An unexpected consensus among diverse ways to measure democracy

Samuel Baltz[1], Fabricio Vasselai[2], and Allen Hicken[3]

Abstract

We show that behind decades of disagreements over how to classify regime types, there is an unexpectedly strong consensus about which countries have been democratic in which years. We match many-valued classifications of democracies and autocracies to binary classifications, by finding the cutpoint that dichotomizes each many-valued measure so that it matches each binary measure as closely as possible. Despite the sharp differences between these datasets, we find identical or nearly identical optimal cutpoints for almost every major classification. This is evidence of a strong underlying agreement between the classifications. We show that the many-valued measures, when dichotomized in the optimal way, can replicate major results in the study of democracy. We also examine which sorts of countries are counted as democracies in this consensus, and we show that countries where national elections have never produced a turnover between parties are a particular focus of disagreement.

[1] PhD candidate, University of Michigan Department of Political Science, sbaltz@umich.edu
[2] PhD candidate, University of Michigan Department of Political Science, vasselai@umich.edu
[3] Professor, University of Michigan Department of Political Science, ahicken@umich.edu

1. Introduction

Surveying the literature on democracy in 1942, Schumpeter lamented that the classical definition of democracy could describe "an infinite wealth of possible forms", evading consensus even after "the legal mind ransacked the lumber room of its constructs in search for tools by which to reconcile" theoretical definitions of democracy with the observed reality of politics.[1] In the nearly eighty years since then, political scientists have largely settled on a theory of democracy that requires the presence of certain observable features. But when it comes to deciding exactly which features are necessary parts of democracy, our lumber rooms have produced such an unruly range of constructs that surveys of the democratization literature report broad disagreement about how to classify a particular country as a democracy or an autocracy in a given year. How different are the measurements that political scientists have produced in our decades of researching exactly how to measure democracy?

In nearly a century of classifying political regimes,[2] no question has been contested more sharply than whether democracy is a matter of type or of degree.[3] Because of this longstanding division, there is no single consensus in classifying democracies and autocracies. Dichotomous classifications of democracy are highly correlated,[4] but it is harder to compare many-valued measures of democracy to each other, and only a few attempts have been made to thoroughly compare many-valued measures of democracy to dichotomous measures.[5] Recently, a technique has arisen for translating many-valued democracy measures into dichotomous ones, mostly aimed at generating classifications for use in empirical analyses.[6] Building on this technique to gauge the level of agreement across datasets, we find deep agreement between dichotomous and many-valued measures of democracy. This agreement holds across very different measures of democracy, and using this consensus classification we can replicate central substantive results from a series of previous analyses.

The traditional view of regime classification was as a largely all-or-nothing proposition, in which a regime can simply be either a democracy or a non-democracy in a given year.[7] But as efforts to chart countries' experiences with democracy over time became more common, researchers began to introduce numerical proxies for democratization, like the percentage of the population voting in national elections.[8] Cutright was among the first to propose a many-valued scale of democracy, which was calculated by combining subjective scores of a country's institutions.[9] In 1980, Bollen proposed a score based on characteristics that are more exogenous to a country's institutional idiosyncrasies, like the presence of free press and free opposition, the fairness of elections, and whether the legislative and executive branches (when existent) were elected.[10] This early work on many-valued measures of democracy was partly a reaction to the perceived limits of dichotomous classification schemes. In later work with Jackman, Bollen summarized the common criticism of dichotomous measures: "Dichotomizing democracy lumps together countries with very different degrees of democracy and blurs distinctions between borderline cases".[11]

And yet, dichotomous classifications of democracy have re-emerged in force. Alvarez et al., offering their own dichotomous classification scheme, responded to Bollen and Jackman that "it is one thing to argue that some democracies are more democratic than others and another to argue that democracy is a continuous feature over all regimes".[12] Even if there are different degrees of democracy and autocracy, these authors argue that there is still a real discontinuity between these types, and that a country in a given year cannot be "half-democratic: there is a natural zero point". Alvarez et al. drew from Schumpeter and Dahl to define democracies as political regimes where governing offices are filled through competitive elections.[13] Relying only on observable historical information about countries, they were able to classify many more countries and years than any previous study, covering a total of 141 countries through 41 years. Their classification became one of the most widely used, has been updated several times,[14] and was a starting point for several subsequent classifications.[15]

As the number of dichotomous classifications has mushroomed, researchers have observed that the agreement between these measures is descriptively quite high. Figure 1 shows the level of agreement among the major, reasonably independent, dichotomous (or easily dichotomized) measures of democracy. The figure shows how often each pair of datasets agrees about whether or not a given country-year is a democracy, among those country-years that appear in both datasets.

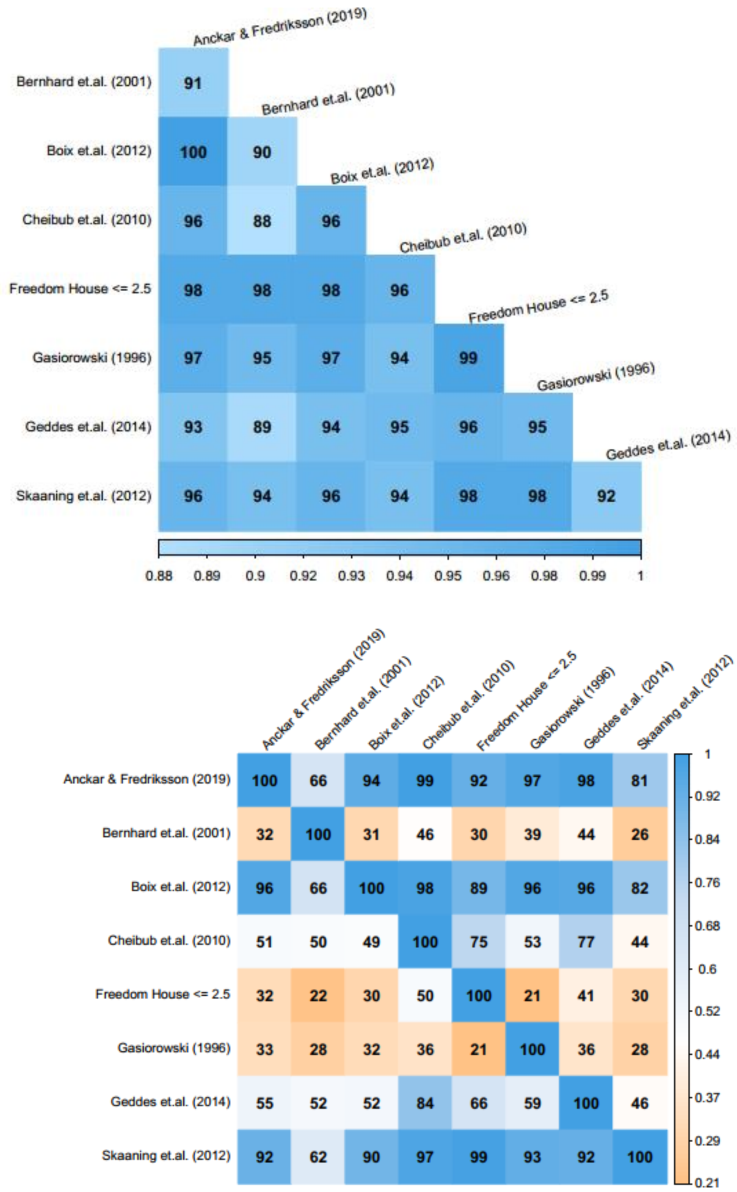

Figure 1: Top: Each cell shows the pairwise agreement between the datasets in that row and column. Bottom: Each cell shows the percentage of the dataset in that column which also appears in the dataset in that row. For example, Geddes et al. has 94% agreement with Boix et al., 96% of the cases in Geddes et al. are in Boix et al., and 52% of the cases that appear in Boix et al. also appear in Geddes et al.[16]

The top of Figure 1 shows the pairwise agreement between each measure of democracy, while the bottom of Figure 1 shows how many country-years are included in both datasets. A typical level of agreement between these datasets is upwards of 90%.

But this high correlation has inspired a subtle conversation about how well different dichotomous datasets agree. Casper and Tufis observed that researchers tended to swap out one measure of democracy for another in their analyses, treating the various dichotomous measures of democracy “as virtually interchangeable”.[17] Because the existing classifications of democracies and autocracies were highly correlated, scholars were lulled into believing that they could freely use whichever variable was more empirically convenient. But of course, pairwise comparisons are only a small part of the story: the bottom part of Figure 1 shows that in some cases fewer than a quarter of the cases in one dataset are included in the other, so pairwise agreement captures only a slice of the cases in each classification. And indeed, Casper and Tufis showed that the same analysis can produce opposite results when conducted on different classifications of democracy.

This problem has prompted deep discussion among authors of dichotomous datasets, who resoundingly agree that many measures are needed to capture a phenomenon like democratization, which “varies enormously from case to case and region to region”,[18] but those differences across classification schemes also present challenges for researchers. Cheibub et al. argue that “differences across regime measures must be taken seriously”.[19] Boix et al. reinforce that differences between the “dozens of distinct measures of democracy” are “not merely academic”, since “empirical results can depend on the specific measure of democracy used”.[20] And Skaaning et al., after surveying a “proliferation of binary indices that identify different defining conditions of democracy” and measurement decisions that are both arbitrary and informative “in strikingly different ways”, note that such endemic disagreements might actually be desirable, since “the discipline is well served by a variety of measures”.[21]

But if the similarity of dichotomous measures has sparked detailed discussions, there has been no corresponding conversation regarding many-valued measures of democracy. Though the creators of these many-valued datasets do analyze their similarity to existing offerings in various ways, there is no one accepted method for checking how similar two many-valued measures of democracy are to one another. And what about the similarity of many-valued measures with dichotomous ones? As these separate but closely related traditions have developed over the decades, with markedly different approaches to the classification of democracies, how do they relate to each other? Are the differences between these two camps merely numeric – say, the biggest difference is that one type is appropriate for logistic and the other for Ordinary Least Squares regression – or are the differences deeper? Could there be country-years that dichotomous coders consider to be democracies but the many-valued classifications suggest are not really democratic, and the same for autocracies?

In recent years work has begun on a method to connect these two approaches, and this connection suggests a way to judge just how much agreement there really is. As many-valued measures of democracy — like the Varieties of Democracy (V-Dem) and Polity measures — have grown in prominence, with very large spatial and temporal coverage, researchers have begun to convert these many-valued variables into dichotomous ones. The main technique is to pick some “cutpoint” along the many-valued measure’s scale, and declare that every value above that cutpoint is a democracy while every value below it is an autocracy.[22] Some researchers, recognizing that features of a many-valued dataset like its coverage or methodology might appeal to authors who wish to use it in analyses that demand a variable with fewer values, have used cutpoints to boil those datasets down in an appropriate way.[23]

An open question in this area, however, is which value should be chosen as the cutpoint. This is not just a matter of taste: in a finding reminiscent of the debates over how to choose dichotomous measures, Bogaards shows that altering the choice of cutpoint for Freedom House and Polity can change the empirical results of published models.[24] A further complication is that every value of the many-valued measures could result from many different combinations of their underlying components,[25] and Boix et al. lament that none of the authors who use cutpoints "offer a concrete reason for the thresholds other than claiming they are intuitive or citing another study that uses the same threshold."[26]

In the following analysis we show that there is good reason to prefer one small range of cutpoints above all others. We build on a simple idea for selecting a cutpoint: we could use whichever cutpoint produces a binary variable that agrees with previous binary measures as much as possible.[27] That is, an optimal cutpoint answers the question: "how can I split a many-valued dataset into two values, such that it matches a dichotomous dataset as closely as possible?" The current thinking is that this requires an unfortunate compromise, since for each dataset to which a researcher wishes to match the many-valued classification, there will be a new and potentially different optimal cutpoint.

We demonstrate that no such problem exists, by finding the best cutpoint for matching the major many-valued measures of democracy to eight dichotomous classifications. For many-valued measures that have as many as 1001 possible cutpoints, we find that *multiple different binary classifications of democracy share exactly the same best cutpoint*, and all the cutpoints are close to each other. Even datasets with extremely different classification rules have extremely similar optimal cutpoints. *This represents a deep form of agreement between classifications of democracy, which has not been captured in previous substantive discussions or pairwise comparisons*. In the Replications section, we use the modal best cutpoint to produce a consensus binary classification, and we replicate a central substantive result from multiple previous analyses.

This article provides a useful tool for dichotomizing the many-valued datasets on democracy. But the cutpoints exercise offers something much deeper than that: it is an opportunity to measure the consensus about democracy. We argue that, since the main many-valued measures of democracy can be boiled down in just one way to a dataset that closely matches existing dichotomous measures, there is an underlying agreement between many-valued measures and dichotomous measures. We also demonstrate agreement between the main many-valued measures of democracy, as well as a tighter agreement between the dichotomous measures than previous studies have been able to identify through pairwise comparisons. Lurking behind the apparent discord in the study of how to classify democracies, there is a robust consensus regarding which countries have been democracies in which years.

2. Analysis

In this section we find the optimal cutpoint for matching the V-Dem polyarchy variable by Coppedge et al. and the Polity5 variable by Marshall and Gurr onto the classifications by Anckar and Fredriksson, Bernhard et al., Boix et al.,[28] Cheibub et al., Gasiorowski , Geddes et al., Puddington et al., and Skaaning et al.[29] For the polyarchy variable, which ranges from 0 to 1 with three digits of precision, we show that of the 1001 possible optimal cutpoints, multiple different binary classifications of democracy share exactly the same best value, and all of the values are close to each other. We observe a similar pattern for polity, though it only has 21 possible values. Crucially, we argue that this represents a form of agreement between these classifications of democracy, and is not just some artifact of the method. This is a highly unexpected level of agreement between datasets that were produced by several groups of authors over decades, with different research questions in mind, using dramatically different methods, aiming to capture substantively distinct notions of democracy, and applying those notions to different cases.

2.1: Finding optimal cutpoints

In Figure 1 we showed that the main existing dichotomous datasets typically have fairly strong pairwise agreement, as previous authors have found.[30] But there are two crucial limitations to pairwise comparisons. First, the number of cases in the intersection of any two datasets shows that pairwise comparisons can rarely tell the whole story; commonly, for large datasets like that of Boix et al., pairwise comparisons can only be made on the basis of a quarter or a third of the dataset, leaving out thousands of country-years. Country-years in the intersection of any two datasets may also systematically tend to be the country-years that the field is most certain about, since these will be the country-years that are more commonly studied and classified, producing a misleadingly high pairwise agreement. Second, pairwise comparisons tell us nothing about how closely these dichotomous measures agree with the many high-quality many-valued measures of democracy.

This motivates the idea of using a many-valued classification of democracy that is part of a very large dataset, like V-Dem or Polity, and identifying the optimal cutpoint for matching it to a binary dataset. But first, it is necessary to describe precisely what we mean by a “cutpoint”. Figure 2 illustrates the basic problem, using a fictional example of a country that became steadily more democratic during the 20th century.

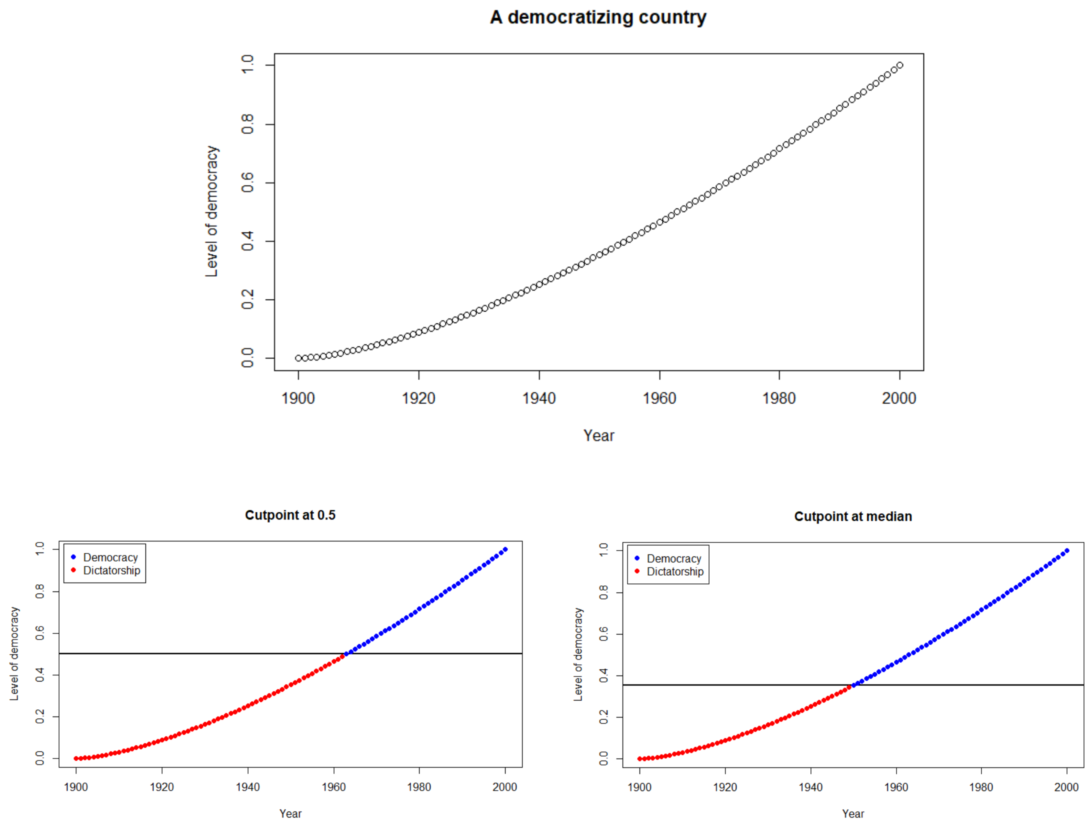

Figure 2: A fictional example to illustrate the idea of cutpoints.

The country in Figure 2 began as a full autocracy in 1900, and then became monotonically more democratic until it attained perfect democracy in the year 2000. The question is: in what year did it become a democracy? Is it when the country became "half democratic", as shown in the bottom-left of Figure 2? Could it be when the country crossed its median democracy value, as in the bottom-right of Figure 2? We could draw the line at — or even in between — any of the 101 points.

Consider a dataset of democracy values **x** that is bounded above some lower bound x_L and below some upper bound x_H. Then a *cutpoint* $\tau \in [x_L; x_H]$ defines a binary classification **b** of democracies ($b = 1$) and autocracies ($b = 0$) as follows: for any $x \in$ **x,** there is a corresponding $b \in$ **b** such that $b = 1$ if $x \geq \tau$ and $b = 0$ if $x < \tau$.[31]

This classification idea makes two claims. First, it claims that if a value $x \in$ **x** is a democracy, then x' is also a democracy $\forall x' > x$, and conversely that if x is an autocracy, then x' is also an autocracy $\forall x' < x$. This is easy to accept whenever x is intended to represent an amount of democracy, so that higher x-values represent a more democratic case. Second, the classification also makes the much stronger assertion that τ is the least democratic democracy. The difficult question in this method is therefore to decide where exactly to set τ.

In a vacuum, picking a cutpoint requires making a contentious (and maybe even incoherent) substantive decision: what proportion of the total quantity of possible democracy is a sufficient amount of democracy to be the least democratic democracy? A common practice is to embrace that the exercise is unavoidably arbitrary, and pick, say, the middle of the range of **x**.[32] But the fact that we have many different measures of democracy also offers a method for picking a cutpoint that is grounded in the literature: as Kasuya and Mori argue, we can pick the cutpoint which best matches a many-valued measurement to a binary measurement as closely as possible. Anybody who has classified countries into either democracies or autocracies has already made a substantive judgment about what is the minimum requirement for a country to be a democracy, so this approach translates the many-valued measure to a binary measure in whichever way matches another researcher's judgments as closely as possible.

Consider the situation in Figure 3. This plot imagines a binary classification for the same burgeoning democracy in Figure 2, produced by a separate author, according to another rule for what constitutes a democracy.

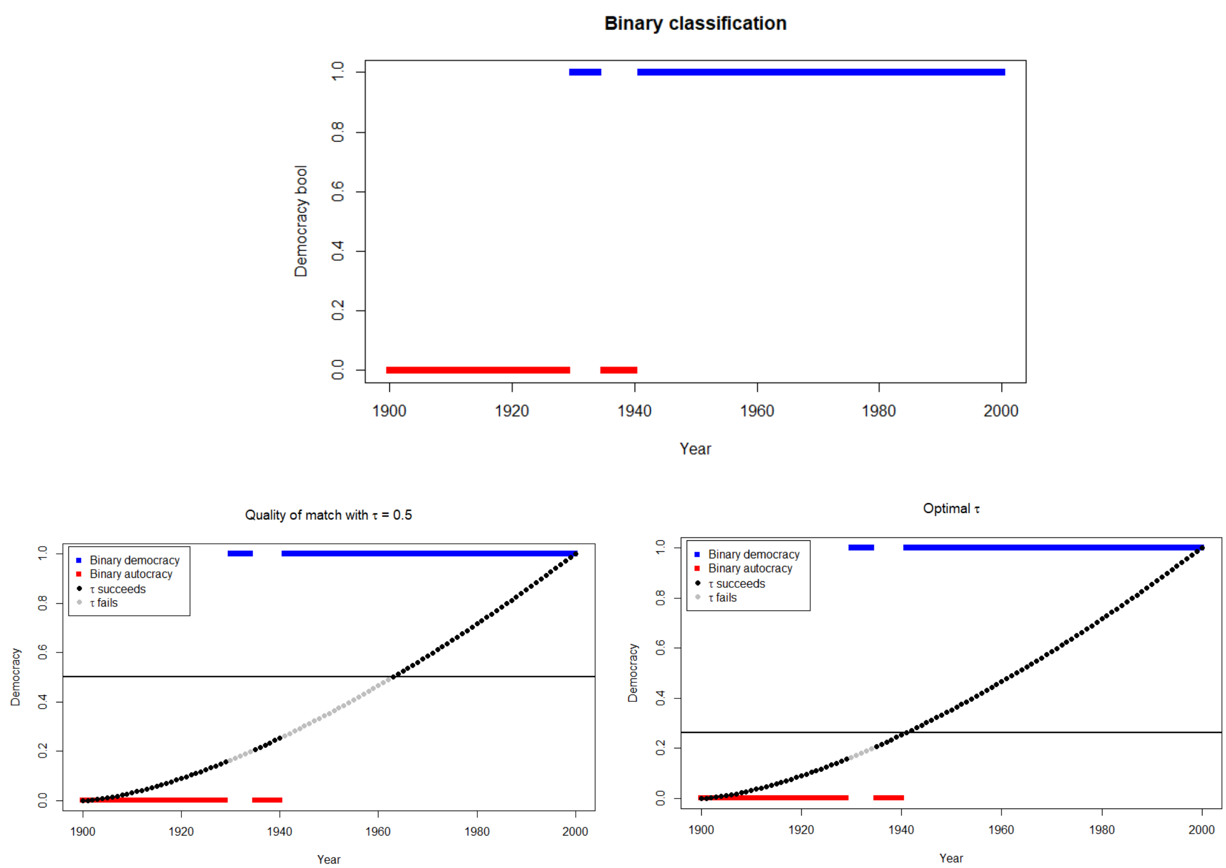

Figure 3: A fictional example of the optimal cutpoint for matching a many-valued classification to a binary classification.

This author has made their own substantive judgments about what exactly is the minimum requirement for a democracy; we can now take advantage of that existing judgment by setting a cutpoint so that our many-valued measure of democracy from Figure 2 matches the binary measure of democracy in Figure 3 as closely as possible. The bottom-left of Figure 3 shows that a naive cutpoint set at $\tau = 0.5$ classifies many cases differently from the binary classification. In the bottom-right panel, we identify the "optimal cutpoint", which we denote τ^*. τ is an optimal cutpoint if changing the value of τ cannot increase the number of cases that match the binary dataset.

An important feature of optimal cutpoints is that they are almost never perfect. The binary classification in Figure 3 illustrates a common situation in which a country becomes democratic, then backslides into autocracy, and then returns to democracy. We can never capture this feature of the binary variable by applying a cutpoint to the many-valued classification in Figure 3, which instead argues that the level of democracy in our fictional country rose every year. So in some situations τ^* might classify every case in the same way as the binary dataset, but in other situations τ^* might only do so for, say, 90% of cases.

But behind this simple picture is a far thornier problem: the extreme diversity in how authors operationalize democracy suggests that if we identify the best cutpoint for matching a many-valued measure of democracy to several different binary measures of democracy, we should expect to find a series of very different optimal cutpoints. The researcher is then faced with a difficult decision about which cutpoint to use for their analysis.[33] But we will now demonstrate a highly convenient fact: the literature on democratic classifications, which includes many different notions of what exactly constitutes a democracy, consistently agrees about the best cutpoint for dichotomizing the main many-valued measures of democracy.

2.2: Different datasets have the same cutpoint

We next find the optimal cutpoint for matching V-Dem's polyarchy variable and the Polity Project's polity variable onto each binary classification of democracy, and we argue that it represents a new form of substantive agreement between the datasets. Figure 4 plots the pairwise agreement between each dichotomous dataset and polyarchy, when polyarchy is dichotomized at every possible cutpoint. Figure 5 shows the same information for polity.

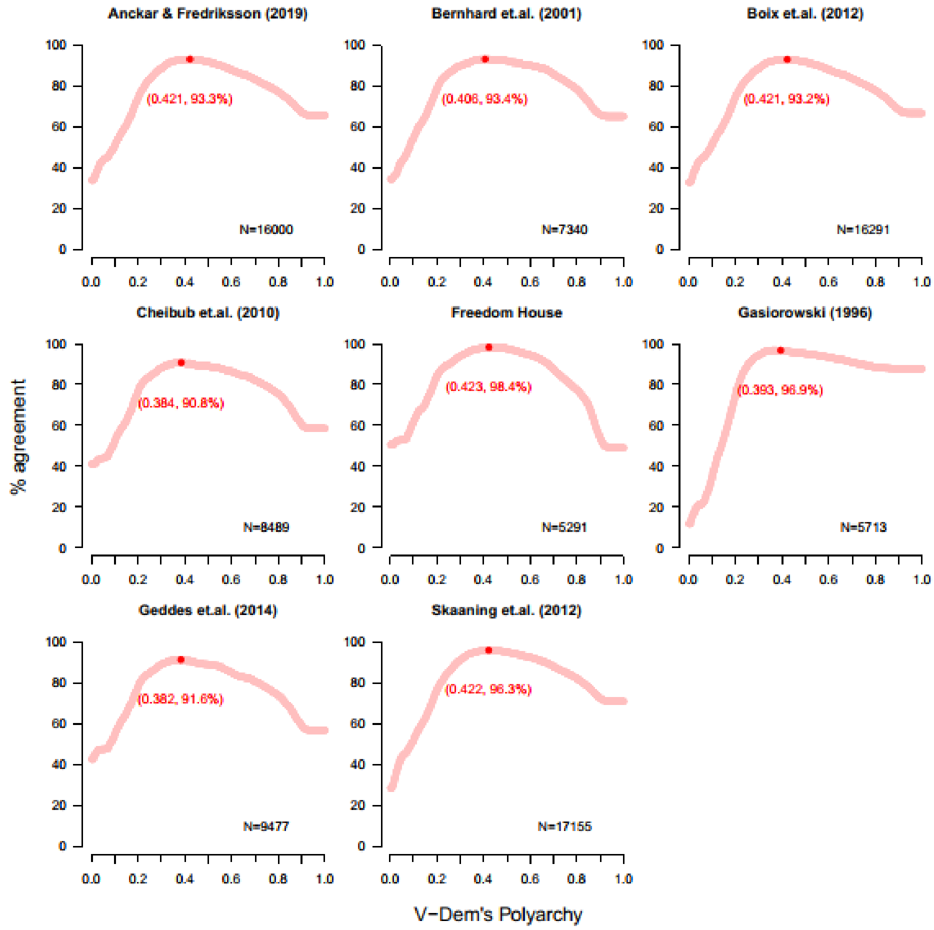

Figure 4: Each panel shows how successful every possible polyarchy cutpoint is at matching a dichotomous dataset. The dot represents the optimal cutpoint and the corresponding agreement level. In the Appendix we vary certain coding decisions.

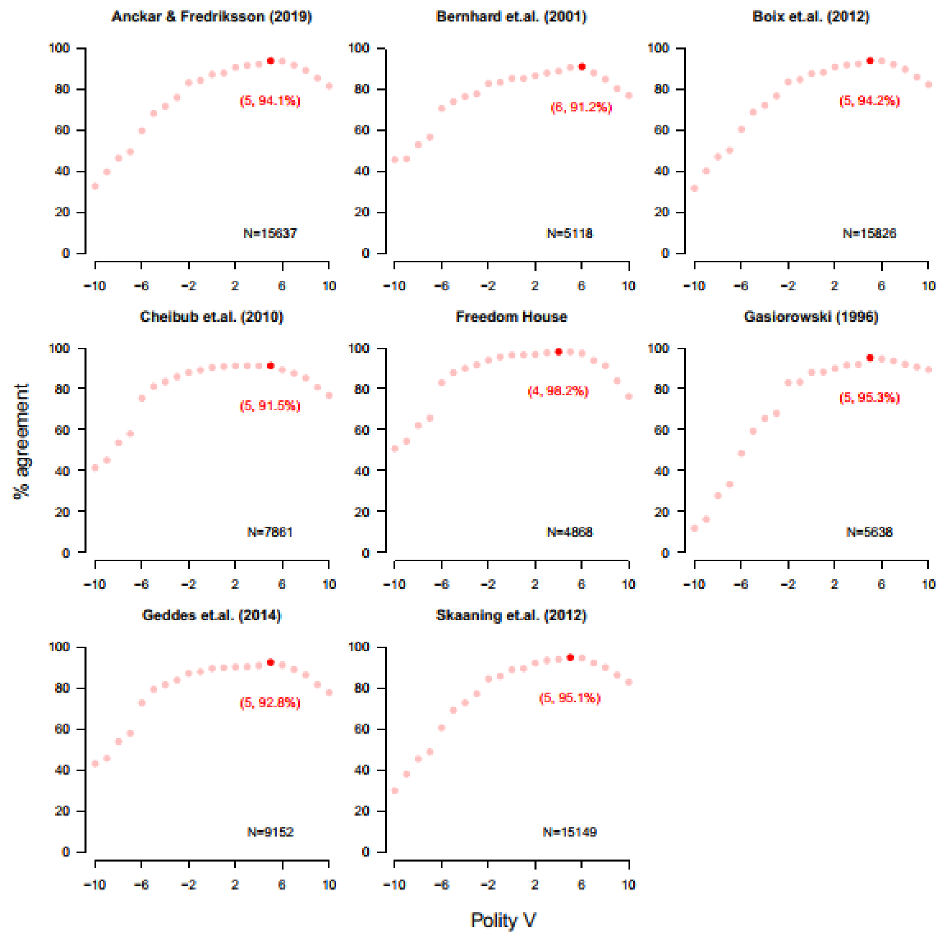

Figure 5: Optimal polity cutpoints.

Figure 4 plots the quality of the match between polyarchy and each binary dataset when τ takes on every possible value in the range of polyarchy, and Figure 5 shows the same data for polity. Figure 4 presents a remarkable regularity: multiple extremely different datasets, with radically different cases and coding methodologies, share the same optimal polyarchy cutpoint. The optimal polyarchy cutpoints for all of the dichotomous measures are in deep agreement with one another, differing by only 0.041, just 4.1% of the total range of polyarchy. We find exactly identical optimal cutpoints between Anckar and Fredriksson and Boix et al., and we are exceptionally close to finding that half of the datasets have the same optimal cutpoint: the optimal cutpoint shared by Anckar and Fredriksson and Boix et al. is just 0.001 less than the optimal cutpoint of Skaaning et al., which in turn is just 0.001 less than the optimal cutpoint for Freedom House. A difference of 0.001 (or 0.1% of the range of polyarchy) is as similar as cutpoints can possibly be without being exactly identical. And the remaining three datasets are also tightly grouped, with optimal cutpoints ranging from 0.382 to 0.406. In this group, Geddes et al. and Cheibub et al. also almost exactly agree with one another, with optimal cutpoints that differ only by 0.002.

Figure 5 shows a similar result. The datasets also closely agree about how to optimally dichotomize polity. Six of the eight datasets have the same optimal polity cutpoint of $\tau^* = 5$, while the other two datasets have adjacent values of $\tau^* = 4$ and $\tau^* = 6$.[34] Another identical cutpoint is mentioned by Pemstein et al., who found that the original Przeworski et al. coding (from which some of these datasets are descended) has an optimal cutpoint "somewhere near 5 on the polity scale".[35]

It is important to remember, though, that polity has only 21 values; up to the precision available in polity, we find the same tight similarity as in polyarchy, but the finding that polyarchy values are often identical or nearly identical is all the more unexpected since it has 1001 possible values.

It is intriguing enough that the values of τ^* are always so similar. But the figures show an even stronger result. Most of the panels in those figures have the same characteristic curve: very low cutpoints are similarly unsuccessful at matching the many-valued measure to each dichotomous measure, the quality of the match becomes sharply better as we choose cutpoints closer to the middle of the range, and then the quality of the match decreases much more gradually as we pick cutpoint values towards the top of the range. The fact that nearly all of these curves look similar means that every cutpoint is similarly successful at mimicking nearly all of the eight dichotomous datasets. So, for example, a cutpoint that matches polyarchy (or polity) to one binary classification with about a 50% success rate will also match polyarchy (or polity) to another dataset with about a 50% success rate.

As strong as the similarity in the optimal cutpoints is, this is an even more intriguing result: not only do the best cutpoints have similar values and similar success rates, but all of the possible cutpoints are similarly successful. As we have discussed, cutpoints add information that pairwise comparisons cannot address, so this finding builds on the conventional wisdom that these datasets tend to be fairly highly correlated with each other.

We have checked the best cutpoints for matching polyarchy to each binary dataset, and we separately did the same for polity. But can the cutpoints method also give us insights into how closely polity and polyarchy relate to each other? So far we have only discussed how to use cutpoints to match a many-valued variable onto a dichotomous variable. But we could also use them to match a many-valued variable onto another many-valued variable. A natural way to match polyarchy onto polity using cutpoints would be to proceed in two steps:

1. Turn polity into a binary variable

2. Find the optimal cutpoint for matching polyarchy onto that binary variable

In Table 1, we show the optimal cutpoint for matching polyarchy onto every possible polity dichotomization; since polity ranges from -10 to 10, there are 19 possible ways to dichotomize polity (excluding the bounds), each with a corresponding optimal polyarchy cutpoint τ^*.

Polity dichotomization	Polyarchy τ*	% Agree
-9	0.04	94%
-8	0.04	88%
-7	0.04	85%
-6	0.18	80%
-5	0.20	81%
-4	0.21	83%
-3	0.24	82%
-2	0.29	86%
-1	0.32	86%
0	0.34	87%
1	0.34	88%
2	0.34	90%
3	0.40	91%
4	0.40	91%
5	0.42	93%
6	0.45	93%
7	0.53	92%
8	0.57	92%
9	0.73	92%

Table 1: Optimal cutpoints for matching polyarchy onto every possible polity dichotomization.

One strategy for dichotomizing polity might be to pick the modal result of polity's optimal cutpoint for matching the binary datasets, since that will produce the binary classification that is the most similar to the literature. In Figure 5, polity's optimal cutpoint was found to be in the range around 5 or 6. The adjacent value 4 is the polity cutpoint suggested by Marshall et al.[36] And once again, we are greeted with the same recurring optimal cutpoint value: when polity is dichotomized in the way that best matches the literature, the optimal polyarchy cutpoint for matching to that binary dataset is almost identical to the recurring optimal cutpoints in Figure 4. And, when we dichotomize polity in the way that most closely agrees with other parts of the literature, we are able to match polyarchy to it better than if we dichotomize it any other way.

This result ties the literature together even more strongly than our previous findings. The best polyarchy cutpoint for matching to any possible dichotomization of polity happens to a) correspond to the optimal polity threshold for matching to the binary datasets, b) correspond to the recommended polity cutpoint, and c) also be the best polyarchy cutpoint for matching to the binary datasets. For all of these things to be true, there must be an exceptional level of agreement between the binary datasets, polyarchy, and polity.[37]

2.3: Similar cutpoints represent substantive agreement

We have shown that different datasets have similar optimal cutpoints, even in cases where they could have more than a thousand different values. What is interesting is not the specific value of the cutpoints that we found, but rather that they appeared over and over again in very different places. Out of 1001 possible cutpoint values (and 21 in the case of polity), we found highly similar — and sometimes exactly identical — cutpoints between eight dichotomous datasets that have:

- **Different authors**
- **Variation in cases** such that four out of every five country-years in one dataset may not even be included in another dataset
- **Empirical disagreements** about as many as one in every eight cases
- **Theoretical disagreements** with each one introducing innovations in substance or measurement specifically designed to differ from previous classifications
- **Different definitions of democracy** as fundamental as whether or not universal adult suffrage is a necessary feature of democracy
- **Different purposes** ranging from datasets aimed at specific research questions to ones constructed for more general use
- **Variation in coding methods** from hand-coding with explicit rules to the output of statistical models

If the datasets shared a common root, or major commonalities among any of these variables, it would be easier to see why the cutpoint which maximizes agreement between polyarchy and one of these datasets should also maximize agreement between polyarchy and another dataset. But this group of datasets includes variation in every important way.

But is it possible that the similarity of the cutpoints, while unexpected, does not tell us anything substantive? We consider three possible objections to our claim that the similarity of the cutpoints represents an unexpectedly strong consensus in the field.

One natural concern is that perhaps there is some methodological reason that this particular cutpoint will keep appearing over and over again. But we already saw strong evidence against this, when we found nearly identical polity cutpoints in addition to the nearly identical polyarchy cutpoints. That suggests that the pattern is not just a special feature of one of those variables, or an unlikely random event. In the Appendix we go further, running simulations which show that the similarity of the cutpoints is not caused by the tool itself, or by the distributions of polyarchy or polity, or by some meaningless feature of the dichotomous variables, or by chance alone. This establishes that their sameness really must represent some level of substantive agreement about democracy.

Another possible objection is that we have placed emphasis on the cutpoints coinciding up to 1, 2, or even 3 decimal places, but it is far from obvious that the third decimal place of a measure of democracy has an interpretable meaning. It is quite reasonable to argue that we should not focus too much on the third digit of polyarchy, and we invite the reader to completely ignore the third digits of our results (in fact, in the following section we will do just that). Then, instead of having two pairs of matching cutpoints, half of the eight datasets have matching cutpoints: rounded to two digits, Anckar and Fredriksson, Boix et al., Skaaning et al., and Freedom House all have an optimal cutpoint of 0.42. So the cutpoints are just as interesting for a reader who only considers two decimal places as they are for a reader who considers three.

One last objection might be that we have taken no account of the fact that we are identifying optimal cutpoints for datasets that develop very different definitions and operationalizations of democracy, and we do not focus on any differences in coding rules. This is certainly true: the differences in how these authors conceptualize democracy *is exactly what makes the similarity of the cutpoints meaningful*. The only commonality we insist on is that they are all classifying levels of democracy.

Crucially, nothing about this exercise asserts that one classification is more correct than another. We should not be understood as claiming that many-valued classifications are superior to dichotomous ones or *vice versa*, or that one dataset is closer than another to some objective ground truth, or that any differences between datasets are examples of one author being right and another author being wrong. There are major differences between different classifications; we have simply shown that the agreement between datasets runs deeper than conventional wisdom suggests.

3. Replications

Is the tremendous similarity of the cutpoints, to borrow the words of Boix et al., "merely academic"? In the spirit of meta-analysis we now pick a central empirical result from two papers that introduced very different and important dichotomous measures, Boix et al. and Cheibub et al., and fully replicate these core results of the optimal polyarchy dichotomization with $\tau^* = 0.42$.[38]

The primary purpose of Boix et al. is to present a dataset, but the authors include descriptive plots of the association between their democratic measure and three core variables: GDP per capita, land equality, and latitude. They find that "economic modernization variables have steadily declined in their correlation with democracy over time". In Figure 6 we show that dichotomized polyarchy successfully replicates every major result.[39]

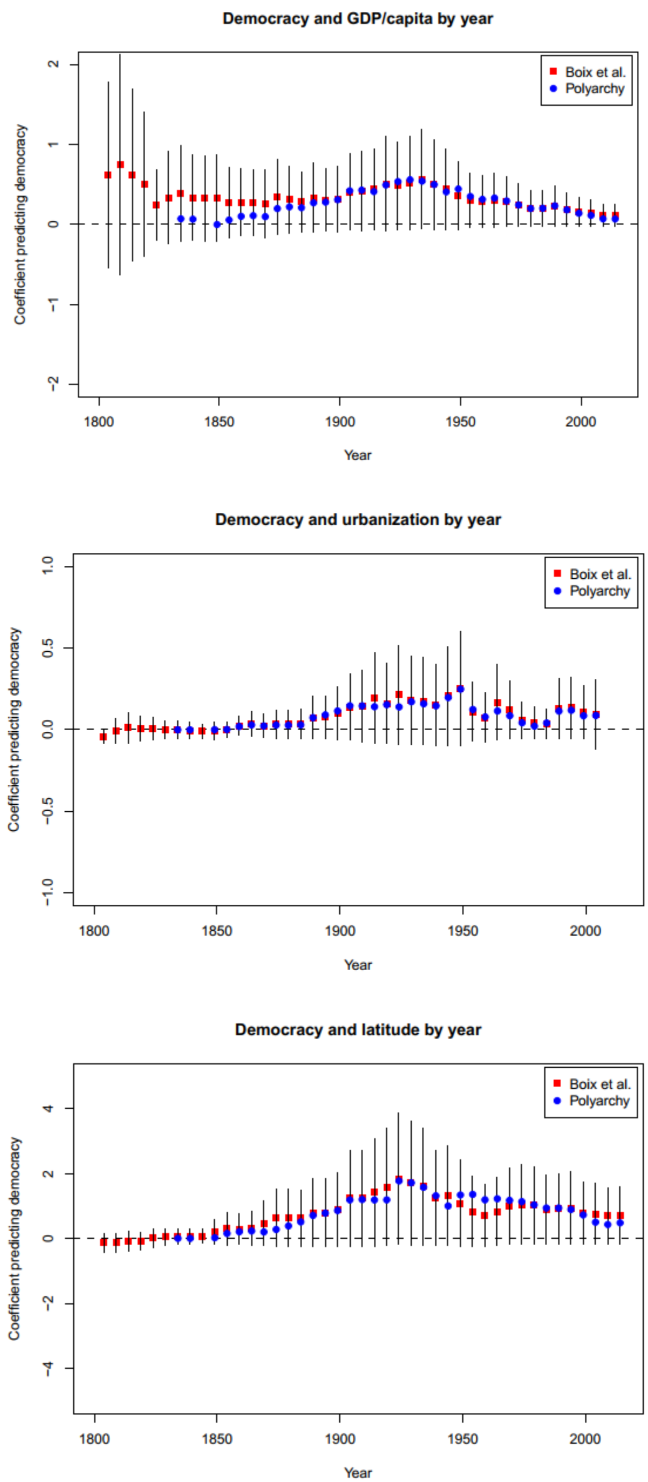

Figure 6: Replication of (respectively, from top to bottom) Figure 3, Figure 4, and Figure 5 in Boix et al., using their democracy coding and polyarchy dichotomized at τ^*=0.42. The years that do not appear have too few observations.

The presentation of the dataset in Cheibub et al. is organized around a series of replications of past work, to illustrate theoretical subtleties of their measure of democracy with respect to previous measures. The first of these is a replication of a main result in Fearon and Laitin.[40] In Table 2 we show that the dichotomized polyarchy dataset behaves exactly as the dataset in Cheibub et al. does: the same variables are significant, with the same signs and nearly identical magnitudes. The consensus classification is exactly as successful as the Cheibub et al. dataset in replicating the results of Fearon and Laitin.

	Fearon and Laitin (2003)	Cheibub *et al*. (2010)	Polyarchy replication
Anocracy	**0.54**		
	(0.24)		
Democracy	0.11		
	(0.31)		
Dictatorship with legislature		-0.36	-0.41
		(0.21)	(0.21)
Instability	**0.53**		
	(0.24)		
Prior war	**-0.84**	**-0.85**	**-0.87**
	(0.31)	(0.32)	(0.32)
GDP/capita	**-0.31**	**-0.34**	**-0.35**
	(0.07)	(0.07)	(0.07)
log(population)	**0.27**	**0.25**	**0.26**
	(0.07)	(0.07)	(0.07)
log(% mountainous)	**0.21**	**0.24**	**0.25**
	(0.09)	(0.09)	(0.09)
Noncontinuous state	0.33	0.37	0.38
	(0.28)	(0.28)	(0.28)
Oil exporter	**0.79**	**0.93**	**0.92**
	(0.28)	(0.27)	(0.27)
New state	**1.63**	**1.52**	**1.59**
	(0.35)	(0.34)	(0.34)
Ethnic fractionalization	0.15	0.10	0.10
	(0.37)	(0.37)	(0.37)
Religious fractionalization	0.43	0.52	0.48
	(0.51)	(0.51)	(0.51)
Constant	**-7.09**	**-6.48**	**-6.48**
	(0.76)	(0.73)	(0.73)
N	6,213	6,251	6,251

Table 2: Cheibub (2010) Table 2 replication. Bolded numbers significant at $p < 0.05$.

Of course, we emphatically are not arguing that a dichotomized version of polyarchy can replace any binary coding of democracy in any analysis and produce the same results. In the Appendix, we probe the limits of this sort of replication by attempting (with mixed success) to replicate results from the very different variable by Svolik.[41] But the replications demonstrate that lying latent in the field is a consensus classification that not only matches existing datasets exceptionally well, but actually reproduces several of the core analyses that motivated previous datasets.

4. What is the consensus?

Knowing that there is a consensus is half the story: we now probe what exactly that consensus is. First, we examine how much the dichotomous datasets agree with one another, and what drives their disagreements. Second, we dichotomize polyarchy at the optimal cutpoint, and we explore how the resulting dataset compares to the majority positions of the dichotomous datasets. Third, we briefly discuss the nature of polyarchy dichotomized at the optimal cutpoint, to better understand the characteristics of the consensus that we have identified in this paper.

4.1: Which countries are in question?

The binary classifications are known to agree closely with each other, but looking at how well they align with polyarchy reveals several open questions in the classification of democracy. For the sake of clarity, we will focus on polyarchy rather than repeating the following analyses for polity (but the high correlation in Table 1 suggests that analyses using polity would produce similar results), and we select a smaller number of dichotomous datasets to study. As shown in Table 3, these datasets represent the range of coding rules and practices among democratic classifications.

Dataset	Requirements for full democracy
Boix *et al*. (2012)	All of: • Executive elected in popular elections • Executive responsible to voters or legislature • Legislature chosen in free and fair elections • Most adult men have voting rights
Cheibub *et al*. (2010)	All of: • Free and fair elections • Minimum level of suffrage
Geddes *et al*. (2014)	None of: • Executive in power without direct, fair, competitive election • Rule change limiting electoral competition • Major parties blocked from elections by military
Freedom House	Expert survey with discussion and review

Table 3: The varying rules for coding a country as a democracy.

We saw, in Figure 1, that these cases typically have fairly strong pairwise agreement. So in the rare cases where they disagree, what drives those disagreements? Of the 186 distinct countries that appear in at least one of the four binary datasets, 105 of them (a clear majority) have no year in which one author considers them democratic while another author considers them autocratic. Among the countries that have a disagreement, there are a total of 806 individual disagreements, and the average number of years per country that feature at least one disagreement is about 4. Only 27 countries (14.5%) have more than 10 years coded as a democracy in at least one dataset while being coded as an autocracy in another. Table 4 records these countries and the years in question.[42]

Country	Number of disagreements	Years
Botswana	47	1966-2010
Lebanon	36	1944-1970, 1975, 1976, 2006-2012
Guatemala	32	1954, 1958-1962, 1966-1981, 1986-1995
Costa Rica	30	1920-1949
Namibia	26	1990-2015
The Netherlands	26	1871-1896
Fiji	25	1970-1986, 1992-1999
The Gambia	22	1972-1993
Lesotho	20	1966-1970, 1994-2008
Argentina	19	Many individual years
Bosnia and Herzegovina	18	1992-2009
Guyana	18	1972, 1992-2008
Thailand	18	Many individual years
Nigeria	17	1966, 1979, 1983, 1999-2012
Ghana	16	Many individual years
Guinea-Bissau	16	1994-1997, 2000-2002, 2004-2012
Haiti	16	1947-1950, 1991, 1995-1999, 2007-2012
El Salvador	15	1972-1975, 1984-1994
South Africa	15	1994-2008
Sri Lanka	15	1977-1982, 1989-1994, 2010-2012
Dominican Republic	14	1963, 1966-1978
Pakistan	14	1948-1957, 1976, 1988, 1999, 2008
Paraguay	14	1989-2002
Zambia	14	1965-1967, 1991-1996, 2008-2012
Peru	12	Many individual years
Panama	11	Many individual years

Table 4: Countries that have more than 10 years which were coded as a democracy in at least one dataset while being coded as an autocracy in at least one other dataset.

4.2: Which countries are in question?

The cutpoints method allows us to extend the venerable question of how well the dichotomous variables agree with one another into new terrain: it allows us to identify the specific cases in which optimally dichotomizing polyarchy produces a different result than the majority of binary datasets.

In Figure 7, for each polyarchy value, we plot how frequently the authors of the dichotomous datasets consider a country-year with that polyarchy value to be a democracy. To obtain this proportion we first check how many cases appear with that polyarchy value in V-Dem and at least one binary dataset. Then, we count the number of times that those cases were classified as a democracy in the binary classifications. Finally, we divide the number of binary classifications as a democracy at that polyarchy value by the total number of country-years at that polyarchy value. That represents how often the country-years at a given polyarchy value are coded as democracies. We should expect an increasing trend, with most cases at low polyarchy values being classified as autocracies, and most cases at high polyarchy values being classified as democracies. The figure confirms this expectation.

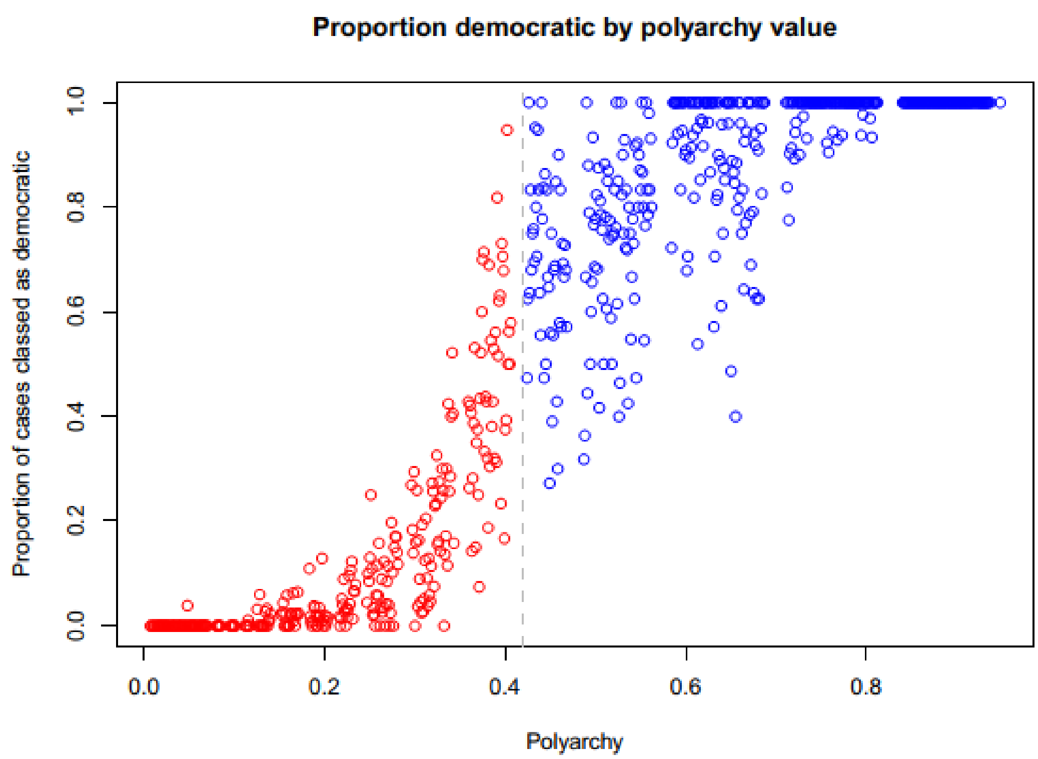

Figure 7: Proportion of cases in all datasets classified as democratic by polyarchy. The cases are coloured according to the cutpoint τ = 0.42, which is marked with a vertical line.

Figure 7 also shows that binary coders are more willing to classify ambiguous cases as democracies than as autocracies. When classifiers encounter a case that is right on the line, they are more likely to bump it up than to bump it down.

But in which specific cases does the cutpoints-based consensus differ from the majority of binary classifications?[43] 16,621 country-years appear in V-Dem and at least one binary dataset, after discounting the 214 exactly tied country-years. Of these, 378 country-years are classified as autocracies in a majority of the binary datasets, but are considered by the cutpoint method to be democracies. 725 country-years are considered democracies by the binary datasets, but the cutpoint method classifies them as autocracies. So the cutpoint is more likely to classify as autocratic cases that most authors consider democratic, and it is less likely to classify as democratic cases that most authors consider autocratic. Similar to the pattern in Figure 7, when most binary datasets consider a case to be autocratic, it is consistently considered autocratic, and most potentially ambiguous cases are classified as democracies.[44]

When the cutpoint method considers a case to be a democracy but most binary datasets call it an autocracy, we will call it a “cutpoint-only democracy”, and we call the reverse case a “cutpoint-only autocracy”. Table 5 shows the ten countries with the most cutpoint-only democracies, and the ten countries with the most cutpoint-only autocracies.

Country	**Cutpoint-only democracies**	**Cutpoint-only autocracies**
Burkina Faso	**23** years: 1979, 1993-2014	
Namibia	**23** years: 1990-2012	
Mozambique	**22** years: 1994-2015	
Tanzania	**22** years: 1994-2015	
Senegal	**20** years: 1979, 1981-1999	
Cyprus	**16** years: 1961-1976	
Côte d’Ivoire	**16** years: 1997, 2001-2015	
Zambia	**16** years: 1993-2008	
Sri Lanka	**15** years: 1979-1990, 2013-2015	
The Gambia	**14** years: 1966-1971, 1981-1988	
United States of America		**96** years: 1800-1833, 1836-1897
Greece		**77** years between 1864 and 1974
Guatemala		**35** years between 1945 and 1992
Chile		**33** years: 1909-1949, 1953
Colombia		**29** years between 1937 and 1982
Lebanon		**25** years between 1946 and 1975
Panama		**19** years between 1949 and 1990
Norway		**17** years: 1886-1897, 1941-1945
Ecuador		**16** years: 1948-1962, 1979
Argentina		**15** years between 1912 and 1983

Table 5: The 10 countries with the most cutpoint-only democracies or autocracies.

Cutpoint-only autocracies are countries that meet the minimalistic criteria used for most binary classifications of democracy, but which have a polyarchy value under 0.42. The most frequent cutpoint-only autocracy is the United States from its founding through the late 1800s. There is a straightforward explanation for this disagreement: the early United States satisfies many of the simple requirements that were used to create most of the binary classifications, like an elected executive and a minimum level of male suffrage, but was also rife with egregiously non-democratic institutions that might lower its polyarchy value.[45]

Cutpoint-only democracies are countries that do not meet the minimalistic criteria used for binary classifications of democracy, while having polyarchy of at least 0.42. This is a much rarer situation than cutpoint-only autocracies. Two particularly instructive cutpoint-only democracies are Namibia and Senegal.

From 1990 until 2012, Namibia is considered a democracy by the cutpoints method, but is considered an autocracy in most binary datasets. However, during these decades its polyarchy value ranges between 0.6 and 0.7, far exceeding the cutpoint of 0.42. Taken together with the prominence of Botswana in Table 4, one feature stands out as the likely reason that the binary classifications are divided on these cases: the “turnover rule”. Can a country be a democracy if it has never experienced a transition from one party to another? Authors who follow Przeworski et al. in answering "no" will classify Namibia and Botswana as autocratic, since they have been governed respectively by the SWAPO Party and the Botswana Democratic Party since independence; authors who answer "yes" may or may not code these cases as democratic.

The prominence of Senegal in Table 5 lends further credence to the idea that alternation of power is the cause of many cutpoint-only democracies. Senegal is classified as an autocracy by most binary datasets through the Senegambia period of the 1980s until the end of the 1990s, but consistently has a polyarchy greater than 0.42. Notably, for a few years after 2000, all four binary datasets classify Senegal as a democracy. So Senegal failed one of the binary datasets’ requirements for democracy throughout the 1980s and 1990s, but suddenly satisfied it in 2000. The substantive interpretation seems clear, since the 2000 election in Senegal was the first time that the country experienced a transfer of power to an opposition party. In this example, V-Dem’s polyarchy variable was a leading indicator of a democratic spell, and the binary datasets could only detect democracy after the change in government in 2000. The many-valued nature of polyarchy may have permitted V-Dem’s country expert coders to record a rise in democracy before the change was observable through strict binary classifications.

There is another noteworthy feature of the cutpoint-only democracies compared to the cutpoint-only autocracies: strong regional groupings. Of the ten cutpoint-only democracies listed in Table 5, eight are in Africa. V-Dem is substantially more likely to assign high polyarchy to African countries than the binary datasets are to classify them as democratic. In contrast, of the ten cutpoint-only autocracies, six are in Latin America, while three represent periods in the distant history of European or North American democracies. These cases are more likely to be classified as democracies in dichotomous datasets than to be assigned a high polyarchy score.

We have seen that there is an unexpectedly strong consensus regarding which countries are democracies and which are autocracies, and we have a sense of the scale and types of disagreements within this consensus. But what is it about democracies and autocracies that makes cutpoints around 0.42 substantively special? What sorts of countries have polyarchy above that cutpoint, and how are they different from the countries below that cutpoint? In the Appendix we discuss the dangers of hunting for substantive drivers of the optimal cutpoint after having already identified it; here we limit ourselves to simple descriptive statements.

The classification that represents a solid consensus of democratization scholars emphasizes the following traits. It firmly insists that democracies always have elections, but countries with elections can be autocracies. Elections in democracies might actually not be fully free or fair, and elections in autocracies are almost never free or fair. Almost all democracies have freedom of association and freedom of expression, and while many autocracies have no free association or expression, a small number of them do. Finally, democracies nearly always have full suffrage, while historically many autocracies did not.

4. Conclusion

It would be easy to read the literature on classifying democracies and conclude that political scientists have very little agreement on one of the oldest and most important topics in our field. For decades we have known that measures of democracy are all highly correlated, which led to a completely justifiable and empirically important focus on why those datasets are not interchangeable. But when we expand the field of comparison to include many-valued datasets with extremely broad coverage, we find agreement that is far deeper than this discussion has tended to reflect.

In this paper we have shown that there is actually very deep agreement on what constitutes a democracy. We addressed one of the major controversies in the literature — whether democracy is a matter of degree or of kind — by dichotomizing many-valued measures of democracy so that they match dichotomous measures of democracy as closely as possible. At the outset it seemed only natural that there should be no single best way to dichotomize a many-valued measure of democracy so that it matches datasets written by different authors, decades apart, about different institutions in different countries during different centuries, with sometimes conflicting definitions of democracy, coded through formal rulesets or discussions or predictive modeling, and which actively disagree about as many as one in every eight cases. And yet, we found that there is one consistent best way to dichotomize multiple many-valued measures of democracy to match these wildly diverse datasets. And that matching is so similar to the existing datasets that it can actually replicate a series of core results in the study of democracies.

While showing that classifications of democracies agree superbly with each other, we have made no attempt to judge how close each of those classifications is to some objective definition of democracy. Nor have we passed judgment on whether some bright line separates democracies and autocracies. These two questions are the subjects of venerable and sophisticated debates, and those debates should and will continue. But they should continue with the awareness that, so far, we agree about much more than we tend to admit.

Different people will have different definitions of democracy; some will be more similar to the consensus of the field, some more at odds with it. And that raises two disparate interpretations of our findings. On the one hand, perhaps our theoretical understanding of democracy is a great deal more harmonious than the conventional wisdom would suggest. So while differing classification rules have represented genuine theoretical disagreements, those disagreements only affect a small number of cases. On the other hand, there might be many tenable definitions of democracy that have yet to be used in a large cross-national classification. Perhaps there is some unseen reason that it is simply too difficult to construct a completely novel and radically different classification of democracy.

Skaaning et al. are far from alone in arguing that "the discipline is well served by a variety of measures". While the identical recurring cutpoints are a marker of how much classifiers of democracies actually agree, they are also a warning flag to anyone who sets out to construct a new measure of democracy. Even if a researcher develops a coding rule that they believe makes major substantive departures from previous classifications, the result is nevertheless very likely to closely resemble the consensus classification.

Notes

1. Schumpeter, *Capitalism, socialism, and democracy*, 247.

2. Popper, *Open Society and Its Enemies*.

3. Boix et al., "data set of political regimes"; Collier and Adcock, "Democracy and dichotomies"; Teorell and Lindberg, "Beyond democracy-dictatorship measures".

4. Casper and Tufis, "Correlation versus interchangeability".

5. Pemstein et al., "Democratic compromise".

6. Boese, "How (not) to measure democracy"; Bogaards, "Measures of democratization"; Kasuya and Mori, "Better regime cutoffs"; Lührmann et al., "Regimes of the world"; Wahman et al., "Authoritarian regime types revisited".

7. Dahl, Polyarchy: *Participation and opposition*; Lipset, *Political Man*.

8. Lerner, *Democracy and development*.

9. Cutright, "National political development".

10. Bollen, "Comparative measurement of political democracy".

11. Bollen and Jackman, "Democracy, stability and dichotomies", 612.

12. Alvarez, "Classifying political regimes", page 21; Bollen and Jackman, "Democracy, stability and dichotomies".

13. Alvarez, "Classifying political regimes"; Schumpeter, *Capitalism, socialism, and democracy*; Dahl, *Polyarchy: Participation and opposition*.

14. Bormann and Golder, "Democratic electoral systems"; Cheibub, Gandhi, and Vreeland, "Democracy and dictatorship revisited"; Przeworski et al., *Democracy and development*.

15. Boix et al., "data set of political regimes"; Geddes et al., "Autocratic breakdown"; Svolik, "Authoritarian reversals".

16. Geddes et al., "Autocratic breakdown"; Boix et al., "data set of political regimes".

17. Casper and Tufis, "Correlation versus interchangeability".

18. Geddes, "What do we know about democratization".

19. Cheibub, Gandhi, and Vreeland, "Democracy and dictatorship revisited", 67.

20. Boix et al., "data set of political regimes", 1525.

21. Skaaning, Gerring, and Bartusevičius, "lexical index of electoral democracy", 1492-4.

22. Bogaards, "Measures of democratization"; Wahman et al., "Authoritarian regime types revisited".

23. Kasuya and Mori, "Better regime cutoffs"; Lührmann et al., "Regimes of the world".

24. Bogaards, "Measures of democratization".

25. Gleditsch and Ward, "international context of democratization".

26. Boix et al., "data set of political regimes", 1528.

27. Boese, "How (not) to measure democracy"; Kasuya and Mori, "Better regime cutoffs".

28. We use the original Boix et al. dataset because we replicate results from that paper. The 2018 update yields the same results.

29. Coppedge et al., "V-dem dataset"; Marshall, Gurr, Jaggers, "Polity project"; Anckar and Fredriksson, "Classifying political regimes 1800-2016"; Bernhard, Nordstrom, and Reenock, "democratic survival"; Boix et al., "data set of political regimes"; Cheibub, Gandhi, and Vreeland, "Democracy and dictatorship revisited"; Gasiorowski, "political regime change dataset"; Geddes et al., "Autocratic breakdown"; Puddington et al., "Freedom in the world"; Skaaning, Gerring, and Bartusevičius, "lexical index of electoral democracy".

30. Casper and Tufis, "Correlation versus interchangeability".

31. According to this rule, when the elements of $\mathbf{x}$ have a precision of $n \in \mathbb{N}$ digits, τ is actually the equivalence class of real numbers that have an identical first *n* digits. When we refer to "the optimal cutpoint" in this paper, what we really mean is "the optimal cutpoint of precision *n*".

32. For one example see Lührmann et al., "Regimes of the world".

33. Kasuya and Mori, "Better regime cutoffs".

34. We also identified the optimal cutpoint for matching polyarchy and polity onto a very different dataset: Svolik's "Authoritarian reversals and democratic consolidation", which classifies consolidated versus transitional democracies rather than democracies versus autocracies. Amazingly, the division between consolidated and transitional democracies in that dataset has the same optimal polyarchy cutpoint as Skaaning et al., $\tau^*=0.422$, and its polity cutpoint of 4 matches that of Freedom House. These are the only global classifications for which we identified optimal cutpoints; we also identified cutpoints for two regional classifications (see the Appendix for more discussion).

35. Pemstein et al., "Democratic compromise", 444. We thank an anonymous reviewer for bringing this to our attention.

36. Marshall and Gurr, "Polity5".

37. In an Appendix, we check the robustness of the similarity of the cutpoints on dramatically different subsets of the data, including temporal subsets, subsets of only the cases with the most disagreements, and even random subsets using *k*-fold cross-validation. We also show that it is highly unlikely that we would find such similar cutpoints unless there is strong underlying similarity among the classifications.

38. This is the optimal cutpoint of half the binary datasets when rounded to 2 digits, and is also the main cutpoint in Kasuya and Mori's "Better regime cutoffs".

39. As we discuss in the Appendix, we strengthen the replication by varying certain details, which makes the numbers sometimes differ from the original results in substantively unimportant ways. These replications are also run on just the subsets of each dataset that appear in V-Dem, to facilitate direct comparison with the replications using the consensus dataset, which also makes some numbers different from those that appear in the original papers. This also changes no substantive findings.

40. Fearon and Laitin, "Civil war".

41. Svolik, "Authoritarian reversals".

42. These disagreements can sometimes arise from coding decisions that might appear arbitrary, such as whether a country-year observation is coded based on its status at the start or end of a year. This particular decision is not empirically important: dropping the first and last year of any mismatches (the only place where this difference in coding rules matters) leaves the analyses in this section substantively identical. But we choose to leave all cases coded exactly as they are in the datasets, because we are interested in capturing consensus in the field as it exists. It is not obvious that any coding rule is purely logistical; even rules to do with what time point a "country-year" represents can have a substantive rationale. Also, users of these datasets do not typically drop such cases, so dropping them would drag our analysis further away from the actual nature of the field. But none of this affects our analysis: no findings change substantially if these country-years are excluded.

43. In the Appendix we break these patterns down to the level of country-years, and discuss the advantages of the cutpoints method over checking the majority judgment of the binary classifiers.

44. Many of these are considered by Lührmann et al.'s Regimes of the World dataset to be electoral autocracies.

45. Mickey, *Paths Out of Dixie*.

Bibliography

Alvarez, Michael, José Antonio Cheibub, Fernando Limongi, and Adam Przeworski. "Classifying political regimes." *Studies in Comparative International Development* 31, (1996): 3-36.

Anckar, Carsten and Cecilia Fredriksson. "Classifying political regimes 1800-2016: a typology and a new dataset." *European Political Science* 18, no. 1 (2019): 84-96. doi:10.1057/s41304-018-0149-8.

Bernhard, Michael, Timothy Nordstrom, and Christopher Reenock. "Economic performance, institutional intermediation, and democratic survival." *Journal of Politics* 63, no. 3 (2001): 775-803.

Boese, Vanessa A. "How (not) to measure democracy." *International Area Studies Review* 22, no. 2 (2019): 95-127. doi:10.1177/2233865918815571.

Bogaards, Matthijs. "Measures of democratization: From degree to type to war." *Political Research Quarterly* 63, no. 2 (2010): 475-488.

Boix, Carles, Michael Miller, and Sebastian Rosato. "A complete data set of political regimes, 1800-2007." *Comparative Political Studies* 46, no. 12 (2012): 1-32. doi:10.1177/0010414012463905.

Boix, Carles, Michael Miller, and Sebastian Rosato. "Boix-Miller-Rosato Dichotomous Coding of Democracy, 1800-2015." Harvard Dataverse ([dataset]; accessed November 22, 2021). https://doi.org/10.7910/DVN/FJLMKT.

Bollen, Kenneth and Robert Jackman. "Democracy, stability and dichotomies." *American Sociological Review* 54, no. 4 (1989): 612-621.

Bollen, Kenneth A. "Issues in the comparative measurement of political democracy." *American Sociological Review* 45, no. 3 (1980): 370-390.

Bormann, Nils-Christian and Matt Golder. "Democratic electoral systems around the world, 1946-2011." *Electoral Studies* 32, no. 2 (2013): 360-369. doi:10.1016/j.electstud.2013.01.005

Casper, Gretchen and Claudiu Tufis. "Correlation versus interchangeability: The limited robustness of empirical findings on democracy using highly correlated data sets." *Political Analysis* 11, no. 2 (2003): 196-203.

Cheibub, José Antonio, Jennifer Gandhi, and James Raymond Vreeland. "Democracy and dictatorship revisited." *Public Choice* 143 (2010): 67-101. doi:10.1007/s11127-009-9491-2.

Collier, David and Robert Adcock. "Democracy and dichotomies: A pragmatic approach to choices about concepts." *Annual Review of Political Science* 2 (1999): 537-565.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, *et al*. "V-dem [Country-Year] Dataset v.9." Varieties of Democracy Project ([dataset]; accessed November 22, 2021). https://doi.org/10.23696/vdemcy19.

Cutright, Phillips. "National political development: Measurement and analysis." American Sociological Review 28, no. 2 (1963): 253-264. doi:10.2307/2090612.

Dahl, Robert Alan. *Polyarchy: Participation and opposition*. Yale University Press, 1973.

Fearon, James D. and David D. Laitin. "Ethnicity, insurgency, and civil war." *American Political Science Review* 97, no. 1 (2003): 75-90.

Gasiorowski, Mark J. "An overview of the political regime change dataset." *Comparative Political Studies* 29, no. 4 (1996): 469-483. doi:10.1177/0010414096029004004.

Geddes, Barbara, Joseph Wright, and Erica Frantz. "Autocratic breakdown and regime transitions: A new dataset." Perspectives on Politics 12, no. 2 (2014): 313-331. doi:10.1017/S1537592714000851.

Geddes, Barbara. "What do we know about democratization after twenty years?" *Annual Review of Political Science* 2, (1999): 115-144. doi:10.1146/annurev.polisci.2.1.115.

Gleditsch, Kristian Skrede and Michael D. Ward. "Diffusion and the international context of democratization." *International organization* 60, no. 4 (2006): 911-933.

Högström, John. "Does the choice of democracy measure matter? comparisons between the two leading democracy indices, freedom house and polity IV." *Government and Opposition* 48, no. 2 (2013): 201-221. doi:10.1017/gov.2012.10.

Kasuya, Yuko and Kota Mori. "Better regime cutoffs for continuous democracy measures." *V-Dem working paper*, 2019.

Lerner, Daniel. *Democracy and development: political institutions and well-being in the world, 1950-1990.* Free Press, 1958.

Lipset, Seymour Martin. *Political Man: The Social Bases of Politic*s. Anchor, 1960.

Lührmann, Anna, Marcus Tannenberg, and Staffan I. Lindberg. 2018. "Regimes of the world (RoW): Opening new avenues for the comparative study of political regimes." *Politics and Governance* 6, no. 1 (2018): 60-77. doi:10.17645/pag.v6i1.1214.

Marshall, Monty G., Ted Robert Gurr. 2018. "Polity5." The Polity Project ([dataset]; accessed November 22, 2021).

Mickey, Robert. *Paths Out of Dixie: The Democratization of Authoritarian Enclaves in America's Deep South, 1944-1972*. Princeton University Press, 2015.

Pemstein, Daniel, Stephen A. Meserve, and James Melton. "Democratic compromise: A latent variable analysis of ten measures of regime type." *Political Analysis* 18, no. 4 (2010): 426-449. doi:10.1093/pan/mpq020.

Popper, Karl. *The Open Society and Its Enemies.* Routledge, 1945.

Przeworski, Adam, Michael Alvarez, José Antonio Cheibub, and Fernando Limongi. *Democracy and development: political institutions and well-being in the world, 1950-1990*. Cambridge University Press, 2000.

Puddington, Arch, Jennifer Dunham, Elen Aghekyan, Christopher Brandt, Rukmani Bhatia, Isabel Linzer, Shannon O'Toole, Sarah Repucci, and Tyler Roylance. "Freedom in the world 2018." Freedom House ([dataset]; accessed November 22, 2021).

Schumpeter, Joseph A. *Capitalism, socialism, and democracy*. New York: Harper & Brothers, 1942.

Skaaning, Svend-Erik, John Gerring, and Henrikas Bartusevičius. "A lexical index of electoral democracy." Comparative Political Studies 48, no. 12 (2015): 1491-1525. doi:10.7910/DVN/29106.

Svolik, Milan. "Authoritarian reversals and democratic consolidation." *American Political Science Review* 102, no. 2 (2008): 153-168. doi:10.1017/S0003055408080143.

Teorell, Jan and Staffan I. Lindberg. "Beyond democracy-dictatorship measures: A new framework capturing executive bases of power, 1789-2016." *Perspectives on Politics* 17, no. 1 (2019): 66-84. doi: 10.1017/S1537592718002098.

Wahman, Michael, Jan Teorell, and Axel Hadenius. "Authoritarian regime types revisited: updated data in comparative perspective." *Contemporary Politics* 19, no. 1 (2013): 19-34. doi: 10.1080/13569775.2013.773200.