To: The Knight Foundation
From: Jonathan Ladd (Georgetown) and Joshua A. Tucker (NYU)
Re: Review of Data Available to Measure the Health of Democracy and Menu of Future Options for Improved Measurement Efforts in the United States
Date: April 19, 2017

****

EXECUTIVE SUMMARY

Measuring democracy is a difficult and often thankless task. For each conceptualization or measurement strategy there is an ample set of (often quite compelling) critiques for why the given approach fails to capture the concept appropriately. And yet, for foundations seeking to strengthen the quality of democracy, having an understanding of where democracy is stronger or weaker – as well as what aspects of democracy may be in particular need of strengthening – is paramount, and this requires having access to measures of democracy.

We were invited to write this report in order to help inform the Knight Foundation's decision of whether to invest in funding new attempts to measure – or to provide access to measures of – the quality of democracy in the United States. The ultimate goal of any such new efforts would be to provide information that could be utilized by program officers seeking to fund projects aimed at strengthening democracy in the United States.

Thus, the contribution of this report is two-fold. First, we provide a comprehensive yet concise overview of the state of field in terms of democracy measures both cross-nationally (where much more direct work on the topic has been done) and within the United States. Second, on the basis of that review, we provide a preliminary menu of options for the Knight Foundation moving forward should it decide to invest resources in improving measures of the quality of democracy in the United States.

Section 1 provides a review of **twelve major cross-national data collections** that either rate countries as democratic or not, rate the prevalence of different democratic criteria, or provide data on political activities closely related to democracy (e.g., elections). For each collection, we describe the data itself, note its relative prominence in the literature, and provide examples of research conducted using the data.

Section 2 provides a review of **seven major existing datasets to measure the quality of democracy in the United States** over time and/or across regions. On key finding is that, while there are many sources of data on the United States, the data are quite different from the international comparative datasets. More specifically, different data collections tend to examine a particular aspect of the American political system, rather than assess it as a whole. That said, for each of these measures, we describe how they have been used to assess the state of American democracy and how prominent they have been among academics and journalists to this point.

Section 3 presents a number of options for future Knight Foundation funded efforts, including two proposals for **improving access to existing sources of data** three proposals for **new data**

collection projects that Knight could support:

3.1: *Expert Surveys* on the quality of democracy in the United States, drawing on best practices and examples from the comparative literature

3.2. New *Mass Surveys* explicitly aimed at measuring the quality of democracy in the United States by asking questions in new ways

3.3. Harnessing *Social Media* data to develop new tools for measuring the quality of democracy in the United States

Our overall conclusion is that there are certainly opportunities for improving measures of the quality of democracy in the United States, important lessons that can be learned from prior efforts to measure democracy cross-nationally, and potentially exciting opportunities for harnessing new sources of data in this effort moving forward.

Section 1: Cross-National Measures of Democracy

We begin by surveying efforts to measure the presence, strength, and quality of democracy cross-nationally. We do so because there has been a tremendous amount of effort to *explicitly* do so in both the scholarly and policy communities (i.e., projects aimed at measuring democracy, as opposed to incidental measures included in studies that can be used for this purpose).

We located thirteen data collections that contain comparative measures or descriptions of democracy, which includes eleven quantitative measures related to democracy and two qualitative collections of descriptions of the state of democracy. Our analysis here is largely focused on the former group of quantitative indexes. The eleven quantitative indexes can loosely be grouped into three categories:

1) Datasets that produces **measures of "democracy"** across countries over time (or something close to that)
2) Datasets that produces measures related to **individual elections**
3) Datasets that produce measures that we would expect to be **related to democracy**, but are not necessarily measures of democracy per se.

Moreover, we have identified four basic ways that these variables are to be generated:

1) Datasets that produce a single set of expert assessments
2) Datasets that survey experts, and then present measures that use various statistical techniques to synthesize the views of experts
3) Datasets that measure "objective" factors, such as the presence of absence of elections
4) Datasets that use statistical techniques to summarize measures from other datasets

Table 1 (below) provides a concise description of each of these 13 data sets.

Table 1. Comparative Datasets and Key Attributes

Index Name	Countries	Years Covered	Measurement Method	Components	Indicators	Scale
Bertelsmann Stiftung	129	2006-Present (Biannual)	Expert survey	17	2	[0, 10]
Democracy Barometer	70	1990-2014	Index aggregation	18	1	[0, 100]
EBRD Reform Index	29	1994-Present	Internal survey	13	2	[1, 5)
Economist Intelligence Unit	167	2006-Present	Internal survey	5	1	[0, 10]
Fragile State Index	178	2006-Present	Internal survey	168	12	[0, 120]
Freedom House	195	1973-Present	Expert survey	25	3	[1, 7]
National Elections across Democracy and Autocracy	200	1945-2012	Internal survey	3	1	{0, 1}
Polity	167	1800–2014	Expert survey	6	6	[-10, 10]
State Department	195	1999-Present	Internal survey	N/A	N/A	Qualitative
UK Ministry of Foreign Affairs	27	2003-Present	Internal survey	N/A	N/A	Qualitative
Unified Democracy Scores	195	1946-Present	Latent variable	10	1	[-2.5, 3.5]
V-Dem	173	1900-Present	Expert survey	30	7	[0, 1]
Perceptions of Electoral Integrity	2012	2012-2016	Expert survey	49	N/A[1]	N/A

In the remainder of the section, we concisely survey each dataset individually, describing how it is put together, providing some assessments of its strengths and weaknesses, and then provide a few illustrative examples of how the measure has been utilized in the literature to date. At the end of the section, we provide some comparative data on the relative impact of each of these collections.

1. Freedom House

Freedom House (FH) is an aggregation of many smaller metrics presumed to be associated with a functioning democracy coded by thousands of experts. The overall Freedom House score ranges from 1 to 7, where 1 would be a highly democratic country and a country with 7 would be extremely autocratic. This index is itself the mean of two ratings – one for civil rights and the other for political rights – that are determined from expert responses to FH's questions. Compared to other indexes, Freedom House has a broad conception of democracy which includes measures such as whether or not the government is free from corruption and unions are allowed to operate freely.

The best thing that can be said for Freedom House is that it incorporates a wide variety of

[1] The Perceptions of Electoral Integrity survey includes 49 expert questions that evaluate the quality of electoral institutions in a given election-year. However, the authors of the survey do not aggregate their results into any indicators or indexes.

measurements to create the FH index. This makes FH one of the more broadly defined measures of democracy, which is valuable to researchers who are interested in democracy defined broadly. Perhaps it is for this reason that FH's index has historically been used at higher rates than any other index.

Freedom House's index has also attracted its fair share of critics. Its aggregation procedure has been accused of having no theoretical justification (Munck and Verkuilen 2002). Freedom House does not make public the components that are used to create their index. Freedom House does not provide a clear guide to its coders as to what the different ordinal values of its various components refer to; thus, it is not clear that FH's coders had a common scale in mind when creating the index. Finally, the design choices FH made for its index – such as making higher values of the FH index indicate lower levels of democracy and scaling it from 1 to 7 – contrast unfavorably with other indexes that have largely agreed that higher values indicate higher levels of democracy and have typically chosen round numbers at the high and low ends of their democracy scales (such as Polity which ranges from [-10, 10]).

The Freedom House data set has been mentioned by more academic articles in Google Scholar than any of the other data sets examined since 2003 (See Figure 1 below). Moreover, the articles in which it has been used are themselves highly cited: the average number of citations for the top 100 article search results for Freedom House is 321, second only to the Polity data sets. Moreover, Freedom Houses' various metrics have been used for a wide variety of purposes in empirical studies, several of which we describe here.

Freedom House's democracy and media indexes have found broad use in the academic literature. Articles seeking to explain how education, growth, geography, executive power sharing, and regime type have affected democratization have all used Freedom House scores as dependent variables in their analysis (e.g., Brinks and Coppedge 2006; Hadenius and Teorell 2007; Norris 2008; Yared et al. 2005). Freedom House's index includes a measure of media freedom that has been used to predict political knowledge, participation, and turnout (Leeson 2008). Additionally, it has been used to determine whether democracy impacts economic growth and repression, (Acemoglu et al. 2005; Poe and Tate 1994; Tavares and Wacziarg 2001). Finally, even when FH's index was not an author's first choice to measure democracy, it was sometimes used as a check to ensure that empirical findings were robust to alternative measures of democracy (Quan Li 2005).

Despite some criticism of Freedom House's methodology (Munck and Verkuilen 2002), it has been widely applied in articles published in some of the most reputable of political science and economic journals; for instance, it has been applied to empirical analyses in multiple articles published in the *American Journal of Political Science, the American Political Science Review, Journal of Economic Perspectives, and the American Economic Review*. Overall, in the last 13 years Freedom House has maintained a position of high visibility and influence in the academic literature.

2. Polity

The Polity data sets are aggregate measures of democracy that rely on hundreds of surveys filled out by experts on democracy and area studies. There are three polity indexes: the Polity2 index, the Democracy Index, and the Autocracy Index. The Polity 2 index is calculated as the

difference between the Democracy Index and the Autocracy Index; it ranges in value from 10 (high democracy) to -10 (high autocracy). The indexes themselves are the result of the weighted sum of three components whose values are determined by expert coders: the degree to which the executive is recruited through free and fair elections; the independence of executive authority; the degree to which political competition is allowed in electoral politics. In contrast to Freedom House, Polity's creators provide clear coding rules to their expert coders; this at least in theory should increase the inter-coder reliability of Polity relative to Freedom House.

However, Polity's critics accuse it of inaccurate and highly volatile measures during periods of regime transition (Skrede Gleditsch and Ruggeri 2010). Also, Polity is described as being problematic for researchers interested in measuring the impact of democracy on incidence of civil war, as political violence is included as a component of the Polity index (Vreeland 2008). Finally, Polity has been accused of aggregating unlike components to create the Polity2 variable with no clear theoretical justification for their aggregation methodology (Coppedge et al. 2011; Munck and Verkuilen 2002).

The Polity data sets have grown dramatically in popularity since the early 2000s (see Figure 1 below). While Polity is mentioned in fewer articles than Freedom House or the EIU, the average polity paper in the first 100 Google Scholar search results has 405 citations – significantly more than any other data set. Thus, it could be said that the polity data is at least as influential, if not more so, in the political science community relative to the other data sets reviewed here.

The data set itself is used in a variety of ways. Most commonly, it is used as a control variable in cross national studies to ensure that the degree of democracy or autocracy of a state does not bias the relationship between an independent and dependent variable of interest (Gehlbach and Malesky 2010; Gerring et al. 2015; Leeson 2008). It is also frequently used as an independent variable to predict outcomes such as economic growth and state consolidation (Carbone and Memoli 2015; Wu 2012). Finally, the polity data is seen as sufficiently reliable that it is often used to measure democratization and democratic consolidation as outcomes of other variables (Eisenstadt, LeVan, and Maboudi 2015).

In summary, the polity data is used as a proxy for democracy and democratization in just about every way possible in the political scientific literature. Moreover, despite the emergence of newer metrics of democracy quality, it remains the most highly cited and arguably most influential current measure of democracy.

3. V-DEM

The Varieties of Democracy (V-Dem) index is an expert-survey-based measure of democracy that in some ways resembles those of Polity, but is at the same time is newer, less well known, and more methodologically sophisticated in their methodology of aggregating indexes. One of its most notable components is that fact that V-Dem explicitly bases each of its seven democracy indexes – the electoral, liberal, majoritarian, consensual, participatory, deliberative, and egalitarian indexes – on a separate theoretical conception of democracy (Coppedge et al. 2011); this is useful for those who may be more interested in one conception of democracy than others. The broader V-Dem index is the mean of these seven indexes.

Like Polity and Freedom House, V-Dem uses expert surveys as its primary source of data.

Unlike those two indexes, rather than using sometimes arbitrary methods for aggregating the indicators derived from expert surveys, V-Dem uses Bayesian factor analysis (BFA) to identify latent features within the data. These features are aggregated to the level of the index to form the V-Dem score. This method has an advantage over using author-defined weights (Polity) or a simple additive approach to combine components to create indexes, because it allows an unbiased algorithm to identify patterns that underlie multiple variables rather than relying simply on human intuition.

In addition to being methodologically sophisticated, V-Dem has adopted several best practices that compare favorably to other expert survey-based indexes. First, unlike Freedom House, V-Dem makes available a disaggregated version of its democracy index which allows other researchers to create their own indexes using V-Dem's many sub components. Also, this allows academics to replicate V-Dem's index to ensure that it has been estimated as stated by its authors. Another advantage offered by V-Dem (and Polity) is that its experts are subjected to tests to ensure that inter-coder reliability is maintained. Overall, V-Dem takes the expert survey and improves upon it with more sophisticated methods of weighting its composite score and by maintaining a clear focus on democratic theory throughout the creation of the V-Dem index.

The V-Dem index is a new arrival in the field of democracy metrics; it has been mentioned in over 600 articles and cited at a rate of 52 times per article. Additionally, it appears that the number of articles that mention V-Dem have rapidly increased since 2011 (Figure 2). However, the vast majority of the articles that mention Varieties of Democracy were self-published by the Varieties of Democracy Institute – the organization that produces the V-Dem index - not any peer reviewed journal.

Despite being self-published, quite a few of the Varieties of Democracy Project's papers have been well cited. In one of those articles, the authors use the disaggregated form of the V-Dem index to measure to what degree various countries fall into the following categories: liberal democracy, deliberative democracy, egalitarian democracy, and participatory democracy (Coppedge et al. 2016) . In another article, Edgell et al (2015) use the V-Dem data to explore whether iterated elections increase the quality of rule of law and protections of civil rights. They use measures of rule of law and civil rights protections from the V-Dem index as their independent variables and the aggregated V-Dem index (excluding the measures used as independent variables) as their dependent variable; overall, they find support for the idea that elections increase civil rights protections (Edgell et al. 2015).

In sum, while the V-Dem Index has not been used in many articles published in major journals, the articles that have used it have been more highly cited than all but one of the newer indexes (excluding established indexes made by Polity, Freedom House, and the Economist Intelligence Unit).

4. Economist Intelligence Unit

The Economist Intelligence Unit's (EIU) measure of democracy in some ways resembles the less sophisticated aspects of the Freedom House democracy index. Like FH, the EIU index is composed of the evaluations of experts; unlike FH, the EIU depends entirely on experts. The index itself is the average of five categories – electoral process and pluralism, civil liberties, the

functioning of government, political participation, and political culture. Each of these categories is itself an aggregation of expert evaluations.

It is worth noting that EIU does not theoretically ground their index in the philosophical or theoretical literature on democracy. Moreover, EIU does not make clear its rationale for any of its methodological decisions. Likewise, -- from the choice of its questions, to its choice of the method by which it weights each question -- the methods by which EIU generates its index are extremely opaque.

Overall, it appears that – like FH – EIU decided to use a maximalist definition of democracy to create the EIU index. Unfortunately, it also appears that EIU has adopted some of the less desirable practices of FH: methodological opacity, arbitrariness of methodological decision-making, and a lack of grounding in the academic theoretical literature. Perhaps that is why the EIU index was described by Bühlmann et al as not serving an academic purpose (Bühlmann et al. 2012).

Moreover, the usage of the Economist Intelligence Unit data is in some ways paradoxical: while the EIU data is frequently mentioned in academic articles (though its popularity appears to have dropped precipitously since 2013), its democracy index is rarely used in empirical models in the academic literature. In cases where it is employed, it generally appears to be done so in articles in European journals and in less well-known American journals. Of course, there are a few exceptions. The EIU data is used to measure the effect of democracy on ecological sustainability in a cross-national empirical study published in Nature (Hauser et al. 2014). The index was also one of many sources used to measure the level of political participation in a paper that explored the relationship between political participation during the formation of constitutions and future level of democracy (Eisenstadt, LeVan, and Maboudi 2015). However, these articles are the exception rather than the rule.

The solution to the aforementioned paradox appears to be the following: EIU data is frequently mentioned in passing for the purposes of conveying a statistic about the economy or politics more generally; it is not, however, commonly used by academics to measure democracy in empirically rigorous analysis.

5. Unified Democracy Scores

The United Democracy Scores (UDS) were developed by Pemstein et al (2015), as an attempt to create a measure of democracy that aggregated popular indexes, but wasn't biased by each index's individual biases and errors. They use a latent variable approach – in which an algorithm aggregates a diverse set of measures of some latent variable that is assumed to underlie each of them – using 10 measures of democracy to generate a score that ranges from 3.5 (high democracy) to -2.5 (low democracy) for each of the 195 countries evaluated in the data set.

This dataset has the advantage of incorporating nearly all the major democracy scoring systems out there, including Polity, V-DEM, and Freedom House. The basic scores thus could be said to provide the most thorough overview/summary of the current received wisdom regarding cross-national variation in democracy. Thus, these scores could be extremely useful for tracking where the US lies relative to other countries over time, as well as how the current trend in democratic development in the United States (e.g., has the country been getting less

democratic in recent years?).

One disadvantage of the UDS is that with latent scores, one cannot observe the contributing factors to the score. This means that the scores in the dataset have no intuitive meaning, which may make them particularly ill-suited for philanthropic organizations interested in targeting particular aspects of democratic backsliding/weakness. Also, if the input measures of democracy used by UDS have some general bias that is common to all inputs, then the UDS will retain that bias.

As a relatively new measure (approximately six years old), the UDS data has not been used in many academic articles, nor have the articles that have used this data received many citations. However, after reviewing the literature, it appears to be the case that UDS is gaining a following among academics and has already been used in several articles that were published by well-known journals.

Perhaps unsurprisingly, UDS is most commonly used to measure overall levels of democracy and autocracy. For instance, Eisenstadt et al (2015) used UDS as one of their two metrics for democracy when they examined the effect of historical political participation on future democratization. Additionally, many new attempts to create democracy indexes have used UDS as a baseline measure, as it effectively captures the latent democracy related attributes of 10 other measures of democracy – including Freedom House and Polity (Coppedge et al. 2011; Gugiu and Centellas 2013; Skaaning, Gerring, and Bartusevičius 2015). Finally, UDS has been used in some less conventional applications: to evaluate the effects of democratization on health outcomes (Safaei 2012) and to measure the effects of democracy on acquisitions and mergers (Guardo, Marrocu, and Paci 2015).

Overall, it appears that UDS data are starting to make a meaningful impact in a reasonably short time. One author commented that the only justification for the continued popularity of larger indexes – such as Polity and FH – relative to UDS is inertia in the field (Marzagao 2014). While that may be an overstatement, the Bayesian latent variable approach employed by the authors ensures that their index retains some of the most salient features of the 10 indexes it was applied to. Thus, UDS or another index using similar methods may see more use in the future.

6. Democracy Barometer

In contrast to the indices that rely primarily on expert surveys, The Democracy Barometer relies heavily on cross-national public surveys, data derived from IGOs – like the World Bank – and other non-expert indicators to estimate their index. The DB largely assumes democracy to be liberal – respecting individual rights – and participatory – allowing and even requiring collective action. The DB index is itself aggregated from over 105 indicators that were presumed to be associated with an underlying component of democracy; these indicators include measurements from the Afrobarometer, the International Labor Organization, the World Bank, and others. The components that are estimated from these indicators include checks and balances of power, freedom of association, freedom of opinion, equality before the law, and many others.

Also, the Democracy Barometer project has the advantage of providing details on a broad range of conceptual issues relevant to democracy. By limiting the study to established democracies,

the DB system does not spend time addressing the difference between autocracies, monarchies, competitive authoritarianism or other non-democratic regimes. The downside is that it only covers democratic countries and does not provide easy combined scores.

As another new measure of democracy, the Democracy Barometer has received only a modest amount of attention from the politics literature, to date being cited in only 379 articles.

There have, however, been a few applications of the Democracy Barometer in the recent academic literature. In an attempt to address the question of whether Western Democracies have been experiencing political decline in recent decades, Merkel (2014) uses the Democracy Barometer to measure differences in political freedom and democracy quality between developed democracies. Overall, he finds that democracy has not declined in any measurable way (Merkel 2014). In a similar article, Baldini (2015) used the Democracy Barometer to measure Britain's level of democracy over time in order to evaluate whether British democracy was in decline; in a conclusion similar to that of Merkel, he found that while there was little apparent decline in the quality of British institutions, public distrust of those institutions had still increased over time (Baldini 2015).

In sum, the Democracy Barometer remains a niche metric that has not been mentioned in many articles or been in many highly-cited articles. However, the fact that it has been used in published studies despite being so new relative to the more established metrics is itself noteworthy.

7. Bertelsmann Stiftung

The Bertelsmann Stifung Transformation Index is designed to be a general gauge of the degree to which a given country's economy has transitioned to a market economy and its political system to a democracy. The index is aggregated from the input of nearly 300 experts from around the world who evaluate over 129 countries. The data for the democracy index is generated according to the following procedure: first, experts are provided with country reports provided by Bertelsmann Stiftung; second, those experts answer questions designed to assess the degree to which the country has democratized; third, a second expert independently evaluates the same country independently of the first expert; the questions are then aggregated to generate five categories: rule of law, stateness, political participation, stability of democratic institutions, and political and social integration.[2] These five components are then condensed into a single measure of democracy that ranges from 1 to 10.

The Bertelsmann Stiftung Transformation Index is a newer data set that has been cited in a modest number of articles (~6,000), although these articles in turn do not seem to have been highly cited in the literature. In fact, only the United Democracy Scores have fewer citations on average than the Bertelsmann Index.

The Bertelsmann Index has seen limited applications in the politics literature. For instance, it has been used by Merkel and Croissant to classify democracies as being electoral, embedded, or defective democracies (Merkel and Croissant 2004). It has also been applied by Møller and

[2] The remaining 12 components measure the degree to which the country's economy is a market economy and how well the country is managed.

Skaaning in a meta-analysis of the many different metrics for rule of law (Møller and Skaaning 2011).

Overall, the Bertelsmann Index appears to not have caught on in a large way in the academic literature. Few of the articles in which the data set was used have garnered much attention from the academic community and it has yet to see a publication in a larger journal. However, it is still a relatively new data set so it remains to be seen whether it will do so in the future.

Election-related Datasets

8. National Elections Across Democracy and Autocracy (NELDA) and Electoral Integrity Project (EIP)

The National Elections Across Democracy and Autocracy (NELDA) is a database of election results that includes a binary measure of whether or not an election was contested. That measure itself depends on three criteria: whether or not opposition was permitted; whether multiple parties are legal; and whether more than one candidate competed in the election. If all three of these conditions are met, then NELDA's index considers the election to be contested; otherwise, it considers the election to be uncontested (Hyde and Marinov 2012). These variables were coded by a collection of undergraduate and graduate research assistants who based their coding choices off of news articles obtained from the ProQuest and Lexis-Nexis archives, the Inter-Parliamentary Union's archives on parliamentary elections, and other sources.

The NELDA index has some advantages relative to other measures of democracy. First, it is highly focused on a single outcome – elections – that are highly relevant to the proper functioning of a democracy. That said, the fact that NELDA's key indicator of whether or not elections were competitive leads to very little variation in the outcome of interest; we would likely see countries like China classified the same as Russia – which obfuscates the fact that while Russia holds national elections, they are neither free nor fair, unlike China which holds no national elections whatsoever. These distinctions matter as there is a substantial literature on the instability of transition regimes, or regimes that lie somewhere between democracy and autocracy (Goldstone et al. 2010).

In terms of usage, NELDA is still quite new. Of the newer data sets, it is perhaps the most widely used in articles that have been published in reputable journals. Still, it has not been mentioned in many articles itself (having been mentioned in about 250 articles) and remains relatively poorly cited (an average of 22 citations). Its relatively low number of mentions and citations can be dismissed due to its newness and its strengths as employing a diverse array of sources to focus on the precise problem of ex ante election quality distinguishes it in a positive way from the competition.

There have been many new articles that have made use of the NELDA data. First, NELDA has been commonly used as a measure of coups and electoral violence as a dependent variable (Daxecker 2012; Hafner-Burton, Hyde, and Jablonski 2013; Marinov and Goemans 2014). Additionally, NELDA data was used to measure whether or not democratization occurred, where democratization was defined as the onset of competitive elections (Donno 2013). It is noteworthy that these articles were published in well-known journals, such as the *British Journal*

of Political Science and the *American Journal of Political Science*.

In sum, it appears as though NELDA is a data set that has effectively carved out a niche for itself among the many sets of data measuring democracy. While other data sets focus on expert surveys of democracy writ large, Hyde and Marinov focus specifically on the quality of elections, which has resulted in it being used in articles that have research questions that involve electoral processes. Perhaps because of its focus, NELDA appears to be making a substantial impact on the political science literature.

We also want to briefly mention the Electoral Integrity Project (EIP), which is similar to NELDA in that it focuses only on elections. Indeed, it is narrowly focused on the procedural aspects of carrying out elections. Unlike NELDA, EIP uses expert surveys to classify elections as being aligned along several relevant dimensions including whether electoral laws were fair, whether electoral procedures were evenhanded, whether electoral boundaries favored one party or another, etc. Overall, there are 11 EIP indexes that are generated based off of the expert surveys gathered by EIP. These 11 indexes were themselves derived from 49 items in the survey completed by 226 specialists (in 2013). To ensure inter-coder reliability, EIP also assessed the degree to which the specialists agreed with one another. The EIP data set includes measures that are broader in scope than those used in NELDA, yet they are similarly restricted to the scope and quality of elections.

The two election-related data sets – NELDA and EIP – represent a highly limited but precise measure of one of the most important features of democracy: the competitiveness and validity of elections. One important question is the following: are NELDA and EIP measuring democracy? To the extent that democracy is contingent upon free and fair elections, NELDA and EIP could be said to be valuable contributors to our understanding of democracy; however, it would be hard to argue that democracy can be defined in total according to the quality of elections. Freedom of press, the degree to which the executive is constrained, etc. Thus, while we might learn from their methods or incorporate them into a broader index, NELDA and EIP are – by themselves – of limited utility to those who wish to define a broader concept of democracy.

Datasets related to democracy, but not producing measure of democracy: *Fragile State Index; EBRD Reform Index; US State Department Human Rights (USSDHR) reports; UK Ministry of Foreign Affairs Countries for Concern*

Finally, we very briefly describe review several datasets that are related to democracy, but do not directly measure democracy or do not do so in a quantitative sense.

The *Fragile State Index (FSI*) assesses the degree to which a state is in a condition of turmoil or is likely to enter into a state if instability; one can think of it as measuring the inverse of state strength. While it is an interesting indicator of state strength, it is not designed to be, nor should it be, used as a measure of democracy. It is worth mentioning that this index has been growing in popularity due to the salience of state weakness following the Arab Spring in 2011.

The *European Bank of Reconstruction and Development (ERBD)* maintains an index – determined by their Chief Economist – that measures the degree to which post-communist states are transitioning into democratic and market-oriented states. This index is limited in its

scope – as it is limited to post-communist countries – and its indexes are primarily concerned on the degree to which the country is building new infrastructure and undertaking economic reforms, not implementing democratic elections; moreover, the fact that ERBD's index appears to be determined by the whims of its Chief Economist leads us to be skeptical of its reliability.

Both metrics have admirable goals, but if our intention is to measure the quality of democratic institutions, neither index accomplishes that goal; nor has either index been used in that capacity in the academic literature.[3]

The *US State Department Human Rights (USSDHR) reports* are an archive of qualitative evaluations of the condition of human rights in 195 countries. While a useful resource for informing expert coders, the USSDHR reports themselves contain no quantitative measures of democracy or human rights conditions. These reports are mentioned in over 8000 articles in Google Scholar and were cited an average of 178 times per article, indicating a reasonably high degree of influence in the political science literature. However, since these reports are themselves highly qualitative, they have been mostly used in qualitative analysis or as information used by human coders to create quantitative metrics of democracy or human rights.

One illustrative example of a qualitative application of the Country Reports can be found in Carleton and Stohl's use of the reports to identify motivations for US foreign aid in a qualitative analysis. More recently, the reports have been used by several authors to create human rights indexes. In 1999, Cingranelli assigned the CRHRP to human coders who then scored the country reports to create a national-level index measuring human rights. Similarly, Sinkkink and Walling use the reports to create a data set on human rights practices and domestic judicial activity (Sikkink and Walling 2007). Finally, Neumayer used the text to create a measure of respect for integrity rights (Neumayer 2003).

Overall, while the CRHRP are reasonably well cited, their qualitative nature appears to have limited the breadth of their applications in the academic literature. As the field has moved away from qualitative analysis (the decline in popularity of these reports is visible in Figure 1), it appears that their use as a source of data for human coders has increased over time.

Similarly, the *United Kingdom Ministry of Foreign Affairs (UKMFA)* produces evaluations of the conditions of human rights for "Countries of Concern" -- countries that have repeatedly violated the human rights of their citizens in the past. As we saw with the USSDHR reports, the UK reports are primarily useful as tools for training human coders; not as data themselves.

In sum, while these data sets – FSI, ERBD, USSDHR, and UKMFA -- have value in certain contexts, they are substantially less relevant to the task of measuring democracy than Polity, Freedom House, V-Dem, and the other quantitative measures of democracy discussed in this section.

Comparative Usage of Existing Data Collections

[3] For applications of the EBRD Reform Index, see Broadman & Recanatini (2002), Eschenbach (2002) and Hoekman (2006) (Broadman and Recanatini 2002; Eschenbach and Hoekman 2006; Gehlbach and Malesky 2010). Likewise, for applications of the Fragile States Index see Sanin (2011), Kolk (2015), Boddam-Whetham (2016), and Werrell et al (2015) (Boddam-Whetham, Gul, and Al-Kobati 2016; Kolk and Lenfant 2015; Sanín 2011; Werrell, Femia, and Sternberg 2015).

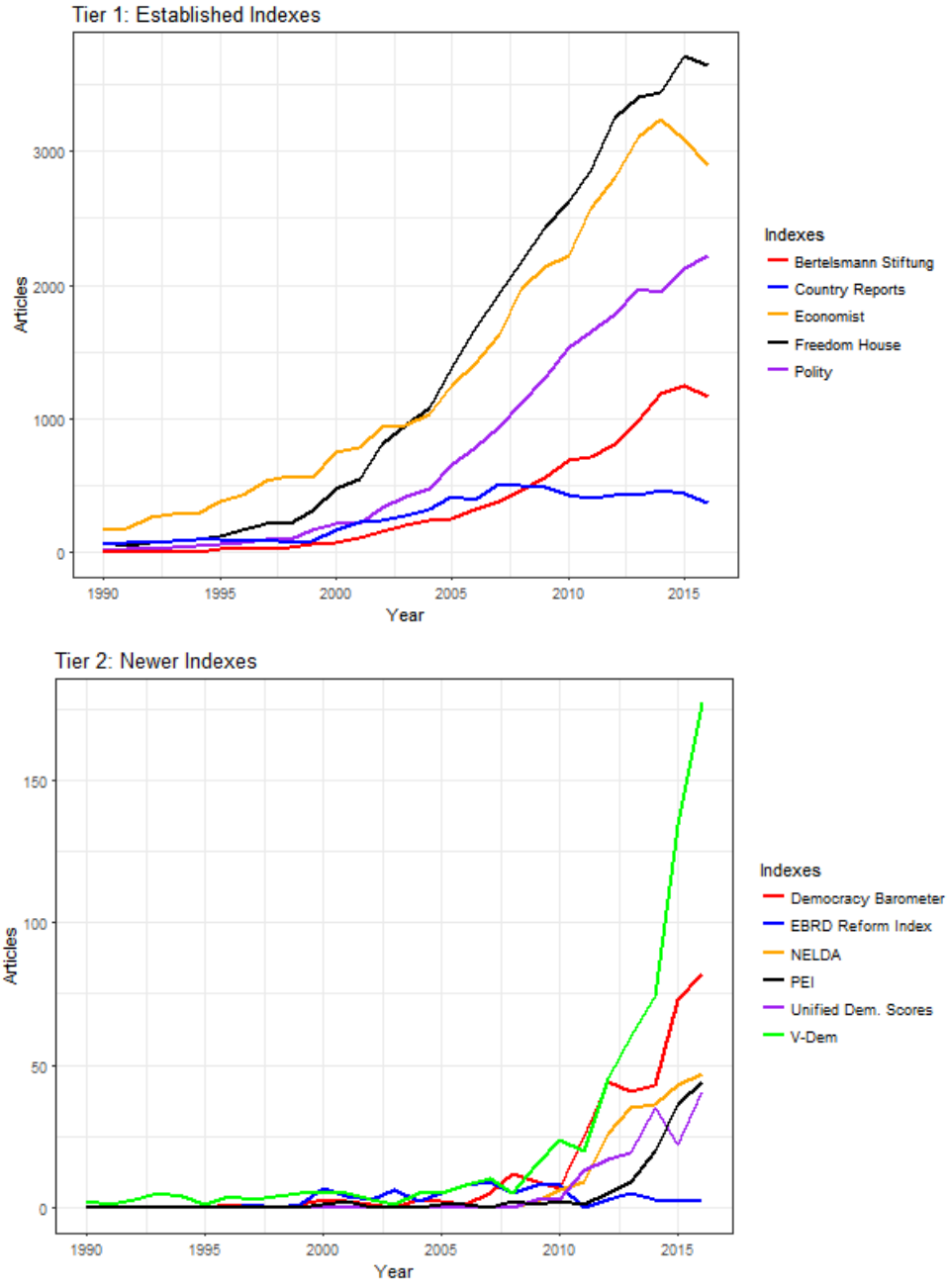
The following table shows the relative popularity of the different indices in the scholarly literature. As is clearly evidence, Freedom House, EIU, and Polity have had by far the greatest impact in the literature to date.

Data Set	Search Query	N Articles	Mean Cites
Freedom House	"Freedom House" data	35,100	321
Economist Intelligence Unit	"Economist Intelligence Unit" data	31,600	122
Polity	"polity I-IV" data	16,000	405
State Department	"Country Reports on Human Rights Practices"	8,070	178
Bertelsmann Stiftung	"Bertelsmann Stiftung" data	6,050	21
V-Dem	“Varieties of Democracy” data	675	52
Fragile States Index	"Fragile States Index" data	575	32
Democracy Barometer	"Democracy Barometer" data[4]	379	71
National Elections across Democracy and Autocracy	"National Elections across Democracy and Autocracy" data	215	22
Unified Democracy Scores	"Unified Democracy Scores" data	158	16
Perceptions of Electoral Integrity	“Perceptions of Electoral Integrity” data	142	14
EBRD Reform Index	"EBRD reform index" data	84	37
UK Ministry of Foreign Affairs	“UK Ministry of Foreign Affairs” and “countries of concern”	0[5]	0

Using Google Scholar data, we also calculated the number of articles that mentioned each data set in each year from the years 1990-2016 (see Figures 1 and 2). To make the changes observed in the popularity of each data set over time clearer, we differentiated between the more established and popular data sets of Tier 1 and the newer data sets of Tier 2.

[4] There are multiple “Democracy Barometers” – such as the Korean Barometer and South African Democracy Barometer - that were removed from our search results to ensure that we were only counting the index calculated by Bühlmann et al (2012).

[5] Despite trying multiple queries we found no articles that referenced the UK’s “countries of concern” as data sources.



Drawing on Figures 1 and 2, it is clear that differing indexes are following different trends over time. While most indexes have seen larger number of mentions since the 1990s (see Polity, Freedom House, V-Dem and Democracy Barometer), there are several that have seen peaks followed by declines (see EBRD Reform Index, the State Department Country Reports, and perhaps the Economist Intelligence Unit). These patterns are likely in part due to fundamental changes in the field – such as moving from more interest in quantitative data rather than qualitative data – and introspection within the field about what constitutes useful measures of democracy.

Section II: Measures of the Health of American Democracy Across Regions and Across Times

U.S. Datasets that Can Be Used to Measure the Health of Democracy Either Over Time or Across Regions

Index Name	Years Covered	Type of Data
American National Election Studies (ANES)	1952-Present (Biannual)	Repeated Surveys
General Social Survey (GSS)	1972-Present	Repeated Surveys
Gallup Confidence in Institutions Time Series	1993-Present (Not Easily Publicly Available)	Repeated Surveys
Pew Research Center's Compilation of Trust in Government Measures (From Pew, *New York Times*, *Washington Post*, CNN and ANES Polls)	1958-present	Repeated Surveys
Richard Cole and John Kincaid's Dataset on Trust in Local Government	1972-2007	Repeated Surveys
NOMINATE (voteview.com)	1789-present	Ideological Summaries of Congressional Roll Call Voting Data
Shor and McCarty's State Legislative Ideal Points	1993-2014	Ideological Summaries of State Legislative Roll Call Voting Data

1. American National Election Studies (ANES)

The American National Election Studies (ANES) is that single most well known and most used dataset by scholars of U.S. politics. It started in its present form in 1952 and, except 2006, has been conducted every 2 years (at the time of the national presidential or midterm congressional election) since then. It measures a wide variety of political attitudes and behaviors over time. These include attitudes on prominent issues of the day as well as perennial issues in American politics that persist on the national agenda for many decades and attitudes about American society that have bearing on politics, such as racial hostility. It also includes questions about participation, vote choice, emotional reactions to candidates, and views about the political system in general (See for example, Campbell et al. 1980; Nie, Verba, and Petrocik 1979; Bartels 1988; Zaller 1992; Miller and Shanks 1996; Kinder and Sanders 1996; Lewis-Beck et al. 2008; Kaufmann, Petrocik, and Shaw 2008).

Some political scientists might argue that almost every question in the ANES in some ways is assessing American democracy. This is surely true to some extent. However, limiting ourselves to questions that allow for over time comparisons because they have been asked for at least several decades and that in some way evaluate the U.S. political system as a whole, rather than

a specific aspect of it, limits things somewhat.

Since 1952, the ANES has asked several questions about what political scientists call political efficacy, e.g. whether people believe that their voice is heard in the political system. These questions are listed below (not every question was asked in every year):

1) "'Sometimes politics and government seem so complicated that a person like me can't really understand what's going on.'" "Agree or Disagree?"
2) "'People like me don't have any say about what the government does.'" "Agree or Disagree?"
3) "'I don't think public officials care much what people like me think.'" "Agree or Disagree?"

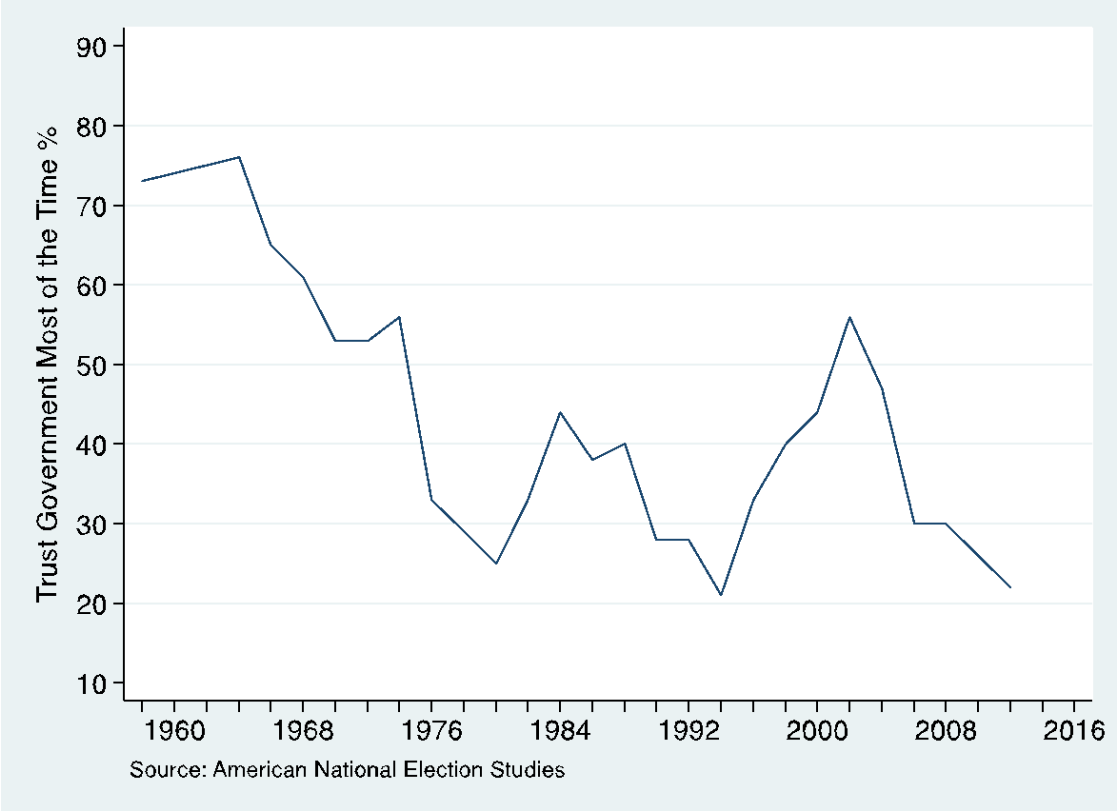
Starting in 1958, the ANES has also asked several questions about people's overall evaluations of the American government, without mentioning any specific politician or party. They are listed below:

1) "How much of the time do you think you can trust the government in Washington to do what is right-- just about always, most of the time or only some of the time?"
2) "Would you say the government is pretty much run by a few big interests looking out for themselves or that it is run for the benefit of all the people?"
3) "Do you think that people in the government waste a lot of money we pay in taxes, waste some of it, or don't waste very much of it?"
4) "Do you think that quite a few of the people running the government are crooked, not very many are, or do you think hardly any of them are crooked?"

Finally, starting in 1964, the ANES has also usually asked two questions about whether, in general, the U.S. government responds to public opinion. These questions are:

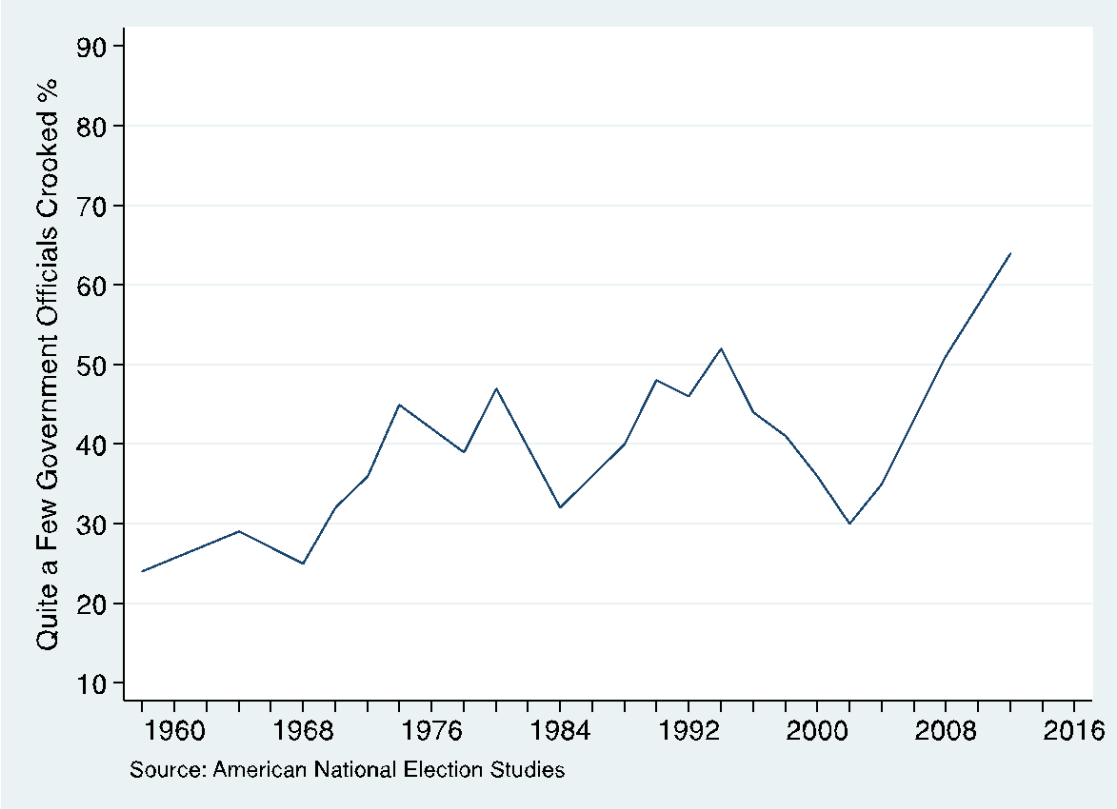
1) "Over the years, how much attention do you feel the government pays to what the people think when it decides what to do -- a good deal, some, or not much?"
2) "And how much do you feel that having elections makes the government pay attention to what the people think, a good deal, some or not much?"

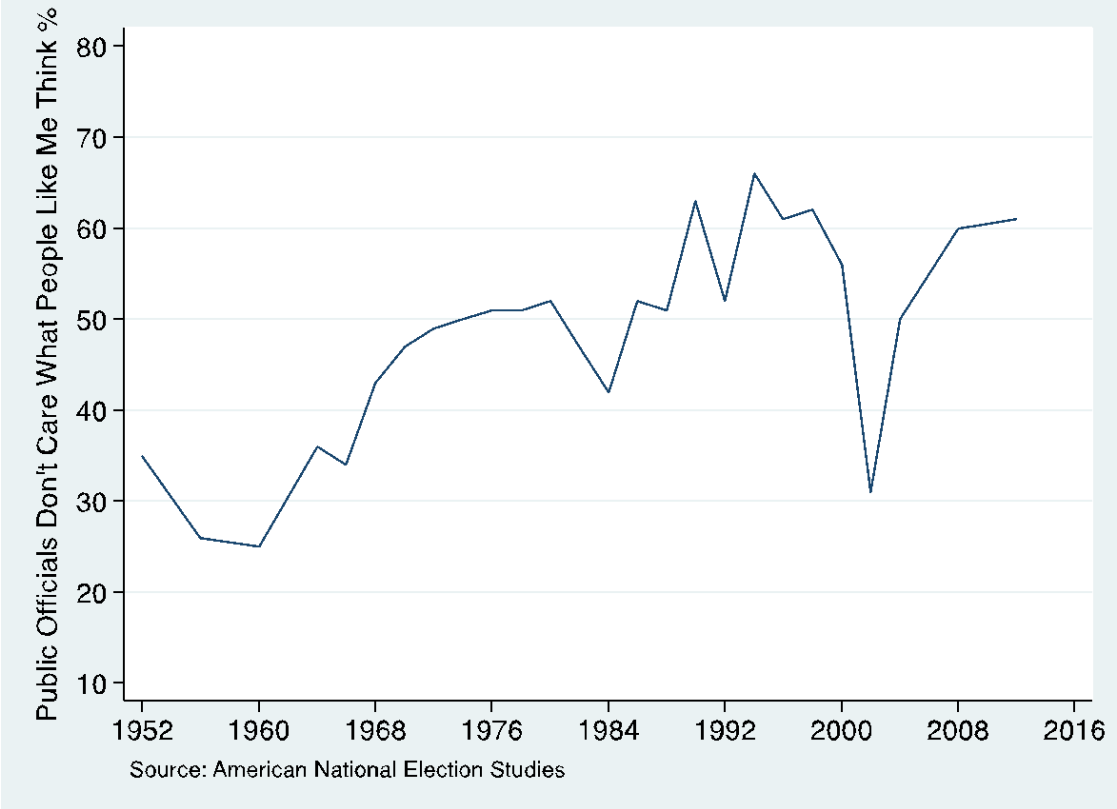
All of these measures have grown more negative between when they were first asked and the last several years. But they have not become more negative in a steady fashion. They have varied over time in response to some major events (such as the Watergate scandal and the 9/11 terrorist attacks) and to smaller events that distinguish every election year. For example, the percentage of respondents trusting the government declined from the low 70s in the 1950s to the low 20s by the 2010s. But there were notable surges in the 1980s and early 2000s.



The next two figures show the percentage of ANES respondents saying that “quite a few” government officials are “crooked” and the percentage agreeing that “public officials don’t care what people like me think.” Again, both measures are substantially worse now than they were in the 1950s, but there have also been seemingly important ups and downs over the intervening years.

This leads us to conclude the following about the ANES as an over-time indicator of the quality of American democracy. First, the ANES is useful because it contains several measures of the public’s overall faith in the democratic process in the United States that go back many decades. Second, all of these questions show not just a steady decline, but notable variation over time in addition to an overall decline. This indicates that the causes of these attitudes are complex. Third, all of these questions are fairly vague. For instance, what exactly does it mean to “trust” the “government.” Both trust and the government could mean many possible things. What does it mean to say that “government officials” are “crooked”? Which officials? What constitutes crookedness? Their vagueness makes these questions weaker as indicators of the strength of the democratic system.

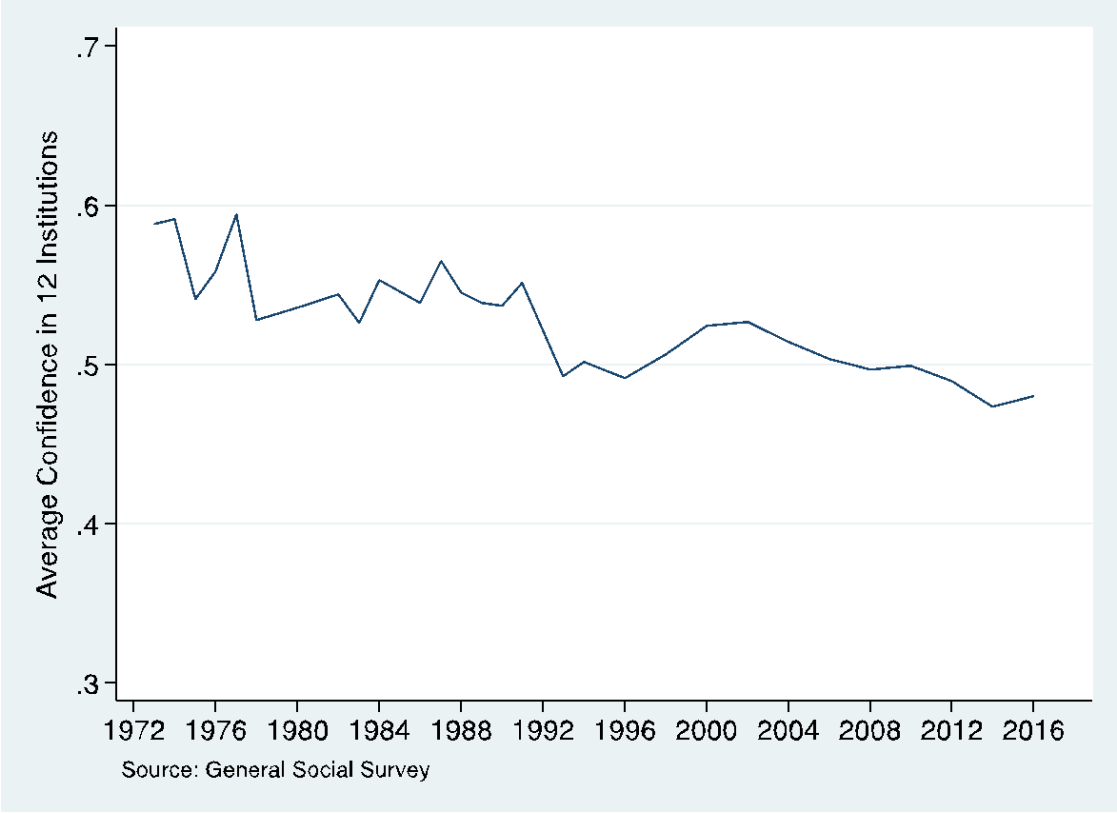

Quite a Few Government Officials Crooked %
90
80
70
60
50
40
30
20
10
1960
1968
1976
1984
1992
2000
2008
2016
Source: American National Election Studies


Public Officials Don't Care What People Like Me Think %
80
70
60
50
40
30
20
10
1952
1960
1968
1976
1984
1992
2000
2008
2016
Source: American National Election Studies

2. General Social Survey (GSS)

The General Social Survey (GSS) is, in many ways, similar to the ANES. It is also funded by the National Science Foundation and overseen by a Board of Overseers drawn from academics across the country. The main differences are that the GSS began in 1972, it is run mainly by sociologists instead of political scientist, it's primary focus is social rather than political attitudes and behaviors, and, while it is conducted approximately every two years, it is not held to correspond with the timing of national elections.

Given its sociological focus, it has asked fewer questions over the decades measuring people's faith in the U.S. democratic process. Applying the same criteria that we used with the ANES (e.g., looking for questions that evaluate the entire political system and are asked repeatedly over time) leads to fewer results. However, there is a question battery that could be used for this purpose. In the second GSS in 1973 and every one since, it has included a series of questions asking respondents whether they have "a great deal," "only some," or "hardly any" confidence in 13 American national institutions. These are: major companies, organized religion, education, the executive branch, organized labor, medicine, television, the Supreme Court, the scientific community, Congress, the press, television, and the military. Overall confidence in these institutions has declined. The loss of confidence has been greatest in the press, television and the three branches of the national government, but it has occurred for almost all institutions to some extent (e.g., Lipset and Schneider 1987; Ladd 2012). The average confidence level across these 13 institutions (calculated by coding "a great deal" as 1, "only some" as 0.5 and "hardly any" confidence as 0) dropped from .59 in 1973 to .49 in 2012.



The confidence question battery has essentially the same strengths and weaknesses as the various ANES questions about faith in American government and politics. They have the benefit of being asked regularly over a long period of time. But they have the drawback of being somewhat vague measures of faith in the political process. It is plausible that, as people lose confidence in all big American institutions, it will lead to them to no longer put faith in the political system to solve the nation's problems. However, it would require further investigation to demonstrate whether that is true. In many ways, the GSS's confidence battery provides new and interesting questions about Americans' beliefs about the democratic process but in and of itself does not answer them.

3. Gallup Confidence in Institutions Time Series

Since 1993, Gallup, Inc. has, in its own polls, asked a confidence question battery that is very similar to that used by the GSS. The question and response options are the same. The only difference is that the institutions are slightly different. Gallup respondents are asked about their confidence in: the military, police, church or organized religion, the medical system, presidency, U.S. Supreme Court, public schools, banks, organized labor, criminal, justice system, television news, newspapers, big business, and Congress.

In contrast to the GSS, Gallup shows a modest increase in average confidence in the early to mid-1990s. Yet both surveys agree that since the mid-2000s average confidence has substantially declined (Norman 2017). Both indicate that confidence is currently quite low.

The Gallup battery gets as much or more attention in the popular press, but the GSS battery is much more frequently used by academic researchers. There are several reasons for this. The GSS battery is asked as part of a large sociological survey. Therefore, one can look at how confidence in these institutions is correlated with a long list of social attitudes, behaviors and demographic characteristics. In contrast the Gallup battery is asked as part of standard 12 minute Gallup polls with a small number of demographic questions and different opinion questions depending on the survey. Finally, and possibly most importantly, it is easy to download all years of the GSS data from the GSS website ("GSS General Social Survey | NORC" 2017). In contrast, old Gallup polls measuring confidence are available from the Roper Archive at Cornell University, but each poll must be searched for in the archive and downloaded individually, then the results combined to calculate time trends. Given no obvious advantages in quality for the Gallup battery, the fact that the GSS is so much easier to use and goes back 20 years earlier leads it to be much more widely used by researchers.

4. Pew Research Center's Compilation of Trust in Government Measures (From Pew, New York Times, Washington Post, CNN and ANES Polls)

In addition to the "trust in government" question on the ANES, various other major polling organizations have asked about trust in the federal government using the exact same or a very similar question wording over the decades. The Pew Research Center has provided a valuable service by collecting the results of all of these polls together into one dataset. This combines polls asking about trust in government conducted by the ANES, the *New York Times*, the *Washington Post*, and the Pew Research Center itself. This combined dataset, which goes back

to 1958, when the ANES first asked the question, allows for a much more detailed understanding of how levels of trust in government have waxed and waned over time (e.g., Pew Research Center et al. 2015)

The weakness of this for the research community is that, like Gallup's confidence in institutions tie series, that data are not easily available to the public or even the social science research community. The results from the combined datasets are regularly published in reports issued by the Pew Research Center on the state of polarization. This compilation of data from all the organizations that have asked about trust in government in polls over the years would be much more useful than the ANES data alone if such a compilation was made more widely available to the public.

5. Richard Cole and John Kincaid's Dataset on Trust in Local versus State versus Federal Government

Richard Cole and John Kincaid have collected surveys on which level of the U.S. government people have the most confidence in and how that has changed over time. The US Advisory Commission on Intergovernmental Relations (ACIR) conducted surveys that compared the public's attitudes toward these three main levels of the US government from 1972 to 1994. Since then, Cole and Kincaid conducted regular surveys using the same questions from 1989 through 2010, allowing one to use all the surveys as a single time series of data.

This is a crucial dataset in the American political system because, unlike in many other nations, all three levels of government have important powers and legislators can change the policy tasks delegated to the different levels over time. Surprisingly, given how central federalism is in the American political system, to our knowledge Cole and Kincaid seem to be the only scholars to make use of these data. However, they have looked at them in a long series of articles (Cole and Kincaid 2000; Cole, Kincaid, and Parkin 2002; Kincaid and Cole 2005; Cole and Kincaid 2006; Kincaid and Cole 2008, 2011).

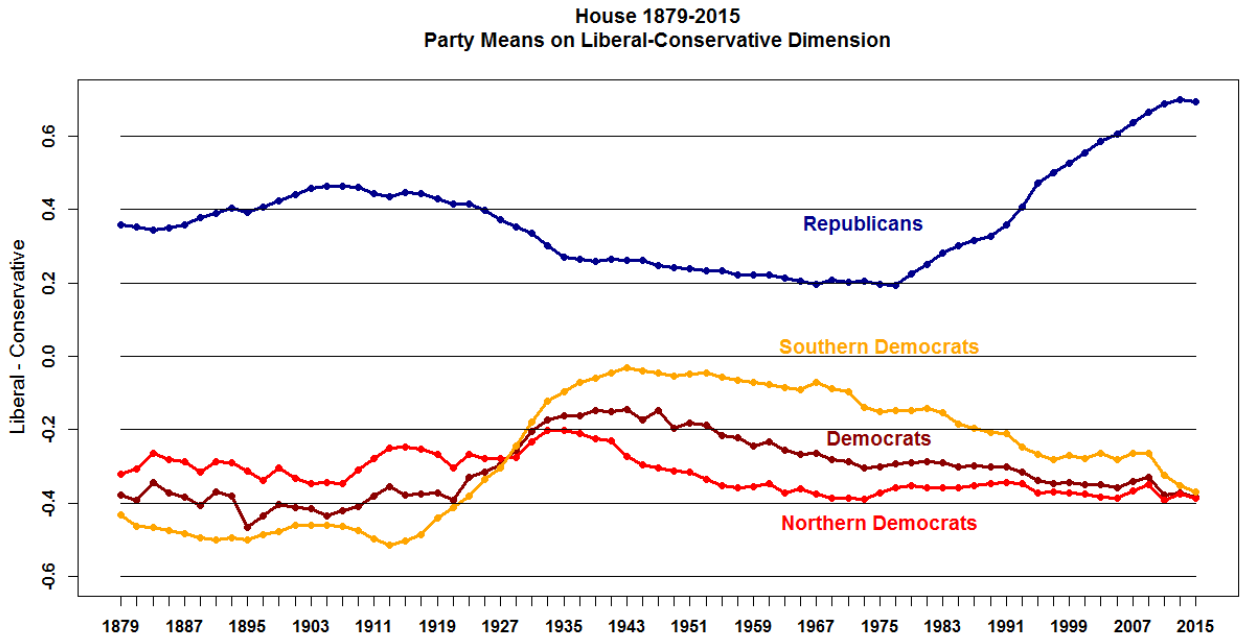
They have found several consistent patterns over time. People persistently think that, compared to the state and local governments, the federal government gives them the least for their money, has the least fair taxes, has too much power and is least trustworthy. White people and the more highly educated are a bit more likely to have positive attitudes toward the federal government compared to the other government levels. Also, there was a notable and historically unusual uptick in support for the federal government after the 9/11 terrorist attacks, but the effect entirely faded after several years.

6. NOMINATE Congressional Ideology Scores

The previous US measures of the health of democracy have been based on surveys. Now we will review two that are based on analyses of legislative roll call data. The first is called NOMINATE. It is a way of calculating ideology scores for members of the U.S. Congress based on all their roll call votes. It was developed by Keith Poole and Howard Rosenthal in the 1980s. The NOMNATE ideology scores are the most widely used data by academic congressional scholars (Poole and Rosenthal 2001).

From the beginning though, it has not just been used to study the ideologies od individual members, but to look at the ideological make-up of political parties (e.g., Poole and Rosenthal 1985, 1991; McCarty, Poole, and Rosenthal 1997; Poole and Rosenthal 2000; McCarty 2001; Poole and Rosenthal 2007; McCarty, Poole, and Rosenthal 2006; Poole and Rosenthal 2001; Carroll et al. 2009; McCarty, Poole, and Rosenthal 2016; Jacobson and Carson 2015). In particular, it is a useful measure of how polarized the two parties are in Congress. The key issue for this is the ability to compare voting scores over time. Poole and Rosenthal developed a version of the scores, called DW-NOMINATE, that can be compared over the course of American history.

The figure (from Keith Poole), shows the NOMINATE voting score of the median Republican, median Democrat overall, median northern Democrat and median southern Democrat in the House of Representatives since Reconstruction. When you look at polarization over time since the modern two party system emerged, NOMINATE scores show big differences between the parties' ideologies during reconstruction and the Gilded Age of the late 1800s. Then, the parties move steadily closer together ideologically until they are fairly close in the 1950s, 1960s and 1970s. Yet in the decades since then, the parties have moved further and further apart. The NOMINATE voting scores of the median Democrat and median Republican in the House are now further apart than at any point since the end of Reconstruction.



Source: Keith Poole's website ("Political Polarization" 2017):
http://pooleandrosenthal.com/political_polarization_2015.htm

The polarization happening in recent decade is also evident, not in party medians, but in how many members of the two parties have overlapping NOMINATE scores. In the decades of low polarization in the middle of the twentieth century, in both the House and Senate, there were several conservative Democrats who had more conservative NOMINATE scores than some liberal Republicans. Yet, by the 2000s, this had ceased. For over a decade, the most

conservative Democrat has been more liberal than the most liberal Republican, and vice versa, as measured by NOMINATE. The parties no longer have any ideological overlap.

As a measure of the overall health of American Democracy, this has strengths and weaknesses. Its main strength is that it is a direct and clear measure of polarization, using a highly respected advanced method for statistically analyzing congressional voting. The downside is that polarization in congressional voting is at best an indirect measure of the overall health of democracy. Extreme polarization may be a symptom of bigger problems, but perhaps not always. More cross national research is needed to determine how frequently ideological polarization in national legislatures predicts broader democratic breakdown in the country suffering from it.

7. Shor and McCarty's State Legislative Ideal Points

Boris Shor and Nolan McCarty have extended the NOMINATE technology to state legislatures. Collecting this data is a painstaking process, because roll call data from state legislatures is not in electronic format in a large number of states. Shor and McCarty had to scan paper records of votes in order to calculate NOMINATE liberal-to-conservative voting scores for every legislator across all 50 states (Shor, Berry, and Mccarty 2010; Shor and McCARTY 2011; McGhee et al. 2014). As a result of the difficulty of doing this, it has only been done for all states for the time period 1993-2014. We cannot make the kind of overtime caparison that we can with NOMINATE scores for the U.S. Congress, which we have going back to 1789. The over-time comparisons we can make are limited, but these scores provide an excellent opportunity to measures partisan polarization across states. The results are often fascinating. Even in states with similar overall ideologies, some have much more legislative polarization than others. For example, the median member of the state House and Senate is quite liberal in both California and Massachusetts. However, there are very different levels of party polarization. In California, the parties vote very differently in the state legislature, the Republican minority votes much more conservatively than the Democratic minority. Yet in Massachusetts, things are different. Both parties are less extreme than in California. Compared to California, in Massachusetts Republicans are much less conservative and Democrats are a bit less liberal. As a result, the gap in voting scores between the Democrats and Republicans is much smaller in Massachusetts.

The limited time period (and huge effort required to collect state data back further in time) limits the usefulness of this data somewhat. Yet they are very useful for determining the state level correlates of political polarization in the last few decades. This is why, in less than a decade since they have been released, they have been cited 120 times according to Google Scholar. As with national NOMINATE scores, it is important to remember that polarization is only one aspect of democratic health. We need more research to understand better how it is correlated with other indicators of weakness in the democratic system. Still, looking at variation across states in polarization is a promising future direction for studying the effects of polarization on the political system.

Section III: New Directions for Knight Foundation

In this section, we present a menu of options for the Knight Foundation should it choose to invest in either trying to come up with new ways to make already existing data more accessible (Part A) or to fund the collection of new data (Part B). In case any of these options prove palatable to the foundation, we also provide suggestions for outreach in initially publicizing the new tools or data (Part C) as well as using the new tools and data to foster a growing discussion of democratic trends (Part D)

A. Proposal for New Multi-Media Tool for Accessing/Sharing Data in Parts I & II.

The point we want to make here is simply that there is **already a great deal of data** currently available measuring the presence, strength, and quality of democracy internationally. However, we are not currently aware of any common gateway to access all of this data; indeed, it has taken us a great deal of time to come up with these lists ourselves, and we cannot even guarantee that they are comprehensive.

One contribution that Knight could make to its own program officers, the policy community, and the scholarly community would be to provide some sort of democracy portal or dashboard that accessed all these data from a common interface.

The "**portal**" idea is probably the simpler of the two, and would involve having a concise list of the types of data available, where to find them, how to access the data, brief summaries of how the data has been used previously, example of utilizing these data, etc. In sum, something very similar to what we have in prepared in this report, but only in an interactive online format. This could become an invaluable resource to anyone interested in conducting their own data analysis.

A "**dashboard**", however, would take this idea one step further, and allow users to not only learn about the availability of data, but would additionally provide simple, user friendly, tools for visualizing and interacting with the data. Thus, one could, for example, seamlessly examine how trends in the United States compare to trends in whatever other countries one selected, and could do so simultaneously across multiple data sets. For instance, general trust in government is measured in several U.S. surveys as well as international surveys. One would be able to select a variable such as government trust, and immediately see in which countries it was measures and over which years. Then, with another click, you would be able to see in a graph trends in government trust in as many of these countries that you wanted to. Such a tool would probably come closest to the original goal of the Knight Foundation of providing quick, accessible information for program officers in the future.

To be clear, a small investment in funds could create either the portal or the dashboard or both utilizing *existing datasets*. However, as we have noted above, there are gaps in the existing data, especially in so far as they pertain to the United States, and therefore a more ambitious effort would involve funding the collection of new data.

B. A Proposal for New Data Collection

While the options for new data collection are of course limitless, we propose three areas of potential emphasis for the Knight Foundation: **expert surveys** on the quality of democracy in the United States, drawing on best practices and examples from the comparative literature; **new mass surveys** explicitly aimed at measuring the quality of democracy in the United States by asking questions in new ways; and **harnessing social media** data to develop new tools for measuring the quality of democracy in the United States.

B.1. New Measures of the Quality of Democracy in the United States Harnessing Methods from Comparative Studies of Democracy

To date, the United States has always been included in the cross-national studies of democracy. Thus, there are people who have been coding the United States as part of these expert analyses. However, all such studies have focused at the aggregate level, providing a single, primarily yearly, measure of "democracy" for the United States.

One potential new direction for the Knight Foundation – and which we suspect could be especially valuable for program officers at Knight and other foundations – would be to try to take the methods – namely, **expert surveys and analysis** -- that have been developed in the context of measuring democracy cross-nationally and apply them to measuring democracy at **a more disaggregated level** in the United States, i.e., to move beyond one score for the entire country for an entire year.

This is an opportune time to engage in such an initiative not just because of the growing interest in the topic, but also because we can harness the previous experiences of the cross-national efforts to establish a set of best practices to use in this endeavor.

Disaggregated Measures

How might we want to disaggregate measures of democracy in the United States beyond a "one score per country per year" approach? We suggest three potential approaches for consideration.

First, and perhaps most obviously, would be to attempt to differentiate quality of democracy **geographically.** We could imagine coming up with democracy measures that vary by region of the country, by state, or even by city. The advantages of doing so would be that – if done properly – this should be of enormous use to foundations trying to figure out where to target spending and programs. We suspect, however, that a robust variable measuring variation in the quality of democracy by state would also attract quite a bit of attention from scholars, and especially from those studying state and local politics (as evidenced by the use of the Shor and McCarty state-level polarization measures). Indeed, we could even imagine this serving to catalyze a new area of study – in which there is probably a great deal of interest in the current climate – among state and local politics scholars.

The challenge, however, to this approach would be finding enough state-level (or city-level) experts to complete the necessary expert surveys. As we discuss below in the best practices discussion, we ideally would want at least multiple coders per unit, which means this approach would require a minimum of multiple people per state with enough expertise to conduct the survey. While not impossible by any stretch, this would involve a robust effort to recruit experts

that goes beyond the normal pool associated with “country experts”. On the other hand, that might also be a benefit of the project (getting more people involved), and perhaps could piggy-back on the efforts of *Bright Line Watch* (described below). The other major question to consider here is whether there actually is sufficient variation in the quality of democracy by state in the United States to warrant this kind of major investment, so perhaps the project would benefit from piloting in a small sub-sample of cases.

Second, one could attempt to measure the quality of democracy for different **socio-demographics** groups in society. Thus, rather than trying to disaggregate conditions geographically, one could try to measure quality of democracy as experienced by older versus younger citizens, rural vs. urban citizens, or by different racial or ethnic groups. We are not aware of previous efforts to measure democracy in this way, but it is possible that it might pick up new sources of trends that more general measures would be less likely to observe. Again, it might help foundations think seriously about how to target grants and projects in the most efficient manner.

To be clear, thinking about the quality of democracy for different components of the population would require instructing experts to focus on topics that vary across socio-demographic groups. While some typical indicators might not vary, others that focus on barriers to participation might. However, this is also the case when we think about disaggregating measures of democracy geographically as well. Still, this type of data collection would likely involve using new questions and instructions for experts that would take time to evaluate.

A *third* approach to disaggregated measures of democracy in the United States would be to increase the **frequency** of assessment. So, while the comparative survey measures mainly produce at most a single measure per year, the Knight Foundation could instead sponsor quarterly, or even monthly, measures. While less novel, such an approach could have three advantages: (1) in a time of fast moving political development, it could make it easier to pick up new trends; (2) it could lead to more accurate measures for program evaluation, as the measure could be more closely tied to when a particular program was implemented; and (3) it would most likely allow for a more direct copy of one of the existing cross-national approaches – just being administered more frequently – thus reducing start costs, especially in terms of evaluating new questions.

Best Practices

Regardless of which of these approaches are adopted, it is a fortuitous time for any expert-survey based approached to measuring democracy more thoroughly in the United States (or elsewhere) because the decades of attempts to do so cross-nationally have yielded a number of clear recommendations for best practices:

1) Definitions of democracy should be based on **clear analytical concepts**, and ideally concepts that are closely linked to existing academic and theoretical literature on democracy. The gold standard here is the new V-Dem project, which has explicitly linked its definitional measures of democracy to different existing theoretical conceptions of democracy. Such an approach has three main advantages: (1) It makes the decision of what to actually code much less *ad hoc* or idiosyncratic; (2) Different experts should be more likely to arrive at the same conclusions if there is a coherent underlying rational for the coding process; and (3) The data

will be generated in a way that makes it ideally suited for theory testing from the get go, thus making it more likely to have a wider impact among academic studies.

2) Instructions to experts/coders should be **clear** and should be **made public.** Again, clarity will be extremely valuable in providing coherence across different experts and coders. Furthermore, publicizing the exact questions that were presented to experts will allow anyone who analyzes the data more of an understanding of how it was generated and, therefore, how best to analyze and interpret it.

3) One thing we have clearly learned from comparative studies is that no matter how good instructions are, there is still likely to be a lot of variability in how different experts view the world. Therefore, no matter what data are collected, we are more likely to uncover ground truth if we use **multiple coders** and report **inter-coder reliability scores** (essentially, how much different coders agree or disagree with each other) as part of making the data public. Indeed, attempts should be made to incorporate the uncertainty across coders into the ultimate scores produced, which leads directly to the final point:

4) There are new and sophisticated **methods for averaging across coders** that allow for more accurate measures of the "latent" democracy variable of interest. Both the V-Dem and Unified Democracy Score datasets rely on such approaches, and should be carefully studied by anyone producing new expert-based democracy score measures.

Taken together, these four guidelines are a very good starting point for any new expert-based democracy scoring method. One important trade off, however, should be noted. Clearly, more coders and more sophisticated methods for combining coder responses can lead to ultimately more accurate measures. However, as methods for computing these scores become more sophisticated, interpreting them intuitively becomes more difficult. Simply asking experts to assess whether quality of democracy is improving, declining, or staying about the same is much easier to interpret than the underlying latent variable produced than knowing that the Unified Democracy score for the United States in 2012 was 1.564765. Thus, careful attention needs to be paid both to what is the ultimate intended take-away from the measures (e.g., do we want identify trends? outlier regions? comparisons to other countries?) in thinking about how to apply what has been learned from best practices.

Existing Efforts

There are two relevant new efforts at US expert-based measures that are worth highlighting here.

The first is the **Electoral Integrity Project,** under the direction of Professor Pippa Norris of Harvard University, which was described previously in Section 1 and has used expert surveys to measure the quality of elections cross-nationally since 2012.[6] We mention it here again because in 2014 and 2016 they replicated their analysis in the United States across all 50 states plus DC.[7] While this is not an attempt to measure democracy *per se* (as discussed previously)

[6] https://www.electoralintegrityproject.com/what-we-do/
[7] https://www.electoralintegrityproject.com/featured-dataset

and the 2016 study has generated a fair amount of controversy[8], this would be a very important resource for anyone attempting a comparative 50 state expert survey to measure quality of democracy.

The second is an even newer project that began as we were in the process of preparing this report, which is the **Bright Line Watch** project, under the direction of four political scientists at Yale, Dartmouth, and the University of Rochester.[9] The project – at the moment – is primarily focused on two efforts. The first is to regularly measure the opinion of political scientists at US universities regarding the quality of democracy in the United States in an effort to provide a baseline measure from which changes in coming years can be measured. Respondents were presented with a battery of democratic characteristics and asked to assess how important they are for democracy and then to rank how well the US is performing on each criterion. In the first wave, 1571 political scientists completed the survey (out of 9820 who were invited to take part). Details regarding the results of the first wave of the survey can be found here http://brightlinewatch.org/?p=167. The second goal of the is convene meetings of experts to discuss trends in democracy in the United States, as well as it seems to publicize other such meetings.[10]

We think it important to note that the Bright Line Watch survey seems to straddle the line between a mass-based survey and an expert survey. Unlike the mass based surveys discussed below and in Section 2, only "experts" were invited to participate in the survey, so it is not intended to be representative of the general population, and indeed the questions seemed phrased in a way to tap into professional expertise. On the other hand, many, many more experts were surveyed than in any of the comparative projects; moreover, the bar for being an expert seemed to be not research expertise on the topic, but rather simply being a professor in a political science department at a university in the United States. We are not exactly sure how often the surveys will be repeated – the next wave is planned for May – but they could be an important step towards what we have called a "disaggregated over time" approach.

B.2: New Democracy Focused Surveys of the United States

Addressing Existing Deficiencies in the Data on the United States

The existing survey measures of assessing democratic health in the American mass public have several strengths, but also some weaknesses. The strengths are that we have a substantially longer time series of survey measures related to democratic health in the United States than in other democracies.

The American National Election Studies (ANES) was first conducted in roughly its current form in 1952 and has with only one exception been conducted every two years since them. Many key

---

[8] https://www.washingtonpost.com/news/monkey-cage/wp/2017/01/23/after-a-professors-op-ed-went-viral-this-is-what-happened-next/.

[9] http://brightlinewatch.org/.

[10] A third goal of the project is to commission essays on specific topics related to the question of democratic quality in the United States.

survey questions were first asked in either the 1950s, 1960s or 1970s. The General Social Survey began in 1972 and its battery measuring confidence in American institutions started in the survey the following year.

In addition, several private organizations asked a wide variety of different survey questions in the 1950s, 1960s, 1970s, and 1980s. These organizations tend not to consistently ask a battery of questions repeatedly in their political polls, creating a detailed time series. But political questions asked many decades in the past by that modern researchers think ae particularly indicative of democratic health can be asked again in new polls to allow for a comparison between then and now. An example of this is the Gallup question from 1960s asking whether people would be upset of their child married someone from the opposite political party. It was not asked in polls for several decades afterwards. But since being rediscovered by Iyengar, Sood and Lelkes (2012), this question has been a popular one for measuring non-ideological (or "affective") animosity between partisans in the current era compared to a baseline of the much less polarized 1960s.

In summary, the fact that major national academic surveys were founded earlier and this started asking questions about democratic performance earlier, plus that fact more commercial firms were conducting more frequent political surveys with a wider variety of questions further in the past, means that there is more opportunity to measure over time changes of the mass public in the U.S. than in other democracies. As a point of comparison, the oldest major European survey the Eurobarometer, began in 1973. The World Values Survey, Comparative Study of Electoral Systems, and European Social Survey started in 1981, 1996 and 2002, respectively. In summary, an advantage of U.S. public opinion datasets is that, there are lot more questions measuring the health of democracy starting in the 1970s and earlier than there are in most other democracies.

However, there is a weakness that goes along with these poll questions asked over a longer period of time. They are often vaguely worded, asking less directly about beliefs in the democratic process than we would like if we were writing the questions from scratch today. However, if one wants historical comparisons, one is stuck with examining questions that were asked of the U.S. public decades ago.

In Section II, we reviewed the useful measures of democratic health contained in ANES surveys, but let us provide a few examples here that illustrate our point. One of the most direct, longest running, and most well-known measures of faith in the political system in the ANES is their trust in government question, which asks: "How much of the time do you think you can trust the government in Washington to do what is right-- just about always, most of the time or only some of the time?" A second similar question in the same ANES battery asks "Would you say the government is pretty much run by a few big interests looking out for themselves or that it is run for the benefit of all the people?" This question battery measuring overall assessments of government, which includes these two questions plus a question about whether "people in the government waste a lot of money and a final question asking if "people running the government are crooked," have been asked in almost every ANES survey since 1958.

The American public's perceptions of government on these 4 questions have gotten dramatically more negative between the late 1950s and now. As noted in Section 2, previous work has speculated about the causes of this declining faith in government (Nye, Zelikow, and King 1997;

Hibbing and Theiss-Morse 2001, 1995; Hetherington and Rudolph 2008; Chanley, Rudolph, and Rahn 2000) and tracked which types of people tended to lose faith in government (Hetherington and Rudolph 2008; Keele 2005; Citrin 1974), and a smaller group of scholarship has even explored possible consequences of this trend (Hetherington 1998, 2004; Hetherington and Globetti 2002; Hetherington and Rudolph 2015; Chanley, Rudolph, and Rahn 2000).

However, a problem in assessing how worried we should be about these indicators of declining faith in the American political system is that the questions are all fairly vague. Each refers to either "the government" or the "people in the government" or "running the government." The government is vast, including three branches of politicians and judges, plus staff in each branch, plus a large sprawling bureaucracy. So it is unclear who is being evaluated (in psychologists' terms the "attitude object" (Eagly and Chaiken 1993)). On top of this, the type of evaluation respondents are making is vague as well. What constitutes "trust" or being "crooked" or "wasting a lot of money." While we understand how much these measures have declined over time and to some degree what other attitudes and behaviors they can help us predict, we don't know exactly what people are thinking when they are answering them. Who are they evaluating when the talk about trust or crookedness? And what types of specific instances or attributes come to mind when people think of those attributes?

The same vagueness issues apply to the ANES's political efficacy (whether you think government will respond to you) 3-question battery and the General Social Survey's (GSS) 12-question confidence in institutions battery. While previous scholars have speculated about what people think about when they answer these questions (Lipset and Schneider 1987; Ladd 2012; Finkel 1985; Pollock 1983; Madsen 1987), the phrasings are general enough that it is not entirely clear which political considerations these questions bring to the top of people's heads when they answer them.

Understanding what people think about when answering these questions is crucial to evaluating the health of American democracy. The American public provides much more negative answers to all of these questions now than they did in the middle of the twentieth century. When people give negative responses (as they increasingly do) does it mean that they are losing faith in the concept of democracy itself, in the particular political system that the U.S now employs, or (much more ephemerally) are they mostly just unhappy about the people currently in charge, while still supporting the underlying system.

Proposal for Open Ended Questions on National Surveys to Get at the Meaning of Long-Existing Questions

To get at this question, we propose conducting a national survey that asks these questions (the ANES's government trust and political efficacy batteries and the GSS's confidence in institutions battery) and then, after each of these batteries asks a memory dump (or "stop and think") style survey question (Zaller 1992; Zaller and Feldman 1992; Ladd 2012). Specifically, this question would ask respondents: "Please tell me all of the thoughts that went through your head when you answered the last four [three/twelve] questions about trust in government [political responsive/confidence in institutions]." They would then be allowed to provide an open-ended response. They would say as much as they wanted on the topic. When they finished, we would give them a final work of encouragement: "Anything else?" After that, they would get one more shot to say anything they wanted on the topic.

Afterward, we will carefully and systematically analyze these open-ended responses in two ways. We would use recently developed automated content analysis software called topic models, which extract automatically from raw text the distinct topics that are contained in it. (This method is called Latent Dirichlet Allocation (Grimmer and Stewart 2013)). But as a robustness check, we will also analyze the responses with human coders, which one of us used to analyze this type of memory dump open-ended questions in the past (see Ladd 2012). The human coders can both look for the prevalence of different topics, to verify that the automated coding is correct, and they can make other observations that the software would not notice. We will interview our human coders after they read through all of the responses to see whether there is anything else memorable about the responses besides the topics, which we should investigate further. If there is, we will read through the open-ended responses ourselves to learn more about what people were thinking when answering these questions.

Overall, the automated topic coding software and the human coders will give us a much better picture of what people are thinking when they answer these questions. This will greatly improve our understanding of what the large declines in support for the American political system in these survey questions over the decades really means. What is it about the American political system that people have lost faith in? What do these low levels of support for the system indicated in these questions tell us about how the American public will respond to threats to democratic institutions that we may face in the future. In short, we know that support for responses to these questions have gotten much more negative over the decades, but because these questions are somewhat vague, we know relatively little about what that means about what precisely the public likes and dislikes about politics right now and which aspects of the system they still strongly support and would defend if necessary.

We also propose a second data new survey data collection effort to rectify the deficiencies in the existing U.S. data. Validating the existing questions which have been asked in the United States for decades, as we proposed above, would be very helpful. A second parallel approach is to ask U.S. national samples some of the questions from international surveys that more explicitly probe support for the concept of democracy. U.S. national surveys have not asked these types of questions because, to be frank, support for the foundations of democracy was often taken as a given, not something that was even worth asking in the United States. While this may or may not be true, never asking these more explicit questions about support for democracy prevents anyone from comparing the United States to other countries around the world on these measures. For instance, the World Values Survey has asked a 7-question battery of questions on support for democratic norms to people in a wide range of countries with differing levels of economic and political development. These 7 questions, listed below, is a set of questions measuring support for democracy that has been asked repeatedly in many countries worldwide as part of the WVS (See Pop-Eleches and Tucker 2017):

"I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good, fairly good, fairly bad or very bad way of governing this country?"

1) "Having a democratic political system" (4-point scale)
2) "Having a strong leader who does not have to bother with parliament and elections." (4-point scale)

3) “Having the army rule” (4-point scale)

“I'm going to read off some things that people sometimes say about a democratic political system. Could you please tell me if you agree strongly, agree, disagree or disagree strongly, after I read each one of them?”

4) “In democracy, the economic system runs badly” (4 point agree-disagree scale)
5) “Democracies aren't good at maintaining order” (4 point agree-disagree scale)
6) “Democracies are indecisive and have too much quibbling” (4 point agree-disagree scale)
7) “Democracy may have problems but it's better than any other form of government” (4 point agree-disagree scale)

Our goal is to ask this battery repeatedly in the United States over the next several years. This will allow us to compare support for democratic norms in the United States to other countries in the modern era of political polarization and possible threats to American democratic norms. It will also allow us to see how support for these democratic norms changes over the next 4 years in response to new events in US and world politics. The beginning of the Trump administration has already been extremely eventful. The remaining 3 and a half years promises to be even more so. We think it is incredibly useful for tracking whether the events of the next few years will or will not alter Americans fundamental levels of faith in the democratic process.

Specifically, we propose a “rolling cross-sectional” survey design, similar to previous surveys like the Annenberg American National Election Survey (Annenberg Public Policy Center of the University of Pennsylvania 2017; Johnston, Hagen, and Jamieson 2004). A rolling cross-sectional survey is constantly interviewing a small number of new respondents. In this way, you can look as any window of dates and get enough respondents interviewed in that window to constitute of standard-sized national survey sample of over 1,00 respondents (See Romer et al. 2006, 2003; Johnston et al. 1992). This is essentially the same method as what commercial survey firms call a “tracking poll.” Academic survey researchers call this a rolling cross-section instead, but they are essentially the same.

Specifically, we propose conducting a rolling cross-sectional poll that interviews 500 respondents every week starting in July 2017 and running through December 2020. This would be a relatively short poll, it would include demographic questions, a few very basic political questions include party identification, self-placement on a liberal-to conservative scale, a presidential job approval rating, and the World Values Survey’s support for democracy question battery. This type of survey will allow us to track changes in support for democracy over the next 3.5 years. The rolling cross sectional design will allow us to take any 14-day window of time and have a standard size poll of 1,00 people to use to measure support for democracy at that time.

In summary, we propose a onetime survey that validates existing major survey items that have measures some aspect of American democracy over many decades. We also propose to, from July 1 2017 through Dec 31 2020, conduct a rolling cross-sectional survey measuring how support for basic democratic principles changes (or stays stable) over this tumultuous time.

B. 3: Harnessing Social Media to Develop New Measures[11]

Our first two suggestions for future data collection have drawn heavily on the methods we have seen employed in previous attempts to measure democracy: expert surveys and mass surveys. In this final section, we consider the possibility of harnessing a new source of data: social media.

Social media seems like an obvious candidate for this task. After all, people are currently trying to use social media to measure all sorts of things,[12] and there are in fact enormous quantities of social media data currently in the public domain that record the musings of billions of people around the world. Moreover, social science has been busy over the past decade developing also sort of tools to turn text into data, thus making the possibility of harnessing social media for quantitative analysis that much more tantalizing, even beyond the fact that social media "posts" are often accompanied by all sort of metadata that already is perfectly suitable for quantitative analysis.

That being said, we are not currently aware of any study that purports to use social media data to provide a "direct" measure of quality of democracy. There are probably good reasons for this, so we begin by explaining exactly this would be so challenging – albeit with tantalizing opportunities as well – but in addition propose three alternative ways that social media data could be used to try to *indirectly* get at the strength of democracy in a country: by measuring the presence (or lack thereof) of mass attitudes associated with support for democracy;[13] by using social media as a tool to measure other indicators that are thought to be correlated with democracy (e.g., tolerance); and finally by assessing patterns of social media *usage* that are presumed to coincide with a healthy democratic society.

The Challenges of Using Social Media Data to Measure Democracy Directly

As we have shown in detail earlier in this report, almost all current efforts to measure democracy involve trying to assign a yearly score for each country in terms of how "democratic" the country is at that point in time, either in an overall "democratic" sense or in a more disaggregated component of democracy. A variety of tools are used to arrive at this score, but they largely involve relying on some form of expert analysis across more or less well defined categories that may or may not be aggregated across multiple experts. So, the question is, could we somehow use social media data to replace the expert analysis, and go straight from the posts of individual politicians and/or citizens to a measure of "democracy"?

The answer is possibly, but it would certainly be a non-trivial task with serious methodological challenges. That being said, the potential payoffs from both a scholarly and policy perspective

[11] Note: this section draws heavily on an in progress requested submission by Tucker to the newsletter of *The Comparative Democratization* section of the American Political Science Association.

[12] To name just a few related to comparative politics, scholars are using social media data to try to measure political polarization (Barberá et al 2015; Barberá 2016), information diffusion in protests (Gonzalez-Baillon et al 2011; Barberá et al 2015; the prevalence and spread of fake news (Allcott and Gentzkow 2017); communication of foreign policy positions (Zeitzoff, Kelly, and Lotan 2015); censorship priorities of authoritarian regimes (King, Pan, and Roberts 2013, 2014); communications strategies of both opposition and pro-regime elites during periods of protest (Munger et al.), etc.

[13] See for example Linz and Stepan (1996), who define democracy in part as being "the only game in town" in the minds of the mass public.

from pulling this off would also be non-trivial.

To come up with a single "how democratic is society" measure from social media would most likely require a three-step process. First, we would need to begin with pre-existing human coded measures of democracy. Next, we would need some collection of social media from each of these countries, say a random sample of tweets from the population over the period of time in questions (here, years). The final step would be to see if we could train a machine learning model to essentially to predict the level of democracy from the presence of social media data.

Such a method would be fraught with difficulties. The data collection (and storage) for such a project would likely be enormous if we chose to rely on data from the mass populace.[14] This should not be underemphasized: the cost and data storage challenges associated with working with something like the Twitter Firehose (i.e., the entire collection of Tweets being collected in real time) would make such a project reliant on serious financial backing. Another major challenge would be that we would need to train the model using text in many, many different languages, which might push scholars to rely more on metadata (which is language invariant), such as the timing of social media posts, or perhaps the geographic diffusion of social media posting. And all this precedes the point that we have no idea whether there are indeed enough "clues" in social media data to distinguish posts in democratic countries from posts in non-democratic countries. So, there would need to be a huge investment in both time and resources before we had any idea of the ultimate viability of the project. Finally, to the extent that we were ultimately interested simply in the kinds of measures we've had previously (how democratic is country X in year Y), then this massive investment in infrastructure and capabilities would simply be to replicate other ongoing studies.

All that being said, however, there are two potential payoffs worth noting here. The first is that if – despite all the challenges identified in the preceding paragraph – we could indeed train a model to produce, for example, a measure equivalent to a Polity (or Freedom House, or even Unified Democracy) Score from social media data, then once that system was in place, there would be an opportunity going forward for enormous cost savings in terms of producing these measures in the future. So if, hypothetically speaking, we were entering a world where funding for Polity and Freedom House were likely to disappear, social media generated democracy scores would be one way we could continue to have these measures going forward.

The second potential advantage is that once we are using social media to generate democracy scores, we would no longer be limited to yearly measures. In fact, we could begin to disaggregate our measures at any interval in which we were interested. Further, we would no longer need to wait for *ex-post* analysis by experts to get our democracy scores; they could, in theory, be produced on a daily basis based on the previous day's data. Combining these two points yields the possibility of the perhaps the most tantalizing aspect of moving to a social media based approach to measuring democracy, which would be some sort of early warning system for countries in which the baseline level of "democraticness" was changing.

[14] One way out of this conundrum would be to train a model based not on social media usage by the population but rather by elites, such as members of the legislature and government ministers. While still a very serious undertaking – finding the social media accounts of members of the government from a single county is a time consuming process, and remember these people (and accounts) are always changing – the data storage and retrieval demands would be potentially more manageable.

However, these potential payoffs also raise a key final caveat of which any effort in this regard must be cognizant, which is that social media itself is dynamic. Training a model on data from 2015 does not guarantee that that model will continue to be relevant in 2017 not so much because the machine learning technology itself might be flawed, but rather because the very signals the model might have learned to pick up in 2015 to identify a more democratic society may no longer be present in the data in 2017. Consider, for example, a world in which emoji are replaced by stickers; a model trained on emojis might simply cease to be relevant. And herein lies the rub: if the benefit of a social media based approach to measuring democracy is that it could replace the very measures that were used to generate the model in the first place (as in the funding scenario above) or that it could produce democracy estimates in real time *but* that the model constantly needed new human coded data to make sure it was still functioning as intended, how valuable would these new contributions turn out to be in practice?

However, even if using social media data to simply produce a "democracy score" proves untenable, there are still other opportunities for using these data to aid in our measurement of democracy more indirectly (or, perhaps more accurately, in terms of producing components of a democracy measure).

Using Social Media to Measure Attitudes Towards Democracy

The first, and perhaps most obvious, opportunity is to use social media data to measure societal support for democracy. While this is not normally considered a measure of democracy in and of itself, this is certainly a component of at least some definitions of democracy.

While minimizing the demands we are placing on social media data as compared to the taller goal considered in the previous section, using social media data to measure public opinion is no small task either. This is a topic one of us has taken up in much greater detail in a chapter in the forthcoming *Oxford Handbook of Public Opinion*[15]; in the remainder of this section we briefly summarize a few of the relevant points made there.

The challenges of using social media data to measure public opinion are by no means trivial. In our chapter, we highlight three in particular. First, there is the difficulty of measuring "opinion" without the structured survey questions to which we have all grown so accustomed. Thus if, for example, we wanted to measure "support for democracy", this would have to be a two-step process: find relevant discussions of democracy; and then run some sort of sentiment analysis that classifiers whether the discussion is positive or negative. Both tasks are challenging, especially if we are trying to disentangle "support for democracy generally" from "support for the performance of the current government" in a democratic regime.[16]

Second, even if we are able to solve the initial challenges of extracting positive or negative attitudes towards democracy from the text of social media posts, we next have to wrestle with the fact that social media users are *not* a representative sample of any population.[17] Thus any attempt to project from "attitudes on social media" to attitudes in the population at large will need to perform some sort of adjustment for non-representativeness. Finally, there other platform

[15] Klašnja et al., forthcoming.
[16] See Pop-Eleches and Tucker (2017), p.49 for more on this distinction.
[17] Occasionally, though, they do represent an entire population of interest, such as in studies of members of the US Congress, where today all members have a social media presence.

specific issues related to aggregating from individual to collective measures of public opinion, including what is made public and what is kept private, as well as limits on data collection.[18]

At the same time, the promise of social media as a means of measuring public opinion are similar to those for measuring democracy: real time data available from around the world and which can be disaggregated to just about any time period. And given the costs of conducting representative surveys cross-nationally, the cost-savings here may be even more significant than those discussed in the previous section.

Using Social Media to Measure Indicators Associated with Democracy

Using social media to measure public opinion is challenging. However, it may be possible to use social media data to measure other indicators that are generally associated with democratic regimes, but which don't have the demands as trying to measure opinions.

One promising candidate could be something like the ratio of tolerant to intolerant posts in a random subsample of social media posts. Such a measure would not face the challenge of trying to identify "opinions towards democracy", but could perhaps serve as an important indicator on its own for the strength of democratic values in a country. There is one such paper in Tucker's lab (the NYU Social Media and Political Participation (SMaPP) lab) that utilizes exactly such a measure, albeit in a single country.[19]

On a grander scale, one could imagine companies such as Facebook publishing indices based on its internal data that might not be available to scholars at the individual level for privacy reasons, but which could be released for public policy/academic outreach efforts as aggregate level indicators. Such measures could play a role in developing next generation democracy indicators.

Social Media Usage Patterns Associated with Democracy

The final "indirect" way we might want to think about using social media to measure democracy would be to eschew analysis of social media data altogether and instead focus on usage patterns that can be measured using traditional surveys. While I have yet to see any attempts to describe particular usage patterns as being more or less conducive to democratic governance, this is not to say we could not come up with such measures in the future.

One way might be to think about the relative proportion of people using social media platforms that are either directly controlled by the state (or else under the thumb of the state) as compared to people using private platforms. In the modern world, we might consider having access to social media platforms that are not subject to state censorship to be an important component of democratic accountability,[20] and this would seem to be something that should be fairly easy to measure using survey data.[21]

---

[18] See Klašnja et al. (forthcoming), p.10-11 for more detail.
[19] See Siegel et al. (2017), which examines the effect of being in more diverse social networks online in Egypt on the propensity to post tolerant – as opposed to intolerant – tweets.
[20] Justifying such a claim is beyond the scope of this essay, but would likely involve similar arguments involving the value of a free press.
[21] See for example the many excellent Pew Social Media Usage surveys in the United States, e.g.

A different type of approach might involve focusing on the segmentation of social media users into more or fewer different social media platforms. For example, if all conservatives use one social media network and all liberals another, this might be seen as worse (or better) for democratic survival than alternative arrangements.[22] A similar argument could be made regarding different ethnic groups in a country.

In both cases, the idea would be to use social media usage patterns to contribute to democracy scores. Validating such claims would of course involve a great deal of *a priori* work, but could provide another way in which social media could inform democracy measures.

Social Media and the United States

We want to close with one final note: the preceding discussion has largely been predicated on the idea of developing *cross-national* measures of democracy. What is probably worth some serious consideration, however, is whether social media might be better suited to exploring within country cross-regional variation in the strength of democracy. This might be in an area in which we have fewer extant measures – thus offering more of a bang for the social media buck – but also might alleviate some of the concerns (e.g., working across different languages, massive data collection and retrieval efforts) in the cross-national context.

C. Outreach I: Publicizing New Tools, Data:

Crucial to either providing new tools or new data is making sure that others – and in particular academics, foundation program officers, and journalists – are aware of their ability. Thus, we propose that any new efforts financed by the Knight Foundation in either of these regards also involve some combination of the following options for outreach:

- Convening conferences of experts who can be introduced to the new tools and data
- Travel to conferences and academic meetings by the principal investigators to present the new tools and data
- The development of a short course to be offered at the *American Political Science Association* (and a commensurate philanthropic) annual meeting to familiarize potential users with the newly available tools and data, including tutorials for reducing barriers to entry to interacting with the tools and data
- A series of short videos related to using tools and data that can be embedded on the website of the Knight Foundation, funded investigators, and interested partner organizations.

D. Outreach II: Utilizing New Tools, Data for Fostering Discussion of Democratic Trends:

After the initial outreach to publicize new tools and data, we recommend that a second round of outreach will be held to encourage scholars to continually scrutinize the questions of measuring and assessing democratic health. This will involve -- but not be limited to -- studies that (a) utilize the new tools produced (Part A) and new data collected (Part B) by the Knight Foundation and will include, among other aims, the goal of developing more overarching sets of "risk

http://www.pewinternet.org/2016/11/11/social-media-update-2016/.

[22] Consider, for example, Gab, a social media platform which *Wired* has called "the Alt-Right's Own Twitter" and "the ultimate filter bubble" (https://www.wired.com/2016/09/gab-alt-rights-twitter-ultimate-filter-bubble/)

factors" for monitoring the health of democracy in both the United States and abroad. This outreach will take two forms: (1) a series of conferences (some invited, some by application) featuring original and commissioned research papers, which could become an annual event hosted by the Knight Foundation; and (2) more fluid, ongoing, online forms of collaboration and communication for foundations and scholars to share research ideas, designs, goals, and results, including, for example, Facebook groups, email list-servers, Wiki-like collaboration platforms, and other related tools.

Works Cited

Acemoglu, Daron, Simon Johnson, J a Robinson, and Pierre Yared. 2005. “Income and Democracy.” *Discussion Paper Series- Centre for Economic Policy Research London* (1959): ALL. http://www.nber.org/papers/w11205.

Annenberg Public Policy Center of the University of Pennsylvania. 2017. “Annenberg Public Policy Center of the University of Pennsylvania.” *National Annenberg Election Survey*. Accessed April 12. http://www.annenbergpublicpolicycenter.org/political-communication/naes/.

Baldini, Gianfranco. 2015. “Is Britain Facing a Crisis of Democracy?” *Political Quarterly* 86(4): 540–49.

Boddam-Whetham, L, X Gul, and E Al-Kobati. 2016. “Vouchers in Fragile States: Reducing Barriers to Long-Acting Reversible Contraception in Yemen and Pakistan.” *Global*. https://pdfs.semanticscholar.org/c78a/531923664175328a4d079af5b8444f2cb9be.pdf (February 12, 2017).

Brinks, Daniel, and Michael Coppedge. 2006. “Diffusion Is No Illusion.” *Comparative Political Studies* 39(4): 463–89.

Broadman, Harry G, and Francesca Recanatini. 2002. “Corruption and Policy: Back to the Roots.” *Journal of Policy Reform* 5(1): 37. https://login.e.bibl.liu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=7273442&site=eds-live&scope=site.

Bühlmann, Marc, Wolfgang Merkel, Lisa Müller, and Bernhard Weßels. 2012. “The Democracy Barometer: A New Instrument To Measure the Quality of Democracy and Its Potential for Comparative Research.” *European Consortium for Political Research* 11(4): 519–36. http://link.springer.com/article/10.1057/eps.2011.46 (February 12, 2017).

Carbone, Giovanni, and Vincenzo Memoli. 2015. “Does Democratization Foster State Consolidation? Democratic Rule, Political Order, and Administrative Capacity.” *Governance* 28(1): 5–24. http://onlinelibrary.wiley.com/doi/10.1111/gove.12056/full (February 5, 2017).

Chanley, Virginia A., Thomas J. Rudolph, and Wendy M. Rahn. 2000. “The Origins and Consequences of Public Trust in Government: A Time Series Analysis.” *Public Opinion Quarterly* 64 (3): 239–56.

Citrin, Jack. 1974. “Comment: The Political Relevance of Trust in Government.” *American Political Science Review* 68 (3): 973–88.

Cole, Richard L., and John Kincaid. 2000. “Public Opinion and American Federalism: Perspectives on Taxes, Spending, and Trust–An ACIR Update.” *Publius: The Journal of Federalism* 30 (1): 189-201.

Cole, Richard L., and John Kincaid. 2006. “Public Opinion on U.S. Federal and Intergovernmental Issues in 2006: Continuity and Change.” *Publius: The Journal of Federalism* 36 (3): 443-459 with Richard L. Cole.

Cole, Richard L., John Kincaid, and Andrew Parkin. 2002. “Public Opinion on Federalism in the United States and Canada: The Aftermath of Terrorism.” *Publius: The Journal of Federalism* 32 (4): 123-147.

Coppedge, M et al. 2011. “Conceptualizing and Measuring Democracy: A New Approach.” *Perspectives on Politics* 9(2): 247–67. http://journals.cambridge.org/article_S1537592711000880 (February 12, 2017).

Coppedge, Michael, Staffan Lindberg, Svend-Erik Skaaning, and Jan Teorell. 2016. “Measuring High Level Democratic Principles Using the V-Dem Data.” *International Political Science Review* 37(5): 580–93. http://ips.sagepub.com/content/37/5/580.short (February 12, 2017).

Daxecker, U. E. 2012. “The Cost of Exposing Cheating: International Election Monitoring, Fraud, and Post-Election Violence in Africa.” *Journal of Peace Research* 49(4): 503–16. http://jpr.sagepub.com/content/49/4/503.short (February 12, 2017).

Donno, Daniela. 2013. “Elections and Democratization in Authoritarian Regimes.” *American Journal of Political Science* 57(3): 703–16. http://onlinelibrary.wiley.com/doi/10.1111/ajps.12013/full (February 12, 2017).

Dubrow, Joshua Kjerulf, Kazimierz M. Slomczynski, and Irina Tomescu-Dubrow. 2008. “Effects of Democracy and Inequality on Soft Political Protest in Europe: Exploring the European Social Survey Data.” *International Journal of Sociology* 38(3): 36–51. http://mesharpe.metapress.com/openurl.asp?genre=article&id=doi:10.2753/IJS0020-7659380302.

Eagly, Alice H., and Shelley Chaiken. 1993. *The Psychology of Attitudes*. New York, NY: Harcourt College Publishers.

Edgell, Amanda B et al. 2015. *When and Where Do Elections Matter? A Global Test of the Democratization by Elections Hypothesis, 1900-2012*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2638285 (February 14, 2017).

Eisenstadt, Todd A., A. Carl LeVan, and Tofigh Maboudi. 2015. “When Talk Trumps Text: The Democratizing Effects of Deliberation during Constitution-Making, 1974–2011.” *American Political Science Review* 109(3): 592–612. http://www.journals.cambridge.org/abstract_S0003055415000222.

Eschenbach, Felix, and Bernard Hoekman. 2006. “Services Policy Reform and Economic Growth in Transition Economies.” *Review of World Economics* 142(4): 746–64.

Finkel, Steven E. 1985. “Reciprocal Effects of Participation and Political Efficacy: A Panel Analysis.” *American Journal of Political Science* 29 (4): 891–913. doi:10.2307/2111186.

Gehlbach, Scott, and Edmund J. Malesky. 2010. “The Contribution of Veto Players to Economic

Reform." *The Journal of Politics* 72(4): 957–75.

Gerring, John, Maxwell Palmer, Jan Teorell, and Dominic Zarecki. 2015. "Demography and Democracy: A Global, District-Level Analysis of Electoral Contestation." *American Political Science Review* 109(3): 574–91. http://www.journals.cambridge.org/abstract_S0003055415000234.

Goldstone, Jack A et al. 2010. "A Global Model for Forecasting Political Instability." *American Journal of Political Science* 54(1): 190–208.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–97. doi:10.1093/pan/mps028.

Guardo, MC Di, E Marrocu, and R Paci. 2015. "The Concurrent Impact of Cultural, Political, and Spatial Distances on International Mergers and Acquisitions." *The World Economy*. http://onlinelibrary.wiley.com/doi/10.1111/twec.12275/pdf (February 12, 2017).

Gugiu, Mihaiela Ristei, and Miguel Centellas. 2013. "The Democracy Cluster Classification Index." *Political Analysis* 21(3): 334–49. http://pan.oxfordjournals.org/content/21/3/334.full (February 12, 2017).

Hadenius, Axel, and Jan. Teorell. 2007. "Pathways from Authoritarianism." *Journal of Democracy* 18(1): 143–57.

Hafner-Burton, Emilie M, Susan D Hyde, and Ryan S Jablonski. 2013. "When Do Governments Resort to Election Violence?" *British Journal of Political Science* 2145: 1–31. http://journals.cambridge.org/abstract_S0007123412000671 (February 12, 2017).

Hauser, Oliver P., David G. Rand, Alexander Peysakhovich, and Martin a. Nowak. 2014. "Cooperating with the Future." *Nature* 511(7508): 220–23. http://www.nature.com/doifinder/10.1038/nature13530%5Cnhttp://dx.doi.org/10.1038/nature13530%5Cnpapers2://publication/doi/10.1038/nature13530.

Hetherington, Marc J. 1998. "The Political Relevance of Political Trust." *American Political Science Review* 92 (4): 791–808.

———. 2004. *Why Trust Matters: Declining Political Trust and the Demise of American Liberalism*. Princeton: Princeton University Press.

Hetherington, Marc J., and Suzanne Globetti. 2002. "Political Trust and Racial Policy Preferences." *American Journal of Political Science* 46 (2): 253–75.

Hetherington, Marc J., and Thomas J. Rudolph. 2008. "Priming, Performance, and the Dynamics of Political Trust." *The Journal of Politics* 70 (2): 498–512.

———. 2015. *Why Washington Won't Work: Polarization, Political Trust, and the Governing Crisis*. Chicago: The University of Chicago Press.

Hibbing, John R., and Elizabeth Theiss-Morse. 1995. *Congress as Public Enemy: Public Attitudes toward American Political Institutions*. New York: Cambridge University Press.

———. 2001. *What Is It about Government That Americans Dislike?* New York: Cambridge University Press, NY.

Hyde, Susan D., and Nikolay Marinov. 2012. "Which Elections Can Be Lost?" *Political Analysis* 20(2): 191–210. http://www.jstor.org/stable/23260172 (February 12, 2017).

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. "Affect, Not Ideology A Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76 (3): 405–31. doi:10.1093/poq/nfs038.

Johnston, Richard, Andre Blais, Henry Brady, and Jean Crete. 1992. *Letting the People Decide: Dynamics of a Canadian Election*. Stanford, CA: Stanford University Press.

Johnston, Richard, Michael G. Hagen, and Kathleen Hall Jamieson. 2004. *The 2000 Presidential Election and the Foundations of Party Politics*. New York: Cambridge University Press.

Keele, Luke. 2005. "The Authorities Really Do Matter: Party Control and Trust in Government." *Journal of Politics* 67 (3): 873–86.

Kincaid, Richard L., and John Cole. 2005. "Public Opinion on Issues of U.S. Federalism in 2005: End of the Post-2001 Pro-Federal Surge?" *Publius: The Journal of Federalism* 35 (1): 169-185.

Kincaid, Richard L., and John Cole. 2008. "Public Opinion on Issues of Federalism in 2007: A Bush Plus?" *Publius: The Journal of Federalism* 38 (3): 469-487.

Kolk, A, and F Lenfant. 2015. "Partnerships for Peace and Development in Fragile States: Identifying Missing Links." *The Academy of Management Perspectives*. http://amp.aom.org/content/29/4/422.short (February 12, 2017).

Ladd, Jonathan. 2004. "Attitudes toward the News Media and the Acquisition of Political Information."

Ladd, Jonathan M. 2012. *Why American Hate the Media and How It Matters*. Princeton: Princeton University Press.

Lauth, HJ. 2013. *Demokratie Und Demokratiemessung: Eine Konzeptionelle Grundlegung Für Den Interkulturellen Vergleich*. https://books.google.com/books?hl=en&lr=&id=u0YhBgAAQBAJ&oi=fnd&pg=PA9&dq=Demokratie+und+Demokratiemessung:+eine+konzeptionelle+Grundlegung+für+den+interkulturellen+Vergleich&ots=agXHe29b9l&sig=QYI_sHCdmIq9msPz0xxNC7LwgKY (March 25, 2017).

Leeson, Peter. 2008. "Media Freedom, Political Knowledge, and Participation." *Journal of*

Economic Perspectives 22(2): 155–169.

Lipset, Seymour Martin, and William Schneider. 1987. *The Confidence Gap*. Rev. Baltimore: Johns Hopkins University Press.

Madsen, Douglas. 1987. "Political Self-Efficacy Tested." *American Political Science Review* 81 (2): 571–81. doi:10.2307/1961970.

Marinov, N, and H Goemans. 2014. "Coups and Democracy." *British Journal of Political Science*. http://journals.cambridge.org/article_S0007123413000264 (February 12, 2017).

Marzagao, Thiago Veiga. 2014. "Measuring Democracy : From Texts to Data." https://etd.ohiolink.edu/!etd.send_file?accession=osu1405085531&disposition=attachment (February 12, 2017).

Merkel, Wolfgang. 2014. "Is There a Crisis of Democracy?" *Democratic Theory* 1(2): 11–25. http://www.berghahnjournals.com/abstract/journals/democratic-theory/1/2/dt010202.xml (February 12, 2017).

Merkel, Wolfgang, and Aurel Croissant. 2004. "Conclusion: Good and Defective Democracies." *Democratization* 11(5): 199–213. http://www.tandfonline.com/doi/full/10.1080/13510340412331304651 (February 12, 2017).

Møller, Jørgen, and Svend-Erik Skaaning. 2011. "On the Limited Interchangeability of Rule of Law Measures." *European Political Science Review* 3(3): 371–94. http://journals.cambridge.org/article_S1755773910000421 (February 5, 2017).

Munck, Gerardo L, and Jay Verkuilen. 2002. 35 Comparative Political Studies *Conceptualizing and Measuring Democracy: Evaluating Alternative Indices*. http://cps.sagepub.com.ezproxy.library.wisc.edu/cgi/content/abstract/35/1/5.

Neumayer, Eric. 2003. "Do Human Rights Matter in Bilateral Aid Allocation? A Quantitative Analysis of 21 Donor Countries." *Social Science Quarterly* 84(3): 650–66. http://onlinelibrary.wiley.com/doi/10.1111/1540-6237.8403010/full (February 12, 2017).

Norris, Pippa. 2008. *Driving Democracy: Do Power-Sharing Institutions Work ?* Cambridge University Press.

Nye, Joseph S., Philip Zelikow, and David C. King, eds. 1997. *Why People Don't Trust Government*. Cambridge, Mass: Harvard University Press.

Pemstein, D, SA Meserve, and J Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis*. http://pan.oxfordjournals.org/content/18/4/426.full (February 12, 2017).

Poe, SC, and CN Tate. 1994. "Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis." *American Political Science Review* 88(4): 853–72. http://journals.cambridge.org/article_S0003055400094521 (February 12, 2017).

Pollock, Philip H. 1983. “The Participatory Consequences of Internal and External Political Efficacy: A Research Note.” *Western Political Quarterly* 36 (3): 400–409. doi:10.1177/106591298303600306

Pop-Eleches, Grigore, and Tucker, Joshua A. 2017. *Communism’s Shadow: Historical Legacies and Political Attitudes*. Princeton, NJ: Princeton University Press.

Quan Li. 2005. “Does Democracy Promote or Reduce Transnational Terrorist Incidents?” *Journal of Conflict Resolution* 49(2): 278–97. http://proxygw.wrlc.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=16618861&site=ehost-live.

Romer, Daniel, Kathleen Hall Jamieson, Christopher Adasiewicz, Kate Kenski, and Paul Waldman. 2003. *Capturing Campaign Dynamics: The National Annenberg Election Survey: Design, Method and Data Includes CD-ROM*. Pap/Cdr edition. New York: Oxford University Press.

Romer, Daniel, Kate Kenski, Kenneth Winneg, Christopher Adasiewicz, and Kathleen Hall Jamieson. 2006. *Capturing Campaign Dynamics, 2000 and 2004: The National Annenberg Election Survey*. Philadelphia: University of Pennsylvania Press.

Safaei, Jalil. 2012. “Post-Communist Health Transitions in Central and Eastern Europe.” *Economics Research International* 2012: 1–10. http://downloads.hindawi.com/journals/econ/2012/137412.pdf (February 12, 2017).

Sanín, Francisco Gutiérrez FG. 2011. “Evaluating State Performance: A Critical View of State Failure and Fragility Indexes.” *European Journal of Development Research* 23(1): 20–42. http://www.palgrave-journals.com/doifinder/10.1057/ejdr.2010.53 (February 5, 2017).

Sikkink, K, and CB Walling. 2007. “The Impact of Human Rights Trials in Latin America.” *Journal of peace research*. http://journals.sagepub.com/doi/abs/10.1177/0022343307078953 (February 12, 2017).

Skaaning, Svend-Erik, John Gerring, and Henrikas Bartusevičius. 2015. “A Lexical Index of Electoral Democracy.” *Comparative Political Studies* 48(12): 1491–1525. http://cps.sagepub.com/content/early/2015/04/23/0010414015581050.abstract (February 12, 2017).

Skrede Gleditsch, K., and A. Ruggeri. 2010. “Political Opportunity Structures, Democracy, and Civil War.” *Journal of Peace Research* 47(3): 299–310.

Tavares, José, and Romain Wacziarg. 2001. “How Democracy Affects Growth.” *European Economic Review* 45(8): 1341–78.

Teorell, J, M Coppedge, SE Skaaning, and SI Lindberg. 2016. “Measuring Electoral Democracy with V-Dem Data: Introducing a New Polyarchy Index.” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2740935 (February 12, 2017).

Vreeland, James Raymond. 2008. “The Effect of Political Regime on Civil War.” *Journal of*

Conflict Resolution 52(3): 401–25. http://journals.sagepub.com/doi/abs/10.1177/0022002708315594 (February 15, 2017).

Werrell, CE, F Femia, and T Sternberg. 2015. “Did We See It Coming?: State Fragility, Climate Vulnerability, and the Uprisings in Syria and Egypt.” *SAIS review of international affairs*. https://muse.jhu.edu/article/582525/summary (February 12, 2017).

Wu, Chin-en. 2012. “When Is Democracy Better for Economic Performance and When Is It Not: The Interaction between Polity and Structural Factors.” *Studies in Comparative International Development* 47(4): 365–388. http://link.springer.com/article/10.1007/s12116-012-9118-3.

Yared, Pierre, Daron Acemoglu, Simon Johnson, and James a Robinson. 2005. “From Education to Democracy?” *American Economic Review* 95(2): 44–49.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. New York, NY: Cambridge University Press.

Zaller, John R., and Stanley Feldman. 1992. “A Simple Theory of Survey Response: Answering Questions versus Revealing Preferences.” *American Journal of Political Science* 36 (3): 579–616.