



Coherence objectivity and measurement: the example of democracy

Sharon Crasnow¹ 

Received: 21 November 2019 / Accepted: 2 July 2020 / Published online: 10 July 2020
© Springer Nature B.V. 2020

Abstract

Empirical research on democracy depends upon data. The need for such data has led to the development of measures of democracy. Measurement models are evaluated in terms of their reliability and validity, both of which may be thought of as related to the objectivity of the measure. Using the Varieties of Democracy Project (V-Dem) as an example, I consider how assessing reliability and validity of measurement models is challenging and argue that democracy might be understood as measured objectively when it is subject to theoretical, empirical, and pragmatic constraints. I call this *coherence objectivity*.

Keywords Democracy · Measurement · Objectivity · Coherence · Values in science · Political science

1 Introduction

Democracy is one of the most extensively studied subjects in political science. Approaches to the topic are normative, empirical, and pragmatic. Normative approaches fall under the rubric of political theory and shape both empirical and policy research since they provide theories through which democracy is conceptualized. Empirical research covers a wide variety of topics including transitions to democracy, backsliding from democracy, the relationship between democracy and peace (the Democratic Peace literature), democracy and economic development, democracy and human rights, and others. In addition, both the theoretical and empirical democracy literature address policy questions—for example, questions about how emerging democracies are best supported by other democratic nations, how threatened democracies could be shored up, and what works (or does not work) for democracy promotion.

✉ Sharon Crasnow
sharon.crasnow@norcocollege.edu

¹ Norco College, Norco, CA, USA

As the theoretical literature makes clear, democracy is not a simple concept. Democracy is a regime type but regimes are themselves complex forms of governance that are characterized by institutions, social/political structures, and various practices. Exactly which institutions, social structures, and practices are involved in the make-up of democracy is disputed and so the concept is contested as well as complex.

Contemporary empirical research in political science and consequently the policy work that it informs depends on data—both their collection and interpretation. Increasingly, as the discipline has become more reliant on quantitative methods, those data are represented as numerical values. The need for such data is one of the motivations for the development of measures of democracy. Such measures appear in the form of indices of democracy that are produced using models of measurement that depend on methods of coding various characteristics believed to be necessary for or indicative of democracy. Given that the concept of democracy is complex and contested, it is not surprising to find a proliferation of measures of democracy. A recent comparison of democracy indices lists 19 (Coppedge et al. 2017, p. 6).

How good are such measures and how are they to be assessed? Are there criteria that can be used to compare measures? Are there conditions under which we might we consider complex concepts such as democracy to be measured objectively? At the most general level, measures in the social sciences are evaluated in terms of their validity and reliability. Measurement is thought to be valid if it yields the correct results. For democracy this means that the measure correctly identifies those regimes that are democratic and the degree to which they are democratic. This characterization is not very helpful, however since it appears to depend on the prior identification of democratic regimes independently of the measures in order to make this judgment. People do have some intuitions about what “democracy” means, and so about which regimes are democratic, but such intuitions are not precise enough to provide standards of the correctness of measures that are good enough to be useful for political science research. Nor, as we shall see, are intuitions about democracy uniformly shared. Consequently, how “correctness” should be understood when assessing the validity of measures of democracy needs explanation.

Reliability is similarly problematic. A measurement procedure is reliable if it *consistently* yields correct results. Consistency can be understood in two ways—as consistent across users—they get the same results when they use the measures—or as producing the same results for the same user in different contexts. In either case, unless this consistency is understood as “consistently valid” the agreement does not ensure the measure is good.

However, in the case of democracy, determining what counts as a valid or correct measure is not obvious. This is a seeming difference between measuring physical properties and measuring a social science concept like democracy and such a difference raises questions about whether it makes sense to talk about measuring democracy. Confidence that we can get valid and reliable physical measures is in part due to confidence in the objectivity of such measures. Because “objectivity” may also be understood in various ways, we may ask in what sense measures of democracy might be objective.

In what follows, I argue that democracy is a value-laden concept and so claims about it, including measurement claims, are mixed claims in the sense that Anna Alexandrova

(2017) uses the term when discussing measures of well-being. Democracy and its measures will inevitably carry value presuppositions and implications. Consequently, if such measures are nonetheless to be considered objective, it cannot be that they are objective in the sense of being value-free. An alternative understanding of objectivity is called for.

This essay focuses on the question of how reliability and validity of measures of democracy are related to the objectivity of claims about democracy. I explore these issues through the case of a particular measurement model—the Varieties of Democracy Project (V-Dem). V-Dem produced its first indices in 2014 and is currently on its 10th iteration.¹ I use V-Dem for a number of reasons: (1) it directly tackles issues that arise for measuring abstract concepts such as democracy; (2) information about how V-Dem indices are constructed and evaluated is readily accessible, giving the project far greater transparency than other democracy measurement projects; (3) it is the most comprehensive measurement project currently available²; and (4) contemporary political scientists who do empirical work on democracy are increasingly using its datasets and so it is emerging as a new disciplinary standard.

I begin with a discussion of different senses of objectivity in order to determine which are relevant. I next use Cartwright, Bradburn, and Fuller’s account of measurement of social science concepts to provide a framework for discussing conceptualizing and measuring democracy (Cartwright, Bradburn, and Fuller 2017). In Sect. 4, I introduce the example through which I explore the question—V-Dem (Varieties of Democracy) Project. The account of V-Dem’s structure and some details of how it provides the values that serve as measures of democracy are offered both as a concrete example of a measurement model and to create a shared point of reference for the sections on evaluation of measurement that follow. In Sect. 5 I explore Alexandrova’s (2017) notion of mixed claims and consider how it might be useful for thinking about democracy. I offer some examples of ways in which different value judgments about what democracy is and how its components are related to each other will produce different measurement outcomes. In Sect. 6 I use Hasok Chang’s (2004) work on measuring temperature and some ideas from more recent work (Chang 2012, 2017, 2018) for thinking about how to evaluate the validity and reliability of a measure.

I conclude by arguing that a measure of democracy might be understood as objective when it is subject to theoretical, empirical, and pragmatic constraints. The effect of these constraints can be seen in the way measures function in the practice of the discipline. It is the success of the measure in the roles that it plays in achieving the theoretical, empirical, and pragmatic goals of political science research that give it objectivity. I call this *coherence objectivity*.

¹ Version 10 was released in March 2020 and most references (except where noted) are to Version 9. The substance of the paper is not affected by the latest version of V-Dem.

² “Countries: 202 countries; Year coverage: 1789–2018; 450 + V-Dem indicators, 81 indices and 5 high-level indices” (V-Dem website <https://www.v-dem.net/en/data/data-version-9/>. Accessed 19 November 2019).

2 Objectivity

As a number of philosophers and historians of science have noted, objectivity has been understood in a variety of ways (for example, Daston and Galison 2007; Douglas 2004; Lloyd 1995). Lloyd (1995) identifies four senses of objectivity. Daston and Galison (2007) offer five main understandings with distinctions within these and overlapping meanings among them. Heather Douglas (2004) discusses three modes of objectivity each with subtypes; objectivity₁—broken into two sub-types that she calls manipulable objectivity and convergent objectivity; objectivity₂—detached objectivity, value-free objectivity, and value-neutral objectivity; and objectivity₃—intersubjectivity which include procedural objectivity, concordant objectivity, and interactive objectivity.

From these various approaches, I distill three main conceptions of objectivity. First, objectivity is often understood through being contrasted to subjectivity. What is objective is free from the biases of individuals—the idiosyncratic beliefs and desires that might affect their judgment. Lloyd describes this sort of objectivity as detachment or disinterestedness. This is also one of Douglas’s understandings in her objectivity₂. Second, objectivity is often used to indicate that we have gotten something right about the object under investigation (the object of inquiry)—we have grasped it, represented it or some aspect of it, accurately. Daston and Galison describe this sense of objectivity as “truth-to-nature”, but also as “right depiction”. Lloyd refers to it as getting at the “Really Real”—the independently existing world. For Douglas this is objectivity₁, manipulable objectivity. Things are objective in that they have causes and effects. A third understanding of objectivity has to do with how research is carried out. Some methods, practices, and techniques provide objectivity whereas others do not. This sense is captured in Daston and Galison’s mechanical objectivity, but also in their account of objectivity as trained judgment. For Lloyd this sort of objectivity demands public accessibility and she identifies observability as one way of interpreting public accessibility. Douglas’s objectivity₃ as procedural objectivity, seems to capture a similar idea, although both her objectivity₃ concepts of concordant and interactive objectivity focus on intersubjectivity as a kind of objectivity.

While it is useful to be aware that there are various ways in which the idea of objectivity enters into our assessment of knowledge—that the term is not always used with the same meaning—it is not immediately clear how one ought to think about objectivity when addressing the issue of measurement in the social sciences. To be more specific, it is not clear in what sense, if any, we could objectively measure democracy. In fact, it is a legitimate question whether it even makes sense to talk about objectively measuring whether and to what the extent social/political organizations—for our purposes, regimes—manifest the concept “democracy”.

Democracy is not like length, mass, or temperature—some other things that we measure—in that it is not a physical property. It is nonetheless real in the sense that democracy has causes and effects. Investigating those causes and effects is one of the main motivations for wanting to classify and ultimately measure democracy. When a regime is classified as or judged to be a democracy there is some sense in which a claim is being made that the social/political object has been grasped. There is quite a bit of agreement about which polities are or are not democracies, at least at the extremes: Sweden, yes; North Korea, no. This agreement indicates some degree of

public accessibility to democracy. Although they are in many ways unlike length, mass, or temperature democracies are (real) configurations of polities, made up of people, institutions, and practices.

But that they are (socially/politically) real does not ensure that it makes sense to measure them or that we can measure them well. Social science measures are considered good insofar as they are *reliable* and *valid*. These features might be thought of as reflecting the objectivity of such measures. If so, the relevant senses of objectivity outlined in this section are those having to do with representing the object of inquiry accurately—grasping it, getting at what is “Really Real” about it, or Douglas’s objectivity₁ (manipulable objectivity)—and intersubjectivity.

That democracy is a social/political object also suggests that social/political values are relevant to its consideration. If our conception of objective science is that it is value-free that understanding creates a problem for thinking of measures of democracy as objective. In fact, some critics have argued that most measures of democracy reflect the social/political values of the global North and so are not adequate to political configurations elsewhere that would be judged democratic under other values—other understandings of democracy. How are we to evaluate this critique? This would seem to be a question of whether the procedure used for measuring is objective.

I argue that the concept of democracy, like many social science concepts, is intrinsically value-laden and so claims about democracy, which regimes are democratic, how many there are, the degree to which they are democratic, or the relationship between democracy and other things (such as the economy, freedoms, well-being) cannot be understood to be objective if objective means value-free. Nonetheless, I do think there is a kind of objectivity we can require of measures of democracy and that this objectivity can be understood through a clarification of what it is to have valid and reliable measures of democracy. Such a conception of objectivity does not require that measures are value-free however.

Recent philosophy of science has challenged the value-free ideal of science and so there are many potential sources for an understanding of objectivity consistent with a science in which values play a role. While there are clearly cases where social or political values have harmed science—examples abound of sexist and racist science—consideration of the role of values in science has shifted from their prohibition to a more nuanced discussion of when they play a legitimate role in good science. Douglas, for instance, argues for an indirect role for values in assessing evidence (although they cannot serve as direct evidence) (Douglas 2009). Sandra Harding makes an even more forceful claim about the role of values when she argues for strong objectivity (for example in Harding 1986, 2015).³

Claims about democracy, and specifically for the purposes of this essay, claims that use measures of democracy as evidence, are what I refer to as “mixed claims” following Alexandrova in her discussion of measuring well-being. For Alexandrova, a claim is mixed if:

1. It is an empirical claim about a putative causal or statistical relation.

³ For an overview of the science and values literature see Elliott (2017).

2. At least one of the variables in this claim is defined in a way that presupposes a moral, prudential, or political value judgement about the nature of this variable (Alexandrova 2017, p. 82).

The first characteristic refers to the sorts of empirical claims mentioned in the introduction—claims about democracy and the economy, democracy and peace, democracy and trade, for example. I argue that the second characteristic is present in the case of democracy through examining how both the notion of democracy itself and the complex of factors that make up democracy are value-laden. As a result, I also argue that when we understand such claims about and measures of democracy to be objective, objectivity is not to be understood as freedom from values. Alexandrova argues that in the case of measuring well-being, although mixed claims abound, a sense of objectivity that does not require that such claims be value-free is possible. I argue this for democracy as well in Sects. 4 and 5 and sketch an alternative path to establishing the objectivity of such measures through their coherence with theoretical, empirical, and pragmatic aspects of the research in which they are employed.

3 Measurement

First, we need a framework through which to discuss measurement of social science concepts. For this purpose, I use the account of measurement developed by Cartwright, Bradburn, and Fuller. They describe measurement as requiring three “steps”: identification of the boundaries of the concept to be measured (characterization), identification of a metric to which that concept will be matched (representation), and rules or procedures through which the matching is to be done (procedures) (Cartwright, Bradburn, and Fuller 2017, p. 78).⁴ For the first, specifying how the concept “democracy” is to be understood proves challenging. Cartwright, Bradburn, and Fuller suggest that such concepts, are “*Ballung*” concepts—cluster concept or family resemblance-like concepts.⁵ Such concepts have no specific core—“different clusterings of features among the congestion (*Ballung*) can matter for different uses” (Cartwright et al. 2017, p. 81). Something about the way we think of democracies indicates that they are all similar in some sense, but any two democracies might be very dissimilar from each other. We might focus on different aspects of democracy as more relevant in a particular situation picking out some properties of democracy as more important than others. We see this in the literature on democracy where, for example, the distinction between liberal democracy and electoral democracy is made through an emphasis on different aspects of the broader notion “democracy”. The claim that a regime is or is not a democracy could, for these reasons, be ambiguous. If a *Ballung* concept is to be measured, a particular characterization of the concept must be chosen and specified. Different characterizations are likely to be suitable for different purposes. Consequently, different measures may be suitable for different purposes.

⁴ The account was presented in 2010 at The Workshop on Advancing Social Science Theory: The Importance of Common Metrics and is summarized in National Research Council (2011). The ideas are summarized in Cartwright and Runhardt (2014) and developed in Cartwright et al. (2017). I rely primarily on the last of these.

⁵ The term *Ballung* is borrowed from Otto Neurath.

Using somewhat different terminology, political scientists Adcock and Collier (2001) describe the process as follows. They refer to what Cartwright, Bradburn, and Fuller call the *Ballung* concept as the “background concept” and call for replacing it with a *systematized* concept, specified through observable indicators in order to make measurement possible. In the case of democracy, the components of democracy—elections, liberties, suffrage, participation, and so on—are also abstract and similarly ambiguous. Decisions about how they are to be operationalized and to what extent these components go together, exist separately, or are crucial to democracy are part of what goes into systematizing the concept. Again, different choices result in different systematized concepts and consequently different concepts being measured.

An important difference between the way that Adcock and Collier approach systematizing a concept and Cartwright et al.’s treatment of *Ballung* concepts is that Cartwright et al. explicitly note the role of values in making decisions about what to include and what to exclude (Cartwright et al. 2017, p. 78). These choices differ depending on interests and other considerations, for example, the influence of some particular theoretical understanding of the term. The result is that a measurement under one characterization or systematization of the concept may not be commensurable with that under another. Since democracy is recognized by researchers as a complex concept of the sort that needs systematization different sorts of democracy are often distinguished.⁶ How to distinguish them—what the components of democracy are—is disputed and so the concept is contested as well as complex.

In addition to systematizing the concept, measurement requires identifying a metric (representation) and the rules for how the indicators are mapped onto that metric (procedures). The type of metric that is desirable also will depend to some extent on what we want the measurement for. In the case of democracy, some indices are dichotomous (democracy/nondemocracy) whereas others provide a scalar measure—either ordinal or interval. While a dichotomous index might be all that is needed if one is counting democracies, it can fail to give information that would be relevant for the exploration of some specific phenomenon, such as the erosion of democracy, or the degree to which being democratic might be related to other characteristics of a polity—such as its economy.⁷ Different research goals may require different approaches to measurement. But notice, that some of what makes sense in terms of a metric depends on what sort of conception of democracy is settled on. If democracy is identified through some essential set of characteristics—holding contested elections, for example—it could be argued that a polity either is or is not a democracy and that degrees do not matter.⁸ Research goals depend on our understanding of the objects of inquiry—how it is conceptualized—and consequently what we imagine we need by way of measurement metric will be thus constrained as well.

Rules for mapping concepts onto metrics include rules for operationalizing indicators—coding rules. Additionally, there are rules for aggregation. Different ways of aggregating the values for indicators provide different measures. For example, values

⁶ Collier and Levitsky’s (1997) “Democracy with Adjectives”.

⁷ Note that counting democracies is one sort of measurement of democracy. The general question of whether democracy is becoming more prevalent might be answered through counting. Even this simple mode of measuring democracy requires specification of the concept.

⁸ And there are approaches that do this (Przeworski et al. 2000).

can be combined additively or multiplicatively depending on whether components of democracy are understood as constituting necessary and sufficient conditions for democracy or are treated as family resemblance characteristics. There also may be differences in how the components are weighted. Should they all receive equal weight or are some more important than others? The answers to questions such as these will give rise to different aggregation formulae. Disagreements about the answers correspond to different theoretical understandings of the core concept.

In this section, I have laid out the basics of measurement, with suggestions about where they might be problematic for measuring democracy. I now turn to a brief account of how these play out in a particular democracy measurement project—V-Dem.

4 The V-Dem project

The project website describes V-Dem in the following way:

Varieties of Democracy (V-Dem) is a new approach to conceptualizing and measuring democracy. We provide a multidimensional and disaggregated dataset that reflects the complexity of the concept of democracy as a system of rule that goes beyond the simple presence of elections. The V-Dem project distinguishes between five high-level principles of democracy: electoral, liberal, participatory, deliberative, and egalitarian, and collects data to measure these principles (<https://www.v-dem.net/en/about/>. Accessed 29 April 2020).

The project claims to offer an approach that differs both in terms of how democracy is conceptualized and measured. Of particular importance is that V-Dem conceptualizes democracy as involving more than elections. The claim that the dataset is multidimensional and disaggregated is a reference to its comprehensive nature but also to its flexibility. It is comprehensive in that it covers many understandings of democracy and it is disaggregated in that the approach links these understandings to a variety of different components of democracy and in turn to discrete indicators of those components. Although the project produces a variety of ready-made indices of different types and different aspects of democracy, the disaggregated indicators included in the dataset make it possible for researchers to explore particular components of democracy separately through looking at each of the discrete indicators. It is therefore possible to explore the relationships among the indicators, as well as their relationship to the broader concept of democracy. Additionally, the various components of democracy could be aggregated differently than they are in the available V-Dem indices.

The source for the five principles V-Dem identifies is the normative (political theory) literature on democracy. These five principles or types of democracy are: electoral, liberal, participatory, deliberative, and egalitarian.⁹ Each principle reflects the core values with which it is associated and as such represents a different understanding

⁹ V-Dem also identifies majoritarian and consensual principles. The former is the idea that rule by the people is rule by the majority and the latter the counterbalance to that idea—that the majority rules with the consent of the minority. However, no suitable operationalization of these principles has been accomplished and while they are discussed they are not measured in the project.

of the core concept democracy (Coppedge et al. 2017, p. 42). These five conceptualizations of democracy are captured through surveying country experts for each of the countries represented in V-Dem. The survey questions are designed to assess the extent to which a state has institutions, structures, and practices that cohere with the associated principles and the answers to the questions are indicators of the extent to which the regime qualifies as democratic under each principle (Coppedge et al. 2017, p. 25). Although crucial for determining the values of many of the indicators, the country experts are not the only source for these values (see the discussion below). The indicator values are aggregated in order to produce the higher-level indices that correspond to the principles. All data, complete information about aggregation formulae, and other information about the methodology used are available on the V-Dem website (<https://www.v-dem.net/en/>).

V-Dem Methodology v. 9 summarizes the structure of the project with the following schema (Coppedge et al. 2019b, p. 12)¹⁰:

- *Core concept (1)*
 - *Democracy Indices (5)*
 - *Democracy Components (5)*
 - *Subcomponents, and related concepts (87)*
 - *Indicators (473)*

The core concept at the top is the background concept of democracy. The five principles give rise to the five indices identified as five components of the background concept of democracy.

To see how this works, consider the example of electoral democracy (one of the five principle). V-Dem understands electoral democracy as polyarchy, an understanding which is widely, although not universally, accepted.¹¹ The following is the V-Dem aggregation formula¹²:

$$\begin{aligned}
 v2x_polyarchy &= .5 MPI + .5 API \\
 &= .5(v2x_elec\textit{off} \\
 &\quad * v2x\textit{el_frefair} * v2x_frassoc_thick * v2x_suffr \\
 &\quad * v2x_free_alt\textit{inf}) + .5(1/8 v2x_elec\textit{off} \\
 &\quad + 1/4 v2x\textit{el_frefair} + 1/4 v2x_frassoc_thick \\
 &\quad + 1/8 v2x_suffr + 1/4 v2x_free_alt\textit{inf}) \\
 &\text{(Coppedge et al. 2019b, 7)}
 \end{aligned}$$

This aggregation formula shows that polyarchy is conceived of as consisting of a number of different components. It also reflects that V-Dem uses both what they refer to as a classical definition in terms of necessary and sufficient conditions (the multiplicative index—*MPI*) and a family resemblance conception of democracy (the

¹⁰ The numbers noted here refer to how many of each of these there are in V-Dem. The 473 indicators far exceed any other measurement project for democracy.

¹¹ Dahl (1971, 1989) proposed understanding electoral democracy as polyarchy.

¹² *Is used to indicate multiplication; the prefix *v2x_* indicates a V-Dem index.

additive index—*API*) as equally legitimate approaches. By assigning each aggregation approach half weight in the total formula (.5 *MPI* +.5 *API* where *MPI* is the Multiplicative Polyarchy Index and *API* is Additive Polyarchy Index) neither is given preference.¹³ A comparison of the V-Dem aggregation formula with each of the alternatives—the multiplicative formula on its own and additive formula on its own finds that they are highly correlated (Coppedge et al. 2019b, pp. 7–8).

Polyarchy aggregates: *v2x_freexp_altinf*= freedom of expression and alternative sources of information index (aggregate of media bias, print/broadcast media critical, print/broadcast media perspectives); *v2x_frassoc_thick*= freedom of association index; *v2x_suffr*= share of the population with suffrage index; *v2el_frefair*= clean elections (free and fair) index; *v2x_elecoff*= elected officials index.

Each of these components is constructed from multiple subcomponents and so is itself an index, for example, the *v2x_freexp_altinf* is an aggregate of:

- Government censorship effort
- Media harassment of journalists
- Media self-censorship
- Media bias
- Print/broadcast media perspectives
- Freedom of discussion for men
- Freedom of discussion for women
- Freedom of academic and cultural

Subcomponents may be either indicators or indices. However, all values ultimately derive from indicator values.

The V-Dem codebook distinguishes five types of indicators and identifies how they are coded. They are: “(A) factual indicators coded by members of the V-Dem team, (B) factual indicators coded by Country Coordinators and/or members of the V-Dem team, (C) evaluative indicators based on multiple ratings provided by experts, and (D) composite indices. ... (E) other democracy measures as well as data on usual correlates of democracy (both factual and subjective)” (Coppedge et al. 2019a, p. 17).¹⁴

As mentioned, country experts are the primary source of values for many indicators of the core values associated with the principles. V-Dem is not unusual in using experts (the two most frequently used indices, Polity IV and Freedom House also use experts). However, V-Dem is unique in using predominantly within-country experts.

Type C: Variables coded by Country Expert

A Country Expert is typically a scholar or professional with deep knowledge of a country and of a particular political institution. Furthermore, the expert is usually a citizen or resident of the country. Multiple experts (usually 5 or more)

¹³ *MPI* treats the components as necessary conditions since should any be missing (receive a value of 0) the aggregation formula yields 0 as well. *API* reflects that any of the components is indicative of polyarchy, and while at least one of the “cluster” must be present, the formula does not treat any individual component as necessary. Recall that polyarchy is only one type of democracy (electoral), although as we shall see that for V-Dem it plays a role in the other four types as well.

¹⁴ I have used categories from v.8 of the codebook to simplify. v.9 of the Codebook gives more detail on which members of the project did the coding. For v. 9, see Coppedge et al. (2019a, p. 27).

code each variable. More information about the Country Experts can be found in the *V-Dem Methodology* document (Coppedge et al. 2019a, p. 27)

V-Dem’s Codebook includes the questions that country experts respond to on a Likert scale. An example is the question about government censorship effort, an indicator for the freedom of expression and alternative sources of information index (*v2x_freexp_altinf*) which is itself a subcomponent of the electoral democracy index (polyarchy).

Question: Does the government directly or indirectly attempt to censor the print or broadcast media? ...

Responses:

- 0: Attempts to censor are direct and routine.
- 1: Attempts to censor are indirect but nevertheless routine.
- 2: Attempts to censor are direct but limited to especially sensitive issues.
- 3: Attempts to censor are indirect and limited to especially sensitive issues.
- 4: The government rarely attempts to censor major media in any way, and when such exceptional attempts are discovered, the responsible officials are usually punished (Coppedge et al. 2019a, p. 185).¹⁵

Since V-Dem takes polyarchy to be a core component of democracy it incorporates the polyarchy index in the aggregation formula for indices of the other principles (liberal, participatory, deliberative, and egalitarian). The methodological justification is that this is consistent with the understanding of democracy as it appears in the theoretical literature—a basic level of democracy must be satisfied before other components contribute to the level of democracy—what matters for democracy once the baseline has been met is the extent to which the other components are present, and assessing the *quality* of democracy presupposes that the baseline has been met (Coppedge et al. 2017, p. 7).

The complete V-Dem dataset includes the five high level indices (corresponding to the five principles), other lower level indices from which the higher levels are constructed, and values for all indicators used to construct the indices.¹⁶ The dataset consequently includes both aggregated and disaggregated data. In this way V-Dem provides a variety of resources—different levels of indices for all included countries and the disaggregated indicators from which these indices have been constructed. This way of presenting the data makes it possible for researchers to use the indices as formulated by V-Dem, to use alternative aggregation formulae to create indices for

¹⁵ All questions include clarification. The clarification for this question: “*Clarification:* Indirect forms of censorship might include politically motivated awarding of broadcast frequencies, withdrawal of financial support, influence over printing facilities and distribution networks, selected distribution of advertising, onerous registration requirements, prohibitive tariffs, and bribery. We are not concerned with censorship of non-political topics such as child pornography, statements offensive to a particular religion, or defamatory speech unless this sort of censorship is used as a pretext for censoring political speech” (Coppedge et al. 2019a, p. 185)

¹⁶ There are a variety of other indices that have been constructed using the V-Dem dataset—for example, the Women Political Empowerment Index which is formed by “taking the average of women’s civil liberties index (*v2x_gencl*), women’s civil society participation index (*v2x_genacs*), and women’s political participation index (*v2x_genpp*)” (Coppedge et al. 2018a, p. 234).

specific purposes, or to look at specific components of democracy individually or in relation to other components should researchers so desire.

5 Value-laden concepts and mixed claims

V-Dem acknowledges that the five principles identified with democracy have normative connotations. The result is that there will be different values emphasized in the five indices (measures) associated with these principles. Additionally, while some of the questions call upon country experts to report “facts” (how many women are in the legislature), others call for the experts to make judgments. The judgments of the country experts are evaluative—based on their interpretations. The empirical and the normative intertwine in the different senses of democracy represented by the V-Dem indices, the choices involved in aggregation formulae, and in the determination of the indicator values used to construct those indices. They may also be entangled in the judgments of experts.

Many other social science concepts include normative and empirical elements in this way. Claims made using such concepts can be thought of as “mixed claims” in Alexandrova’s sense—claims that mix the moral and the empirical (Alexandrova 2017, p. 80). She identifies well-being as one such concept because it incorporates normative presuppositions—presuppositions about what constitutes a good life. Alexandrova argues that mixed claims do not in themselves destroy the possibility of objectivity where they are ineliminable, however they do require an account of how such claims can be objective. I argue that the same is true for democracy and when turning to the question of measurement an account of objectivity is called for there as well. I shall return to this issue in Sect. 6.

First, let us take a closer look at how values permeate measures of democracy. Democracy is often understood as a good in itself and a good insofar as it incorporates other goods. The latter judgment is one way in which particular conceptions of democracy will incorporate particular values. Consider, for example, the judgment made by V-Dem that polyarchy (electoral democracy) is a core component of all five types of democracy. This choice prejudices the question of the role of liberties in democracies since some liberties are included in polyarchy (*v2x_freexp_altinf* = freedom of expression and alternate sources of information). Not all indices presuppose liberties as so integral to democracy. For example, Przeworki et al. (2000) propose a minimalist understanding of democracy as any regime in which executive and legislative offices are filled by contested elections. This conception of democracy does not incorporate such an assumption about the presence or absence of political and civil liberties.

Because of the presumption that democracy is itself a good, valuation of regimes as less than democratic is not only a descriptive judgment of these regimes but also carries with it a negative value judgment when regimes fall short. Different conceptions of democracy also aggregate components differently in addition to aggregating different components and in so doing reflect different values. When such normative judgments have ramifications for policy decisions, we need to know what has gone into the measure in order to assess the appropriateness of the role of those values in the measure

for the goals of research. The transparency of V-Dem makes such an investigation possible.

Polities often aspire to at least appear democratic because it can affect financial (and other) support they may receive. Larry Diamond has suggested that one of the reasons for the increase in hybrid or semi-autocratic regimes—usually regimes that have some form of election but do not exhibit core freedoms—is understandings that identify democracies solely through elections (Diamond 2002).¹⁷ In such competitive authoritarian systems elections are often only for show.

Evaluative connotations are also noteworthy when polities are described as “falling away from democracy”. The frequently used metaphor of backsliding is one such description. “Backsliding” occurs when some components of democracy are eroded or eliminated—but which ones? The answer depends, in part, on how democracy is characterized (conceptualization). Different conclusions might be drawn given different understandings of what democracy involves, but empirical research is also relevant to investigating the relationship between various components of democracy.

For Alexandrova concerns about mixed claims arise because of the way that values are incorporated into measurement models of well-being. She notes that if they are not explicit and examined, we may be inattentive to their inappropriateness or impose them on those who do not share those values. Concerns like these are relevant for measures of democracy as well. While democracy is itself thought valuable (as is well-being), what is understood to be valuable about it may vary depending on how it is conceptualized since different conceptualizations may incorporate different values.¹⁸ For example, it has been a commonplace to think of democratization as occurring in three waves with the most recent occurring during the period beginning in the 1970s and continuing to the 1990s—the so-called “third wave” of democracy. Huntington’s work is closely associated with this idea (Huntington 1991). Diamond has noted that this understanding is contingent on an understanding of democracy that allows semi-authoritarian regimes to be classified as democracies—a classification that he questions (Diamond 2002).¹⁹ In a similar critique, Pamela Paxton has argued that the three waves of democracy disappear if suffrage—a key component of democracy—is understood as including women rather than interpreted as adult male suffrage, an interpretation implicit in most indices (Paxton 2000). Diamond’s point depends on what characteristics are included in our understanding of democracy, whereas Paxton’s is about how a component of democracy is understood. These conceptual differences indicate that such claims about democracy are mixed claims that result from differences in classifying and counting democracies—one sort of simple measure.

¹⁷ This would be a minimalist definition of democracy focused solely on electoral democracy, not understood as polyarchy.

¹⁸ That democracy is valued universally is thought to be justified through seemingly robust evidence that the majority of people answer that they would prefer to live in a democratic state when asked. However, recent research suggests that such responses are not wholly reliable given that people may not share a common conception of democracy (Ulbricht 2018). See the discussion of a Chinese conception of democracy later in this section.

¹⁹ Diamond points to Freedom House as classifying a number of such regimes as democratic (Diamond 2002, p. 22).

Worries about how values are incorporated into measures of democracy have led to criticisms that democracy indices are biased towards an understanding of democracy as it exists in the global North. V-Dem's use of country experts who reside in the countries they code was, in part, a response to such criticisms. Other frequently used indices (e.g., Polity IV and Freedom House), have been created primarily by scholars from the global North and coded by in-house experts. The claim has been that such indices are not adequate to the forms of democracy that exist elsewhere in the world.

To assess this criticism and to evaluate measures requires making judgments about what counts as a good measure. Standardly, this is a matter of reliability and validity but these in turn can be understood as related to some notion of objectivity. I have argued that measures of democracy incorporate values and so claims based on such measures will be mixed claims.

If this is correct, then in order for measures of democracy to be considered objective in a sense that supports an assessment of them as reliable and valid, that sense of objectivity cannot require that the measures are value-free.

I propose that we consider a hard case for reliability and validity of measures of democracy: the claim that China is a democracy. Surveys indicate many Chinese perceive their form of government as democratic and, even more surprising, do not take elections to be a key component of democracy.²⁰ The understanding of democracy underlying these views is that a regime is democratic if it governs in a way that is consistent with the well-being of the people (Lu and Shi 2015; Zhang and Meng 2018).²¹ Lu and Shi argue that this understanding of democracy among the Chinese has been shaped by a deliberate policy on the part of the Chinese Communist Party (CCP) in which the party used Confucian and Leninist ideas already prevalent in the culture to mold a public understanding of democracy that supports the current regime. They make the claim that the regime promotes a false conception of democracy for its own ends. This claim depends on a belief that a particular understanding of democracy—one consistent with the sort of democracy indices I have been discussing (including V-Dem) on which China is not measured to be democratic—is accurate.

Zhang and Meng describe a similar pattern of belief about democracy and China based on their survey of Chinese elites. However, they treat the Chinese understanding as an alternate conception of democracy rather than a false one. The difference here is directly relevant to considerations about the objectivity of claims about democracy. Is the Chinese conception merely one among a variety of legitimate conceptions reflecting different values or are there (objective) grounds for rejecting the claim that China is a democracy?

V-Dem's approach is flexible, in that allows for different aggregation formulae, different configurations of the components of democracy to adapt to different conceptions. However, given the way that the values for these components are based on the values of indicators and what V-Dem takes to be the indicators for those components

²⁰ “When asked to evaluate the nature of the CCP regime, only 1.3% of the respondents believed that China was not a democracy; 18.2% thought China was a full democracy; and 59.7% regarded China as a democracy with minor or major problems” (Lu and Shi 2015, p. 30).

²¹ “We find that many Chinese officials understand democracy according to the Confucian tradition of *minben* and the CCP's political heritage of mass line, that officials should listen to the people and make benevolent policies for them” (Zhang and Meng 2018, p. 656).

(the aggregation formulae), it is hard to see how China would count as a democracy by any measure developed through the use of V-Dem datasets and consistent with their five principles. In fact, China ranks near the bottom on all five indices of democracy (2019 V-Dem Annual Democracy Report). It ranks near the bottom on Polity IV and Liberty House indices as well.

Given the nature of the critique of measures of democracy—that they reflect the values of the global North—the mere agreement among these indices seems insufficient grounds for concluding that they are assessing China correctly. In fact, it could be argued that this uniformity merely shows shared value assumptions that the Chinese reject.

And yet there does seem to be a meaningful sense in which China is correctly, and objectively, judged not to be a democracy. While the conceptualization of democracy is value-laden this does not mean that it is completely determined by values. Many V-Dem indicators are empirically grounded—some in publicly observable features (the make-up of the legislature) and some in the evaluations of country experts.

But this sort of empirical grounding cannot be the only means through which we answer, since judgments about what to include in the concept (and hence the measure) also have value implications. Theory also plays a role. The construction of the V-Dem measurement model depends on a conceptual breakdown of the principles, components, sub-components, and indicators stemming from that theoretical work on democracy. Additionally, the responses of the country experts to the survey questions through which the coding of indicators is done are informed by theory. These factors are constraints on the measures and thinking through how they constrain them is necessary for determining how to assess the validity of those measures.

The goodness of measures is also assessed through their reliability. Scholars affiliated with V-Dem have conducted reliability tests to determine to what extent various properties (for example, gender) of those experts might affect their reliability. Evidence of both intercoder reliability (intersubjective objectivity) and consistency with other indices (concordant objectivity) support the reliability of V-Dem. This kind of agreement among measurers does not in itself support the validity of the measures however. To clarify how measures of democracy can be meaningfully evaluated depends on spelling out the constraints that support validity. In what sense can claims about whether and to what extent any polity is democratic be judged correct? In what sense are there objective standards for doing so?

6 Epistemic iteration and objectivity as coherence

In this section, I offer an account of “coherence objectivity” as a way of answering the questions posed at the end of the previous section. The core feature of this account is the idea that measures are objective when their use coheres with theory, empirical knowledge, and successful application. More generally, they are objective when they function successfully to support the aims of research. I turn to Chang’s discussion of measuring temperature to assist in this project.

Chang’s (2004) *Inventing Temperature: Measurement and Scientific Progress* gives a historical and philosophical account of the path through which measurement of tem-

perature was standardized. Out of this account, he develops a framework for thinking about problems of measurement, the relationship between measurement and scientific progress, and the nature of scientific progress. The philosophical analysis in his account centers around a notion of “epistemic iteration”—a kind of bootstrapping process through which provisionally accepted knowledge can lead to further understanding, which in turn leads to the revision or even rejection of the original starting point.

Democracy is *prima facie* quite different from temperature. Temperature is a measure of a physical phenomenon and as such it has both a basis in physical sensation—heat—and is connected to a system of physical laws. In both of these it differs from democracy, which names a social/political concept neither directly connected to physical sensation or directly observable as such. Neither is democracy embedded in a system of laws. Nonetheless, democracy is real in that it has causes and effects. Its social (political) reality is connected to intuitions about fairness in social/political governance, but more formally, to historically developed theories of political organization and governance, and how those forms of organization and governance were produced and have effects.

The differences between democracy and temperature matter in a variety of ways, but for now I focus on how they matter for judging the validity of measures of democracy. Chang’s case study of temperature reveals that validation was not completely straightforward in the case of temperature. The sensation of heat—the empirical observation—is not consistent or reliable. It varies from person to person, it differs depending on context. Warm water will feel hot if we immerse our hands in it after being out in a blizzard without gloves, and different “amounts” of heat have different effects on different materials. In other words, the immediate empirical basis of the sensation of heat does not give us what is needed if we are to standardize temperature in the way measurement requires. In Adcock and Collier’s terms, the background concept (heat) needs to be systematized (temperature)—and so linked to indicators. For temperature, the indicator can be the thermometer, but thermometers need to be invented and standardized, which, in turn, depends on identifying other indicators of temperature—boiling water, for example.

In addition, although the physical sensation of heat provides a place to start when measuring temperature, that start sustains an ordinal scale and not an interval scale. The scientific aims for which measurement of temperature is sought require an interval scale since the physical laws into which temperature is integrated involve mathematical operations not possible with an ordinal scale. For example, Boyle’s law gives us the relationship between pressure, volume, and temperature. Standard numerical measures of temperature can be used in calculations with this law and this coherence assures us that we have it right—that we have a successful measure. Of course, we need to be able to measure volume and pressure as well to determine that there is this fit. This leads to a dilemma. Thermometers work because there is a relationship between volume, temperature, and pressure—measuring temperature turns out to depend on Boyle’s law, and so it appears that our fixed point cannot be established until we can measure temperature. This is the problem of “nomic measurement” (Chang 2004, p. 59).

Chang characterizes the process whereby the qualitative, ordinal scale can serve as the epistemic basis for an interval scale as a process of “epistemic iteration”. Briefly,

in order to get the temperature measurement project off the ground it is necessary to establish a fixed point—a melting point or boiling point of some substance that will serve as an indicator of temperature—but to establish such a fixed point requires already having a fixed point against which to standardize it. We identify and re-identify the melting point in order to establish that it is indeed a fixed point. In order to identify and re-identify that point, it must be measured—for which we need a measurement instrument, but we do not have a measurement instrument until we have identified the fixed point from which we will measure. In spite of this seemingly insurmountable problem, the story of measuring temperature is a success story.

Chang offers an account of this episode of scientific progress as a case of epistemic iteration:

Epistemic iteration is a process in which the successive stages of knowledge, each building on the preceding one are created in order to enhance the achievement of certain epistemic goals. In each step, the later stage is based on the earlier stage, but cannot be deduced from it in any straightforward sense. Each link is based on the principle of respect and the imperative of progress, and the whole chain exhibits innovative progress within a continuous tradition (Chang 2004, p. 226).

Scientists (researchers) start from some existing body of knowledge using what Chang calls the “principle of respect.” They adopt and use a body of knowledge—what is currently accepted as known—and in starting from it they “respect” it. However, this respect does not require that they wholeheartedly embrace it or steadfastly hold it to be true. “The initial affirmation of an existing system of knowledge may be made uncritically, but it can also be made while entertaining a reasonable suspicion that the affirmed system of knowledge is imperfect” (Chang 2004, p. 225). The principle of respect may sometimes be in tension with what Chang refers to as “the imperative of progress”. We build on what we know and we do so in order to improve our knowledge and in the process we may revise or even discard what we thought we knew.

Chang’s account begins with the observable—the sensation of heat. The account that he gives of epistemic iteration does not require starting with the observable, however. Epistemic iteration may begin with some other accepted bit of knowledge—theory, observation, or perhaps even knowledge of how to do something. The starting point is not a foundation, but rather a point from which to begin.

I argue similar processes of iteration and the generation of coherence are visible in the V-Dem project. V-Dem starts with theory as a means of identifying features of democracy. “A thorough search of the literature on this protean concept reveals seven key principles that inform much of our thinking about democracy: electoral, liberal, majoritarian, consensual, participatory, deliberative, and egalitarian” (Coppedge, et al. 2019b, p. 4). Normative theories of democracy constrain measures of democracy in this project by providing a theoretical framework—a starting point and a touchstone. These theories are not purely normative. They are also empirically informed through the study of regime types as they have appeared around the globe over the course of history. Consequently, theories of democracy are shaped by both normative considerations and actual democracies—iteratively. Both the connection to theory and the realities of the political world (at any particular moment) serve as (temporally) fixed points.

They are part of the (current) body of knowledge that we affirm (principle of respect). The reliability (consistency) of the measure with theory provides a means to construct valid measures—measures which cohere with theory.

Debates within the discipline of political science about the nature of democracy—for instance, the question of whether democracies can be identified simply through contested elections as Przeworski, et al. claim, or whether a regime must exhibit other features to count as a democracy—rely on theoretical resources. But they also look at the consequences for knowledge projects, the aims of research, and the policies they inform. While conceptualizing democracy is a starting point for developing measures, concepts may alter in response to changes in regimes around the world, the results of empirical research on the causes and effects of democracy, debates about the values democracies exemplify, the effects on policies of different ways of conceiving of and measuring democracy, and on the goals of research. If we believe that democracies are to be supported and our measures count semi-autocratic regimes as democracies, we may find ourselves supporting regimes that hold values quite different from those we are committed to. This tension may be a reason to reconsider our conceptualizations and measures.

Chang's account of epistemic iteration for temperature begins with the principle of respect but what moves knowledge-building processes along is the imperative of progress—the need to both know and use what we know. Measures of democracy are needed for a variety of empirical research programs. When a measure of democracy (or a democracy index) is used to successfully identify empirical regularities, that result provides support for the validity of the measurement model.

Empirical research of this sort covers a wide variety of topics as already noted: transitions to democracy as well as reversions; the relationship between democracy and peace (the Democratic Peace literature); democracy and economic development; democracy and human rights; democracy and inequality; and others. In addition, research may directly address or later be used to address policy questions, such as, how emerging democracies might be supported by other democratic nations, what actions might be taken to shore up threatened democracies, or what works (or does not work) for democracy promotion.

Empirical generalizations produced by such research do not typically take the form of laws in political science, and so the tight fit that we see with temperature and laws is not present. Nonetheless results of research using measures of democracy may produce robust empirical generalizations or middle-range theories. The way these empirical generalizations, the measures produced through the model, and the overall success of our knowledge when it is put into practice all fit together provides a form of validation of the measures that the research relies on. It is the coherence among previously accepted knowledge, the results of research using the measures, and the application of knowledge produced through their use that produce objectivity—an objectivity that is more holistically assessed than the forms of objectivity discussed in Sect. 2, although not unrelated to them. This objectivity, while similar to Alexandrova's notion of procedural or pragmatic objectivity, is somewhat broader in the coherence sought. It is a coherence of measures with theory, empirical knowledge, and practical knowledge that provides support for assessing the validity of the measures. This is what I refer to as *coherence objectivity*.

My use of “coherence” is both inspired by and resonates with Chang’s use of the term as an adjective applied to scientific practice. He gives the following description: “The coherence of a system goes beyond mere consistency between the propositions involved in its activities; rather, coherence consists in various activities coming together in an effective way toward the achievement of the aims of the system” (Chang 2012, p. 16).²² He notes that coherence is a matter of degree and it is to be judged relative to the aims of a specific scientific practice. I use coherence as a modifier for objectivity in order to indicate the way a measurement model is evaluated by how it is integrated within the practice of science—for the measure of democracy, within political science. The discipline is shaped by theoretical commitments and the goals of both empirical understanding and practical application of the knowledge gained through research. Consequently, that a measurement model is implicated in practices with those aims and that the measurement model supports some measure of success at achieving them serves as a means of evaluating its validity. The coherence of the measurement model with the successful practice in which it plays a role serves as a means of objectively evaluating its validity.

Coherence objectivity provides a way of assessing validity that is signaled by but is more than intercoder reliability understood more narrowly as agreement among coders (intersubjectivity) with respect to the understanding of particular concept. Without an account of validity, intercoder reliability would provide the only evidence of how good a measure is. As a consequence, evaluation of measurement would depend almost entirely on intercoder reliability—agreement among the coders—and indeed, the literature frequently treats validity as reliability, collapsing the two concepts. But such an understanding is too narrow.

While intercoder agreement is relevant to assessing validity, I have argued that validity is also gauged by how the measure functions in a variety of other ways—thus its coherence objectivity. This point harkens back to the Cartwright, Bradburn, and Fuller theory of measurement used as a framework in Sect. 3. As they note, with a *Ballung* concept, different specifications of the boundaries of the concept—different “clusters”—may be relevant for different uses. Consequently, evaluating the validity of measures depends not only on intercoder reliability—agreement about the concept—but also on its appropriateness for its intended use. This includes its relation to already accepted empirical regularities, how well it works in successful knowledge production, and how successful the knowledge produced is for negotiating our way in the world.

We can see this broader sense of coherence operating implicitly in the V-Dem project. V-Dem conducts various reliability tests assessing validity by assessing intercoder reliability in different contexts. For example, intercoder reliability is examined through bridge coding—country experts coding their own and another country. Other

²² Chang gives a similar description in the context of offering an account of truth: “Coherence as I intend it goes beyond consistency between propositions; rather, it consists in various actions coming together in an effective way towards the achievement of one’s aims (Chang 2017, p. 107). Note that here and in the previous quotation “consistency” means logical consistency, not the sorts of consistency that I am discussing in relationship to assessing reliability. Also, Chang uses the notion of coherence as part of his account of realism (see Chang 2018). As I am using “coherence objectivity” it is not tied to any particular stance in the realism/antirealism debates and would be compatible with a variety of views, including (not surprisingly) my own as offered in Crasnow (2000).

experts engage in lateral coding—coding (responding to questions about) a specific attribute or set of attributes at some point in time across countries. Since the 2015/2016 update V-Dem has also used anchoring vignettes. These are imaginary cases for which the same set of questions are asked as for real cases. The vignettes are designed so as to require no specific country expertise and in this way they are intended to check for consistent understandings of key concepts. V-Dem measures have also been compared with those of other indices (Polity IV and Freedom House) and give similar results where they code the same countries.

The important point to note here is that while these exercises are treated by V-Dem as various kinds of reliability tests, in fact, they suggest that validity is not exhausted by agreement among coders. Validity as coherence of knowledge in its theoretical, empirical, and pragmatic manifestations is reflected in the variety of other ways that V-Dem tests for reliability, although not fully specified as such. The sorts of reliability tests described above do not guarantee validity but depend on it.

Checks on intercoder reliability are, in part, checks on familiarity with theory given that the five principles associated with the core concept “democracy” are all identified through theory. The construction of idealized anchoring vignettes as a way of judging intercoder reliability illustrates this point. Vignettes are crafted through theory and do not require knowledge of any particular country. When country experts complete a questionnaire for a vignette, their responses provide information about how they are interpreting the Likert scale as they answer the questions and inconsistencies, if they present, will be revealed.

Post-survey questionnaires have also been used to gather information about whether coder reliability is affected by other characteristics; again, these exercises aim at a broader conception of coherence than that suggested by intercoder reliability. Coders who rate as having less knowledge appear to be less reliable—that is less accurate in the sense that they do not use the concept correctly. “Less knowledge” is interpreted here as indicated through the way coders rank countries as democratic on a scale of 1–100 with higher scores being more democratic. Those who rate non-democratic countries, such as North Korea, as democratic, or rate clearly democratic countries like Sweden as less than fully democratic are identified as having lower awareness of the concept of democracy—less knowledge. There is some indication that coders that do this are less reliable, although the evidence is inconclusive (Marquardt et al. 2018).

These various reliability tests indicate that V-Dem uses theory (in anchoring vignettes) and previous shared background knowledge (North Korea is not democratic; Sweden is) in approaches that are examining the validity of the measurement model through coherence with theory and shared background knowledge. That is, they are assessing its coherence objectivity not merely consistency among the coders (objectivity as intersubjectivity).

7 Conclusions: What about China?

In the previous section I offer a conception of reliability and validity consistent with an understanding of objectivity that looks to the coherence of the measurement model with theoretical, empirical, and pragmatic knowledge—coherence objectivity. This is

an objectivity which grounds intersubjectivity but it is not equivalent to it. A return to the China example can clarify how these constraints operate to generate objectivity.

It is possible to argue that China does not count as a democracy for theoretical reasons—the Chinese concept of democracy is not compatible with the body of democratic theory. Even though there are differences among theories of democracy, these differences are mostly due to differences in the *weighting* of components, not the very presence of core components such as free and fair elections or rights. While it is possible to argue that these theories of democracy are incomplete in failing to include the understanding of democracy that Lu and Shi and Zhang and Meng identify as Chinese, the argument would have to be given. Such an argument would require providing evidence of the coherence of this alternative conception—evidence that would be accepted by the international community and that would effectively alter the way democracy is characterized (and measured). Theoretical constraints are reflected in the currently used measures and this is why V-Dem and other indices rank China as undemocratic.

Appealing to coherence objectivity also offers a second way for assessing the claim that China is democratic. Robust empirical generalizations provide empirical constraints as well. While broad generalizations are admittedly rare, they are not entirely absent. The best known and perhaps also the most robust of these is the Democratic Peace—the generalization that democracies do not go to war with each other. A China that went to war with nations classified as democratic would call for an explanation—either a rejection of the idea that China was democratic or overturning the generalization of the Democratic Peace. Which avenue is the right one to take would depend on a better understanding of the underlying mechanisms that sustain peace among democracies. Current research approaches the question through investigating the relationships among various components of democracy and in this way is linked to questions of measurement iteratively, in a pattern much as Chang describes. The disaggregated nature of V-Dem makes it suitable for such research. Processes through which measurement is validated and new empirical knowledge is produced are thus intertwined.

Finally, there are pragmatic constraints that contribute to coherence objectivity. The degree to which a particular understanding and the measures that result from it successfully produce new knowledge—their usefulness—is also a desideratum for objectivity in this sense. As a measurement model V-Dem has a number of virtues that contribute to making such assessments. Because it is transparent in its construction and explicit about its conceptualization of democracy, it facilitates debates about what should be included in the understanding of democracy by making it clear what is at stake, thus it is suited for use in research on topics such as democracy promotion where there is an extensive literature on why attempts at democracy promotion has failed. Because different conceptions of democracy are precisely defined for V-Dem and how they are understood is spelled out in their links to indicators it is possible to examine the relationships among components in a way that facilitates such research. This transparency also allows for reassessment and revision of our understanding of democracy. Measuring democracy is unlike measuring temperature in that we have not settled on standards and debates about standards will always depend on values,

but it is like measuring temperature in that the process of standardizing is iterative and implicated throughout knowledge production and use.

Theoretical and pragmatic constraints reflect the value-laden nature of the concept of democracy and claims made about it. Theoretical accounts are normative, describing democracy as an ideal, even when theories are informed by consideration of democracies as they have appeared in practice. Pragmatic considerations are aspirational. The judgment of how successful the production of new knowledge is when using particular measures depends in large part on the ends that knowledge is to serve. The assessment is not merely instrumental, but also requires considering the worth of those ends—a further value judgment.

That China promotes a conception of itself as democratic both within its borders and to the international community speaks to a desire to engage and prevail in the debate about values. The question of whether democracy should be conceptualized and measured differently than it currently is depends on theory, how theoretical and empirical accounts hang together, but also on international debates about what political values should be promoted and what obligations regimes have to their citizens. Whether China can make the case to the international community consequently depends of these factors as well. Is it possible for the Chinese leadership to govern in a way that is consistent with the well-being of the people without protecting human rights and individual liberties? To what extent are human rights and liberties required in a democracy? While these questions require engagement with values, they also have an empirical component. Their answers will be mixed claims.

Using V-Dem measures, it is clear that China is not a democracy, but the account offered here indicates that the debate is not closed. Measurement of democracy is not standardized and some of the considerations about values and the aims of political science research suggest that more than one conception of democracy is likely to be needed.

Nonetheless, I have suggested an approach here under which measurement of complex social phenomenon can be considered objective. There are grounds upon which to make the claims about which regimes are democratic and to what extent, and there are ways to evaluate those claims. Our specification of the concept through its social/historical/theoretical sources, the body of knowledge in which measures are embedded, and the practices through which the measures are produced and used provide the needed constraints and a coherence objectivity for such claims.

Acknowledgements I am grateful to two anonymous reviewers who pushed me for clarification. The participants in Cambridge HPS seminar, CHESS at Durham University, the Philosophy of Social Science Roundtable, and a workshop on the concept of democracy at UCSD were all provided feedback that helped me to focus the project (all in Spring 2019). Special thanks to Mary Morgan, Nancy Cartwright, and Stephan Haggard for conversation and concrete advice on both the framing and details. All errors are mine.

References

- Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, 95(3), 529–546.
- Alexandrova, A. (2017). *A philosophy for the science of well-being*. Oxford: Oxford University Press.

- Cartwright, N., Bradburn, N., & Fuller, J. (2017). A theory of measurement. In L. McClimans (Ed.), *Measurement in medicine: Philosophical essays on assessment and evaluation*. London: Rowman & Littlefield International Ltd.
- Cartwright, N., & Runhardt, R. (2014). Measurement. In N. Cartwright & E. Montuschi (Eds.), *Philosophy of social science: A new introduction* (pp. 265–287). Oxford: Oxford University Press.
- Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. Oxford: Oxford University Press.
- Chang, H. (2012). *Is water H₂O? Evidence, realism, and pluralism*. Dordrecht: Springer Publishing.
- Chang, H. (2017). Operational coherence as the source of truth. In *Proceedings of the aristotelian society* CXVII, Part 2, pp. 103–122.
- Chang, H. (2018). Realism for realistic people. *Spontaneous Generations*, 9(1), 31–34.
- Collier, D., & Levitsky, S. (1997). Democracy with adjectives. *World Politics*, 49(3), 430–451.
- Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., et al. (2019a). V-Dem codebook v9. In *Varieties of democracy (V-Dem) project*. Gothenburg, Sweden: Varieties of Democracy Institute.
- Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Marquardt, K. L., et al. (2019b). V-Dem methodology v9. In *Varieties of democracy (V-Dem) project* Gothenburg, Sweden: Varieties of Democracy Institute.
- Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S., & Teorell, J. (2017). V-Dem: Comparisons and contrasts with other measurement projects. In *Varieties of democracy (V-Dem) project*, Gothenburg, Sweden: Varieties of Democracy Institute.
- Crasnow, S. (2000). How natural can ontology be? *Philosophy of Science*, 67(1), 114–132. <https://doi.org/10.1086/392764>.
- Dahl, R. (1971). *Polyarchy: Participation and opposition*. New Haven: Yale University Press.
- Dahl, R. (1989). *Dilemmas of pluralist democracy*. New Haven: Yale University Press.
- Daston, L., & Galison, P. (2007). *Objectivity*. New York: Zone Books (MIT Press).
- Diamond, L. (2002). Thinking about hybrid regimes. *Journal of Democracy*, 13(2), 21–35.
- Douglas, H. E. (2004). The irreducible complexity of objectivity. *Synthese*, 138, 453–473.
- Douglas, H. E. (2009). *Science, policy, and the value-free ideal*. Pittsburgh: Pittsburgh University Press.
- Elliott, K. C. (2017). *A tapestry of values: An introduction to values in science*. Oxford: Oxford University Press.
- Harding, S. (1986). *The science question in feminism*. Ithaca NY: Cornell University Press Harding.
- Harding, S. (2015). *Objectivity and diversity: Another logic of scientific research*. Chicago: University of Chicago Press.
- Huntington, S. (1991). *The third wave: Democratization in the late twentieth century*. Oklahoma City: University of Oklahoma Press.
- Lloyd, E. (1995). Objectivity and the double standard for feminist epistemologies. *Synthese*, 104, 351–381.
- Lu, J., & Shi, T. (2015). The battle of ideas and discourses before democratic transition: Different democratic conceptions in authoritarian China. *International Political Science Review*, 36(1), 20–41.
- Marquardt, K. L., Pemstein, D. Seim, B., & Wang, Y. (2018). *What makes experts reliable?* Working paper series 2018, Vol. 68, Gothenburg, Sweden: Varieties of Democracy Institute.
- National Research Council. (2011). *The importance of common metrics for advancing social science theory and research: A workshop summary*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/13034>.
- Paxton, P. (2000). Women's suffrage and the measurement of democracy: Problems of operationalization. *Studies in Comparative International Development*, 43, 92–111.
- Przeworski, A., Alvarez, M. E., Cheibub, J. A., & Limongi, F. (2000). *Democracy and development: Political institutions and well-being in the world, 1950–1990*. Cambridge: Cambridge University Press.
- Ulbricht, T. (2018). Perceptions and conceptions of democracy: Applying thick concepts of democracy to reassess desires for democracy. *Comparative Political Studies*, 51(11), 1387–1440.
- Zhang, K., & Meng, T. (2018). Political elites in deliberative democracy: Beliefs and behaviors of Chinese officials. *Japanese Journal of Political Science*, 19(4), 643–662.