Civic Mobilization in Civil-Resistance Transitions Dataset

Version 1.0

CODEBOOK

Bob Lamb and **Arpitha Peteru**
Foundation for Inclusion
Co-Principal Investigators

Research Assistants:
M.K. Penn Boris, Mariel Cooksey, Marcelo Guerra,
Devon Cantwell, Breonna Stanton, and Adebanke Ilori

August 2021

1. Introduction

The Program on Nonviolent Action at the United States Institute of Peace (USIP) tasked the Foundation for Inclusion (FFI) with leading an eight-month initiative to collect and visualize data on civic mobilization in political transitions initiated through nonviolent action, or civil-resistance transitions (CRTs). The project's overall goal, according to USIP, was "to increase understanding of the role of nonviolent action in promoting sustainable peace and democracy."[1] This initiative sought to create a cross-national events dataset on civic mobilization and factors influencing democratic transition and consolidation, for use by researchers interested in understanding why civic mobilization continues through some transitions and not others, as well as the types of mobilization that are most effective in promoting a peaceful democratic transition and consolidation.

Version 1.0 of the *Civic Mobilization in Civil-Resistance Transitions (CM-CRT) Dataset* is based on original research identifying civic mobilization (CM) events since 1945 that took place during 72 of the 83 CRTs that USIP's Jonathan Pinckney had identified in prior research.[2] The dataset is provided in three separate comma-separated value (CSV) files:

- **crtCases (crtCases.csv)** includes 83 observations, one for each CRT case, along with basic data (e.g., start and end dates, description, key information about sources, etc.) and identification codes. A total of 72 of these cases were included in the data-collection process (see Section 2: Process).

- **crtEvents (crtEvents.csv)** is the main dataset, with data for 1,446 civic mobilization events, including dates, location, size, scale, actors, types of mobilization, targets, and a range of other attributes (e.g., the event's spontaneity, its coherence, its diversity, etc.). Events represent 72 of Pinckney's 83 CRTs cases, although only 67 distinct cases appear in this dataset because 5 cases were found during data collection not to have any CM events at all (Cameroon 1961, Central African Republic 1993, Madagascar 1993, Lesotho 2002, and East Timor 2002). An additional 145 valid CM events (across 22 cases) are excluded because their variables were not coded (due to a mix of human error and time and resource constraints) but are included in the crtCandidates dataset.

[1] USIP Program on Nonviolent Action, "Request for Proposals: Civic Mobilization in Civil Resistance Transitions Data Collection Project," United States Institute of Peace, October 16, 2020.

[2] Jonathan Pinckney, *From Dissent to Democracy: The Promise and Perils of Civil Resistance Transitions,* Oxford University Press, 2020; Jonathan Pinckney, *When Civil Resistance Succeeds: Building Democracy After Popular Nonviolent Uprisings,* International Center for Nonviolent Conflict, 2018.

- **crtCandidates (crtCandidates.csv)** includes 59,334 observations, each of which is a "candidate event," generally a news article or report found via queries of news and events datasets.[3] Because some information sources identified more than one candidate event, a total of 54,374 unique sources were manually screened (and hundreds more informally consulted) over five and a half months to find the CM events that would be coded and included in crtEvents. Candidate events for 82 of the 83 CRTs appear in this file (Cameroon 1961 had no candidate events).

Each event in the crtEvents dataset represents an instance of *civic mobilization* (CM) by some group of people. For the purposes of this dataset, a CM event was defined as an Actor taking an Action, that is, some entity that could reasonably be described by one of the Actor categories undertaking some set of activities that could reasonably be described by one of the Action categories (for descriptions of these categories, see Event Variables in Section 4: Variables in the Events and Candidates Datasets).

- The Actor ontology is derived from a mix of logical categories and the inputs of participants in a USIP workshop involving experts pursuing research on nonviolent movements and political transitions.[4]

- The Action ontology is based on logical categories of influence derived from the field of strategy, with subcategories deriving from inputs of workshop participants and a review of the ontologies of Gene Sharp and the CAMEO codebook, which is the basis for Phoenix 1.3, NAVCO 3.0, and other events datasets in this field. The CM-CRT 1.0 dataset does not directly use the CAMEO categories or codes, but its subcategories are designed to be easily mapped onto CAMEO.[5]

Building the dataset required defining the unit of analysis precisely (e.g., Pinckney's CRTs had start and end *years,* but we needed start and end *dates),* validating and revising the data-collection protocol, querying existing news and events datasets to find candidate events, screening those candidates to identify CM events, collecting

[3] The main sources were LexisNexis and the Phoenix events dataset; see Scott L. Althaus, Joseph Bajjalieh, John F. Carter, Buddy Peyton, Dan A. Shalmon, *Cline Center Historical Phoenix Event Data Variable Descriptions (v1.3.0),* University of Illinois, May 4, 2020.
[4] "Civil-Resistance Transitions Expert Workshop," USIP-FFI Video Conference, February 8, 2021.
[5] Gene Sharp, *The Politics of Nonviolent Action,* Porter Sargent Publishers, 1973; Philip A. Schrodt, *CAMEO: Conflict and Mediation Event Observations Event and Actor Codebook, Version 1.1b3,* Pennsylvania State University, March 2012; Althaus et al, *Cline Center Historical Phoenix Event Data*; Orion A. Lewis, Erica Chenoweth, and Jonathan Pinckney, *Nonviolent and Violent Campaigns and Outcomes 3.0: Effects of Tactical Choices on Strategic Outcomes Codebook (N.A.V.C.O. 3.0),* University of Denver, July 28, 2016.

and entering data each of the CM events, then validating and cleaning the data to present it in its final form.

Challenges

Completing all of these steps in the limited time available was not without its challenges. The budget allowed for three research assistants (RAs) working full-time for about four months, which turned out not to be nearly enough staffing to identify and code events for all 83 cases.

Creating an events dataset begins with an unavoidable unknown: how many events will ultimately need to be coded. We could not know at the beginning of the project whether the 83 cases would have an average of 10 CM events, an average of 50 CM events, or perhaps more. Nor could we know what the range would be. We knew some cases would have no CM events, but not *how many* cases would. And we knew other cases would have many CM events, but not what the upper limit would be (e.g., whether the largest case would have 40 events or 400 events). Finally, we could not know, before validating the protocol, how complex each case would be and therefore how much time it would take to screen candidate events to identify then code the CM events. Some cases would clearly be more complex than others, as when a large number of events were happening in multiple places at once or over multiple days or weeks, or when most Actors were peaceful but a subset were involved in rioting or street fights.

At the end of the project, those figures are known.

The research team was able to complete an average of 18 events per day once training and protocol validation was complete. But the range was wide. Some cases took less than a day to complete (e.g., Cameroon 1961), while others took a single RA well over two weeks (e.g., Nigeria 1993).

Accounting for the 1,446 CM events that were fully coded and the 145 CM events that were identified but not coded, the cases in the dataset averaged about 22 CM events each, ranging from no events (in five cases) to 307 events (in Egypt 2011).

It remains unknown how many CM events would need to be coded for the 11 cases excluded from the dataset, but we do know they were excluded specifically because they were among the cases with the largest media coverage. The dataset has an average (mean) of about 621 candidate events per case and on average only about 3.6 percent of them were identified as CM events. By contrast, the cases excluded from the dataset averaged 1,327 candidate events (more than double the dataset's average). At the normal rate, we could expect those 11 cases would have an average of 47 CM events each (again, more than double the average), or at least 500

additional events to code. At 18 events per day, more than 28 additional work days (or 84 full-time-equivalents, for just the RAs) would have been needed in the budget to complete the dataset.

In addition to the uncertainties related to the scale of the project, the validation phase took significantly longer than expected, with a total of nine rounds of validation tests required before the team reached an acceptable level of inter-coder reliability. The main challenge was that a CRT is a new category of knowledge, introduced just a few years ago. While Pinckney's work has contributed a great deal of new understanding to the field, it took some time for the team to become conversant with the subtleties of what counts as civic mobilization, how a civil-resistance movement differs from a civil-resistance transition, and how to determine if two events are distinct or extensions of each other.

One of the greatest challenges was defining a specific end date. Pinckney had identified start and end years for all 83 CRTs, but to query the news and events datasets to collect candidate events, we needed dates down to the day (for our solution to that, see Section 2: Process).

Finally, this research was taking place in the context of an ongoing global pandemic, which has proven to be highly deadly, disruptive, and disproportionate. A majority of the staff working on this project personally experienced significant disruptions requiring them to reduce hours or leave the team entirely. Hiring and training new team members midway through the project disrupted an already tight schedule and heavy workload.

Additional challenges are noted throughout this codebook as they relate to specific processes (see Section 2: Process) or specific variables (see Sections 3 and 4).

Biases

The challenges described above reduced the number of cases and events that could be ultimately included in the dataset. The most direct implication of that missing data is a bias in the dataset, in favor of events associated with smaller cases over events associated with the much larger cases excluded from data collection.

Even for the cases and events included in the dataset, however, these challenges complicated the task of addressing biases in media coverage to the degree we knew would be important to address from the start.

Most events datasets suffer from the same set of biases: they tend to derive from primarily English-language news articles and other information sources, published in North America and Europe, targeted to middle class readers and economic or

policy elites, and written predominantly by white authors of European ancestry. The experiences, realities, and political contexts that these authors are rooted in, therefore, tend to differ markedly from those of the communities and individuals they feature in many of their stories, giving way to reporting bias through an observer lens.

Biases, therefore, can be found both in *what events* are selected for coverage and in *how* those events are covered:

- There is much more coverage of events in North America, Europe, and countries that represent strong strategic interests for North American or European countries than in politically less connected or economically poorer countries.
- A well-documented implicit bias drives a tendency for many media reports to frame actions by historically marginalized individuals and communities in negative ways (e.g., riots, looting) and similar actions by white individuals and communities in positive ways (e.g., protesting, celebrating).[6] This affects the interpretation of articles when determining what type of action is taking place.
- More generally, certain types of events tend to be covered more, or more sensationally, than others. Local reporters or those with strong familiarity with the given context can also be subject to such influences, as well as local prejudices, a bias favoring or opposing the status quo, personal experiences or connections to certain industries or political organizations, ideological commitments, and social networks. Consider, for example, how the Hong Kong protests that started in 2019 were covered by journalists living in Western democracies, compared to those from mainland China, compared to coverage from Hong Kong media.[7]

As a key example of how this applies to the CM-CRT dataset: the Action subcategory "24 Intervene" (nonviolent) is defined as "physically obstructing, disrupting, preventing, occupation, and other forms of physical noncooperation" while "61 Seize or Damage Property" (violent) involves "taking, damaging, or destroying property, or otherwise rendering it unusable" (see Section 4: Variables in the Events and Candidates Datasets). It is not difficult to imagine two different reporters covering the same event: one frames it as a blockade by peaceful

[6] See, for example, Travis L. Dixon, *A Dangerous Distortion of Our Families: Representations of Families, by Race, In News and Opinion Media,* Color of Change, December 2017.
[7] Beatrice Gao, "Journalistic Bias In Hong Kong Protest Reporting: Q&A With Stanley Rosen," USC U.S.-China Institute, December 5, 2019, https://china.usc.edu/journalistic-bias-hong-kong-protest-reporting-qa-stanley-rosen.

demonstrators getting attacked by counterprotestors; the other frames it in terms of vandalism and street fighting. Users are cautioned to consider how such implicit biases and framing might have affected the coding in this dataset, for these two subcategories in particular and for others as well.

Some of these biases could have been partly overcome by directly reviewing archival event footage. For the reasons cited earlier, however, we did not have access to video footage for most events, beyond some videos of some fairly recent events that were available on YouTube.

In short, the CM-CRT dataset suffers from the same biases as other events datasets because its events derive from the same biased sources, primarily LexisNexis and Phoenix. LexisNexis has articles from hundreds of newspapers, newswires, broadcast transcripts, magazines, blogs, and journals from around the world, but most are in English and European languages carrying a heavy Western gaze. Phoenix derives its events from LexisNexis, the *New York Times,* the *Wall Street Journal,* BBC Monitoring's *Summary of World Broadcasts,* and the Central Intelligence Agency's *Foreign Broadcast Information Service*, which also have an unambiguous Western bias. (For our efforts to at least partially offset these biases, see the "key sources" subsection of Section 2: Process.)

Cases and Events

The following table summarizes the dataset. The 83 CRTs are listed in rough chronological order by start year (here and in the dataset they are sorted by *ffiID;* see Identification Codes in Section 4: Variables in the Cases Dataset). The "CRT Case" column is the main identifier used to distinguish cases from one another (some countries have more than one transition in the dataset, as indicated by their start and end dates). The "Included CRT Case" column indicates whether the case made it into the dataset (72 are included). "Included CM Events" indicates the number of civic-mobilization events coded for each case (1,446 total), while "Excluded CM Events" indicates the number of CM events that did not get coded for that case (145 total). The last column, "Candidate Events," shows the total number of events the research team screened for the purpose of identifying which ones counted as actual CM events (59,334 total).

CRT Case	Country	Start Date	End Date	Included CRT Case	Included CM Events	Excluded CM Events	Candidate Events
GTM1945	Guatemala	1945-03-15	1946-12-31	Yes	16	0	31
IND1947	India	1947-08-15	1947-12-31	Yes	12	0	112
HTI1956	Haiti	1956-12-12	1957-10-22	Yes	32	1	48
GHA1957	Ghana	1957-03-06	1957-12-31	Yes	6	0	22

COL1958	Colombia	1958-05-04	1959-12-31	Yes	23	1	37
VEN1958	Venezuela	1958-01-23	1959-02-13	Yes	16	1	73
COD1960	Democratic Republic of Congo	1960-06-30	1961-08-02	Yes	3	0	47
KOR1960	South Korea	1960-04-26	1961-05-16	Yes	29	0	46
CMR1961	Cameroon	1961-10-01	1961-12-31	Yes	0	0	0
DOM1962	Dominican Republic	1962-01-16	1966-06-01	Yes	36	1	172
ZMB1964	Zambia	1964-10-24	1965-11-11	Yes	1	0	8
MWI1964	Malawi	1964-07-06	1965-12-31	Yes	2	0	3
SDN1964	Sudan	1964-11-15	1965-05-08	Yes	3	0	5
MDG1972	Madagascar	1972-05-18	1975-12-21	Yes	3	0	6
THA1973	Thailand	1973-10-15	1975-01-26	Yes	25	0	61
PRT1974	Portugal	1974-04-25	1976-06-27	Yes	44	4	173
GRC1974	Greece	1974-07-24	1976-12-31	Yes	18	0	63
BOL1979	Bolivia	1979-08-08	1985-07-14	Yes	112	11	1075
IRN1979	Iran	1979-02-11	1981-10-02	No	NA	NA	2044
ARG1983	Argentina	1983-12-10	1984-12-31	Yes	12	0	574
URY1984	Uruguay	1984-11-25	1986-12-31	Yes	6	7	211
BRA1985	Brazil	1985-03-15	1988-10-05	Yes	8	0	174
SDN1985	Sudan	1985-04-06	1986-04-12	Yes	37	0	1725
HTI1986	Haiti	1986-02-07	1995-12-31	Yes	15	9	821
PHL1986	Philippines	1986-02-25	1988-12-31	Yes	12	0	511
KOR1987	South Korea	1987-05-25	1989-12-31	No	NA	NA	427
PAN1989	Panama	1989-12-20	1991-01-27	Yes	22	0	1566
CHL1989	Chile	1989-12-14	1990-03-11	Yes	11	0	190
DDR1989	East Germany	1989-12-01	1991-03-15	Yes	21	0	193
POL1989	Poland	1989-06-04	1991-10-27	Yes	35	0	801
CSK1989	Czechoslovakia	1989-12-10	1991-06-21	Yes	11	2	260
HUN1990	Hungary	1990-03-25	1990-08-03	Yes	3	0	383
BGR1990	Bulgaria	1990-06-10	1991-10-13	Yes	56	0	1360
BEN1990	Benin	1990-12-02	1992-12-31	Yes	15	0	194
MNG1990	Mongolia	1990-03-12	1993-06-06	Yes	26	1	734
BGD1990	Bangladesh	1990-12-06	1991-02-27	Yes	12	0	110
NPL1990	Nepal	1990-04-08	1991-05-12	Yes	40	1	288
ALB1991	Albania	1991-03-31	1992-03-22	No	NA	NA	1270
SVN1991	Slovenia	1991-07-07	1991-12-23	Yes	2	0	300
SUN1991	Soviet Union	1991-12-21	1992-03-31	Yes	5	0	191
EST1991	Estonia	1991-08-20	1992-09-20	Yes	7	0	447
LVA1991	Latvia	1991-08-21	1993-06-06	Yes	11	0	621
LTU1991	Lithuania	1991-09-06	1992-11-15	Yes	12	0	871

BLR1991	Belarus	1991-08-25	1997-12-31	No	NA	NA	1439
GEO1991	Georgia	1991-04-09	1995-11-05	No	NA	NA	1188
MLI1991	Mali	1991-03-26	1993-12-31	Yes	23	0	619
NER1991	Niger	1991-10-26	1993-03-27	Yes	22	0	1202
ZMB1991	Zambia	1991-11-02	1992-12-31	Yes	17	0	408
KGZ1991	Kyrgyzstan	1991-08-31	1992-12-31	Yes	2	0	121
GUY1992	Guyana	1992-10-05	1993-12-31	Yes	4	0	102
THA1992	Thailand	1992-05-24	1995-07-02	Yes	27	2	2033
NGA1993	Nigeria	1993-01-04	1993-11-17	Yes	38	2	1294
CAF1993	Central African Republic	1993-09-19	1993-12-31	Yes	0	0	5
MDG1993	Madagascar	1993-03-27	1994-12-31	Yes	0	0	74
MWI1994	Malawi	1994-05-17	1995-05-17	Yes	11	0	255
ZAF1994	South Africa	1994-05-10	1995-12-31	Yes	28	0	279
NGA1999	Nigeria	1999-02-27	1999-05-29	Yes	23	2	2078
IDN1999	Indonesia	1999-10-20	2000-12-31	No	NA	NA	1619
MEX2000	Mexico	2000-07-02	2000-12-01	Yes	49	3	2325
PER2000	Peru	2000-11-20	2001-06-03	Yes	16	1	2055
HRV2000	Croatia	2000-01-03	2000-02-18	Yes	1	0	592
SRB2000	Serbia	2000-10-07	2002-12-30	Yes	27	1	3570
SEN2000	Senegal	2000-03-19	2000-04-01	Yes	2	0	89
GHA2000	Ghana	2000-12-28	2000-12-31	Yes	1	0	66
LSO2002	Lesotho	2002-05-25	2002-06-04	Yes	0	0	66
MDG2002	Madagascar	2002-07-05	2002-12-15	Yes	6	0	302
TMP2002	East Timor	2002-05-20	2002-05-20	Yes	0	0	7
GEO2003	Georgia	2003-11-23	2004-01-04	Yes	3	0	883
LBR2003	Liberia	2003-08-11	2006-01-16	Yes	21	1	3659
UKR2004	Ukraine	2004-12-26	2006-08-04	No	NA	NA	1844
LBN2005	Lebanon	2005-04-13	2005-06-19	Yes	10	0	649
KGZ2005	Kyrgyzstan	2005-04-04	2005-08-14	Yes	21	0	473
NPL2006	Nepal	2006-05-18	2008-07-23	No	NA	NA	1126
PAK2008	Pakistan	2008-02-18	2008-09-09	No	NA	NA	1205
TUN2011	Tunisia	2011-01-14	2012-12-31	Yes	32	3	416
EGY2011	Egypt	2011-02-11	2014-06-08	Yes	218	89	4013
YEM2011	Yemen	2011-11-23	2012-02-25	Yes	43	1	1302
UKR2014	Ukraine	2014-02-24	2014-06-07	No	NA	NA	1788
BFA2014	Burkina Faso	2014-11-01	2016-05-22	Yes	23	0	1197
THA2014	Thailand	2014-05-07	2014-08-21	Yes	10	0	50
GMB2016	The Gambia	2016-12-01	2018-04-12	Yes	5	1	31
COD2016	Democratic Republic of Congo	2016-11-14	2016-12-31	Yes	3	0	6

ARM2018	Armenia	2018-05-08	2019-01-14	No	NA	NA	656
			TOTAL	**72**	**1,446**	**145**	**59,334**

Codebook Structure

This codebook is structured as follows. The next section describes the overall process used to create the dataset. After that are the two main sections, which discuss all of the variables in the dataset along with their codes, categories, and criteria. The first introduces the variables in the Cases dataset (crtCases.csv), whose observations are at the case level. The second introduces the variables in the Events (crtEvents.csv) and Candidates (crtCandidates.csv) datasets, which have the same data structure. An Appendix provides the complete data-collection instrument (the data-entry form used to enter detailed data about each event).

Accompanying the CM-CRT dataset is a set of "lab notebooks" (technically, "R markdown notebooks") describing the coding, validation, and cleaning processes used to generate the dataset. These notebooks come in two file formats: HTML for casual users to read on a Web browser, and Rmd for users who want to test the embedded code (R, Python, and Bash) in the RStudio integrated development environment. The `00_intro_READ_ME.nb.html` file gives a more detailed introduction to the notebooks.

2. Process

The overall data-collection protocol was as follows.

1. Identify start and end dates for all cases

A civil-resistance transition begins on the day a civil-resistance movement succeeds. For example, General Marcos Pérez Jiménez was one of the leaders of a 1948 military coup against Venezuela's democratically elected government and rose to dictatorial rule by the early 1950s. Opposition movements begun underground in 1956 began leading public demonstrations against his regime in 1957. After a failed coup attempt in early 1958 revealed deep divisions within the military, the opposition movement gained momentum with the public and within the military, which ultimately led to Jiménez stepping down on January 23, 1958. The civil-resistance *transition* began on that day, so that is the start date of the Venezuela 1958 case.

Many cases have that same pattern; it is not that the transition *from* an autocratic regime is always due exclusively to a nonviolent movement (many were, but some began with a coup, for example), but that the nonviolent movement played a significant role in setting the conditions for regime change. The start date for just about every case can be identified by the day the leader or leaders of a repressive regime leaves power (voluntarily or otherwise). At that moment, there is a *possibility* of a transition to democracy. More specifically, the process of identifying start dates began with the assumption that the start year Pinckney had identified for each CRT was the correct year, and the team reviewed that year's political history in that country to identify the event that could most credibly be interpreted as the end of an autocratic regime; when found, that day was considered the start date.[8]

The end dates were significantly more difficult to identify. While some transitions clearly succeeded and others clearly failed, many fell into a sort of stasis, never quite becoming a democracy, never quite backsliding to autocracy. In some cases, that stasis lasted many years or even decades.

In Pinckney's work, he identified end *years* based on specific events or on patterns in the country's polyarchy scores. To find articles describing potential CM events, we needed a clear end day (or at least month) that we could use in automated queries of LexisNexis and Phoenix. Early experiments with different approaches led to very different choices of end date by different research assistants (RAs). Reasonable inter-coder reliability was achieved by using the following protocol:

1. Determine if the CRT ended abruptly or violently. The main resource for this was Chapter 4 of the Geddes *Autocratic Regimes Code Book* by Barbara Geddes, Joseph Wright, and Erica Frantz (2014), which provided clear dates of events indicating a CRT has failed (due to a coup, self-coup, assassination, bloodless coup, etc.).[9] Other sources included the Phoenix dataset and general research to find events clearly indicating an abrupt end. The day the incumbents (i.e., the individuals, party, or regime who took power at the beginning of the CRT) definitively lost power was considered the end date of the CRT.

2. For a CRT that did not end abruptly, determine if it was generally a successful case of democratic consolidation. The main resource for this was the *Varieties of Democracy* (V-Dem) dataset, which provided clear dates of

[8] One particularly useful source was the *Global Nonviolent Action Database,* Swarthmore College, https://nvdatabase.swarthmore.edu.
[9] Barbara Geddes, Joseph Wright, and Erica Frantz, *Autocratic Regimes Code Book, Version 1.2,* Pennsylvania State University, 2014.

democratic elections, as well as the Phoenix dataset and general research.[10] The day of the last major democratic election that took place during Pinckney's end year was considered the end date of the CRT.

3. For any CRT with no clear end, use December 31 of the first year of a three-year period of relative stability in that country's polyarchy scores (in the V-Dem and Polyarchy 2 datasets[11]), with a bias toward using December 31 of Pinckney's end year for that case.

Users of this dataset might reasonably disagree with some of the end-date selections for these cases (or some start dates, for that matter). A series of R markdown notebooks is included with the dataset detailing all of the processes and code used to query LexisNexis and Phoenix to identify candidate events to screen for CM events to include in the dataset.[12] Using different specifications to replicate the queries and event-selection process should be a straightforward matter for users adept with the R and Python programming languages.

2. Identify two "key sources" for each CRT

In an effort to at least partially balance the biases discussed earlier (see Biases in Section 1: Introduction), we endeavored to identify at least two or three rich sources of information about each CRT case from online sources so that gaps in events data could be manually filled through desk research.

An attempt was made to find sources by authors from the country or region (either in English or in one of the local languages spoken by FFI staff) or at least from credible Western authors who based their work on field research and direct knowledge of the region. Ideally, such sources would include reports from (local) academic research centers or think tanks, books or journal articles by credible (local) scholars, primary-source documents from archival collections, or publications targeted toward diaspora communities, all with coverage of the time period of the CRT (rather than the nonviolent movement leading up to it).

---

[10] Michael Coppedge, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, Agnes Cornell, M. Steven Fish, Lisa Gastaldi, Haakon Gjerløw, Adam Glynn, Allen Hicken, Anna Lührmann, Seraphine F. Maerz, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Pamela Paxton, Daniel Pemstein, Johannes von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundtröm, Eitan Tzelgov, Luca Uberti, Yi-ting Wang, Tore Wig, and Daniel Ziblatt, *V-Dem Codebook v11.1,* Varieties of Democracy (V-Dem) Project, 2021.
[11] Ibid.; Tatu Vanhanen, *Polyarchy 2,* International Peace Research Institute (PRIO), 2000, https://www.prio.org/Data/Governance/Vanhanens-index-of-democracy.
[12] The R markdown notebooks total about 35 pages of text plus more than 3,000 lines of R code (plus some in Python and Bash).

Many of the "key sources" documents we identified did meet these criteria, but most did not. Several challenges made it difficult to identify independent sources capable of counterbalancing biases in our main sources.

- For most cases, there was significantly more research on the revolution than on the transition that followed, reflecting the interests of scholars, policy advisers, and funders.
- More generally, the research team had to rely exclusively on materials available online. Data collection took place during a global pandemic, when lockdowns made it impossible to physically visit libraries to engage in book or archival research as we would have preferred (e.g., the Library of Congress closed to visitors in March 2020 and had not opened by mid-2021), forcing the team to rely almost exclusively on sources that had been digitized and made available on the Internet. Physical access to books and archival holdings (e.g., of local diaspora newspapers) would have made this research significantly richer.
- Finally, time and resource constraints (see Challenges, above) made it difficult to spend as much time as the team would have preferred on finding less biased sources of information.

The co-PIs are certain that a number of relevant events are not included in this dataset (especially for events before 1977, but more recently as well) and a number of variables will have missing values, as a result of these challenges.

3. Query Phoenix and LexisNexis to identify candidate events

With the start and end dates identified, we were now able to query the main information sources.

- For the Phoenix events dataset, we wrote R scripts that selected, for each CRT, all events that took place between the start and end dates and for which Phoenix's *target_root* variable was the country in question but excluded events labeled as "irrelevant" under the *conflict* variable.
- For LexisNexis, we edited Python scripts provided by the company that returned all articles that were published between the start and end dates in the news, regional, or international sections of any news publication and that named the country itself in the title of the article.

The results were combined into a single list of "candidate events" for each case and shared with the RAs in a series of Google Sheets (referred to in this codebook as the "include-exclude sheets" or "IE sheets"). The IE sheets were offered in batches

(several cases in one Google Sheet) and each sheet contained multiple worksheets (one CRT case per worksheet, or tab). Each worksheet included basic data, including the title of each article, a link to the full text for any events where reading the title alone would not be enough to determine if the event was a civil-mobilization event, and a link to a Google Form where all the remaining data could be entered for any article/event selected for inclusion in the dataset.

4. Review the candidate events quickly

The process for screening the candidate events to identify which were civic mobilizations was as follows.

- The first RA *(coder1)* started by selecting a case in one of the include-exclude sheets, putting their staff ID in the *coder1* column, and reading the information in the title variable and other supplemental variables. If these variables clearly identified a domestic actor engaging in an act of civic mobilization, they were to code the event as a candidate for inclusion *(include1 = 1)*. If it was clearly irrelevant, they coded it for exclusion *(include1 = 0)*. If it was ambiguous, they either coded it as ambiguous *(include1 = 99)* and returned to it later, or they immediately read the full text of the article to determine whether it described an example of civic mobilization. If it was still unclear after doing so, or the RA was not fully confident in their suspicion, they were instructed to think a little more about it and decide on either a 1 or a 0 for *include1,* then put a 1 in the *check* column to signal to the other RAs that they needed a second opinion.

- A second RA *(coder2)* then scanned through the *check* column to pick a row with a 1, entered their staff ID in the *coder2* column for that row, then reviewed the title, other variables, and article and potentially also did some additional independent research, then decided whether they agreed with the first RA's include-exclude decision (i.e., the value in *include1)*. If the second RA agreed, they simply deleted the 1 in the check column (but left their staff ID code in the *coder2* column to indicate that event had initially been an edge case). If the second RA disagreed with the value in *include1,* they simply changed the 1 in the *check* column to a 2 (again keeping their staff ID code in *coder2)*. They were instructed never to change anything in the *include1* column. The co-PIs then took responsibility for reviewing any articles with a 2 in the *check* column to make a final decision, which was made by putting the co-PI's staff ID code in the *coder1* column and either 1 or 0 under *include1* as the final decision on whether to include that event.

- Finally, all RAs were instructed to consider the possibility that any real-world event might be represented by more than one row (i.e., multiple articles about the same event; more than one *eventID* for the same event; etc.). In that case, they were to use the *combineWith* field to enter the *eventID* of the first article about that real-world event (usually the row with the lowest *eventID* number). In addition, if any event took place on more than one day (e.g., recurring events, such as an every-Thursday protest), they were to use the *endYear, endMonth,* and *endDay* columns to indicate the end date of the event (i.e., the last day on which the event occurred). If they encountered an event in which the title suggests the state repressed citizens, political opponents, etc., they were to determine whether that state action was in fact a reaction to an instance of civic mobilization; if so, they were to create a new event about that civic mobilization, code it as a 1 in *include1,* and link it to the state response event via the *combineWith* and *repression* variables.

The output of this step was a fully coded set of include-exclude sheets indicating which candidate events would be included in the dataset. Implicitly, this step identified all of the false positives (the candidate events identified by the Phoenix and LexisNexis queries that turned out not to be CM event (i.e., coded for exclusion *(include1 = 0)* and true positives (the candidate events coded as *include1 = 1.*

5. Review "key sources" to identify false negatives

As the previous step identified and excluded false positives, the final step in event identification was to identify and include false negatives: CM events that had not been captured by the LexisNexis and Phoenix queries.

The first RA *(coder1)* was instructed to review the "key sources" documents (see Step 2, above) and any additional sources as needed to see if any additional real-world events took place that had not been captured in Steps 3 and 4. If any were discovered, RAs were instructed to create a new row in the worksheet for that CRT, enter values for all variables, and create a new *eventID* variable (incrementing the highest *eventID* by 1) to ensure all events had a unique identifier.

The second RA *(coder2)* was instructed to quickly scan the "key sources" documents as well to determine whether they agree with the first RA's "include" determination. If so, they simply put their staff ID in *coder2.* If not, they were to put a 2 in the check column so a co-PI could review it and make a final determination.

6. Come to consensus on final list of events

The RAs were instructed to coordinate to validate the final list of *include1* events for each CRT, with the exception of the events tagged with a 2 under *check,* which were the responsibility of the co-PIs. Validation meant that they agree the protocol was followed (i.e., these instructions) and the final list is a fair and complete representation of the civic mobilization events that happened during that CRT, given the sources available and the process required.

7. Collect data for each included event

Finally, for each event identified and coded for inclusion in the dataset, an RA would select a custom link to the data-entry form (a Google Form set to pre-fills basic data for each event) where they were then required to make decisions about the values for every variable in the dataset. They were permitted to do this step at any point during Steps 4, 5, or 6; they did not need to wait until all events were finalized before doing so, but most made the include-exclude decisions first before coding the included events.

The data-entry form contained the entire codebook (see Appendix), so the criteria and categories for all variables were easily and immediately accessible. RAs were instructed to look for additional articles (via LexisNexis or a Google search) for any variables the initial article provided no information about, and they were given plenty of space to identify any additional sources. They were instructed not to spend a lot of time looking for additional sources, however, because they found dramatically diminishing returns on time spent hunting for additional information: their experience was that if they didn't find information for a variable after searching for 5 to 10 minutes, they were highly unlikely to find that information at all.

As each form was completed and submitted, the data was saved on a separate Google Sheet that, along with the include-exclude sheets, provided the raw data for the dataset. These were processed, cleaned, and saved as three separate comma-separated value (CSV) files: `crtCases.csv` with 83 observations (one for each CRT case identified for Pinckney), `crtEvents.csv` with 1,446 observations (one for each CM event), and `crtCandidates.csv` with 59,334 observations (one for each candidate event or CM event).

3. Variables in the Cases Dataset (crtCases)

The first of the three CSV files that make up the dataset is **crtCases.csv**, which has 83 observations, each of which is an individual civil-resistance transition. Its main

unique identifier (i.e., it is not used for more than one observation) is *crtID* (e.g., East Germany's 1989 transition is DDR1989). It shares some variables in common with the other two CSV files. As the variables are stored as CSV files, data types (e.g., string, integer, double, etc.) are not formally defined here, but in some cases the variable descriptions below do mention data type.

Identification Codes

Several schemes were used, for different purposes, to identify different cases, countries, and events and are included in the dataset to facilitate matching with key external datasets and internal identification purposes.

cowID
The Correlates of War country identification code is a two- or three-digit code (e.g., Guatemala is 90, Tunisia is 616) that Jonathan Pinckney's original CRT dataset used for country codes.[13]

ffiID
This is an eight-digit code FFI devised to sort cases in rough chronological order. The first four digits represent the year the CRT started (e.g., Bolivia's transition began in 1979); the second four digits are the *cowID* with one or two leading zeros (e.g., Bolivia's ffiID is 19790145).

phxID
This is the three-letter country code used in the Phoenix dataset (e.g., Thailand is THA, Liberia is LBR). For events deriving from the Phoenix dataset, this variable picks up the value in Phoenix's *target_root* variable.

crtID
This is the main identification code used to identify *cases* (individual civil-resistance transitions) in the CM-CRT dataset. It is a seven-character alphanumeric code in which the first three characters are the phxID and the last four are the year the CRT started (e.g., Ghana's 1957 transition is represented by GHA1957, while its transition in 2000 is GHA2000).

country
This is a character string standardizing the names of the countries appearing in the dataset. Some countries experienced more than one CRT since 1945 (e.g., Haiti in 1956 and 1986; Madagascar in 1972, 1993, and 2002).

[13] "COW Country Codes," The Correlates of War Project, https://correlatesofwar.org.

Dataset Variables

included

This is a Boolean variable (1 or 0) that indicates whether the CRT case is or is not included in the dataset (1 if included, 0 if not). Of the 83 cases identified by Pinckney, 72 are included in the dataset. Some cases had an unexpectedly large number of events for which data needed to be collected, so not all cases were able to be coded given the time and resources available. The excluded cases are almost entirely extremely large cases (i.e., the number of candidate events initially found via LexisNexis exceeded 4,000 observations), so that bias needs to be accounted for in any analysis of the dataset.

partial

This is a Boolean variable (1 or 0) that indicates which CRT cases (that are included in the dataset) do *not* have all of their CM events coded (1 if only partially coded, 0 if all CM events are coded). Two of the cases in the dataset (Egypt 2011 and Haiti 1986) do not have data for all of the events that had been identified for inclusion in the dataset; because of the unexpectedly large number of events for some cases (see *included,* above), time and resources ran out before all CM events for these two cases could be coded. Of the cases included in the dataset, Egypt has by far the largest number of events (307 CM events) but only 218 events were able to be included in the final dataset. Haiti's 1986 transition had 24 CM events but only 15 were able to be included.

lxnRemoved

This is a Boolean variable (1 or 0) that indicates which post-1977 CRT cases have event data based on candidate events drawn only from the Phoenix dataset (1 indicates cases for which LexisNexis candidate events were removed from event selection; 0 represents the normal protocol). According to the protocol, these are cases whose candidate events should have been drawn from queries of both Phoenix and LexisNexis. But the queries for these cases had produced so many results that event selection and data collection would have taken more time than was available for the project. So near the end of the project when time was running short, it was decided that, in order to include as many *cases* in the final dataset as possible, we would remove the candidate events derived from LexisNexis for a subset of the largest cases, which dramatically reduced the number of candidate events that needed to be screened. This variable is included to enable users, as needed, to account for any biases that might have arisen from that decision.

lxnNA

This is a Boolean variable (1 or 0) that indicates which cases do not have any events identified via LexisNexis, not because LexisNexis results were excluded (as in

lxnRemoved) but because LexisNexis does not have data before 1977. The value is the same as the value as *pre1977* and the opposite value from *post1977.*

pre1977 and ***post1977***
These are Boolean variables (1 or 0) that indicate which cases started before 1977 and which started after 1977. For pre1977, the value is 1 if before 1977 and 0 if after; post1977 is the mirror image (0 if before, 1 if after). This is relevant only because 1977 is the first year for which LexisNexis has data.

order
This is an integer (1, 2, or 3) used for countries that had more than one civil-resistance transition: 1 for that country's first CRT, 2 for its second, and 3 for its third. (Only Madagascar and Thailand had three CRTs.)

Case Variables

caseDescription
This is text with two sentences describing how the transition began and ended (e.g., for Dominican Republic 1962: *"Starts with the resignation of Balaguer. Ends with the 1966 elections that Balaguer wins."*).

outcome
This is a very blunt Boolean variable (1 or 0) that indicates whether the case represented a *clear* democratic transition—that is, the transition ended with an election or the establishment of a coalition government much more broadly representative than the regime it had displaced (value of 1). About two-thirds of the cases in the dataset have a clear democratic outcome like this. That figure might in fact be higher, however, because many of the cases have end dates based not on an event that represents a clear success (e.g., election) or failure (e.g., coup) but on the stabilization of that country's polyarchy score (see *endDate,* below). Some polyarchy scores stabilized above their pre-CRT levels, some close to it, and some below. It is left to the user of the dataset to determine their own criteria for democratic consolidation (e.g., thresholds to be considered a democratic success). This variable is not intended to be used for analysis, only for convenience when a rough (rather than robust) estimate of outcomes is desired.

startDate
This is a character string that represents the beginning of the civil-resistance transition by a four-digit year, a two-digit month, and a two digit day, separated by hyphens and with leading zeros as needed. The start date is defined as the first day of the civil-resistance transition, that is, on the day the nonviolent resistance movement succeeded (e.g., a leader's resignation or ouster, an opposition

candidate's election, a declaration of independence, etc.). This variable is built using the *startYear, startMonth,* and *startDay* variables (below).

endDate
This is a character string that represents the end of the transition, using the same scheme as *startDate*. The end date turned out to be very tricky to identify, because for about a third of the cases there was not a clear moment of success or failure to transition to democracy. The end dates were selected according to the following criteria: (1) for cases that ended violently or abruptly, the date of the violent or abrupt event; (2) for cases that are clearly successful transitions to democracy, the date of the last election that took place on the year the CRT ended (according to the Pinckney dataset); or (3) for ambiguous cases (e.g., those that never became full democracies), December 31 of the year the CRT ended (according to the Pinckney dataset) or of the first year in which that county's V-Dem polyarchy score remained stable for three years in a row. This variable is built using the *endYear, endMonth,* and *endDay* variables (below).

startYear
A four-digit integer representing the year the CRT began (see *startDate,* above).

startMonth
A two-digit integer representing the month the CRT began (see *startDate,* above).

startDay
A two-digit integer representing the day the CRT began (see *startDate,* above).

endYear
A four-digit integer representing the year the CRT ended (see *endDate,* above).

endMonth
A two-digit integer representing the month the CRT began (see *endDate,* above).

endDay
A two-digit integer representing the day the CRT began (see *endDate,* above).

4. Variables in the Events and Candidates Datasets

The second two CSV files that make up the dataset, **crtEvents.csv** and **crtCandidates.csv,** have the same variable structure.

Identification codes

cowID
Same as in Variables: Cases.

crtID
Same as in Variables: Cases.

ffiID
Same as in Variables: Cases.

phxID
Same as in Variables: Cases.

eventID
This is the main identification code used to identify individual *events* in the CM-CRT dataset (for both *civic-mobilization events* included in the events dataset and *candidate events* included in the candidate-events dataset). It is a 13-character code in which the first eight characters are the *ffiID* (start year and *cowID* for the case), the ninth character is a hyphen, and the remaining four characters are a four-digit serial number (with leading zeros as needed) that is unique to each individual event. For example, the event *"Nigerian government workers go on strike, demanding the government provide salary increases and collective bargaining deregulation promised previously"* began on February 5, 1993, and ended on February 12, 1993, and has the unique *eventID* code 19930475-0068.

country
Same as in Variables: Cases.

Dataset variables

include1 and ***uncoded***
The *include1* variable is a Boolean (1 or 0) that indicates whether the research assistant who coded that case determined that the candidate event is a bona fide example of an act of civic mobilization: if so, the value is set to 1; if not, 0. As events were defined as an Actor taking an Action (see Event Variables, below), a CM event would have to feature some entity that could reasonably be described by one of the Actor categories undertaking some set of activities that could reasonably be described by one of the Action categories. Most events selected for inclusion in the dataset were later coded via the data-entry Google Form. Some events selected for inclusion were not coded (and therefore do not appear in crtEvents). For two cases (Egypt 2011 and Haiti 1986), this was because of time and resource constraints: progress was being made on coding the events but time ran out. For others, it was

due to human error: events were selected for inclusion in the dataset *(include1 = 1)* and a staff ID code was entered in the *staffIDform* variable, but then the data-entry form was either not completed or not submitted, so the data was not saved. The *uncoded* Boolean variable captures these errors. In crtCandidates, *uncoded* has a value of 1 if *include1* is 1 and the event does not appear in crtEvents; a value of 0 if *include1* is 1 and the event *does* appear in crtEvents (whose events are all, by definition, fully coded); and a value of NA for everything else.

combineWith

Any time an RA encountered a candidate event that seemed to be describing the same real-world event as a different event in the list of candidate events, the RA was instructed to enter the *eventID* code of that different candidate event in the *combineWith* field. Only that different candidate event was then coded via the data-entry Google Form.

coder1, coder2, and ***check***

The *coder1* and *coder2* variables are categorical variables (10, 11, 12, 13, 14, 15, 16, or 17) indicating the identity of the staff who made include-exclude decisions for that event. In most cases, *coder1* simply identifies the RA who made the include-exclude determination for that event. Some candidate events were ambiguous, in the sense that different people could come to different conclusions about whether it counts as a civic mobilization event. For events where the RA either was not fully confident in their *include1* determination or believed an argument could be made for a different determination, they would set the *check* variable (an integer: 1 or 2) to 1 so other RAs would know a second opinion was needed. The RA who then offered that second opinion would put their staff ID code in *coder2,* review the event, and either delete the 1 under *check* if they agreed with the first RA's selection (although this practice was not followed consistently) or change *check* to 2 if they disagreed. The 2 would signal to the co-PIs that a tie-breaking decision was needed, and a co-PI would review the event to break the tie. The co-PI would change *coder1* to their own staff ID code then either change or keep the value in *include1,* depending on what the final decision is. If this set of variables is needed to account for coder biases, it is recommended to first look for any value in *coder2:* if there is a staff ID code there, it means the first RA had asked for a second opinion and the second RA *(coder2)* gave it. In that case, if *check* has either a 1 or an NA, that means *coder2* agreed with the original value of *include1* so no further change was made to the dataset; if *check* has a 2, that means *coder2* disagreed and the co-PI whose staff ID code is in *coder1* (usually 11) had triple-checked that event and made the final determination of what the value of *include1* should be. The user of this dataset might reasonably disagree with the final decision.

staffID, staffIDform, form, and ***Timestamp***
These are categorical variables (10, 11, 12, 13, 14, 15, 16, or 17) indicating the identity of the staff who entered data for that event through the data-entry form. The value for *staffIDform* comes from the include-exclude sheet: RAs were instructed to put their staff ID code in this field before opening the form to enter data. The value of *staffID* comes from the data-entry form itself, which was coded to automatically pick up the value of *staffIDform* from the include-exclude sheet. The values of both variables should be identical, but there are events where one RA began the form but another completed it. These staff ID variables are included to allow users to test and control for any coder biases. The *form* variable is a simple Boolean (0, 1) indicating that *staffIDform* has a value of some sort (i.e., that someone indicated an intention to complete the data-entry form for that event). The *Timestamp* variable derives from Google Forms and reflects the moment the data-entry form was submitted (if NA, the data was not entered via the data-entry form, which was the case at the very beginning of data collection before the form was created).

falseNeg, truePos, and ***falsePos***
These Boolean variables (1 or 0) indicate the relationship between the candidate events (derived from automated searches of LexisNexis or Phoenix) and the civic-mobilization events (selected for inclusion in crtEvents). The true positives *(truePos* = 1) are the candidate events that were selected for inclusion in the dataset (i.e., the value of *include1* is 1 for any events that were initially included in the include-exclude sheet). The false positives *(falsePos* = 1) are the candidate events in the IE sheets rejected for inclusion (i.e., the value of *include1* is 0). Some events were manually added (to either the include-exclude sheet or to the data-entry form) after searching sources outside of LexisNexis and Phoenix. (LexisNexis does not cover the period before 1977, so all pre-1977 cases come from Phoenix or manual searches.) The CM events added from sources discovered manually, therefore, are the false negatives *(falseNeg* = 1). (True negatives are not tracked, as they would include all non-CM events that did not appear in our queries of LexisNexis or Phoenix.)

Event Variables

For the purposes of this dataset, a civic-mobilization event is defined as an Actor taking an Action—that is, some entity that could reasonably be described by one of the Actor categories (see *actor01* variable) undertaking some set of activities that could reasonably be described by one of the Action categories (see *action1* and *action01* variables). Clarifications, ambiguities, and edge cases are called out in the variable descriptions below.

eventDescription
This free-text variable is a brief summary of the event, using the name or description of the Actor as the subject of the sentence, a description of the Action as the verb, and the name or description of the Target whose behavior the Actor wants to change.

year, month, monthIE, day, dayIE, and ***eventDate***
These variables refer to the date the event in question took place. For events that took place over more than one day, they refer to the date on which the event started. The *year* is a four-digit number; *month* and *day* are each two digits, with leading zeros as needed. Media reports for some events did not specify the day the event happened (and in a couple of events, the month was also not specified); in those cases the *day* or *month* variable has a value of 00. There might be some ambiguities in the dataset for events that took place on one day but the news article reporting on it was published the following day. This is especially true of events that originally derived from the Phoenix dataset, which assumes (often incorrectly) that the event took place on the day the story was published. RAs were instructed to change the date to the day the event actually occurred, but in practice that was not always possible to ascertain. When a date was changed, the RAs sometimes changed it in the data-entry form and sometimes in the *year, month,* and *day* fields of the include-exclude sheet. Further, for some events, an RA might have recorded an event date during the initial include-exclude screening but, during the more detailed data collection that took place while completing the data-entry form, discovered a better source for the dates and entered the better dates in the form (in some cases forgetting to go back to the IE form to correct it). For that reason, the *month* and *day* variables from the IE sheets were changed to *monthIE* and *dayIE* in post-processing to preserve data entered at all stages of the project. The dates with the IE suffix should therefore be treated as incorrect; they are included in the dataset for transparency purposes only. It should be assumed that the correct dates are the non-IE variables: *year, month, day,* and *eventDate* (which is based on the non-IE variables).

endYear, endMonth, endMonthIE, endDay, and ***endDayIE***
While most events in the dataset took place on a single day, about a tenth of the events in the dataset took place over more than one day. These variables are NA for single-day events, and when they have values they indicate the last day of any multi-day events. Those with IE suffixes were derived from the include-exclude sheets in the same way as the *monthIE* and *dayIE* variables described above.

actor and ***actor01***
The *actor* variable is free-form text indicating the name of the Actor (if they are an identifiable entity or person) or describing the Actor (if they have no discernible

name), to add specificity beyond the subcategory codes of the *actor01* variable. Technically there should be an *actor1* variable to describe the top-level Actor category, but there are only two such categories—state actors and nonstate actors—and the first digit of the two-digit *actor01* subcategory variable adequately differentiates them (1 for state, 2 for nonstate). The subcategories are:

- **10 State Other.** Used only if no other state subcategory is appropriate. It includes members of any political faction, party, or other position who are currently in some position of authority but not otherwise covered by another subcategory.
- **11 Incumbents.** The new actors who have been selected, self-selected, appointed, or elected to positions of authority in the state as a result of nonviolent movement that initiated the civil-resistance transition; sometimes the incumbents are leaders of the earlier nonviolent movement; sometimes they are not.
- **12 Former Incumbents Included.** Former regime actors who are still in (or are returning to) positions of authority in the state, working with or under the new incumbents as included members of the new or transitional regime.
- **13 Opposition Actors.** Elected, appointed, or employed officials of political parties (or any allied organizations) who are not in the majority in government or who represent minority factions in government.
- **14 Civil Servants.** Any non-elected, non-appointed, civilian government workers for the state at the national or federal level.
- **15 Military.** Members of the military services. Repressive actions carried out by the military in response to any other civic-mobilization event were not generally included as CM events themselves; such actions were instead covered by the *repression* variable for the event to which they were responding. Combat operations and raids were not, overall, coded as CM events. Non-standard actions, however, were considered CM events (e.g., coup attempts, protests or strikes by troops, etc.).
- **16 Police.** Members of a local or national police force or other non-military security services. Like the military, repressive actions, mass arrests, etc. were not generally treated as a CM event but were instead covered by the *repression* variable in other events. Most CM events led by police were protests and strikes.
- **17 Local State Actors.** Any elected, appointed, or employed officials or civil servants at the subnational level (regional, provincial, county, local, etc.).
- **18 Foreign State Actors.** Any foreign governments. RAs were discouraged from coding events initiated from foreign actors overall, because the volume of such events would have been much more than could have been coded given the time and resources available. Foreign involvement was instead

covered by the *foreignInfluence* and *foreignSupport* variables coded in other events. Users are therefore encouraged to rely on those variables rather than this subcategory for any analysis of foreign involvement (although very events actually have values for those variables).

- **20 Nonstate Other.** Used only if no other nonstate subcategory was appropriate. This can be any unarmed nonstate actor relevant to a civic mobilization not covered by another subcategory. In most cases, this was used for protests and demonstrations by everyday residents and civilians.
- **21 Former Incumbents Excluded.** Former regime actors no longer in positions of authority in the state.
- **22 Political Factions or Candidates.** Members of any political party or non-party faction not currently in power and not covered by another category.
- **23 Nonstate Armed Actors.** Armed actors not categorized as state military or police and *not* acting spontaneously. Includes parastate, paramilitary, guerrilla, insurgent, organized criminal, gang, bandit, and other actors.
- **24 Businesses or Business Associations.** This can include business owners or managers, acting in their individual capacities or organized via coalitions or associations.
- **25 Workers or Unions.** This can include individual workers, whether unionized or not, as well as worker collectives, unions and union leaders, and so on.
- **26 Students or Youth.** This includes any group that is predominantly (even if not exclusively) students, young adults, teenagers, or younger; actions by teachers would be classified here if they are acting with or on behalf of students (but would be categorized under *25 Workers or Unions* otherwise).
- **27 Religious Communities.** Includes organized institutions, groups of believers or adherents, sects, factions, etc., whether acting on their own initiative or through an institution.
- **28 Indigenous Communities.** Includes individuals, organizations, native governments, communities, etc. who are or represent people whose heritage traces to the native, indigenous, or aborignal peoples of the territory where the Actors live or where the Actions take place.
- **29 Foreign Nonstate Actors.** NGOs, IGOs, foreign businesses, and other foreign nonstate actors. As with *18 Foreign State Actors,* RAs were generally discouraged from coding events led by foreign actors; the key exceptions were actions undertaken by refugees.

actorSize and ***actorPrecision***

The *actorSize* variable is an integer representing the precise or approximate number of people represented by the Actor at the start of the event. This figure is

included only if a figure or estimate was provided by the sources of information. RAs were provided the following instructions to address ambiguities:

- Use a precise number if a source gives a precise number.
- Use the estimate cited by a source ("about 350 people" = 350).
- Multiply any order-of-magnitude estimates by 3 ("thousands of protesters" = 3000; "hundreds of thousands" = "300000").
- If a range is given, use an approximate midpoint ("between 300 and 800" = 550).
- If a maximum or minimum is given, use the number cited ("at least 20,000 people" = 20000; "up to fifty" = 50).

The *actorPrecision* categorical variable can take one of three values, according to the judgment of the RA who coded the actorSize variable:

- 0 = imprecise (e.g., "thousands"; midpoint of wide range; etc.)
- 1 = estimate (e.g., "about 150"; midpoint of a narrow range; average; "up to"; "more than"; etc.)
- 2 = precise (exact number used)
- Empty: There was no evidence in the source materials

actorConfidence and ***actorNotes***
These variables are the first in a series of *xConfidence* and *xNotes* variables intended to help users gauge the validity and robustness of the dataset. The *actorConfidence* categorical variable had three potential values:

- **0** = **Low confidence.** Evidence in the source texts was implicit or ambiguous, but with additional contextual knowledge could reasonably be interpreted to support the selected value for this variable.
- **1** = **Medium confidence.** Evidence in the source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge.
- **2** = **High confidence.** Evidence in the source texts was explicit and unambiguous; this was the preferred standard for all values.

The actorNotes free-text variable was intended to give RAs the opportunity to make note of any ambiguities, clarifications, or dilemmas they believed a user of this dataset would need to know about the values chosen for this variable.

action, action1, and ***action01_x***
The *action* variable is free-form text indicating the nature of the action that took place, to add specificity beyond the *action1* categories and *action01_x* subcategories.

There are six top-level *(action1)* Action categories represented by a one-digit code. With illustrative examples of such actions, these categories and their codes are:

- **1 Communicate ("talk").** Offer, Persuade, Reject, Approve, Disapprove
- **2 Demonstrate ("show").** Protest, Strike, Boycott, Intervene; nonviolent noncooperation
- **3 Contribute ("give").** Mediate, Assist, Yield, Investigate, Defect, or Promote Democracy, Civility, or Justice
- **4 Exchange ("trade").** Engage in Dialogue, Negotiate, Bribe
- *5* **Coerce ("threaten").** Threaten, Exhibit Force Posture, Reduce Relations, Suspend Rights
- **6 Force ("overpower").** Seize or Damage Property, Detain or Deport, Repress or Abuse, Engage in Violent Combat, Use Unconventional Mass Violence

There are 29 subcategories of Action represented by six variables whose names have the structure `action01_x` (where `x` is the *action1* top-level category number) and whose values are represented by a two-digit code (the first digit corresponds to the associated *action1* category). Some of the terminology and categories in our ontology diverges from the CAMEO ontology (e.g., we define coercion differently; most of CAMEO's "coerce" subcategories fall under our "force" category), but almost all of CAMEO's second-level categories map directly onto our second-level categories (with just a few exceptions), which will make this dataset compatible with other datasets using CAMEO (Phoenix, NAVCO, etc.). The 29 subcategories of this dataset and their two-digit codes are:

- *action01_1:* **10 Communicate (unspecified).** Used only if no other "Communicate" subcategory was found to be appropriate. To communicate means to make a public statement, verbally or in writing, of facts, position, beliefs, etc.
- *action01_1:* **11 Offer.** Make an offer of cooperation, dialog, material goods, friendship, alliance, etc. This does *not* include actions that represent a carrying through with the offer (see the "Contribute" category).
- *action01_1:* **12 Persuade.** Request something in a friendly or unfriendly manner; attempt to persuade someone to do something; appeal (friendly request); demand (unfriendly request); insist; encourage (an action or compliance with a demand or rule); or employ facts, evidence, or rhetoric as part of a persuasive argument.
- *action01_1:* **13 Reject.** Turn down an offer of cooperation, material goods or services, etc.
- *action01_1:* **14 Approve.** Express positive sentiments toward or about someone or something (a policy, an idea, etc.), including defending, apologizing, recognizing, etc. using words only; actually signing a policy

(approving a policy) or otherwise responding to a request would use a different category (e.g., one of the "Contribute" categories).

- *action01_1:* 15 **Disapprove.** Express negative sentiments toward or about someone or something (a policy, an idea, etc.) using words only; includes shaming and some categories of social noncooperation as well as expressions of contempt or hate, mockery, insults, disapproval, etc.
- *action01_2:* **20 Demonstrate (unspecified).** ["show"] Used only if no other "Demonstrate" subcategory was found to be appropriate. Any category of social, political, or economic noncooperation not otherwise specified below (this can include expressions of support or celebration that do not fit under the "protest" subcategory below); noncoercive acts only.
- *action01_2:* **21 Protest.** Expression of disapproval, demand, request, insist, etc. by actions, usually by physically gathering a mass of people in a public location.
- *action01_2:* **22 Strike.** Stop showing up to work; slow down work; do work incompetently on purpose; etc.
- *action01_2:* **23 Boycott.** Stop purchasing goods or services, canceling subscriptions, etc.
- *action01_2:* **24 Intervene.** Physically obstructing, disrupting, preventing, occupation; and other forms of physical noncooperation. Does *not* include rioting or vandalism (see "61 Seize or Damage Property").
- *action01_3:* **30 Contribute (unspecified).** Used only if no other "Contribute" subcategory was found to be appropriate. Provide or give one's time, resources, attention, skills, engaging in material cooperation, etc. with no expectation of directly receiving something of value in return; includes actual actions only (participate, collaborate, cooperate, volunteer, etc.); for *offers* of action, use "11 Offer."
- *action01_3:* **31 Mediate.** Mediate a dispute, offer one's good offices, meet with several parties, etc. This category is for an Actor helping *other* parties with a dispute of some sort; for an Actor or Target who is party to the dispute and participating in a mediation, "41 Engage in Dialogue" or "42 Negotiate" were used instead.
- *action01_3:* **32 Assist.** Provide aid, donate time or money to help someone directly; if assistance is only *offered* but not carried out, it was code as "11 Offer"; if the Actor is a state, assistance meant assistance offered through policy or the state; if Actor is nonstate, assistance meant assistance offered personally, or as a grant, donation, volunteer work, etc.
- *action01_3:* **33 Yield.** Give up; concede, accede, reverse, revert, change behavior in response to a request of some sort.
- *action01_3:* **34 Investigate.** Find information (i.e., giving one's time and resources to research something relevant to the transition).

- *action01_3:* **35 Defect.** Change sides in a dispute; this is broader than NAVCO's "defect" category, which focuses only on state actors; this can also include nonstate actors switching sides (e.g., in shifting alliances between movements).
- *action01_3:* **36 Promote Democracy.** Promoting (or changing one's behavior to engage in) voting (e.g., get-out-the-vote efforts), running for election, registration drives, electoral observer, or changing one's behavior to become more democratic (policy change, attitude change, personal behavior, etc.); normal acts of democracy (e.g., actually voting) were *not* included unless doing so was clearly a significant act of (and evidence for) mobilization of a population.
- *action01_3:* **37 Promote Civility.** Promoting civility (treating others with respect and dignity, speaking nicely, etc.) or changing one's behavior to become more civil; this focuses on individual behavior and interpersonal communication, not policy, institutional, or systemic change.
- *action01_3:* **38 Promote Justice.** Promoting justice (equitable opportunity and treatment in rights, benefits, outcomes, etc.) or changing one's support for actions that improve justice (e.g., supporting new policy, reforming institutions, etc.); justice focuses on institutions, systems, and society-level outcomes, in contrast with civility, which focuses on individual and interpersonal behavior.
- *action01_4:* **40 Exchange (unspecified).** Used only if no other "Exchange" subcategory was found to be appropriate. Giving time, resources, attention, skills, etc. (purchasing, trading, compromising, etc.) with an explicit expectation of receiving something of value in return, immediately or in the future.
- *action01_4:* **41 Engage in Dialogue.** Efforts to come to a mutual understanding
- *action01_4:* **42 Negotiate.** Efforts to come to a mutual agreement.
- *action01_4:* **43 Bribe.** Giving money or something else of value in exchange for cooperation, policy change, political support, etc.
- *action01_5:* **50 Coerce (unspecified).** Used only if no other "Coerce" subcategory was found to be appropriate. Implicit or explicit threats of either violence or the loss of something valued; this is the most general category (and is different from the NAVCO3 "Coerce" category, which is much more like "6 Force," below).
- *action01_5:* **51 Threaten.** Explicit threats of violence or threats of the destruction or removal of something valued.
- *action01_5:* **52 Exhibit Force Posture.** Displaying or mobilizing weapons, troops, fighters, equipment, etc. in a threatening manner; does *not* include the actual use of force (use a "6 Force" category for that); this *can* include

actions by state *or* nonstate actors but applies to organized groups only (if an individual or spontaneous group does this, use "51 Threaten" instead).

- *action01_5:* **53 Reduce Relations.** Withhold or stop providing benefits or cooperation in an effort to coerce someone to do something; this can include withdrawing from negotiations, imposing embargoes or sanctions, breaking diplomatic relations, withdrawing personnel, etc.
- *action01_5:* **54 Suspend Rights.** Suspend laws, the constitution, elections, etc. or impose curfews, censorship of media, or other restrictions of freedom.
- *action01_6:* **60 Force (unspecified).** Used only if no other "Force" subcategory was found to be appropriate. Acts of physical or psychological violence, or the *direct taking* of something valued.
- *action01_6:* **61 Seize or Damage Property.** Taking, damaging, or destroying property, or otherwise rendering it unusable; this includes rioting and vandalism; this *can* be done by state *or* nonstate actors.
- *action01_6:* **62 Detain or Deport.** These are actions that can also be done by nonstate actors (e.g., taking hostages, kidnapping, or "disappearing" people) as well as by state actors. (See NAVCO3 subcategories for 173, 174, and 181.)
- *action01_6:* **63 Repress or Abuse.** This is direct physical violence that physically harms, traumatizes, or kills people.
- *action01_6:* **64 Engage in Violent Combat.** This is direct physical violence involving at least two groups of people who physically harm, traumatize, or kill people in the opposing groups.
- *action01_6:* **65 Use Unconventional Mass Violence.** This is direct physical violence, mass expulsions, ethnic cleansing, genocide, etc. against unarmed people.

actionConfidence and ***actionNotes***

Similar to *actorConfidence* and *actorNotes,* but for the Action variables: *actorConfidence* is a categorical variable (0 means the evidence for the Action subcategory selected was implicit or ambiguous, 1 means the evidence required some interpretation, and 2 means the evidence was explicit) and *actorNotes* is a free-text variable offering additional insights on any of the Action variables.

target, target01, targetConfidence, and ***targetNotes***

The Target of the actions—that is, the entity whose behavior the action was intended to change in some way—was not technically part of the definition of the CM event but it was required nonetheless. These four variables are directly analogous to the Action variables: *target* is a free-text variable naming or describing the Target of the CM event; *target01* uses the same subcategories as actor01 to describe the Target; *targetConfidence* is a categorical variable (0, 1, 2) describing how explicit evidence the article was with regard to the identity of the target; and

targetNotes is a free-text variable where RAs could not any observations they considered important for users to know.

eventLocation and ***eventScale***

The *eventLocation* variable is simply the name of the location where the event took place. If the event was concentrated within a town or city, the RAs were instructed to name the area (e.g., "Tahrir Square in downtown Cairo") if it was mentioned in any article; otherwise, they were to name the main town, city, province, region, etc. If it took place in multiple locations throughout the country, the name of the main location (if identifiable) was identified and *eventScale* was coded as 1 (city or region) or 2 (entire country), as appropriate. (The NAVCO3 codebook, Section 4 "Geo_scope," was a key reference for this.) The *eventScale* categorical variable indicated the geological scale or scope of the event:

- 0 = specific named place (area within a town or city)
- 1 = town, city, metropolitan area including suburbs, province, or subnational region of country
- 2 = entire country or most of country

title and ***titleIE***

These variables capture the title of the main information source for the event. The value for *titleIE* was taken directly from the data entered via the include-exclude sheets (in those sheets, this variable was called *title* but in post-processing it was changed to *titleIE* to differentiate it from the data entered via the data-entry forms). The value for *title* was taken directly from the data-entry form. The path the values of these variables took was as follows: The queries of both LexisNexis and Phoenix returned one article per observation, so all candidate events deriving from them were described by a single news article. The titles of those articles therefore appeared, unedited, as the value of the *title* variable in the include-exclude sheets; this became *titleIE* in the dataset. The data-entry form was designed to automatically import the value of the *title* (now *titleIE)* variable from the include-exclude sheet; this became *title* in the dataset. In most cases, the value of *title* and *titleIE* are therefore identical. But RAs were able to change the value of title in the data-entry form if it made sense to do so and in those cases the values differ from each other. Some events, for example, did not derive from LexisNexis or Phoenix, and in most of those instances the title in the include-exclude sheets was initially NA. Either during include-exclude screening or during data entry via the forms, RAs were able to change the title of the event. In most cases, they simply used the title of the alternative source of information (e.g., an article from the *New York Times Archives).* In a small handful of the pre-1977 cases, the information about the event came from a longer report or book whose title would not reasonably

reflect the content of the event; in those events, the value of *title* is simply a description of the event, written by the RA in question.

url, urlIE, urlForm, and ***broken***

These are the URLs (Internet addresses) of the information source associated with the value in title or titleIE. In most cases, users with a LexisNexis account (and in some cases a New York Times subscription) should be able to follow these links directly to the appropriate article. In events deriving from non-LexisNexis sources, the links are to a source that was available on the Internet as of June 2021. The path the values of these variables took was similar to the path for *title* and *titleIE*. It began with queries to LexisNexis and Phoenix, and the articles they returned all had either LexisNexis URLs or (in the case of some Phoenix results) NAs in the results. These values were transferred directly to the *url* variable in the include-exclude sheets, for which we additionally created the *urlG* variable (renamed *urlIE* in post-processing to clarify where its values came from). The only difference between *url* and *urlG/urlIE* was that url used an ampersand (`&`) in the address while *urlG/urlIE* instead used `%26` (the URL percent-encoded version of an ampersand) in the address. This was done because at some point early in data collection it became clear that URLs with an ASCII ampersand were not resolving to the correct Internet addresses in some contexts for some RAs, and URLs with a percent-encoded ampersand were not resolving correctly in others. Having both forms of the URL available in the include-exclude sheets was therefore a matter of convenience. (In post-processing, all percent-encoded ampersands were changed to ASCII ampersands.) The data-entry form also had *url* as a variable, and it was set to automatically pick up whatever value was in *url* in the include-exclude sheet. As with *title* and *titleIE,* RAs were able to revise what appeared in any of the URL variables, in either the include-exclude sheet or in the data-entry form; this was particularly true of pre-1977 events and those derived from the Phoenix dataset that had no original URL. All three variables were therefore preserved: *url* from the include-exclude sheets remains *url* in the dataset; *urlG* from the include-exclude sheets is now *urlIE* to indicate its value could only have come from its original inclusion in the IE sheets or from subsequent revision in the IE sheets; and *url* from the data-entry form is now *urlForm,* and its value could have come either from *url* in the IE sheets (which the form automatically picked up) or from an RA changing its value in the data-entry form. In a handful of events, the URLs in one of more of these variables could not resolve to a valid address on the LexisNexis server, so a *broken* Boolean variable was added to the include-exclude sheet so RAs could flag them (the only values are 1 to indicate a link not working or NA).

contentItem and ***eid***

All articles in the LexisNexis database have a unique ID code with the structure xxxx-xxxx-xxxx-xxxx-00000-00, where x is an alphanumeric character. This

variable stores that unique code. In the URL to the articles at LexisNexis, this code appears after `?resultid=urn:contentItem:` and before `&DocumentIdType=news`, at least in FFI's account. In cases where a user's LexisNexis account differs (or if LexisNexis changes its URL scheme, as it did midway through this project), it will be convenient for the user to be able to auto-generate a new URL by concatenating the URL prefix (e.g., https://solutions.nexis.com/wsapi/news-and-directories/document), followed by `?resultid=urn:contentItem:`, followed by the value in *contentItem*, followed by the appropriate suffix (e.g., `&DocumentIdType=news`). The *eid* variable is the unique ID code that the Phoenix dataset uses to identify its sources of information. It is included to facilitate connecting additional variables from Phoenix to the events in this dataset.

sourceEventYN, titleSourceXevents, and ***idSourceXevents***
The main sources of information for the event variables described above can be found in *title, titleIE, url, urlIE,* and *urlForm.* The *sourceEventYN* Boolean variable (Yes or No) indicates whether any additional sources of information were used for any of the event variables). RAs could enter the title (optional) or the identifier of up to three additional sources of information *titleSource1events* and *idSource1events* for the first additional source, or *titleSource2events* (and so on) for additional sources. Because data collection was taking place in the context of a global pandemic and widespread lockdowns, staff did not have access to physical books in any library and had to rely entirely on sources of information available online; for that reason, the identifier variables (starting with *idSource)* are almost entirely URLs.

eventSource, sourceLxn, sourcePhx, and ***sourceOther***
These variables indicate which events ultimately derived from which main set of information sources. The *eventSource* categorical variable indicates the three general sources of information: "LexisNexis", "Phoenix", and "Other" (key documents or other sources). In general, any event with a value in *eid* but not *contentItem,* or with a value in both *eid* and *contentItem,* derived from Phoenix; any event with a value in *contentItem* but not *eid* derived from LexisNexis; and any event with NA in both eid and contentItem derived from a different source. The other two variables are Boolean (0 or 1): *sourceLxn* is 1 if *eventSource* is "LexisNexis"; *sourcePhx* is 1 if *eventSource* is "Phoenix"; *sourceOther* is 1 if *eventSource* is "Other."

countryPhx, placename, statename, and ***cameoCode***
These, along with the *eid* variable (above), all derive from the Phoenix dataset, included here primarily for convenience: *countryPhx* is the three-character country code of the Phoenix dataset's *country* variable, which in many cases is different from this dataset's *phxID* variable (which derived from Phoenix's *target_root* variable). The *placename* and *statename* variables are pulled directly from Phoenix

without modification, and *cameoCode* is pulled directly from Phoenix's cameo variable without modification.

notes and ***eventNotes***
Staff had two opportunities to enter general notes regarding any ambiguities, clarifications, or dilemmas they they believed a user of this dataset would need to know about about the event as a whole: in the *notes* variable of the include-exclude sheet (generally having to do with considerations for why the event was or was not selected for inclusion in the dataset) and in the *eventNotes* variable of the data-entry form (where more general clarifications were offered).

Attribute Variables

More specific attributes of each event are captured in the variables in this section. Data for these variables was based on the content of the articles that were used to define the event plus some additional research. Early in data collection, the RAs discovered that more time spent searching the "key sources" documents, LexisNexis, the New York Times Archives, and Google search results did not generally produce significantly better results for any attribute variable, so they were instructed to spend no more than 5–10 minutes to find additional sources. The instructions were to enter a value only if at least one source made clear reference to it or, at worst, the value could reasonably be inferred (hence the importance of the associated *xConfidence* variables). Due to how such events are covered, no attribute value had values for more than half of the events (a couple were as low as 5 percent).

alignment, alignmentConfidence, alignmentNotes, and ***alignmentSource***
The *alignment* categorical variable describes how aligned the Actor is to the new state incumbents:

- 1 = Actor is aligned or associated with new incumbent
- 0 = Actor is not aligned with either the new incumbent or the old regime
- –1 = Actor is aligned or associated with old regime
- Empty: There was no evidence in the source materials one way or the other

Note that if the dataset's CSV file is opened in some spreadsheet programs, such as Microsoft Excel, the program might generate an error (e.g., Excel generates a `#NAME?` error) for the third category because it begins with `-1`. This variable is intended to help users know the degree to which the information sources were clear about "who's side" the Actor was on, if any (e.g., about twice as many events featured Actors aligned with the new incumbents as those aligned with the old regime). The *alignmentConfidence* categorical variable had three potential values:

- **0** = **Low confidence.** Evidence in the source texts was implicit or ambiguous, but with additional contextual knowledge could reasonably be interpreted to support the selected value for this variable.
- **1** = **Medium confidence.** Evidence in the source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge.
- **2** = **High confidence.** Evidence in the source texts was explicit and unambiguous; this was the preferred standard for all values.

The *alignmentNotes* free-text variable was intended to give RAs an opportunity to make note of any ambiguities, clarifications, or dilemmas they believed a user of this dataset would need to know about the values chosen for this variable. The *alignmentSource* free-text variable provided the URL of any source of information that was used that was not referenced by any of the sources for the event variables. All of the remaining attribute variables have *xConfidence, xNotes,* and *xSource* variables that function the same way as these, so the descriptions will not be repeated. About half of the CM events have values for this variable.

coherence, coherenceConfidence, coherenceNotes, and ***coherenceSource***
The *coherence* categorical variable describes how internally coherent the Actor group was. The categories are:

- 0 = Actor is fragmented, with contradictory claims on leadership or authority
- 1 = pluralistic, with multiple sources of recognized leadership or authority
- 2 = unitary, with a single source of recognized leadership or authority
- Empty: There was no evidence in the source materials one way or the other

The other *coherence* variables listed here operate the same way as those in *alignment.* Nearly half of the CM events have values for this variable.

inclusivity, inclusivityConfidence, inclusivityNotes, and ***inclusivitySource***
This is a categorical variable describing how democratically inclusive the intent or tendency of the Action was. The categories are:

- 1 = Action attempted expansion of democratic "winning coalition"
- 0 = Action was meant neither to expand nor to contract democratic inclusion
- −1 = Action was maximalist or represented a contraction toward factions
- Empty: There was no evidence in the source materials one way or the other

Note that the third category might open with a `#NAME?` error in Excel because of the `-1` at the beginning. The other variables listed here operate the same way as

those in *alignment*. Only about a quarter of the CM events have values for this variable.

spontaneity, spontaneityConfidence, spontaneityNotes, and ***spontaneitySource***
This is a categorical variable indicating how spontaneous the Action was. The categories are:

- 0 = Action was planned, premeditated, or organized by Actor
- 1 = Action included a mix of organized and uncontrolled elements
- 2 = Action emerged spontaneously or without organization
- Empty: There was no evidence in the source materials one way or the other

The other variables listed here operate the same way as those in *alignment*. Nearly half of the CM events have values for this variable.

diversity, diversityConfidence, diversityNotes, and ***diversitySource***
This is a categorical variable indicating the degree to which the Actor reflected the identity distribution of society. The categories are:

- 0 = Actor group does not reflect the ethnic and other identity distribution of population
- 1 = Actor group includes some ethnic diversity but not equitably
- 2 = Actor group reflects ethnic and identity diversity of the population
- Empty: There was no evidence in the source materials one way or the other

For some events, RAs were able to find photographs or videos where they could visually scan the Actors to get an overall sense of the event's diversity. Only about 7 percent of the CM events have values for this variable.

gender, genderConfidence, genderNotes, and ***genderSource***
This is a categorical variable indicating the degree to which the Actor entity was gender-balanced. The categories are:

- 0 = Actor group does not reflect gender balance
- 1 = Actor group includes some gender balance
- 2 = Actor group reflects gender balance
- Empty: There was no evidence in the source materials one way or the other

For some events, RAs were able to find photographs or videos where they could visually scan the Actors to get an overall sense of the event's gender balance. Only about 6 percent of the CM events have values for this variable.

repression, repressionConfidence, repressionNotes, and ***repressionSource***
This is a categorical variable indicating whether or how the state incumbents responded to the event with any degree of repression. This variable refers *only* to

actions taken by the state incumbent. For example, in events where the Actor is nonstate, this variable refers to the state incument's response to the Actors' Actions. For events where the Actor is state but not subcategory "11 Incumbents", this variable still refers to the state incumbent's response to the events. The categories of *repression* responses are:

- 0 = State incuments did not respond with repression (used only when a source says so explicitly)
- 1 = State incumbents responded using non-physical repression
- 2 = State incumbents responded using physical repression with no fatalities
- 3 = State incumbents responded using physical repression with fatalities
- Empty: There was no evidence in the source materials one way or the other

About a third of the CM events have values for this variable.

foreignInfluence, foreignInfluenceConfidence, foreignInfluenceNotes, and ***foreignInfluenceSource***
This is a categorical variable indicating whether or how a significant foreign actor attempted to influence the event. The *foreignInfluence* variable refers to any involvement by a significant foreign actor attempting to influence the event; it does not ask about who the foreign actor is supporting (for that, see *foreignSupport,* below). The categories are:

- 0 = no involvement
- 1 = statement
- 2 = economic support
- 3 = diplomatic support
- 4 = military support
- 5 = military intervention
- Empty: There was no evidence in the source materials one way or the other

Less than 5 percent of the CM events have values for this variable.

foreignSupport, foreignSupportConfidence, foreignSupportNotes, and ***foreignSupportSource***
This is a categorical variable indicating whether or how the foreign influence (see *foreignInfluence)* benefited or was intended to benefit the Actor. The categories are:

- 1 = benefited Actor
- 0 = no benefit or harm
- −1 = harmed Actor or benefited opponents of Actor
- Empty: There was no evidence in the source materials one way or the other

Less than 5 percent of the CM events have values for this variable.

sourceAttributeYN, titleSourceXattributes, and ***idSourceXattributes***
The main sources of information for the attribute variables described above can be found in *title, titleIE, url, urlIE,* and *urlForm.* The *sourceEventYN* Boolean variable (Yes or No) indicates whether any additional sources of information were used for any of the attribute variables). As with the *titleSourceXevents* variables (above), RAs could enter the title (optional) or the identifier of up to three additional sources of information: *titleSource1attributes* and *idSource1attributes* for the first additional source, or *titleSource2attributes* (and so on) for additional sources, and the source identifiers are generally URLs.

Appendix

CM-CRT Event Data Entry Form

1. Metadata: Do not revise (skip to section 2)

Please do not edit any of the pre-filled responses in this section. But PLEASE VERIFY that they refer to the correct event you are coding.

staffID
Short answer text

cowID
Short answer text

crtID
Short answer text

ffiID
Short answer text

phxID
Short answer text

eventID
Short answer text

country
Short answer text

year
Short answer text

month
Short answer text

day
Short answer text

title
Short answer text

url
Short answer text

2. Event Data

2.1. Describe the event concisely.

One-sentence summary of the event, using the name or description of the Actor as the subject of the sentence, a description of the Action as the verb, and the name or description of the Target whose behavior the Actor wants to change.

eventDescription
Long answer text

2.2. Where did the event take place?

The name of the location where the event took place. If concentrated within a town or city, name the area (e.g., "Tahrir Square in downtown Cairo") if it's mentioned in any article; otherwise, name the main town, city, province, region, etc.. If it took place in multiple locations throughout country, name main location (if identifiable) and code `eventScale` (below) as 1 (city or region) or 2 (entire country) as appropriate. See NAVCO3 codebook, Section 4 "Geo_scope", for guidance.

eventLocation
Short answer text

2.3. Across what scale did the event take place?

Code for categories of the geological scale or scope of the event (Actor and Action).

eventScale
- 0 = specific named place (area within a town or city)

- 1 = town, city, metropolitan area including suburbs, province, or subnational region of country
- 2 = entire country or most of country

2.4. If this is a multi-day event, when was the end date?

endYear
Short answer text

endMonth
Short answer text

endDay
Short answer text

2.5. General notes on this event.

Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about this event.

eventNotes
Long answer text

3. Actors and Actor Categories

3.1. Name (or describe) the Actor.

Free-form text indicating the name of the Actor (if it's an identifiable entity or person) or describing the Actor (if no discernible name), to add specificity beyond the category codes (below).

actor
Short answer text

3.2. How many people were included in the "Actor" above?

An integer representing the precise or approximate number of people represented by the "Actor" at the start of the event (include only if provided by the source). Do not use commas or letters; for large numbers, be very careful about the number of zeros.

- Use a precise number if a source gives a precise number.

- Use the estimate cited by a source ("about 350 people" = 350).
- Multiply any order-of-magnitude estimates by 3 ("thousands of protesters" = 3000; "hundreds of thousands" = "300000").
- If a range is given, use an approximate midpoint ("between 300 and 800" = 550).
- If a maximum or minimum is given, use the number cited ("at least 20,000 people" = 20000; "up to fifty" = 50).

actorSize
Short answer text

3.3. How precise is the Actor size estimate above?

How precise is the `actorSize` estimate above.

actorPrecision

- 0 = imprecise (e.g., "thousands"; midpoint of wide range; etc.)
- 1 = estimate (e.g., "about 150"; midpoint of a narrow range; average; "up to"; "more than"; etc.)
- 2 = precise (exact number used)
- Empty: There was no evidence in the source materials

3.4. What subcategory was the Actor?

STATE ACTOR

- 10 State Other: Use only if you absolutely cannot identify an appropriate category below. This includes members of any political faction, party, or other position who are currently in some position of authority but not covered by another category
- 11 Incumbents: The new actors who have been selected, self-selected, appointed, or elected to positions of authority in the state as a result of nonviolent action
- 12 Former Incumbents Included: Former regime actors who are still in (or returning to) positions of authority in the state, working with or under the new incumbents
- 13 Opposition Actors: Elected, appointed, or employed officials of political parties (or any allied organizations) who are not in the majority in government or represent minority factions in government
- 14 Civil Servants: Any non-elected, non-appointed, civilian government workers for the state at the national or federal level
- 15 Military: Members of the military services

- 16 Police: Members of a local or national police force or other non-military security services
- 17 Local State Actors: Any elected, appointed, or employed officials or civil servants at the subnational level (regional, provincial, county, local, etc.)
- 18 Foreign State Actors: Any foreign governments

NONSTATE ACTOR

- 20 Nonstate Other: Use only if you absolutely cannot identify an appropriate category below. This can be any unarmed nonstate actor relevant to a civic mobilization not covered by another category
- 21 Former Incumbents Excluded: Former regime actors no longer in positions of authority in the state
- 22 Political Factions or Candidates: Members of any political party or non-party faction not currently in power and not covered by another category
- 23 Nonstate Armed Actors: Armed actors not categorized as state military or police and NOT acting spontaneously. Includes parastate, paramilitary, guerrilla, insurgent, organized criminal, gang, bandit, and other actors.
- 24 Businesses or Business Associations: Can include business owners and managers, acting in their individual capacities or organized via coalitions or associations
- 25 Workers or Unions: Can include individual workers, whether unionized or not, as well as worker collectives, unions and union leaders, etc.
- 26 Students or Youth: Any group that is predominantly (even if not exclusively) students, young adults, teenagers, or younger; actions by teachers would be classified here if they are acting with or on behalf of students (but under 25 Workers or Unions otherwise)
- 27 Religious Communities: Includes organized institutions, groups of believers or adherents, sects, factions, etc., whether acting on their own initiative or through an institution
- 28 Indigenous Communities: Includes individuals, organizations, native governments, communities, etc. who are or represent people whose heritage traces to the native peoples of the territory where the Actors live or where the Actions take place
- 29 Foreign Nonstate Actors: NGOs, IGOs, foreign businesses, and other foreign nonstate actors

actor01

- 10 STATE Other
- 11 Incumbents
- 12 Former Incumbents Included

- 13 Opposition Actors
- 14 Civil Servants
- 15 Military
- 16 Police
- 17 Local State Actors
- 18 Foreign State Actors
- 20 NONSTATE Other
- 21 Former Incumbents Excluded
- 22 Political Factions or Candidates
- 23 Nonstate Armed Actors
- 24 Businesses or Business Associations
- 25 Workers or Unions
- 26 Students or Youth
- 27 Religious Communities
- 28 Indigenous Communities
- 29 Foreign Nonstate Actors

3.5. How confident are you in your responses to the questions about the Actor?

actorConfidence

- 2 = High confidence: evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = Medium confidence: evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge
- 0 = Low confidence: evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value

3.6. Notes on responses to Actor questions.

Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about these variables or the values you chose for them.

actorNotes
Long answer text

4. Actions and Action Categories

4.1. Describe the Action

Free-form text describing the Action that took place, to add specificity beyond the category codes.

action
Short answer text

4.2. What was the predominant nature of the Action?

action1

- 1 Communicate ("talk"): Offer, Persuade, Reject, Approve, Disapprove
- 2 Demonstrate ("show"): Protest, Strike, Boycott, Intervene; nonviolent noncooperation
- 3 Contribute ("give"): Mediate, Assist, Yield, Investigate, Defect, or Promote Democracy, Civility, or Justice
- 4 Exchange ("trade"): Engage in Dialogue, Negotiate, Bribe
- 5 Coerce ("threaten"): Threaten, Exhibit Force Posture, Reduce Relations, Suspend Rights
- 6 Force ("overpower"): Seize or Damage Property, Detain or Deport, Repress or Abuse, Engage in Violent Combat, Use Unconventional Mass Violence

4.2.1. COMMUNICATE: What or how did the Actor try to communicate to the Target?

- 10 Communicate (unspecified): ["talk"] Use only if you absolutely cannot identify an appropriate category below. Make a public statement, verbally or in writing, of facts, position, beliefs, etc.
- 11 Offer: Make an offer of cooperation, dialog, material goods, friendship, alliance, etc.; does NOT include actions that represent a carrying through with the offer (see Contribute category)
- 12 Persuade: Request something in a friendly OR unfriendly manner; attempt to persuade someone to do something; appeal (friendly request); demand (unfriendly request); insist; encourage (an action or compliance with a demand or rule); employ facts, evidence, or rhetoric as part of a persuasive argument
- 13 Reject: Turn down an offer of cooperation, material goods or services, etc.
- 14 Approve: Express positive sentiments toward or about someone or something (a policy, an idea, etc.), including defending, apologizing,

recognizing, etc. using words only; actually signing a policy (approving a policy) or otherwise responding to a request would use a different category (e.g., one of the Contribute categories)
- 15 Disapprove: Express negative sentiments toward or about someone or something (a policy, an idea, etc.) using words only; includes shaming and some categories of social noncooperation as well as expressions of contempt or hate, mockery, insults, disapproval, etc.

action01_1
- 00 NONE (Action was a different top-level category)
- 10 Communicate (unspecified)
- 11 Offer
- 12 Persuade
- 13 Reject
- 14 Approve
- 15 Disapprove

4.2.2. DEMONSTRATE. What form did the Actor's demonstration take?

- 20 Demonstrate (unspecified): ["show"] Use only if you absolutely cannot identify an appropriate category below. Any category of social, political, or economic noncooperation not otherwise specified below (this can include expressions of support or celebration that do not fit under the "protest" subcategory below); noncoercive acts only
- 21 Protest: expression of disapproval, demand, request, insist, etc. by actions, usually by physically gathering a mass of people in a public location
- 22 Strike: Stop showing up to work; slow down work; do work incompetently on purpose; etc.
- 23 Boycott: Stop purchasing goods or services, canceling subscriptions, etc.
- 24 Intervene: Physically obstructing, disrupting, preventing, occupation; and other forms of physical noncooperation. Does NOT include rioting or vandalism (code those as 61 Seize or Damage Property)

action01_2
- 00 NONE (Action was a different top-level category)
- 20 Demonstrate (unspecified)
- 21 Protest
- 22 Strike
- 23 Boycott
- 24 Intervene

4.2.3. CONTRIBUTE. How or what did the Actor contribute to the Target?

- 30 Contribute (unspecified): ["give"] Use only if you absolutely cannot identify an appropriate category below. Provide or give one's time, resources, attention, skills, engaging in material cooperation, etc. with no expectation of directly receiving something of value in return; includes actions only (participate, collaborate, cooperate, volunteer, etc.); for offers of action, use 11 Offer.
- 31 Mediate: Mediate a dispute, offer one's good offices, meeting with several parties, etc.; this category is for an Actor helping other parties with a dispute of some sort; for an Actor or Target who is party to the dispute and participating in a mediation, use 41 Engage in Dialogue or 42 Negotiate
- 32 Assist: Provide aid, donate time or money to help someone directly; if assistance is only offered but not carried out, code as category 11 Offer; if the Actor is a state, assistance means assistance offered through policy or the state; if Actor is a nonstate, assistance means assistance offered personally, or as a grant, donation, volunteer work, etc.
- 33 Yield: Give up; concede, accede, reverse, revert, change behavior in response to a request of some sort
- 34 Investigate: Find information (i.e., giving one's time and resources to research something relevant to the transition)
- 35 Defect: Change sides in a dispute; this is broader than NAVCO, which focuses only on state actors; this can also include nonstate actors switching sides (e.g., in shifting alliances between movements)
- 36 Promote Democracy: Promoting (or changing one's behavior to engage in) voting (e.g., GOTV), running for election, registration drives, electoral observer, or changing one's behavior to become more democratic (policy change, attitude change, personal behavior, etc.); normal acts of democracy (e.g., actually voting) should not be included unless doing so is clearly a significant act of (and evidence for) civic mobilization
- 37 Promote Civility: Promoting civility (treating others with respect and dignity, speaking nicely, etc.) or changing one's behavior to become more civil; this focuses on individual behavior and interpersonal communication, not policy, institutional, or systemic change
- 38 Promote Justice: promoting justice (equitable opportunity and treatment in rights, benefits, outcomes, etc.) or changing one's support for actions that improve justice (e.g., supporting new policy, reforming institutions, etc.); justice focuses on institutions, systems, and society-level outcomes, in contrast with civility, which focuses on individual and interpersonal behavior

action01_3

- 00 NONE (Action was a different top-level category)
- 30 Contribute (unspecified)
- 31 Mediate
- 32 Assist
- 33 Yield
- 34 Investigate
- 35 Defect
- 36 Promote Democracy
- 37 Promote Civility
- 38 Promote Justice

4.2.4. EXCHANGE. What form did the Actor's efforts to exchange with the Target take?

40 Exchange (unspecified): ["trade"] Use only if you absolutely cannot identify an appropriate category below. Giving time, resources, attention, skills, etc. (purchasing, trading, compromising, etc.) with an explicit expectation of receiving something of value in return, immediately or in the future.

41 Engage in Dialogue: Efforts to come to a mutual understanding

42 Negotiate: Efforts to come to a mutual agreement

43 Bribe: Giving money or something else of value in exchange for cooperation, policy change, political support, etc.

action01_4

- 00 NONE (Action was a different top-level category)
- 40 Exchange (unspecified)
- 41 Engage in Dialogue
- 42 Negotiate
- 43 Bribe

4.2.5. COERCE. How did the Actor coerce the Target?

- 50 Coerce (unspecified): ["threaten"] Use only if you absolutely cannot identify an appropriate category below. Implicit or explicit threats of either violence or the loss of something valued; this is the most general category (and is different from the NAVCO3 "Coerce" category)
- 51 Threaten: Explicit threats of violence or the destruction or removal of something valued

- 52 Exhibit Force Posture: Displaying or mobilizing weapons, troops, fighters, equipment, etc. in a threatening manner; does NOT include the actual use of force (use a 6 Force category for that); this CAN include actions by state OR nonstate actors but applies to organized groups only (if an individual or spontaneous group does this, use 51 Threaten instead)
- 53 Reduce Relations: Withhold or stop providing benefits or cooperation in an effort to coerce someone to do something; this can include withdrawing from negotiations, imposing embargoes or sanctions, breaking diplomatic relations, withdrawing personnel
- 54 Suspend Rights: Suspend laws, the constitution, elections, etc. or impose curfews, censorship of media, or other restrictions of freedom

action01_5

- 00 NONE (Action was a different top-level category)
- 50 Coerce (unspecified)
- 51 Threaten
- 52 Exhibit Force Posture
- 53 Reduce Relations
- 54 Suspend Rights

4.2.6. FORCE. How did the Actor use force against the Target?

- 60 Force (unspecified): ["overpower"] Use only if you absolutely cannot identify an appropriate category below. Acts of physical or psychological violence, or the direct taking of something valued
- 61 Seize or Damage Property: Taking, damaging, destroying property or otherwise rendering it unusable; this includes rioting and vandalism; and this CAN be done by state OR nonstate actors
- 62 Detain or Deport: These are actions that can also be done by nonstate actors (e.g., taking hostages, kidnapping, or "disappearing" people) as well as by state actors. See NAVCO3 subcategories for 173, 174, and 181.
- 63 Repress or Abuse: This is direct physical violence that physically harms, traumatizes, or kills people
- 64 Engage in Violent Combat: This is direct physical violence involving at least two groups of people who physically harm, traumatize, or kill people in the opposing groups
- 65 Use Unconventional Mass Violence: This is direct physical violence, mass expulsions, ethnic cleansing, genocide, etc. against unarmed people; the Actors can be state or nonstate (which is the case with ALL subcategories)

action01_6

- 00 NONE (Action was a different top-level category)

- 60 Force (unspecified)
- 61 Seize or Damage Property
- 62 Detain or Deport
- 63 Repress or Abuse
- 64 Engage in Violent Combat
- 65 Use Unconventional Mass Violence

4.3. How confident are you in your responses to the questions about Action?

actionConfidence

- 2 = High confidence: evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = Medium confidence: evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge
- 0 = Low confidence: evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value

4.4. Notes on responses to Action questions.

Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about these variables or the values you chose for them.

actionNotes
Long answer text

5. Targets and Target Categories

5.1. Name (or describe) the Target.

Free-form text indicating the name of the Target (if it's an identifiable entity or person) or describing the Target (if no discernible name), to add specificity beyond the category codes (below).

target
Short answer text

5.2. What subcategory was the Target?

STATE TARGET

- 10 State Other: Use only if you absolutely cannot identify an appropriate category below. This includes members of any political faction, party, or other position who are currently in some position of authority but not covered by another category
- 11 Incumbents: The new actors who have been selected, self-selected, appointed, or elected to positions of authority in the state as a result of nonviolent action
- 12 Former Incumbents Included: Former regime actors who are still in (or returning to) positions of authority in the state, working with or under the new incumbents
- 13 Opposition Actors: Elected, appointed, or employed officials of political parties (or any allied organizations) who are not in the majority in government or represent minority factions in government
- 14 Civil Servants: Any non-elected, non-appointed, civilian government workers for the state at the national or federal level
- 15 Military: Members of the military services
- 16 Police: Members of a local or national police force or other non-military security services
- 17 Local State Actors: Any elected, appointed, or employed officials or civil servants at the subnational level (regional, provincial, county, local, etc.)
- 18 Foreign State Actors: Any foreign governments

NONSTATE TARGET

- 20 Nonstate Other: Use only if you absolutely cannot identify an appropriate category below. This can be any unarmed nonstate actor relevant to a civic mobilization not covered by another category
- 21 Former Incumbents Excluded: Former regime actors no longer in positions of authority in the state
- 22 Political Factions or Candidates: Members of any political party or non-party faction not currently in power and not covered by another category
- 23 Nonstate Armed Actors: Armed actors not categorized as state military or police and NOT acting spontaneously. Includes parastate, paramilitary, guerrilla, insurgent, organized criminal, gang, bandit, and other actors.
- 24 Businesses or Business Associations: Can include business owners and managers, acting in their individual capacities or organized via coalitions or associations

- 25 Workers or Unions: Can include individual workers, whether unionized or not, as well as worker collectives, unions and union leaders, etc.
- 26 Students or Youth: Any group that is predominantly (even if not exclusively) students, young adults, teenagers, or younger; actions by teachers would be classified here if they are acting with or on behalf of students (but under 25 Workers or Unions otherwise)
- 27 Religious Communities: Includes organized institutions, groups of believers or adherents, sects, factions, etc., whether acting on their own initiative or through an institution
- 28 Indigenous Communities: Includes individuals, organizations, native governments, communities, etc. who are or represent people whose heritage traces to the native peoples of the territory where the Actors live or where the Actions take place
- 29 Foreign Nonstate Actors: NGOs, IGOs, foreign businesses, and other foreign nonstate actors

target01

- 10 STATE Other
- 11 Incumbents
- 12 Former Incumbents Included
- 13 Opposition Actors
- 14 Civil Servants
- 15 Military
- 16 Police
- 17 Local State Actors
- 18 Foreign State Actors
- 20 NONSTATE Other
- 21 Former Incumbents Excluded
- 22 Political Factions or Candidates
- 23 Nonstate Armed Actors
- 24 Businesses or Business Associations
- 25 Workers or Unions
- 26 Students or Youth
- 27 Religious Communities
- 28 Indigenous Communities
- 29 Foreign Nonstate Actors

5.3. How confident are you in your responses to the questions about the Target?

targetConfidence

- 2 = High confidence: evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = Medium confidence: evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge
- 0 = Low confidence: evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value

5.4. Notes on responses to Target questions.

Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about these variables or the values you chose for them.

6. Event Data Sources

6.1. Did you use any sources beyond what is listed in Section 1 for the above responses?

Respond "yes" ONLY IF the source in Section 1 did not have enough explicit evidence to answer ALL of the questions in Sections 2-5, AND you therefore relied on one or more additional sources that DID have explicit evidence.

sourceEventYN

- Yes
- No

6.2. What was the most important additional source for your responses above?

Respond ONLY IF you answered YES to the question 6.1. Enter the title and the URL below (if no URL, enter an ID code or citation information: author, journal, etc.)

titleSource1events
Short answer text

idSource1events
URL, ID code, or citation information
Short answer text

6.3. Add one or two additional sources (ONLY if actually used).

Respond ONLY IF you answered YES to the question 6.1. Do not add ANY sources unless they contain explicit evidence that you actually used in the responses to Sections 2-5. Enter the titles and the URLs below (if no URL, enter an ID code or citation information: author, journal, etc.)

titleSource2events
Short answer text

idSource2events
URL, ID code, or citation information
Short answer text

titleSource3events
Short answer text

idSource3events
URL, ID code, or citation information
Short answer text

You're almost done!

Take a deep breath. Stretch your legs. Drink some water. You're doing great!

7. Attribute Data

7.1. ALIGNMENT. How aligned is the Actor to the new state Incumbents?

This will often be obvious for some Actor categories (e.g., Actor category "11 Incumbent" will obviously be coded here as 1), but useful to specify for others where alignment isn't necessarily predictable.

alignment
- 1 = Actor is aligned or associated with new incumbent
- 0 = Actor is not aligned with either the new incumbent or the old regime

- –1 = Actor is aligned or associated with old regime
    - Empty: There was no evidence in the source materials one way or the other

alignmentConfidence
How confident are you in your response to the question above?

- 2 = evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge
- 0 = evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value
- Empty: There was no evidence in the source materials

alignmentNotes
Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about this variable or the value you chose for it.
Long answer text

alignmentSource
Add a source (URL, ID, etc.) ONLY IF it was used to inform this variable and no other variable; any source used for two or more variables for this event should be listed in the last section of this form. No title is needed.
Short answer text

7.2. COHERENCE. How internally coherent is the Actor group?

coherence

- 0 = Actor is fragmented, with contradictory claims on leadership or authority
- 1 = pluralistic, with multiple sources of recognized leadership or authority
- 2 = unitary, with a single source of recognized leadership or authority
- Empty: There was no evidence in the source materials one way or the other

coherenceConfidence
How confident are you in your response to the question above?

- 2 = evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge

- 0 = evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value
- Empty: There was no evidence in the source materials

coherenceNotes
Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about this variable or the value you chose for it.
Long answer text

coherenceSource
Add a source (URL, ID, etc.) ONLY IF it was used to inform this variable and no other variable; any source used for two or more variables for this event should be listed in the last section of this form. No title is needed.
Short answer text

7.3. INCLUSIVITY. How democratically inclusive was the intent or tendency of the Action?

inclusivity
- 1 = Action attempted expansion of democratic "winning coalition"
- 0 = Action was meant neither to expand nor to contract democratic inclusion
- −1 = Action was maximalist or represented a contraction toward factions
- Empty: There was no evidence in the source materials one way or the other

inclusivityConfidence
How confident are you in your response to the question above?
- 2 = evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge
- 0 = evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value
- Empty: There was no evidence in the source materials

inclusivityNotes
Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about this variable or the value you chose for it.
Long answer text

inclusivitySource
Add a source (URL, ID, etc.) ONLY IF it was used to inform this variable and no other variable; any source used for two or more variables for this event should be listed in the last section of this form. No title is needed.
Short answer text

7.4. SPONTANEITY. How spontaneous was the Action?

spontaneity

- 0 = Action was planned, premeditated, or organized by Actor
- 1 = Action included a mix of organized and uncontrolled elements
- 2 = Action emerged spontaneously or without organization
- Empty: There was no evidence in the source materials one way or the other

spontaneityConfidence
How confident are you in your response to the question above?

- 2 = evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge
- 0 = evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value
- Empty: There was no evidence in the source materials

spontaneityNotes
Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about this variable or the value you chose for it.
Long answer text

spontaneitySource
Add a source (URL, ID, etc.) ONLY IF it was used to inform this variable and no other variable; any source used for two or more variables for this event should be listed in the last section of this form. No title is needed.
Short answer text

7.5. DIVERSITY. How diverse was the Actor: does it reflect the identity distribution of society?

diversity

- 0 = Actor group does not reflect the ethnic and other identity distribution of population

- 1 = Actor group includes some ethnic diversity but not equitably
- 2 = Actor group reflects ethnic and identity diversity of the population
- Empty: There was no evidence in the source materials one way or the other

diversityConfidence
How confident are you in your response to the question above?

- 2 = evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge
- 0 = evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value
- Empty: There was no evidence in the source materials

diversityNotes
Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about this variable or the value you chose for it.
Long answer text

diversitySource
Add a source (URL, ID, etc.) ONLY IF it was used to inform this variable and no other variable; any source used for two or more variables for this event should be listed in the last section of this form. No title is needed.
Short answer text

7.6. GENDER. How gender-balanced was the Actor?

gender

- 0 = Actor group does not reflect gender balance
- 1 = Actor group includes some gender balance
- 2 = Actor group reflects gender balance
- Empty: There was no evidence in the source materials one way or the other

genderConfidence
How confident are you in your response to the question above?

- 2 = evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge
- 0 = evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value

- Empty: There was no evidence in the source materials

genderNotes
Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about this variable or the value you chose for it.
Long answer text

genderSource
Add a source (URL, ID, etc.) ONLY IF it was used to inform this variable and no other variable; any source used for two or more variables for this event should be listed in the last section of this form. No title is needed.
Short answer text

7.7. REPRESSION. Did the state incumbents respond to this event with any degree of repression?

This variable refers in all cases ONLY to actions taken by the state incumbent (State Actor subcategory "11 Incumbents"). For example, in events where the Actor is nonstate, this variable refers to the state incument's response to the Actors' Actions. For events where the Actor is state but not subcategory "11 Incumbents", this variable still refers to the state incumbent's response to the events.

repression
- 0 = State incuments did not respond with repression (use only when a source says so explicitly)
- 1 = State incumbents responded using non-physical repression
- 2 = State incumbents responded using physical repression with no fatalities
- 3 = State incumbents responded using physical repression with fatalities
- Empty: There was no evidence in the source materials one way or the other

repressionConfidence
How confident are you in your response to the question above?
- 2 = evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge
- 0 = evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value
- Empty: There was no evidence in the source materials

repressionNotes
Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about this variable or the value you chose for it.
Long answer text

repressionSource
Add a source (URL, ID, etc.) ONLY IF it was used to inform this variable and no other variable; any source used for two or more variables for this event should be listed in the last section of this form. No title is needed.
Short answer text

7.8. FOREIGN INFLUENCE. Did a significant foreign actor attempt to influence the event? If so, how?

This variable refers to any involvement by significant foreign actor attempting to influence the event; it does not ask about who the foreign actor is supporting (that's the next question).

foreignInfluence
- 0 = no involvement
- 1 = statement
- 2 = economic support
- 3 = diplomatic support
- 4 = military support
- 5 = military intervention
- Empty: There was no evidence in the source materials one way or the other

foreignInfluenceConfidence
How confident are you in your response to the question above?
- 2 = evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge
- 0 = evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value
- Empty: There was no evidence in the source materials

foreignInfluenceNotes
Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about this variable or the value you chose for it.

Long answer text

foreignInfluenceSource
Add a source (URL, ID, etc.) ONLY IF it was used to inform this variable and no other variable; any source used for two or more variables for this event should be listed in the last section of this form. No title is needed.
Short answer text

7.9. FOREIGN SUPPORT. Was the foreign support (above) intended to benefit the Actor?

Code for categories of how foreign influence benefited or was intended to benefit the Actor

foreignSupport
- 1 = benefited Actor
- 0 = no benefit or harm
- -1 = harmed Actor or benefited opponents of Actor
- Empty: There was no evidence in the source materials one way or the other

foreignSupportConfidence
How confident are you in your response to the question above?
- 2 = evidence in source text was explicit and unambiguous (this is the preferred standard for all values)
- 1 = evidence in source texts was either ambiguous or not fully explicit and required some interpretation but no additional contextual knowledge
- 0 = evidence in source texts is implicit or ambiguous, but with additional contextual knowledge can reasonably be interpreted to support selected value
- Empty: There was no evidence in the source materials

foreignSupportNotes
Please note any ambiguities, clarifications, or dilemmas you feel a user of this dataset would need to know about this variable or the value you chose for it.
Long answer text

foreignSupportSource
Add a source (URL, ID, etc.) ONLY IF it was used to inform this variable and no other variable; any source used for two or more variables for this event should be listed in the last section of this form. No title is needed.
Short answer text

8. Attribute Data Sources

8.1. Did you use any sources beyond what is listed in Section 1 for your responses to the attributes questions?

Respond "yes" ONLY IF the source in Section 1 did not have enough explicit evidence to answer ALL of the questions in Sections 7, AND you therefore relied on one or more additional sources that DID have explicit evidence.

sourceAttributeYN
- Yes
- No

8.2. What was the most important additional source for your responses above?

Respond ONLY IF you answered YES to the question 8.1. Enter the title and the URL below (if no URL, enter an ID code or citation information: author, journal, etc.)

titleSource1attributes
Short answer text

idSource1attributes
URL, ID code, or citation information
Short answer text

8.3. Add one or two additional sources (ONLY if actually used).

Respond ONLY IF you answered YES to the question 8.1. Do not add ANY sources unless they contain explicit evidence that you actually used in the responses to Sections 7. Enter the titles and the URLs below (if no URL, enter an ID code or citation information: author, journal, etc.)

titleSource2attributes
Short answer text

idSource2attributes
URL, ID code, or citation information
Short answer text

titleSource3attributes
Short answer text

idSource3attributes
URL, ID code, or citation information
Short answer text

That's it!

Thank you so much for your speed and attention to detail!

Are you ready to submit (you can return to review later)?
If you are ready to submit, please enter your staff ID code here
Short answer text

* * *