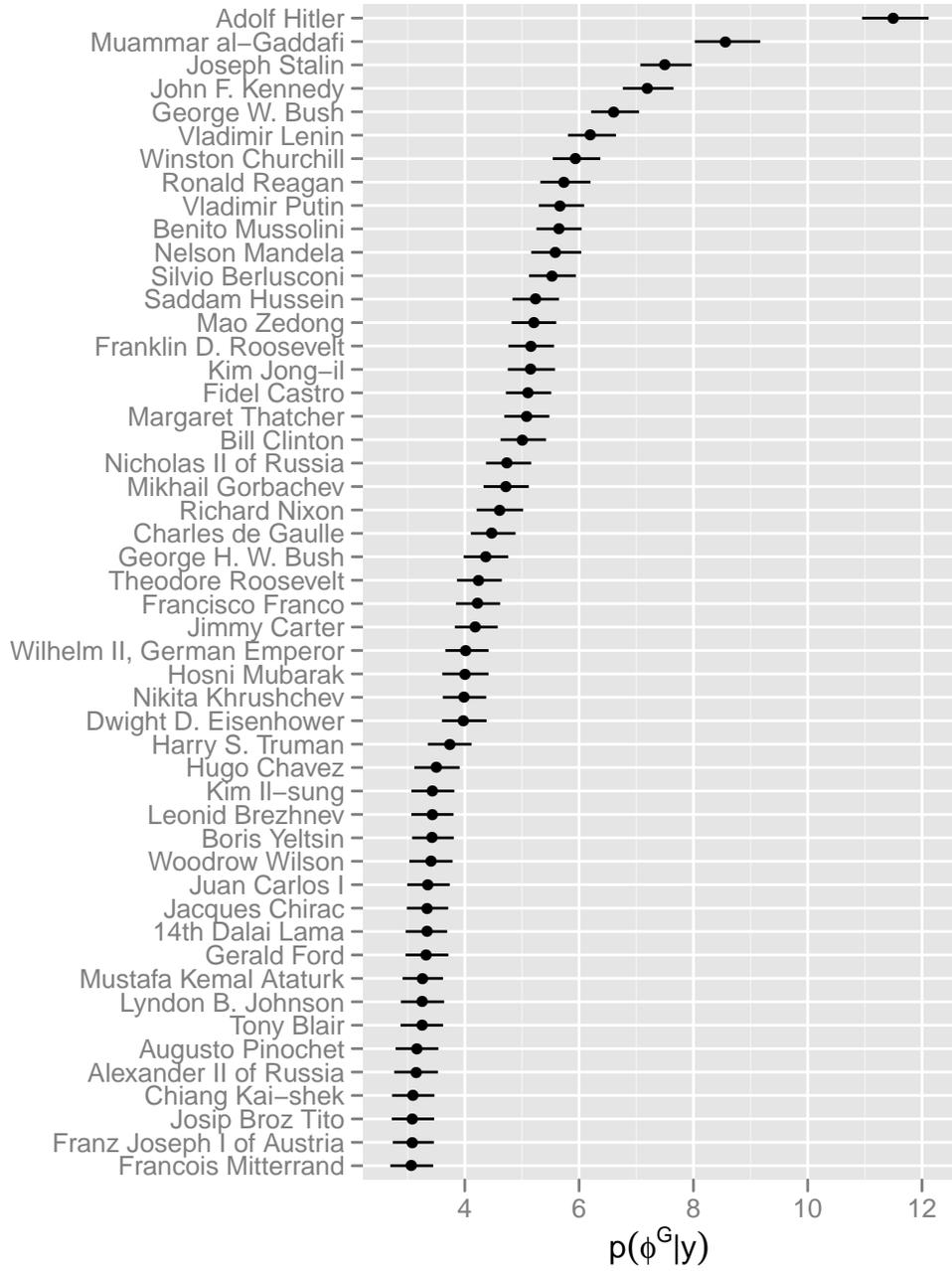Leaders with the highest global prominence are plotted in the figures below.

Figure 5: Posterior distribution of global prominence ϕ_i^G for the leaders of the leaders with the 50 highest mean prominence. Higher values of ϕ^G represent a higher latent prominence.
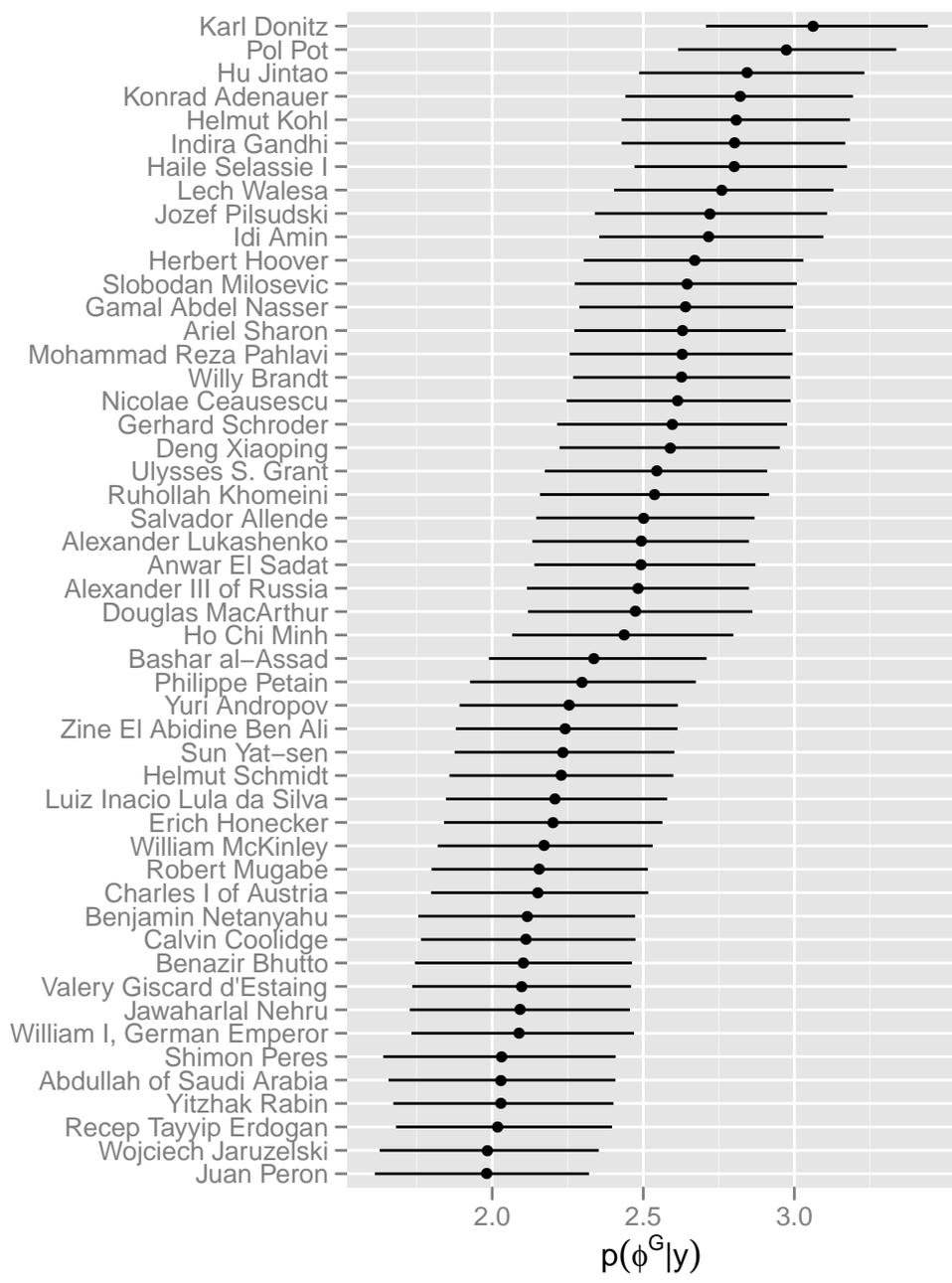
Figure 6: Posterior distribution of global prominence ϕ_i^G for the leaders of the leaders with mean prominence ranked 51 through 100. See Figure 5 for those ranked 1-50. Higher values of ϕ^G represent a higher latent prominence.

5.2 Factor Loadings

The loadings for each indicator on the global factor and its language-specific factor are plotted in Figure 7.

Figure 8 plots the variance explained for each language's indicators by the global factor and the language-specific factor. The variance explained by the global factor for a language's indicators represents how a language's measures of leader's prominence agree with the consensus. For all but Spanish and Portugese languages, the global factor explains more than half the variance in the indicators. The global factor explains 76 percent of the variance in the English language indicators, and the English-language factor explains only 17 percent of the variance. For the Spanish language indicators, the global factor explains only 38 percent of the variance, while the Spanish-language factor explains 51 percent of the variance.

It is not surprising that the English indicators are most representative of the global prominence estimates, since the English Wikipedia is the *lingua franca* of Wikipedias. Although the overall traffic to the English Wikipedia is dominated by the United States, United Kingdom, and Canada, for any given country, the English Wikipedia is the most visited Wikipedia after the Wikipedias for the primary language(s) of that country.[16] Among editors across all Wikipedias, 76 percent "contribute in some form" to the English Wikipedia, and 93 percent read it. Since individuals from many languages contribute or read the English Wikipedia, the content of the English Wikipedia is less likely to diverge from other Wikipedias because individuals from that language will likely adjust either the English Wikipedia or their own language Wikipedia to fill in the oversights. [17] The German Wikipedia is also very similar to the global prominence is harder to explain. The German Wikipedia has the second highest number of articles and also has a reputation for having higher quality articles than the English Wikipedia (citation needed).

The Spanish indicators have the lowest variance explained by the global factor. The Spanish Wikipedia has the third highest number of views and articles. We conjecture that this may be due to the large number of countries that speak Spanish, and by extension, contribute to the Spanish Wikipedia, see 1. Seven countries each have greater than a 5 percent shares of the page views for the Spanish Wikipedia, while no other Wikipedia among those considered, has more than three countries with more than a 5 percent share of page views. While for most languages, only leaders from one or two countries get a home language bias term, for Spanish, leaders from many countries get this bias term, meaning more leaders are further from the global prominence. Additionally, South American countries do not interact as much with Europe or Asia, so likely less is written about leaders from these other countries.

---

[16]http://stats.wikimedia.org/wikimedia/squids/SquidReportPageViewsPerCountryBreakdown.htm (Retrieved March 17, 2012).

[17]See   http://commons.wikimedia.org/wiki/File:WP_2011_Editor%27s_Survey_-_Topline.pdf.
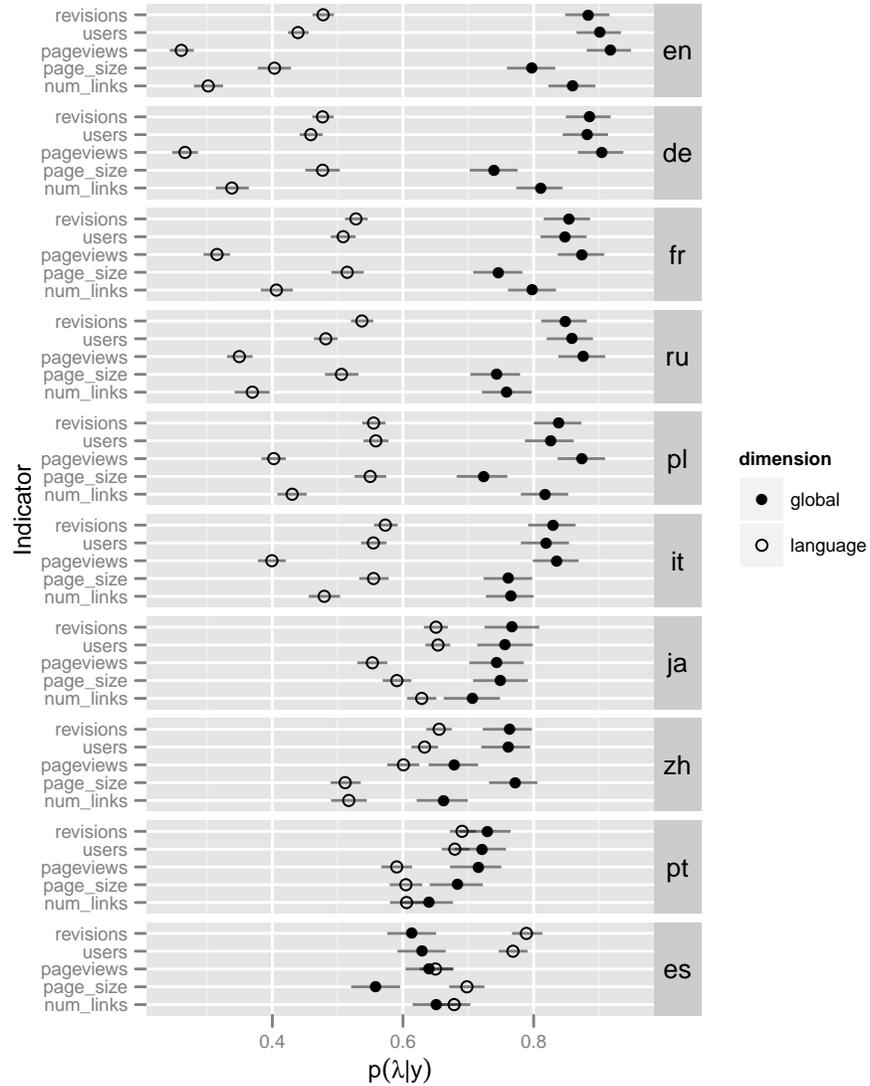
Figure 7: Posterior distributions of the loadings for each indicator on the global (λ^G) and language-specific dimensions (λ^w). The point-ranges summarize the posterior distributions with the mean, 2.5 percentile, and 97.5 percentile. The indicators are ordered by the language, with languages ordered by the percent of the variance explained by the global factor. Thus, the global factor explains the most variation in the English language indicators, and the least variance in the Spanish language indicators.
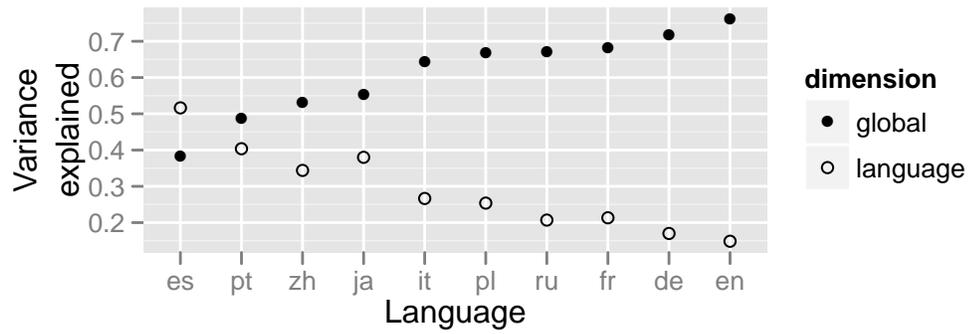
16

Figure 8: Fraction of variance explained of each language's indicators explained by the global factor and by its language-specific factor. The higher the variance explained by the global factor the more similar that language is to the consensus estimates of prominence. The variance explained by a dimension is the sum of the squared factor loadings.

6 Explaining Prominence

In order to understand why some leaders are more prominent than others, we estimate regression models of the prominence scores on variables related to the leader and his tenure(s). The dependent variables in these models are the mean of the posterior distributions of the prominence scores in section 5, $\bar{\phi}_i^d = \mathrm{E}\, p(\phi^d | y)$. We estimate eleven models, one for the prominence score in each dimension: the global prominence, and ten language-specific prominences. For a given dimension, the outcome variable is the estimated prominence of the leader in that dimension. For each dimension $d \in \{G, \mathtt{de}, \ldots, \mathtt{en}\}$, we estimate the following model,

$$\bar{\phi}_i^d = \kappa_{j[i]} + \beta' x_i + \epsilon_i \tag{4}$$

where i indexes leaders, $\kappa_{j[i]}$ is a country fixed effect with $j[i]$ indicating the country of leader i, X is the design matrix, and ϵ are i.i.d errors. The design matrix X consists of the following columns,[18]

> years_out Number of years since the end of the last tenure.
> Major power Fraction of tenure that the country of the leader was a major power (COW).
> Civil War Fraction of tenure that the state was in a civil war, (COW Intra-State War Data (v4.0) Sarkees and Wayman 2010).
> War_duration Total number of days in inter-state war (COW Inter-State War Data (v4.0) Sarkees and Wayman 2010).
> War_wins_3 Number of inter-state wars that ended in victory for the leader's country, with each war weighted by the fraction of the war duration that the leader was in power; COW Inter-State War Data (v4.0) (Sarkees and Wayman 2010).
> War_losses_3 Number of inter-state war losses. Defined similarly to war_losses_3.
> War_draws_3 Number of inter-state wars ending in neither a victory nor a defeat. Defined similarly to war_losses_3.
> bat_death Number of battle deaths in inter-state wars (COW Inter-State War Data (v4.0) Sarkees and Wayman 2010).
> pecpc_growth Average per capita energy consumption (Singer, Bremer, and John 1972, National Military Capabilities (v4.0)).
> Presidential Fraction of tenure that the country is both a democracy, polity score greater than 6, and a presidential system; (Marshall and Jaggers 2005, Polity IV version 2010).
> Democracy Fraction of tenure that the country has a polity score greater than 6 (Polity IV).
> Autocracy Fraction of tenure that the country has a polity score less than -6 (Polity IV).
> Ungoverned Fraction of tenure that the country has a polity score of -66, -77, or -88 (Polity IV).

---

[18] If not otherwise indicated, the variable is from Archigos 2.9 (Goemans, Gleditsch, and Chiozza 2009).

`ntenures` Number of distinct tenures.
`Male` Leader is male.
`Alive` Leader is alive as of 2012.
`Age` Age of leader at death or in 2012.
`duration` Duration of all leaders tenures.
`Leader2004` Leader still in power as of 2004-12-31, when Archigos 2.9 ends.
`Irregular_entry` Leader entered power irregularly.
`Irregular_exit` Leader exited power irregularly.
`Bad_Post_ten` Leader had a bad post-tenure fate.

Table 2 presents the results of the model in equation (4). For ease of interpretation, we only list signs for the effect of each of these variables when they are significant at the 5% level. The first column seeks to explain global prominence, the remaining columns show whether these variables increase or decrease the prominence of leaders in that specific language Wikipedia *relative to global prominence*. A + sign thus indicates that in this specific language Wikipedia, this variable would generate a higher prominence score than found in the global prominence score. Likewise, a − sign indicates that in this specific language Wikipedia, this variable would generate a *lower* prominence score than found in the global prominence score.

Table 2: The Prominence of Leaders in Cultural Perspective

	global	en	ja	es	de	ru	fr	it	pl	pt	zh
Years_out	−			+	−		−		−		
Major Power	+		−		+		−	−	−		+
Civil War		+	+			+					+
War_duration	+	−					−	−			−
War_wins_3	+				−		+	+	+		
War_losses_3	+						+		−	−	
War_draws_3	+	+			+					+	+
bat_death	+				+	+					−
pecpc_growth						+			−		
Presidential		+		−		+	+				
Democracy			+	+				+		+	+
Autocracy	+					+		+		+	
Ungoverned						+					
ntenures							+				
Male	−				+		−				
Alive	+	+		+		−					
Age	−			−			−	+			
duration	+		+	+	+		+	+	+	+	
Leader_2004	+		+	+	+	+	+	+		+	+
Irregular_entry	+					+					
Irregular_exit	+					+					−
Bad_Post_ten		+		+		+	+		+		
Constant	+					+	+				

Table 2 shows some remarkable cultural variation in how Wikipedia users evaluate recent versus historical leaders, leaders of major powers, the relevance of the number of days this leader was at war, the outcome of war, gender, age and, for leaders of the Chinese Wikipedia, the irregular exit of leaders.

Let us take one striking finding: users of the German (`de`) Wikipedia attach significantly less weight to victories in war compared to the global prominence score. This is surprising to us, for without Hitler's initial victories, it is highly doubtful he would have gained the prominence he did both globally and among the German-speaking users of Wikipedia, who rank him the most prominent leader overall. In other words, Hitler's victories would lead German-speaking users of Wikipedia to assign him a *lower* prominence score than global users do. Similarly, global users seem to 'care more' about how a leader lost power than do Chinese-speaking users of Wikipedia, as suggested by the negative coefficient for *irregular exit* for the users of the Chinese (`zh`) Wikipedia.

While we do not want to make too much of these initial results, we are fairly confident that the specific historical circumstances that figure 'prominently' in the consciousness of users of the different Wikipedias *do* indeed significantly affect the prominence they assign to leaders. German and Chinese Wikipedia users care more about the leaders of major powers than does the average global user, while Japanese, French, Italian and Polish-speaking users appear to care less about such leaders. These initial results thus point to some striking cultural variation, which we will explore in our future research.

7   Conclusion

This is very preliminary work. We appreciate any comments and suggestions the reader may have.

8 Appendix

8.1 Top Leaders By Country

Tables 3 and 4 list the 25 leaders in each language with the highest mean prominence [19]

These lists are intrinsically interesting, useful for assessing the external validity of this measure, and gaining some intuition about the correlates of prominence.

These lists have both commonality and variation. Each language privileges leaders from countries that speak that language. In fact, the top ranked leader of each language is from a country which speaks that language. More recent leaders also appear highly ranked in these lists.

However neither the home-language nor the recency bias completely determines the prominence rankings. Many leaders appear in rankings outside the language of their country, and many historical leaders appear in these rankings. The leaders: Adolf Hitler, George W. Bush, Joseph Stalin, Benito Mussolini, John F. Kennedy, Vladimir Lenin, Saddam Hussein, Fidel Castro, Muammar al-Gaddafi, Winston Churchill, Hugo Chavez, Mao Zedong, Vladimir Putin, Nelson Mandela, and Nicholas II of Russia all appear in the top 25 most prominent leaders in five or more of the lists. This list includes several historical leaders, e.g. Nicholas II of Russia and the leaders of the major powers during World War II.

References

Ballard, Elizabeth J. and Peter Suedfeld (1988). "Performance Ratings of Canadian Prime Ministers: Individual and Situational Factors". English. In: *Political Psychology* 9.2, pp. 291–302. ISSN: 0162895X. URL: http://www.jstor.org/stable/3790958 (cit. on p. 5).

Bueno de Mesquita, Bruce et al. (2003). *The Logic of Political Survival*. MIT Press (cit. on pp. 2, 5).

Cohen, Jeffrey E. (2003). "The Polls: Presidential Greatness as Seen in the Mass Public: An Extension and Application of the Simonton Model". In: *Presidential Studies Quarterly* 33.4, pp. 913–924. ISSN: 1741-5705. DOI: 10.1046/j.0360-4918.2003.00091.x. URL: http://dx.doi.org/10.1046/j.0360-4918.2003.00091.x (cit. on p. 5).

Curry, Jill L. and Irwin L. Morris (2010). "The Contemporary Presidency Explaining Presidential Greatness: The Roles of Peace and Prosperity?" In: *Presidential Studies Quarterly* 40.3, pp. 515–530. ISSN: 1741-5705. DOI: 10.1111/j.1741-5705.2010.03784.x. URL: http://dx.doi.org/10.1111/j.1741-5705.2010.03784.x (cit. on p. 5).

---

[19] These prominence estimates come from a different factor model than that in Section 5.1. For this, we estimated a separate single-dimension factor model for each language.

Gleditsch, Kristian S. and Michael D. Ward (1999). "A revised list of independent states since the congress of Vienna". In: *International Interactions* 25.4, pp. 393–413. DOI: `10.1080/03050629908434958`. eprint: `http://www.tandfonline.com/doi/pdf/10.1080/03050629908434958`. URL: `http://www.tandfonline.com/doi/abs/10.1080/03050629908434958` (cit. on p. 6).

Goemans, H.E. (2008). "Which Way Out?: The Manner and Consequences of Losing Office". In: *Journal of Conflict Resolution* 52.6, pp. 771–794. DOI: `10.1177/0022002708323316`. eprint: `http://jcr.sagepub.com/cgi/reprint/52/6/771.pdf`. URL: `http://jcr.sagepub.com/cgi/content/abstract/52/6/771` (cit. on pp. 2, 5).

Goemans, Henk E., Kristian Skrede Gleditsch, and Giacomo Chiozza (2009). "Introducing Archigos: A Dataset of Political Leaders". In: *Journal of Peace Research* 46.2, pp. 269–283. DOI: `10.1177/0022343308100719`. eprint: `http://jpr.sagepub.com/content/46/2/269.full.pdf+html`. URL: `http://jpr.sagepub.com/content/46/2/269.abstract` (cit. on pp. 6, 18).

Horowitz, Michael, Rose McDermott, and Allan C. Stam (2005). "Leader Age, Regime Type, and Violent International Relations". In: *Journal of Conflict Resolution* 49.5, pp. 661–685. DOI: `10.1177/0022002705279469`. eprint: `http://jcr.sagepub.com/cgi/reprint/49/5/661.pdf`. URL: `http://jcr.sagepub.com/cgi/content/abstract/49/5/661` (cit. on p. 4).

Horowitz, Michael C. and Allan C. Stam (2012). "Leader Experiences and International Conflict". In: *Book Manuscript* (cit. on p. 2).

Kenney, Patrick J. and Tom W. Rice (1988). "The Contextual Determinants of Presidential Greatness". English. In: *Presidential Studies Quarterly* 18.1, pp. 161–169. ISSN: 03604918. URL: `http://www.jstor.org/stable/27550541` (cit. on p. 5).

Macintyre, Ben (May 5, 2010). "Ben Macintyre: My top 50 Prime Ministers". In: *The Times*. URL: `http://www.timesonline.co.uk/tol/news/politics/election_2010/article7117478.ece` (cit. on p. 2).

Marshall, Monty G. and Keith Jaggers (2005). "Polity IV Project: Dataset Users' Manual". URL: `http://www.cidcm.umd.edu/polity/` (cit. on p. 18).

Martin, Andrew D., Kevin M. Quinn, and Jong Hee Park (2011). "MCMCpack: Markov Chain Monte Carlo in R". In: *Journal of Statistical Software* 42.9, p. 22. URL: `http://www.jstatsoft.org/v42/i09/` (cit. on p. 11).

McCann, Stewart J. H. (1995). "Presidential Candidate Age and Schlesinger's Cycles of American History (1789-1992): When Younger Is Better". English. In: *Political Psychology* 16.4, pp. 749–755. ISSN: 0162895X. URL: `http://www.jstor.org/stable/3791890` (cit. on p. 5).

Michel, Jean-Baptiste et al. (2011a). "Quantitative Analysis of Culture Using Millions of Digitized Books". In: *Science* 331.6014, pp. 176–182. DOI: `10.1126/science.1199644`. eprint: `http://www.sciencemag.org/content/331/6014/176.full.pdf`. URL: `http://www.sciencemag.org/content/331/6014/176.abstract` (cit. on pp. 3, 6).

— (Mar. 11, 2011b). "Supporting Online Material for Quantitative Analysis of Culture Using Millions of Digitized Books". documentation. URL: `http:`

//www.sciencemag.org/content/suppl/2010/12/16/science.119964 4.DC1/Michel.SOM.revision.2.pdf (cit. on p. 9).

Sarkees, M.R. and F.W. Wayman (2010). *Resort to war: a data guide to interstate, extra-state, intra-state, and non-state wars, 1816-2007*. Correlates of war series. CQ Press. ISBN: 9780872894341. URL: http://books.google.com/books?id=jjIxOAAACAAJ (cit. on p. 18).

Schlesinger, Arthur M. (Nov. 1, 1948). "Historians Rate the U.S. Presidents". In: *Life*, pp. 65–66, 68, 73–74 (cit. on p. 5).

Sheppard, Simon (1998). "Ranking New Zealand's Prime Ministers". In: *Political Science* 50.1, pp. 72–89. DOI: 10.1177/003231879805000104. eprint: http://pnz.sagepub.com/content/50/1/72.full.pdf+html. URL: http://pnz.sagepub.com/content/50/1/72.short (cit. on p. 5).

Simonton, Dean K. (1992). "Presidential greatness and personality: A response to McCann (1992)." In: *Journal of Personality and Social Psychology* 63.4, pp. 676–679. ISSN: 1939-1315 (Electronic); 0022-3514 (Print). DOI: 10.1037/0022-3514.63.4.676 (cit. on p. 5).

Simonton, Dean Keith (1984). "Leaders as eponyms: Individual and situational determinants of ruler eminence". In: *Journal of Personality* 52.1, pp. 1–21. ISSN: 1467-6494. DOI: 10.1111/j.1467-6494.1984.tb00546.x. URL: http://dx.doi.org/10.1111/j.1467-6494.1984.tb00546.x (cit. on pp. 5, 6, 11).

— (1986). "Presidential Greatness: The Historical Consensus and Its Psychological Significance". English. In: *Political Psychology* 7.2, pp. 259–283. ISSN: 0162895X. URL: http://www.jstor.org/stable/3791125 (cit. on p. 5).

— (1991). "Predicting Presidential Greatness: An Alternative to the Kenney and Rice Contextual Index". English. In: *Presidential Studies Quarterly* 21.2, pp. 301–305. ISSN: 03604918. URL: http://www.jstor.org/stable/27550719 (cit. on p. 5).

— (1993). "Putting the Best Leaders in the White House: Personality, Policy, and Performance". English. In: *Political Psychology* 14.3, pp. 537–548. ISSN: 0162895X. URL: http://www.jstor.org/stable/3791713 (cit. on p. 5).

— (2001). "Predicting Presidential Performance in the United States: Equation Replication on Recent Survey Results". In: *The Journal of Social Psychology* 141.3, pp. 293–307. ISSN: 0022-4545. URL: http://www.informaworld.com/10.1080/00224540109600552 (cit. on p. 5).

— (2006). "Presidential IQ, Openness, Intellectual Brilliance, and Leadership: Estimates and Correlations for 42 U.S. Chief Executives". In: *Political Psychology* 27.4, pp. 511–526. ISSN: 1467-9221. DOI: 10.1111/j.1467-9221.2006.00524.x. URL: http://dx.doi.org/10.1111/j.1467-9221.2006.00524.x (cit. on p. 5).

Simonton, D.K. (1987). *Why presidents succeed: a political psychology of leadership*. Yale University Press. ISBN: 9780300038361. URL: http://books.google.com/books?id=8XF2QgAACAAJ (cit. on p. 5).

Singer, David J., Stuart Bremer, and Stuckey John (1972). "Capability Distribution, Uncertainty, and Major Power War, 1820-1965". In: *Peace, War, and Numbers*. Ed. by Bruce Russett. Sage, pp. 19–48 (cit. on p. 18).

Sorokin, Pitirim (1925). "Monarchs and Rulers: A Comparative Statistical Study". English. In: *Social Forces* 4.1, pp. 22–35. ISSN: 00377732. URL: http://www.jstor.org/stable/3004376 (cit. on p. 5).

Theakston, Kevin and Mark Gill (2006). "Rating 20th-Century British Prime Ministers". In: *The British Journal of Politics & International Relations* 8.2, pp. 193–213. ISSN: 1467-856X. DOI: 10.1111/j.1467-856X.2006.00220.x. URL: http://dx.doi.org/10.1111/j.1467-856X.2006.00220.x (cit. on p. 5).

	English (en)	Japanese (ja)	Spanish (es)	German (de)	Russian (ru)
1	George W. Bush	Junichiro Koizumi	Hugo Chavez	Adolf Hitler	Joseph Stalin
2	Adolf Hitler	Adolf Hitler	Adolf Hitler	Gerhard Schroder	Vladimir Putin
3	Ronald Reagan	Hideki Tojo	Francisco Franco	Wilhelm II, German Emperor	Vladimir Lenin
4	Bill Clinton	Joseph Stalin	Augusto Pinochet	Hugo Chavez	Adolf Hitler
5	John F. Kennedy	Yoshiro Mori	Joseph Stalin	George W. Bush	Nicholas II of Russia
6	Saddam Hussein	George W. Bush	Juan Peron	Konrad Adenauer	Boris Yeltsin
7	Joseph Stalin	John F. Kennedy	Vicente Fox	Joseph Stalin	Mikhail Gorbachev
8	Tony Blair	Ito Hirobumi	Jose Luis Rodriguez Zapatero	Helmut Schmidt	Alexander Lukashenko
9	Theodore Roosevelt	Saddam Hussein	George W. Bush	John F. Kennedy	Saddam Hussein
10	Jimmy Carter	Roh Moo-hyun	Salvador Allende	Mustafa Kemal Ataturk	Leonid Brezhnev
11	Fidel Castro	Yasuhiro Nakasone	Nelson Mandela	Helmut Kohl	Muammar al-Gaddafi
12	Winston Churchill	Kim Jong-il	Jose Maria Aznar	Saddam Hussein	Alexander II of Russia
13	Richard Nixon	Nobusuke Kishi	Alan Garcia	Fidel Castro	Nikita Khrushchev
14	Mao Zedong	Ryutaro Hashimoto	Benito Mussolini	Silvio Berlusconi	George W. Bush
15	Franklin D. Roosevelt	Vladimir Putin	Alberto Fujimori	Willy Brandt	Alexander III of Russia
16	Gerald Ford	Kakuei Tanaka	Vladimir Lenin	Winston Churchill	Mikheil Saakashvili
17	Benito Mussolini	Tomiichi Murayama	Carlos Menem	Muammar al-Gaddafi	Benito Mussolini
18	George H. W. Bush	Eisaku Sato	Nestor Kirchner	Vladimir Lenin	Winston Churchill
19	Vladimir Lenin	Shigeru Yoshida	Porfirio Diaz	Erich Honecker	Mustafa Kemal Ataturk
20	Woodrow Wilson	Kim Il-sung	Alvaro Uribe	Benito Mussolini	Fidel Castro
21	Ulysses S. Grant	Kiichi Miyazawa	Juan Carlos I	Mao Zedong	Yuri Andropov
22	Vladimir Putin	Keizo Obuchi	John F. Kennedy	Ludwig Erhard	Nursultan Nazarbayev
23	Dwight D. Eisenhower	Mao Zedong	Jose Lopez Portillo	Vladimir Putin	Kim Jong-il
24	Hugo Chavez	Chiang Kai-shek	Jorge Rafael Videla	Bill Clinton	Carl Gustaf Emil Mannerheim
25	Nelson Mandela	Douglas MacArthur	Ricardo Lagos	Karl Donitz	John F. Kennedy

Table 3: Leaders with the top 25 mean famescores, $\bar{\phi}_{i,w}$, for each Wikipedia.

	French (fr)	Italian (it)	Polish (pl)	Portuguese (pt)	Chinese (zh)
1	Charles de Gaulle	Silvio Berlusconi	Jozef Pilsudski	Getulio Vargas	Mao Zedong
2	Adolf Hitler	Adolf Hitler	Lech Walesa	Luiz Inacio Lula da Silva	Chiang Kai-shek
3	Francois Mitterrand	Bettino Craxi	Aleksander Kwasniewski	Adolf Hitler	Chen Shui-bian
4	Jacques Chirac	Joseph Stalin	Joseph Stalin	Fernando Henrique Cardoso	Sun Yat-sen
5	Joseph Stalin	Romano Prodi	Wojciech Jaruzelski	George W. Bush	Deng Xiaoping
6	Philippe Petain	Giulio Andreotti	Adolf Hitler	Antonio de Oliveira Salazar	Hu Jintao
7	George W. Bush	Vladimir Lenin	George W. Bush	Pedro II of Brazil	Lee Teng-hui
8	Hugo Chavez	Benito Mussolini	Vladimir Putin	Fernando Collor de Mello	Jiang Zemin
9	Nelson Mandela	George W. Bush	Boleslaw Bierut	Joseph Stalin	Adolf Hitler
10	Vladimir Putin	Muammar al-Gaddafi	Vladimir Lenin	Juscelino Kubitschek	14th Dalai Lama
11	Muammar al-Gaddafi	Massimo D'Alema	Augusto Pinochet	Nelson Mandela	Joseph Stalin
12	Abdelaziz Bouteflika	Aldo Moro	14th Dalai Lama	Fidel Castro	Chiang Ching-kuo
13	14th Dalai Lama	John F. Kennedy	Edward Rydz-Smigly	Jose Sarney	Yuan Shikai
14	Valery Giscard d'Estaing	Nelson Mandela	Francisco Franco	Vladimir Lenin	Empress Dowager Cixi
15	John F. Kennedy	Nicholas II of Russia	John F. Kennedy	Hugo Chavez	Kim Jong-il
16	Fidel Castro	Fidel Castro	Ronald Reagan	Janio Quadros	Franklin D. Roosevelt
17	Benito Mussolini	Francesco Cossiga	Muammar al-Gaddafi	Muammar al-Gaddafi	Margaret Thatcher
18	Vladimir Lenin	Victor Emmanuel III of Italy	Mao Zedong	Saddam Hussein	Wang Jingwei
19	Margaret Thatcher	Hugo Chavez	Edward Gierek	Itamar Franco	George W. Bush
20	Winston Churchill	Josip Broz Tito	Nicholas II of Russia	Washington Luis	Benito Mussolini
21	Nicholas II of Russia	Carlo Azeglio Ciampi	Winston Churchill	Deodoro da Fonseca	Harry S. Truman
22	Georges Clemenceau	Saddam Hussein	Fidel Castro	John F. Kennedy	Muammar al-Gaddafi
23	Silvio Berlusconi	Giovanni Giolitti	Alexander Lukashenko	Humberto de Alencar Castelo Branco	Vladimir Lenin
24	Saddam Hussein	Augusto Pinochet	Benito Mussolini	Nicholas II of Russia	Winston Churchill
25	Mao Zedong	Winston Churchill	Charles de Gaulle	Benito Mussolini	Saddam Hussein

Table 4: Leaders with the top 25 mean famescores, $\bar{\phi}_{i,w}$, for each Wikipedia.