

Recent Events and the Coding of Cross-National Indicators

Comparative Political Studies
2024, Vol. 57(6) 921–937
© The Author(s) 2023



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: [10.1177/00104140231193006](https://doi.org/10.1177/00104140231193006)
journals.sagepub.com/home/cps



Nils B. Weidmann¹ 

Abstract

Much research in political science relies on datasets produced by human coders. Many variables included in these datasets are not based on observable facts but rather require a considerable level of human judgment. This project studies the extent to which this judgment is affected by availability bias and how it influences the retrospective coding of historic cases. The analysis uses coder-level data from the V-Dem project, one of the few datasets collecting and releasing codings tagged with timestamps when they were produced. The results show that recent dramatic events in a country just prior to the coding have a small, but visible impact on coder ratings, but primarily for those variables that are directly related to the observed events. The magnitude of this effect, however, is small. This alleviates concerns that prominent events in world politics around the time of coding significantly affect the reliability of cross-national indicators.

Keywords

quantitative methods, democratization and regime change, political regimes

¹Department of Politics and Public Administration, University of Konstanz, Germany

Corresponding Author:

Nils B. Weidmann, Department of Politics and Public Administration, University of Konstanz, Universitätsstr, 10, Konstanz 78457, Germany.
Email: nils.weidmann@uni-konstanz.de

Data Availability Statement included at the end of the article

Much empirical work in political science, and in particular in the comparative study of political regimes, relies on data produced by humans. Using human-coded cross-national datasets, comparative political scientists can study the spread (or decline) of democratic regimes (Lührmann & Lindberg, 2019), monitor the state of human rights worldwide (Fariss, 2014), or gauge the extent of media freedom across a global sample of countries (Kellam & Stein, 2016). Without human-coded cross-national datasets, much research in comparative politics, democratization, and development would be impossible to conduct. Many of the variables included in these datasets constitute subjective assessments made by the coders. For example, when coding the corrupt activities of the legislature (variable *v2lgrcrrpt*) for the “Varieties of Democracy” dataset (V-Dem, Coppedge et al., 2020), coders need to decide how one could possibly recognize corruption of members of the legislature in everyday politics, and how these observed outcomes translate into the different levels of the coding scale, which ranges from “commonly” to “never, or hardly ever.” These codings constitute difficult decisions for the coders: There are few if any precise guidelines (i) which empirical facts to take into account for the assessment and (ii) how to translate them into quantitative scores, leaving both decisions up to the coder. Additional complexity arises due to fact that the coding of many indicators is historic and oftentimes goes back several years if not decades. In essence, the coding of these indicators is a decision process with a considerable level of uncertainty.

Given that a coding process of this kind provides few guidelines for coders, it is likely to be affected by different biases in human decision making. Recent work is increasingly trying to understand these cognitive processes. Colgan (2019), for example, argues that human-coded datasets in IR may have an “American” bias, since they are produced by coders holding American values and views of the international system. Arnon et al. (2023) however, examining quantitative human rights scores produced by humans, find no evidence of bias. So far the strongest critique of human bias in political ratings has been put forward by Little and Meng (2023), who argue that findings of democratic decline are the result of changes in human perceptions but not in political institutions. Comparing subjective and objective democracy indicators, they show that democratic backsliding can be detected in the former but not the latter.

This paper examines yet another type of bias that could affect human coders. Given that coding decisions for historic cases are particularly difficult, it is reasonable to assume that coders draw on particular cognitive shortcuts. The availability bias is one of the most researched among them. It means that when confronted with a difficult decision-making task, humans use readily available and closely related information to solve it. In the coding of historic, cross-national variables, this could mean that human coders are influenced by recent dramatic events in the country in question, even if the coding applies to

a time period many years ago. Using time-stamped coder-level data from the V-Dem project, the analysis tests how these recent events affect coder ratings. The results show that there is some evidence for an availability bias in coder ratings, but the magnitude remains generally small and is unlikely to affect research done with these indicators in a major way.

Cognitive Biases in Human Coding

In comparative political science research, many human-coded datasets exist that cover a variety of political variables. A (non-exhaustive) list shows that these data cover many “soft” variables such as the rule of law ([World Justice Project, 2020](#)), conditions for independent media ([International Research & Exchanges Board, 2016](#)), press freedom ([Reporters without Borders, 2020](#)), political rights and civil liberties ([Freedom House, 2020a](#)), institutional characteristics of countries ([Bertho, 2012](#)), and, of course, democracy ([Marshall et al., 2018](#); [Economist Intelligence Unit, 2020](#); [Coppedge et al., 2020](#)). Producing these country ratings is a difficult task for human coders. Some require completely subjective assessments, such as, for example, V-Dem’s variable capturing the degree of a “rigorous and impartial public administration.”¹ Others are at least in principle based on factual outcomes or events, but it is very difficult for coders to collect all the information and/or aggregate them to a single country rating according to clear and well-specified coding rules. For example, V-Dem codes corruption in the legislature on a 0–4 scale. The facts supporting this coding may exist, but they are unlikely to be available to the coder.²

Decisions under uncertainty have been the subject of research for a long time, mostly in psychology and related disciplines. [Tversky and Kahneman \(1974\)](#) describe three heuristics that humans have been found to employ when they have insufficient knowledge as a basis for a decision. One of them is the well-known “availability heuristic.” Availability means that when making a decision, humans draw on those pieces of information that are more easily available to them. In this paper, we explore availability on a temporal dimension. When coding variables for a particular country for a time period that has long passed, it is well possible that coders are influenced by recent, dramatic events in the country in question.

What dramatic events are most likely to affect coding decisions? Existing research has produced conclusive evidence of a “negativity bias” in news, where readers are most likely to perceive and remember negative events ([Soroka et al., 2019](#)). In the cross-national study of democracy, where coders rate countries along a normative dimension between “closed autocracy” and “liberal democracy,” these negative shifts correspond to setbacks away from democracy, which can manifest themselves in military coups or the violent repression of the opposition. Consequently, if coders are affected by

availability bias, their retrospective ratings of a country should be reduced if a dramatic negative event immediately precedes the coding.

A short example serves to illustrate this. Imagine two coders rate the level of a repression in a given country, ten years ago. This is the typical coding task when producing cross-national datasets. The first coder produces this rating in year t , the second coder a year later at $t + 1$. In between t and $t + 1$, the country experiences a large wave of protest that is violently repressed by the government, with several casualties. Availability bias arises if these recent events influence the second coder's rating such that it is lower than the first coder's rating, despite the fact that both coders rate the *same* historic case—but they do so at different points in time.

If coders of comparative datasets are affected by the availability heuristic, there are different ways in which this can happen. In particular, we distinguish which coding decisions the available information (=recent events) will be used for. In a "narrow" version of availability bias, coders use this available information for decisions that affect the same type of phenomenon they are supposed to code. The above example illustrates this: If a country recently experienced a dramatic event that clearly shows a high level of repression (such as a violent crackdown against the opposition), coders will use this information when rating the same outcome—the level of violent government repression in this country. This narrow version of availability bias therefore suggests that recent dramatic events of governmental repression affect the retrospective coding of repression-related variables.

However, availability bias can also play out in a broader sense. Rather than influencing the coding of only those variables that are directly related to observed dramatic events, these events could also affect other variables. For example, having observed government violence against protesters, coders may implicitly downgrade their normative assessment of a regime and therefore assess this regime not just as repressive but also as "corrupt," "clientelistic," "undemocratic," and "illiberal." If this broader version of availability holds, we should see an effect of recent dramatic events also on a broad range of normative variables, where coders change their retrospective assessment of these countries and rate them as more illiberal after these events. In the following analysis, we put the narrow and the broad version of availability bias to an empirical test.

Research Design

The empirical analysis uses time-stamped coder ratings for variables related to democracy and democratic institutions. In the following, we describe the data source and the research design for the analysis.

V-Dem Coder-level Data

The empirical analysis relies on the coder-level data from Version 10 of the V-Dem project (Coppedge et al., 2020), a large data collection effort that captures many different aspects of “democracy” and is one of the leading datasets in the cross-national analysis of regimes. V-Dem distinguishes between factual (Types A and B) and more subjective (Type C) variables (Coppedge et al., 2020, p. 28). The latter constitute V-Dem’s key contribution and are the main focus of this project. Type C variables capture subjective ratings of political variables, produced by a number of coders with particular expertise about a country or region. Each of these coders answers a set of questions about the country and time period they have been assigned. Answers to the questions are recorded using different scales; few questions have a binary response (yes/no), while most others have an ordinal scale.

The V-Dem coder-level dataset contains the individual ratings produced by the coders before they are processed and aggregated further. In the version made available for this research, each coding has a time stamp associated to it, indicating the year in which it was produced. V-Dem codings are usually collected in January to cover the previous (and oftentimes also earlier) years. To maximize the amount of data for this analysis while retaining comparability between variables, only variables with an ordinal scale of 0–4 (the most common scale in V-Dem) were selected.³ Low values of these variables indicate illiberal political practices, and high values correspond to liberal-democratic features or outcomes. The abovementioned example of the *v2lgcrrpt* variable covering corruption in the legislature is an example for this.

Figure 1 illustrates the structure of the coder-level data from V-Dem. On the timeline along the x-axis, there is the actual case to be coded (here, the *v2lgcrrpt* variable for Egypt in 2016). This variable is then coded by three coders in early 2018 and two coders in 2019. We use the term “coding time” to denote the time (year) in which a variable is coded. In the example, *v2lgcrrpt* has two coding times: 2018 and 2019.

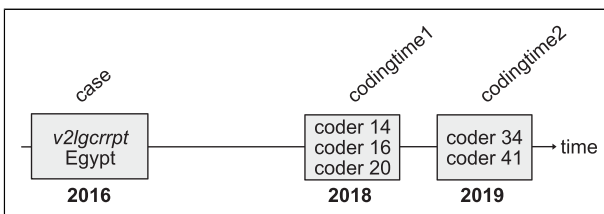


Figure 1. Data structure of the V-Dem coder-level data. A variable (*v2lgcrrpt*) is coded for a particular case (Egypt in 2016) at two different coding times (2018 and 2019).

Unit of Observation

In almost all cases, a particular variable for a given case is only coded *once* by a particular coder, which is why we cannot analyze changes within the same coder's ratings over time. For that reason, all ratings for a given variable and case are averaged by coding time; in the above example, this would give us two values for *v2lgcrrpt* in Egypt 2016: at coding time 2018, the average of coders 14, 16, and 20; and at coding time 2019, the average of coders 34 and 41. Oftentimes, there is only a single coder rating for a particular coding time.⁴

For the analysis, we make use of the fact that in the V-Dem project, codings for particular country-years are oftentimes generated at different points in time. These repeated observations allow us to examine changes in the average coding decisions over time, between the different coding times. For the analysis, we analyze *pairs* of codings made in consecutive years, in other words, where coding time 1 and coding time 2 are exactly one year apart. While other comparisons are possible (e.g., we analyze changes in coder ratings between a given year and three years later), this approach allows us to better attribute coding changes to the events that happened in between the two coding times. For all pairs provided in the coder-level data, we examine how shifts in the FH score in between the two coding times affect downgrades in the coding.

Measuring Dramatic Events

The purpose of the empirical analysis is to test whether recent dramatic political developments in a country between *codingtime1* and *codingtime2* affect coders' subjective ratings of past cases. We start by using different event datasets to capture specific events that indicate a worsening and increasingly illiberal political situation. The first of these indicators is the occurrence of events with violent repression of protest in the coded country. This variable is coded from the "Mass Mobilization Dataset" by [Clark and Regan \(2021\)](#), selecting protest events where the government responded with "killings" or "shootings." The second event-based predictor is the number of coups or coup attempts, coded from the "Cline Center Coup D'état Project Dataset" ([Peyton et al., 2020](#)). Both types of events are usually associated with high levels of political violence to suppress political opposition or to oust a democratic government.

These event-based predictors, however, cover only a small subset of developments that indicate whether a country is becoming increasingly illiberal. In fact, there is a broad range of events that indicate the deterioration of democracy and could therefore influence coders' retrospective assessment. For example, the struggle surrounding Poland's constitutional court in 2015–2016 was a clear and visible indicator of democratic decline but is

obviously not captured by the two event datasets. Rather than expanding the set of event-based predictors (which may not even be feasible due to limited data availability), we use an aggregate indicator that captures a broad range of shifts away from liberal democracy: the well-known Freedom House “Freedom in the World” rating (FH henceforth). The FH scores quantify a country’s level of political rights and civil liberties with an annual score between 0 and 100 (Freedom House, 2020a). This score is computed as the sum of several constituent indicators, capturing the electoral process, political pluralism and participation, the functioning of government, the freedom of expression and belief, associational and organizational rights, the rule of law, as well as personal autonomy and individual rights.

In many cases, Freedom House ratings change slowly. These changes are unlikely to be visible to coders and therefore not expected to influence their coding as they do not constitute “dramatic” events. We therefore select major changes from Freedom House in two ways. First, we select cases where a country was downgraded by at least 5 points in the FH scale within a single year, to capture major political shifts away from democracy. These major shifts happen rarely. Since we can only exploit variation in the coding times, we can only use changes during the years 2016–2020. During this time, only about 3% of the country-years experienced FH drops of 5 or more. A second way to identify major downgrades is by using the three FH status categories. FH classifies a country as “free,” “partly free,” or “not free,” depending on the values of the underlying political rights and civil liberties scores (Freedom House, 2020b). A status downgrade happens if a country drops from “free” to “partly free” or from “partly free” to “not free.” These status downgrades are major events that are discussed in the annual FH news release, so they are likely to be visible to the coders. In the analysis below, we test whether from any coding year to the next, a drop in the FH score of at least 5 points or a status downgrade leads to an adjustment of the coder ratings downward.

Results

In the empirical analysis, we assess the extent to which recent, dramatic events in a country affect the coders’ retrospective ratings. To do so, we proceed in a stepwise fashion. We start with a pooled analysis using sets of V-Dem variables, before testing each variable independently. The analysis focuses on the change in the coder ratings between two consecutive years. More precisely, the dependent variable is the difference of the average V-Dem coder rating between *codingtime1* and *codingtime2*. Overall, coder assessments are relatively stable. Out of the 117,817 observations in the main dataset, in 31,482 cases (about 27%) the coding remains unchanged, while upgrades (44,288, about 37%) and downgrades (42,047, about 36%) occur almost at the same rate. The distribution of changes in coder ratings from one year to the

next is visualized in [Figure 2](#). It shows that while the vast majority of ratings do not change (large peak at the center), downgrades and upgrades are similar in magnitude and almost perfectly balanced. [Appendix A2](#) presents summary statistics at the level of the variables, distinguishing between variation between the individual cases (the standard deviation of the average coder ratings for each country/year) and the average variation within a case (the mean across the standard deviations of the coder ratings for each country/year).

For the analysis, we use OLS models with the first difference of the coder ratings (the rating at *codingtime2* minus the rating at *codingtime1*) as the dependent variable. The pooled models include fixed effects for the different V-Dem variables, since the scaling of each of them could entail particularly low or high changes from year to year. In addition, the models cluster the standard errors at the level of cases (country/years), since these observations are subject to the same political developments and therefore not independent.

Changes in Coder Ratings

In line with the theoretical discussion above, we first test a narrow version of the availability heuristic, where coders are influenced by recent events in the country to be coded but use this information only to code variables that are closely related to these events. The first regression uses the event-based indicators for dramatic violent events (repressed protest and coups) to see if these events affect the coding of V-Dem variables related to violence and repression. In V-Dem, there are five of these variables; these include *v2cltort* (freedom from torture), *v2clkill* (freedom from political killings), *v2csreprss* (governmental repression of civil society organizations), *v2csrlgref* (governmental repression of religious organizations), and *v2meharjrn* (physical harassment of journalists).

The first two models in [Table 1](#) test how the violent repression of protest (Model 1) and the occurrence of a coup (Model 2) in between the two coding times affect the changes in the coder ratings. In line with our expectation, both

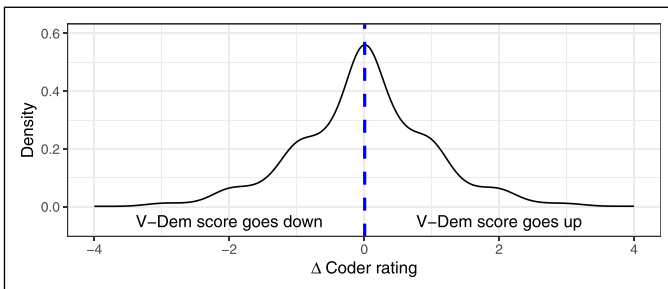


Figure 2. Distribution of differences in the coder ratings.

Table 1. Effect of Recent Events in Coded Country on Changes in Coder Ratings. OLS Models With Variable FEs and Standard Errors Clustered by Case (Country/Year).

	<i>Dependent variable</i>					
	Δ Coder rating					
	Repression-related variables			All variables		
	(1)	(2)	(3)	(4)	(5)	(6)
Repressed protest	-.148*** (.030)		-.075*** (.016)			
Coup or coup attempt		-.616*** (.079)		.072 (.046)		
FH downgrade ≥ 4					.023 (.028)	
FH status downgrade						.083 (.059)
V-Dem variable FEs	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SEs (ctr/year)	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,431	13,431	117,817	117,817	117,817	117,817
Adjusted R2	.008	.008	.002	.002	.002	.002

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

receive a negative effect, which indicates that the occurrence of these events is associated with a drop in the coder ratings. The magnitude of the effect is lower for repressed protest, where an event of this kind is associated with a drop in the coder assessments by only about .15 on the 0–4 scale. The effect of a coup is considerable, indicating that retrospective ratings for repression-related V-Dem variables are on average .6 points lower on the 0–4 scale if a country experienced a coup.

In Models 3–6 in [Table 1](#), we expand the analysis such that it includes all V-Dem variables as well as the alternative indicators of dramatic events based on Freedom House. Models 3 and 4 estimate the impact of the event-based predictors (repressed protest and coups) on changes in the ratings for *all* variables, Models 5 and 6 use the occurrence of an FH downgrade of at least 5 points (Model 5) or a status downgrade (Model 6) as predictors. The coefficients in these models are much smaller throughout. Only repressed protest is significant, but again in the direction we hypothesized. The estimate shows that the occurrence of physical repression against protesters leads to a

downgrade in the coder rating of about 0.08 on the 0–4 scale, which is not a lot given the average level of disagreement among coders (see Section A2 in the Appendix for the standard deviation within a case, which ranges between 0.5 and 0.8). In general, however, averaging over all V-Dem variables in the analysis, we fail to detect a strong and consistent effect of recent dramatic events on coder ratings beyond the repression-related variables.

Are there particular variables that are primarily affected by recent events? To provide a more fine-grained picture, the following repeats the above regression analysis, but using only one V-Dem variable at a time. For each of these variables, we fit a linear model using different operationalizations of recent events but without the variable fixed effects. Figure 3 visualizes the coefficients estimated in these models. To correct for multiple comparisons for each predictor, a Bonferroni correction is applied and the significance threshold is adjusted accordingly.

The results in Figure 3 present a mixed picture. Repressed protest seems to have a negative effect on many of the V-Dem variables, consistent with the result from the pooled analysis above. For coups, coefficients fluctuate considerably between positive and negative values, without a clear trend. This is also the case for the FH-derived predictors, although the coefficients are less volatile. Overall, together with the pooled analysis above, the analysis of the coder-level data provides some evidence for a narrow definition of availability bias, where recent events affect those variables closely related to these events. However, there is no conclusive evidence for a broader definition of availability bias, where recent events also affect other types of variables.

One concern with the first result is that it may be due to Bayesian updating by coders, and not necessarily due to availability bias. In other words, the new information that coders receive between *codingtime1* and *codingtime2* could reveal previously unknown knowledge about past characteristics of the country they are coding; for example, the occurrence of a military coup could lead coders to (rightfully) update their assessment of civilian control of the military. It is difficult to test whether availability or Bayesian updating drives the results we found. One way to do this is to look at the time that has passed between the case to be coded and the time the coding was actually carried out. Availability bias would suggest that codings for cases a long time ago should be more difficult, which is why coders should be influenced more by recent events *the more time has passed*. Bayesian updating suggests the opposite effect. Recent information about a country or a regime will be relevant primarily to assess dynamics that occurred *in the near past*, and much less so for those that occurred a long time ago.

Additional analyses in Appendix Section A3 test these competing predictions. We re-estimate the models in Table 1 but include an interaction effect with an additional variable that counts the number of years between the case (country/year) and the year in which the coding was done. Again, as

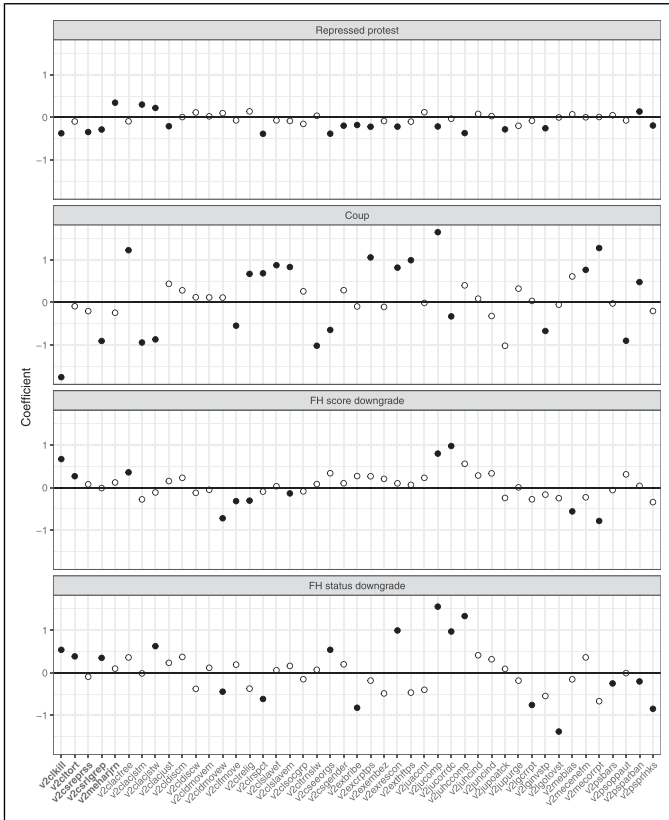


Figure 3. Effect of recent events on changes in coder ratings, estimated separately for each V-Dem variable. The plot shows the regression coefficients for the respective predictor (repressed protest, coup, score downgrade, and status downgrade), and solid dots indicate significant coefficients (Bonferroni correction applied for $m = 45$ comparisons). Repression-related variables on the left are highlighted in bold.

above, we see that the event-based predictors negatively affect the rating of repression-related variables. However, the interaction effect is negative, which indicates that the effect of recent dramatic events *increases* the more time has passed since the case to be coded. This is evidence consistent with availability bias and speaks against the conjecture that the results could be due to Bayesian updating by coders.

Changes in Democracy Indicators

According to the results in the previous section, the effects of recent political events on coder assessments are discernible only for some of the variables included in V-Dem. Does this matter for the aggregated democracy indicators generated from these coder ratings? Do these effects percolate through the aggregation process from the individual ratings to the final indicators? To find out, we repeat the first part of the analysis but use the V-Dem *democracy scores* instead of the raw coder ratings. The V-Dem project creates five “high-level” democracy indicators from the original coder ratings, through an aggregation process that is described in [Coppedge et al. \(2020\)](#). The indicators are “polyarchy,” “liberal democracy,” “participatory democracy,” “deliberative democracy,” and “egalitarian democracy.” Many cross-national analyses rely on these aggregate indicators.

As above, we observe changes in these indicators for the same country and year, which were produced in two consecutive years. We analyze how they are related to political events in the countries in question that occurred during this time. To see how this research design works, it is necessary to explain that V-Dem ratings for a particular case can change between versions. V-Dem releases annual versions of their data, each of which uses the most recent codings available for all the cases in the data (not just the ones added in the latest release). This is why the scores for a given country/year can change between versions, due to changes in the V-Dem measurement model or because additional coders have been added.

As already mentioned, there are almost no repeated codings for cases prior to 2005, which is why we restrict the analysis to cases for 2005 and later. We start with V-Dem version 6, which was coded in January 2016, and include all later ones until version 11, coded in 2021. As mentioned, changes in the democracy scores can happen for several different reasons, as, for example, adjustments to the V-Dem measurement model ([Pemstein et al., 2020](#)). For codings of 2005 and later, however, these changes can also be affected by new coders being recruited for new V-Dem versions, such that the aggregated scores are composed of the previous and the new codings. If later codings are affected by recent events as we have seen above, this should be reflected in the changes of the aggregated indicators.

To test this, we use the same research design as above and analyze pairs of V-Dem democracy scores, with scores for the same country and year produced at different points in time (in consecutive V-Dem versions). For example, one of the pairs in this dataset consists of the polyarchy scores for Egypt in 2016, released in 2018 (V-Dem version 8, value of .214) and in 2019 (V-Dem version 9, value of .202). We test whether the decrease in this indicator is related to dramatic events in between the two coding times. The analysis

Table 2. Effect of Events in Coded Country on Changes in Democracy Ratings.

	Dependent variable			
	Δ Democracy score			
	(1)	(2)	(3)	(4)
Repressed protest	-.003*** (.001)			
Coups or coup attempts		-.001 (.002)		
FH downgrade ≥5			-.002 (.002)	
FH status downgrade				-.003 (.002)
Dem. indicator FEs	Yes	Yes	Yes	Yes
Clustered SEs (ctr/year)	Yes	Yes	Yes	Yes
Observations	41,604	41,604	41,604	41,604
Adjusted R ²	.007	.006	.006	.006

Note. **p* < .05; ***p* < .01; ****p* < .001.

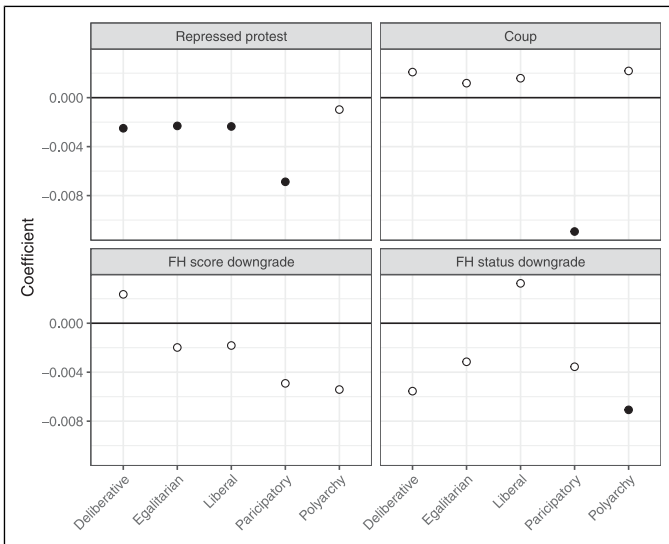


Figure 4. Effect of recent events on changes in democracy scores, estimated separately for each V-Dem democracy indicator. The plot shows the regression coefficients for the respective predictor (repressed protest, coup, score downgrade, and status downgrade), and solid dots indicate significant coefficients (Bonferroni correction applied for *m* = 5 comparisons).

includes variable (=democracy indicator) FEs, and robust standard errors are again clustered by case (country/year).

Model 1 in [Table 2](#) shows a similar result as in the pooled coder-level analysis above, with only repressed protest being associated with a downgrade in the democracy score. This means that the effect on the original coder ratings we found in the coder-level analysis is preserved in the aggregation process, despite the considerable amount of data processing that happens to transform the original scores into the high-level democracy indicators. However, this effect is very small in relation to the range of the democracy indicators (about .3% on the 0–1 scale). We then test the individual democracy indicators, with the results shown graphically in [Figure 4](#). Repressed protest has a consistent negative effect on four of the five indicators, in line with availability bias, but small in magnitude. The only other significant effects are found for coups and FH status downgrades, but only for one of the democracy indicators each, and again small in magnitude.

Conclusion

Given the proliferation of human-coded datasets in cross-national research, more attention has been paid to the potential biases that human coding can introduce ([Colgan, 2019](#); [Aron et al., 2023](#)). The coding of these datasets is a challenging task for humans, corresponding to decision making under uncertainty. One of the biases that is known to arise in these situations is the availability heuristic, where people draw on more easily available information when making decisions. For the coding of subjective assessments in cross-national datasets, this means that recent events in the country in question can affect coder decisions for historic cases. Using time-stamped coder-level data from the V-Dem project, this article tests this assumption. The analysis reveals evidence for a narrow version of the availability bias, where recent repressive events affect the coding of repression-related variables. There is little evidence for a broader version of this effect on variables unrelated to repression. The paper also showed that the narrow effect found in the coder-level analysis can be detected in the aggregated democracy indicators; however, the magnitude of this effect is too small to meaningfully affect research done with these indicators.

Overall, this is good news for the large number of cross-national coding projects that exist in the discipline, at least when it comes to the particular form of bias tested here. Still, while V-Dem is probably the only project that due to its time stamps makes analyses of this sort possible, it is at the same time a least likely case for this bias to occur. Unlike many other projects, V-Dem employs a number of mechanisms that reduce the amount of cognitive biases as much as possible. For once, V-Dem scores are aggregations of coder ratings recorded at different points in time. This means that if coders were affected by

recent events in different ways, they would likely not be systematic. Also, at least for more recent years covered in the dataset, the coding takes place immediately after the end of a year, which means that coders can draw on information that is still readily available to them. For the many instances of datasets that rely on subjective assessments *without* a V-Dem-like coding approach, there may be more reason to worry. Indicator creation has become more frequent primarily in the policy world, where many foundations, think-tanks, and NGOs collect data to rank countries according to their respective mission, worldview, and policy interest. Here, coders could respond to recent events in ways consistent with availability bias (but also several others), which could systematically affect ratings they produce. In a world where scientific evidence—in particular in political science—is coming under increased scrutiny from different political camps, it would be important to be fully transparent about the origins of supposedly “objective” ratings and how they could be driven by potential biases in the humans producing them.

Acknowledgments

The author acknowledges support from the V-Dem project (in particular, Staffan Lindberg, Johannes von Roemer, and Garry Hindle) by making their coder-level data available for the purpose of this project. The paper greatly benefited from comments at the V-Dem Brown Bag workshop, the Virtual Workshop on Authoritarian Regimes (organized by Quintin Beazer and Holger Kern, with Christian Houle, Michael Kenwick, Marcus Kurtz, and Arthur Spirling), the Annual Meeting of German Peace and Conflict Studies Association (Roos van der Haer and Clara Neupert-Wentz), the International Studies Seminar Series at the University of Konstanz (in particular, Michael Herrmann) as well as suggestions from Anne Meng and Andrew Little. The author also wishes to thank the German Research Foundation (DFG) for financial support (Research Grant #402127652).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Deutsche Forschungsgemeinschaft (grant number: 402127652).

ORCID iD

Nils B. Weidmann  <https://orcid.org/0000-0002-4791-4913>

Data Availability Statement

Replication materials and code can be found in the Comparative Political Studies Dataverse at <https://doi.org/10.7910/DVN/FKQI6M>.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Variable v2clrspct, question: “Are public officials rigorous and impartial in the performance of their duties?” (Coppedge et al., 2020, 164).
2. Variable v2lgcrrpt, question: “Do members of the legislature abuse their position for financial gain?” (Coppedge et al., 2020, 138).
3. To ensure compatibility with other annual codings, variables that do not follow an annual coding pattern (e.g., those related to elections) were also excluded. We also exclude variables that do not constitute normative assessments, such as the question about the dominant chamber in bicameral legislatures. See the appendix for the complete list of variables included.
4. This results in (non-systematic) measurement error, which is unlikely to bias the results we obtain from the aggregated codings.

References

- Arnon, D., Haschke, P., & Park, B. (2023). The right accounting of wrongs: Examining temporal changes to human rights monitoring and reporting. *British Journal of Political Science*, 53(1), 163–182. <https://doi.org/10.1017/s0007123421000661>
- Bertho, F. (2012). “Presentation of the institutional profiles database 2012.” http://www.cepii.fr/institutions/doc/IPD_2012_cahiers-2013-03_EN.pdf
- Clark, D., & Regan, P. (2021). *Mass mobilization protest data*. Harvard Dataverse. <https://massmobilization.github.io>
- Colgan, J. D. (2019). American bias in global security studies data. *Journal of Global Security Studies*, 4(3), 358–371. <https://doi.org/10.1093/jogss/ogz030>
- Coppedge, M., Gerring, J., Lindberg, S. I., Jan, T., Altman, D., Bernhard, M., Steven Fish, M., Glynn, A., Allen, H., Anna, L., Marquardt, K. L., Kelly, M. M., Paxton, P., Pemstein, D., Sigman, R., Staton, J., Wilson, S., Cornell, A., Alizada, N., & Medzihorsky, J. (2020). *V-dem dataset V10 codebook*. Varieties of democracy (V-dem) project.
- Economist Intelligence Unit. (2020). “Democracy index 2019: A year of democratic setbacks and popular protest.” Economist Intelligence Unit. <https://www.eiu.com/topic/democracy-index>
- Fariss, C. J. (2014). Respect for human rights has improved over time: Modeling the changing standard of accountability. *American Political Science Review*, 108(2), 297–318. <https://doi.org/10.1017/s0003055414000070>
- Freedom House. (2020a). *Freedom House Country Ratings*. <http://www.freedomhouse.org>.

- Freedom House. (2020b). *Freedom in the world 2020 methodology*. Methodological Documentation. <https://freedomhouse.org/reports/freedom-world/freedom-world-research-methodology>
- International Research & Exchanges Board. (2016). *Media sustainability index (MSI)*. <https://www.irex.org/resource/media-sustainability-index-msi>
- Kellam, M., & Stein, E. A. (2016). Silencing critics: Why and how presidents restrict media freedom in democracies. *Comparative Political Studies*, 49(1), 36–77. <https://doi.org/10.1177/0010414015592644>
- Little, A., & Meng, A. (2023). *Subjective and objective measurement of democratic backsliding*. SSRN. <https://www.ssrn.com/abstract=4327307>
- Lührmann, A., & Lindberg, S. I. (2019). A third wave of autocratization is here: What is new about it? *Democratization*, 26(7), 1095–1113. <https://doi.org/10.1080/13510347.2019.1582029>
- Marshall, M. G., Ted, R. G., & Keith, J. (2018). *Polity IV project: Political regime characteristics and transitions, 1800-2018*. Technical Report. <https://www.systemicpeace.org/inscr/p4manualv2018.pdf>
- Pemstein, D., Marquardt, K. M., Tzelgov, E., Wang, Y., Medzihorsky, J., Krusell, J., Miri, F., & Johannes, V. R. (2020). *The V-dem measurement model: Latent variable analysis for cross-national and cross-temporal expert-coded data*. Technical report V-Dem Project. https://v-dem.net/static/website/files/wp/wp_21_5th.pdf
- Peyton, B., Joseph, B., Shalmon, D., Martin, M., & Bonaguro, J. (2020). *Cline center coup D'état project dataset cline center for advanced social research. V.2.0.0 November 16*. University of Illinois Urbana-Champaign. https://doi.org/10.13012/B2IDB-9651987_V3
- Reporters without Borders (2020). *World press freedom index*. <https://rsf.org/en/ranking>
- Soroka, S., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences of the United States of America*, 116(38), 18888–18892. <https://doi.org/10.1073/pnas.1908369116>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- World Justice Project. (2020). *Rule of law index*. World Justice Project. <https://worldjusticeproject.org/our-work/research-and-data/wjp-rule-law-index-2020>

Author Biography

Nils Weidmann is Professor of Political Science and head of the “Communication, Networks and Contention” Research Group at the University of Konstanz. His research interests include political protest and violent conflict, the political impacts of new communication technology, and political methodology.