Constellations of Fragility: an Empirical Typology of States

Sebastian Ziaja¹ · Jörn Grävingholt¹ · Merle Kreibaum¹

Published online: 24 May 2019
© The Author(s) 2019

Abstract

We present a typology of states that distinguishes constellations of state fragility based on empirical patterns. State fragility is here defined as deficiencies in one or more of three core functions of the state. These functions include violence control, implementation capacity, and empirical legitimacy. Violence control refers to the state's ability to manage the uses of violence within society. Implementation capacity refers to the state's ability to provide basic public services. Empirical legitimacy refers to the population's consent to the state's claim to rule. Employing three to four indicators per dimension for 171 countries over the period 2005–2015 and finite mixture model clustering, we find six dominant constellations that represent different types of state dysfunctionality.

Keywords Fragile states · Typology · Measurement · State performance

Introduction

The debate on state fragility poses a challenge to the academic community. Addressing the inability of states to perform for the benefit of their population, it brings issues to the table that have been scrutinized by political science for decades and longer. At the same time, state fragility is a term heavily propagated by the policy community in recent times, creating a high demand for scholarly advice (e.g., OECD 2015; The World Bank 2011). Unfortunately, the concepts of state fragility employed in the

The data presented in this article and visualization tools are accessible at <http://statefragility.info/>.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12116-019-09284-3>) contains supplementary material, which is available to authorized users.

✉ Sebastian Ziaja
sebastian.ziaja@die-gdi.de

¹ German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE), Bonn, Germany

policy arena proved hardly suitable for analytic use. For example, a large-N study on the origins of state fragility by Bertocchi and Guerzoni (2012) followed the World Bank Development Report 2011 (The World Bank 2011) and relied on the Country Policy and Institutional Assessment (CPIA) as a measure of state fragility. The CPIA, however, includes assessments of liberal economic policies that would not even fit the broadest definitions of state fragility discussed in the literature; it is thus not suitable for analyzing state fragility (Ziaja 2012: 48). One could of course discard the concept altogether and try to address the substantial problems it refers to with more established theories instead. We argue, however, that the broader concept of state fragility has an added value compared to narrower concepts, such as those employed in the literature on “limited statehood” (Krasner and Risse 2014): it mimics attempts of policy makers in the development and security community to bring order into the world of weak states and thus allows investigations into the impact of these constructs on policy and outcomes. We also believe that the concept of state fragility can be of analytic utility for explaining the perpetuity and diversity of underdevelopment and violent conflict in many parts of the world if it is operationalized with more rigor than in the policy debate (cp. Besley and Persson 2011).

The concept of fragility is “bringing the state back in” once more—to borrow the famous phrase coined by Evans et al. (1985). Their work focused on the state as an organizational structure that fares war, fosters economic growth, and must come to terms with civil society. This resonates well with issues addressed under the labels of “authority, capacity and legitimacy”—albeit with varying definitions—in the fragility debate (Brinkerhoff 2011; Call 2011; Carment et al. 2010; Grävingholt et al. 2015; Tikuisis and Carment 2017). But the empirical base of the earlier debate was largely restricted to studies of particular sectors in particular countries or regions (e.g., Wade 1990). The concept of fragility provides new impetus for two ideas that had been somewhat neglected: the interdependence between the three core functions of the state, and the systematic comparison of how states deal with this interdependence.

With this article, we aim to contribute to the conceptual and analytical debate on state fragility by providing an empirical typology of statehood. Starting from three core functions of the state (violence control, implementation capacity, and empirical legitimacy), we derive constellations of fragility using unsupervised learning techniques that help uncover groups in unlabeled data. Most traditional approaches towards classifying fragile states distinguish countries by their degree in fragility, expressed in a continuous index (e.g., Marshall and Cole 2014; The Fund for Peace 2014). They usually measure fragility across various dimensions but fail to acknowledge this multidimensionality in the aggregation step. We agree with Gisselquist (2015, p. 1270) that “unpacking state fragility and studying its dimensions and forms can help us to better develop and examine policy-relevant hypotheses about how aid-recipient states become more resilient.” This is where most of the analytical novelty and utility of the state fragility concept lies: maintaining its multidimensionality in the aggregate invites scholars to consider the joint effects of these dimensions. We suggest that empirically derived *constellations of fragility* provide an additional perspective on state fragility that previous measurement efforts have not been able to offer.

The multidimensional concept of state fragility used in this article builds on Call (2011). We draw on both observational and expert coded data to measure the three dimensions, and employ a model-based clustering approach to identify groups of similar countries, as proposed by us in Grävingholt et al. (2015). Tikuisis and Carment (2017) have also recently provided a typology based on some of these principles; their approach, however, has limitations in conceptualization, measurement, and aggregation, which we seek to overcome. Our aggregation method allows us to identify empirical groups (i.e., constellations of fragility) based on observable, continuous traits—in the absence of any direct indicator of group membership and without the need to manually (and to a certain extent arbitrarily) define thresholds. Similar techniques are being successfully used in biology to cluster species by outer appearance or genetic structure (e.g., Hausdorf and Hennig 2010) and in medicine to identify diseases based on symptoms (e.g., Martis et al. 2009). Applications in political science include a test of the varieties-of-capitalism typology (Ahlquist and Breunig 2012), the clustering of legislative speeches (Quinn et al. 2010), and the discovery of new rhetoric strategies of politicians (Grimmer and King 2011). Insights gained with these approaches show that introducing a level of detail between “one-size-fits-all” approaches and individual case studies can offer analytical advantages. In the case of state fragility, it allows to identify a finite set of constellations. These constellations are not just different in degree, but in kind, as they represent different types of societal and political equilibria. In the world of international politics, both time and resources are usually too scarce to start policy design for each individual case from scratch. Instead, the sheer variety of contexts that need to be addressed calls for a limited number of templates from which policy makers can pick the best-suited one in order to then tailor it to the individual case at hand. In that sense, typologies are not “old-fashioned,” as propagated by parts of the linear orthodoxy, but useful exploratory devices (cp. Collier et al. 2012).

In developing a typology at the nation-state level, we are aware that we need to rely on cross-national data—the quality of which is often unsatisfactory. The principal value of this exercise, however, is not in the precise measurement of any one country’s fragility characteristics at a given point in time. Instead, we are primarily interested in identifying general patterns of fragility that can be observed across the full sample of cases over time.

The Concept of State Fragility

State fragility is a concept that emerged in the 1990s when scholars noted that, with the demise of the Soviet Union, many countries experienced a power vacuum that threatened to destabilize whole world regions. Failed interventions in Haiti and Somalia showed that the USA and its allies were not able or willing to provide sufficient resources to stabilize these “collapsing” or “failing” states (Gros 1996; Helman and Ratner 1992; Zartman 1995). Policy makers worried even more about the dangers connected to state collapse after the terror attacks of September 11, 2001. Fear spread that “ungoverned” territories in Central Asia and elsewhere would serve as safe havens for international terrorists. The 2002 *National Security Strategy* of the

USA exemplified this new perception when stating: “America is now threatened less by conquering states than we are by failing ones” (The White House 2002, p. 1).

An appropriate response was certainly hampered by political and bureaucratic obstacles, but another central hurdle were conceptual disagreements (Faust et al. 2015). What was the nature of the dysfunction affecting countries such as Somalia or Haiti? In an attempt to better trace the issue, several international and non-governmental organizations set out to gauge fragility (e.g., Carment et al. 2010; Marshall and Cole 2014, The Fund for Peace 2014). These approaches provided one-dimensional indices or dichotomous indicators of state fragility. They failed to demonstrate, however, that cases with similar index scores represented homogeneous groups (Ziaja 2012, p. 52). This failure led some scholars to question the concept of fragility altogether (e.g., Bøås and Jennings 2005). Yet, given the persistence of protracted crises around the world, further efforts in understanding these dynamics and how one can react to them are of fundamental importance. In addition, the fragility of states continues to attract major policy attention. The European Union, for instance, emphasized in its 2016 Global Strategy that “[f]ragility beyond our borders threatens all our vital interests” (European Union 2016, p. 23).

Pospisil and Kühn (2015) have recently argued that among aid agencies the concept of “fragile states” had lost traction and had been superseded by a focus on “fragility and resilience” in a less state-centered manner. But their interpretation that donors have come to regard the state as less relevant is not convincing. Instead it appears that the new focus of aid agencies on state-society relations represents a changed view on what constitutes a strong state. Thus, the latest OECD Policy Guidance for donor support in situations of conflict and fragility, which has canonized resilience in the donor literature, is still centered on the axiom that “[e]ffective states matter for development” (OECD 2011, p. 11).

The concept of state fragility is a reminder that states are organizations whose survival depends on the fulfillment of critical functions. Understanding how the realization of these functions enables states to cope with internal unrest and with external shocks is crucial to development policy and beyond (OECD 2008). A useful approach towards measuring state fragility, we argue, would derive the conceptual foundations from the existing literature, strive for an empirical implementation of the concept that maintains its core assumptions, and employ a replicable and robust aggregation technique (cp. Gisselquist 2014).

Many authors writing on state fragility since the early 2000s have disaggregated the phenomenon into several dimensions: usually two (e.g., Kaplan 2014), three (e.g., Carment et al. 2010), or four (e.g., Rice and Patrick Rice and Patrick; see Ziaja 2012 for an overview). But a strong case has been made in favor of the three-dimensional approach: Brinkerhoff (2011, pp. 136–7) shows that three dimensions are necessary (and sufficient) to understand the central societal cleavages that tend to affect fragile states.¹ Based on this literature, we suggest to conceptualize fragility as constituted of deficiencies in three distinct, though interdependent, dimensions: *violence control*,

¹Brinkerhoff’s dimensions, labelled “effectiveness,” “legitimacy,” and “security,” are similar to those adopted in this article.

implementation capacity, and *empirical legitimacy*. As we argued in Grävingholt et al. (2015), each dimension represents a particular type of state-society relation and can be traced back to complementary strands of political theory.²

Violence control refers to the demonstrated ability of the state to manage the use of physical violence within its territory. A state that condones unauthorized violence risks losing its monopoly of violence against competitors. The process of taking control was described by Olson (1993) as a “roving bandit” becoming stationary in order to better extract taxes. From this perspective, the state is a corporate actor maximizing profit (cp. Tilly 1985). But such an arrangement must be of benefit not only for the ruler. Thomas Hobbes justified this idea of the state as the Leviathan as a means of ending anarchy, and thus protecting the population. Citizens can rely on the state to guarantee their physical integrity and to enforce set rules, thus laying the foundation for socio-economic activities. Looking at related approaches, both Eizenstat et al. (2005, p. 136) and Call (2011, p. 307) adopt such a human security perspective and define a lack of violence control as a “security gap.” Note that this definition does not exclude the possibility that states with high levels of violence control abuse this power against their populations.

Implementation capacity denotes the demonstrated ability of the state to provide basic services to its population. The idea of obliging the state to cater for the public may be attributed to John Locke, one of the fathers of the contractualist argument, binding both the state and society by a hypothetical contract (cp. Brinkerhoff 2011, p. 134). The scope of basic services provided by real states (and expected by their populations) varies substantially (Fukuyama 2004). In a largely enlightened and increasingly globalized world, however, there exists an almost universal set of minimal services that any state—even the most authoritarian or libertarian one—is expected to provide. In political philosophy, the emergence of these expectations has been explained with reference to Rawls’ (1971) “veil of ignorance” or Buchanan’s (1975) “postconstitutional contract.” A minimal set of services encompasses those that improve the life chances on a very basic level, such as primary education and rudimentary health care. Call (2011, p. 306), using a very similar definition of capacity, refers to these as “core public goods.” This makes his and our definitions much narrower than the one employed by Carment et al. (2010), who include economic, demographic, and environmental features in this dimension. Restricting our definition to the core features, we argue that failure to perform in one or more of these areas diminishes life chances for large parts of the population, and would thus threaten the abovementioned contract. Note that the definition of implementation capacity presented here differs from that employed in the classical “state capacity” literature (Saylor 2013 provides an overview). It may thus seem unfortunate to employ a similar term with a different meaning. But it is employed so disparately in the academic literature (ranging from extractive capacity to implementation and service provision)

²The insistence on the state as a prerequisite for development has made critics argue that the concept of state fragility is promoting a Western-based blueprint of the “democratic and capitalist state governed by the rule of law” (Krasner and Risse 2014: 548). However, we explicitly avoid to define fragility in terms of institutions and opt for a perspective of “functionalist equivalence” (Draude 2007): It matters *whether* the state performs but *not how* exactly it gets there.

and at the same time so established in the policy-oriented literature we build upon (e.g., Call 2011; Eizenstat et al. 2005) that we decide to maintain it.

Empirical legitimacy, the third dimension of state fragility, refers to the degree to which the state enjoys the consent of the population to its holding and exercising political power.³ While a state's legitimacy may be less tangible than violence control or implementation capacity, its importance from a perspective of fragility is well-established not only in the academic literature but also in the policy world.⁴ In line with Max Weber's ([1919] 2010) classical distinction of traditional, charismatic, and rational-legal types of rule, we conceive of legitimacy as a resource that can be derived in various ways. Yet, it depends crucially on the belief among the ruled in the rightfulness of the fact that the state bodies claim the right to rule. By this definition, although political representation and empirical legitimacy certainly correlate, both democratic and undemocratic regimes can be legitimate—a claim that again differentiates our approach from that taken by Carment et al. (2010). And even if there is discontent with the current condition of a political regime, the state itself may still retain a certain amount of legitimacy, based on the idea of the nation. The nation can forge a sense of identity that has been explored by the constructivist literature (Anderson 1991). Like for the other dimensions, we focus exclusively on the demonstrated ability of the state to fulfill a function internally. Tikuisis and Carment (2017) include “international recognition” in their definition of legitimacy, blurring the distinction between states that lack support within (e.g., Syria) and those that lack support in the international community (e.g., Taiwan). While international recognition is undoubtedly an interesting field of study, it is separate from the focus on the domestic functionality of a state that guides our analysis in line with most of the academic and policy-oriented state fragility literature.

More recently, efforts have been made to analyze violence control, implementation capacity, and empirical legitimacy of governance systems with respect not only to the nation state but also to sub-national, international, and non-state actors in a given territory (Krasner and Risse 2014; Risse and Stollenwerk 2018). This literature is an important and welcome addition to the research on fragile or limited statehood. It is important to note, however, that its object of investigation differs from the one addressed by this study. In taking a narrower focus on the state at the national level, we do not deny the importance of other actors but merely limit the contribution of this research to the character and degree of fragility of the nation state.

Obviously, all three dimensions of statehood interact in various ways. It seems particularly plausible to assume that substantial deterioration in any one dimension will sooner or later lead to concomitant deterioration in the other two. Moreover, and as a mirror image to the logic of a vicious cycle, an argument could be made

³In this, we follow Levi et al. (2009, p. 354): “Legitimacy denotes popular acceptance of government officials’ right to govern”; and Gilley (2006, p. 500): “a state is more legitimate the more that it is treated by its citizens as rightfully holding and exercising political power.” We do not follow these authors’ operationalizations, though, for reasons explained in the following section.

⁴In a recent literature review, Risse and Stollenwerk (2018, p. 404) argue that “(e)mpirical legitimacy in terms of social acceptance [...] constitutes a key condition for effective governance in areas of limited statehood.” On the policy side, a major OECD study found that “[a] lack of legitimacy is a major contributor to state fragility because it undermines authority, and therefore capacity” (OECD 2010, p. 15).

for a virtuous cycle: improve on one dimension and the others will follow soon. In fact, however, recent empirical research suggests that the dyadic interaction effects between dimensions of statehood may be more complicated than the vicious and virtuous cycle arguments predict. McLoughlin (2015), e.g., has demonstrated that better service delivery—a typical element of implementation capacity—does improve the empirical legitimacy of a fragile or conflict-affected state under certain circumstances; but not necessarily and by far not always. Instead she found that “there is no straightforward alignment between objective service outputs and legitimacy gains” (McLoughlin 2015, p. 352). Legitimacy, she argues, is too high a degree socially and normatively constructed to merely be a function of state performance.

Generating the Dimension Scores

Any attempt to operationalize these three dimensions over space and time requires a considerable amount of compromise. One particular challenge in choosing indicators is the interdependence of different functions of statehood. We thus need to be careful to avoid these interdependencies in our operationalization. For example, the ability of a state to tax its population requires a high degree of violence control (enforcing compliance) or empirical legitimacy (equivalent with voluntary compliance) and is at the same time the prerequisite for projecting implementation capacity (providing public goods).

What increases the attribution problem even further is that some states are unwilling to strictly enforce laws (including taxation) in the hope of obtaining popular consent (Holland 2016), and the occasional availability of alternative revenues from natural resources or foreign aid. Our selection of indicators to measure state functions must strike a careful balance between conceptual fit (validity), measurement precision (reliability), and availability (coverage). We discuss our choices below.

In order not to truncate our sample artificially, we include all independent countries with at least 250,000 inhabitants in our universe of cases—and not only those suspected to be “fragile” based on whatever prior knowledge available. We aim at measuring our latent state functions directly, e.g., the demonstrated ability of a state to implement policy (implementation capacity), but we often have to resort to observable outcome variables to imperfectly proxy these latent concepts, e.g., using the level of public service provision. Where available, expert assessments complement such proxy indicators, allowing us to capture dysfunctionalities that do not show up in observable indicators, such as the latent inability of states to control their territory in the absence of actual violence.

The *violence control dimension* represents the state’s ability to mute competing claims to the monopoly of violence and excessive manifestations of violence. We draw on two proxy variables to measure the level of violence control at the disposal of the state. One is *battle-related deaths*.⁵ This includes all casualties directly related to combat occurring within the territory of a country. The measure reflects the

⁵The [Supplementary File](#) provides a complete list of data sources for this and all following indicators.

intensity of internal and external attacks on the integrity of a state and thus the degree to which the state faces organized (but only acute) challenges to its monopoly of violence. Whereas war size is usually defined by absolute battle deaths, we employ battle deaths per 100,000 inhabitants because this better mimics the impact violent conflict has on a country's population. The second observational indicator of violence control is *homicides*, i.e., “unlawful death purposefully inflicted on a person by another person” (UNODC 2013, p. 9). Individual instances of homicide do—in the vast majority of cases—not stem from explicit challenges to the dominance of the state. But widespread lethal crime can be considered an indicator of organized crime in conflict with governing authorities, i.e., a systemic malfunction affecting the state's claim for dominance. In addition to these observable count measures of lacking state control, the Bertelsmann Transformation Index (BTI) provides a direct expert assessment which is better able to detect latent conflict: the BTI *monopoly of violence* indicator (BTI 2016: 16).

The *implementation capacity dimension* represents the state's ability to carry out policies. While the classical state capacity literature is agnostic about what this capacity is being employed for in detail, the state fragility literature is explicit about the state's obligation to provide something to the people in return for their obedience. This something may range from the minimalist “night-watchman” state to an extensive welfare state. We opt for a rather minimalist definition that is restricted to assisting citizens with basic life chances. These include the protection from (relatively easily) avoidable harmful diseases, a basic education that allows for an active participation in social and economic activities, and a basic administration that regulates social and economic activities sufficiently to increase collective gains and avoid massive negative externalities. Our proxies for disease control are the share of the population with access to improved drinking *water* sources and under-five mortality per 1,000 births, hereafter *child mortality*. Our education proxy is the rate of *primary school enrollment*. These are all outcome measures that may also be influenced by other actors, so we require an additional corrective to assess whether the state's bureaucracy itself is actually less capable than it seems. BTI *basic administration* provides such a corrective. It is an expert-based assessment on the existence of fundamental structures of a civilian administration, such as a basic system of courts and tax authorities (BTI 2016, p. 17). Other approaches to measure core implementation capacity rather than public good outcomes have been proposed, but none of these is available with global coverage over a sufficient number of years (e.g., Lee and Zhang 2017).

Legitimacy is notoriously difficult to measure (von Haldenwang 2017; Weatherford 1992). In line with our conceptualization of *empirical legitimacy* as the acceptance of state rule, we are explicitly not aiming for assessing normative legitimacy, i.e., the extent to which the state's claim to rule conforms to a predefined set of norms. Unfortunately, no valid and reliable survey data of sufficient coverage on perceived legitimacy exists (cp. Call 2011: 308). The World Values Survey (WVS) provides data only about seven percent of country-years covered by our sample.⁶ It

⁶The [Online Appendix](#) demonstrates the lack of coverage in detail.

would require an imputational overstretch to use this data. Nonetheless, Gilley (2006) has used the WVS to present one of the few rationalizations of empirical legitimacy across a significant number of countries. Yet even his study does not cover more than 72 countries.⁷ Levi et al. (2009) have used Afrobarometer survey data to analyze the effect of trustworthiness of government and procedural justice on legitimacy. However, Afrobarometer and its siblings in other continents do still not provide sufficient coverage across time and space due to insufficient survey frequency. In addition, Levi, Sacks, and Tyler limited their analysis to “[c]ountries involved in a transition to democracy” (2009, p. 370), rightly assuming that in these contexts survey data would yield a reliable representation of respondents’ actual beliefs. Under conditions of a repressive government with an elaborate system of surveillance and control, by contrast, such an assumption would be more than daring. In the absence of reliable survey data, our second best option is thus to draw on indirect indicators of legitimacy. One of these is repression expressed in state-sponsored human rights violations. Due to its high cost, outright repression is a state’s last resort. It can thus serve as a proxy indicator, as Dogan (1992, p. 120) notes: “Theoretically, the lower the degree of legitimacy, the higher should be the amount of coercion. Therefore, in order to operationalize the concept of legitimacy it is advisable to take into consideration some indicators of coercion, such as the absence of political rights and of civil liberties.” We employ a new, continuous meta index of *human rights* protection developed by Fariss (2014). A similar reasoning applies to the cost of restricting *press freedom*. It will only be attempted when free media would undermine the state’s ability to claim the support of the wider population. We employ Freedom House’s “Freedom of the Press Data” to measure press freedom. Finally, a more legitimate state can be expected to drive fewer citizens into emigration, e.g., through political persecution. Even if people have no possibility of expressing their discontent publicly, they usually still have the option of “exit” (Hirschman 1970). The number of *asylums granted* in other countries per 100,000 inhabitants in the sending country is a good indicator for politically (rather than economically) motivated exit. To be sure, none of these indicators measures empirical legitimacy directly, and none of them is a perfect representation of the underlying concept. Yet, they jointly represent conditions of which at least one can be expected to be present in any state struggling with achieving domestic legitimacy. Hence, for want of better options, we consider this set of indicators the best approximation of empirical legitimacy available with sufficient coverage.

Some of our indicators do not report data for every country year in our sample. In the case of homicides, for example, reporting is incomplete for many poor countries. BTI data is only published biannually. To close these gaps, we linearly interpolate missing data points within countries. Where data at the beginning or the end of a time series are missing, we extrapolate the latest available score. Table 1 shows what share of observations is imputed for each indicator, and how many years we extrapolate, if necessary. Note that there may still be missing data for some country years if missing

⁷In addition, some of Gilley’s results, such as a massive contrast between Azerbaijan (one of the highest scores) and Russia (lowest of all), suggest serious issues with the reliability of his measures in certain types of authoritarian contexts (Gilley 2006, pp. 512–513).

Table 1 Imputation, truncation, and transformation of the indicators

Indicator	Imputed	Years extrapol.	Lower	Upper	Logged	Inverted
Violence control						
– Battle deaths per 100,000 inh.	0%	0	0	2	Yes	Yes
– Homicides per 100,000 inh.	30%	7	0	90	Yes	Yes
– BTI monopoly of violence	39%	0	0	10	No	No
Implementation capacity						
– BTI basic administration	40%	3	0	10	No	No
– Child mortality per 1000 births	0%	7	2	300	Yes	Yes
– Primary school enrollment rate	20%	3	0.3	1	No	No
– Access to improved water source rate	1%	3	0.3	1	No	No
Empirical legitimacy						
– Asylums granted per 100,000 inh.	0%	0	0	20	Yes	Yes
– FH freedom of the press	0%	3	5	110	No	Yes
– Human rights protection score	0%	0	-3	3	No	No

observations lie outside the extrapolation ranges we define, and if countries have no single data point for a particular indicator (e.g., most OECD countries for the BTI indicators). Table 2 shows the number of observations available after imputation. The [Supplementary File](#) provides full details on our imputation procedure and discusses its justification.

Table 2 Summary statistics: dimension scores and imputed and transformed indicators

Dimension and indicator	<i>N</i>	Median	Mean	SD	Min.	Max.	Impact
Violence control	1885	0.60	0.58	0.22	0.00	1.00	
– Battle deaths	1885	1.00	0.94	0.19	0.00	1.00	4
– Homicides	1882	0.62	0.61	0.20	0.00	1.00	83
– Monopoly of violence	1626	0.80	0.78	0.22	0.10	1.00	14
Implementation capacity	1885	0.55	0.53	0.25	0.00	0.99	
– Basic administration	1649	0.70	0.71	0.24	0.10	1.00	9
– Child mortality	1885	0.55	0.55	0.24	0.08	0.99	86
– Primary enrollment	1671	0.93	0.86	0.18	0.01	1.00	2
– Water access	1867	0.90	0.80	0.23	0.00	1.00	3
Empirical legitimacy	1885	0.49	0.50	0.22	0.00	0.96	
– Asylums granted	1885	0.93	0.83	0.22	0.00	1.00	7
– Press freedom	1885	0.56	0.57	0.23	0.10	0.96	47
– Human rights	1885	0.57	0.58	0.20	0.07	1.00	45

SD, standard deviation; *Impact*, percentage of dimension scores determined by indicator

Temporal coverage: 2005–2015; Scale ranges from 0 (low performance) to 1 (high performance)

In order to combine the information across the indicators into dimension scores, we transform all raw data to scores ranging from 0 to 1, where higher values imply better outcomes. This is done by first truncating the raw variable scores at pre-defined lower and upper bounds. This step is necessary to avoid that extremely large values dwarf the differences between other countries in this dimension. We calibrated these extremes so that variables that best represent each dimension determine the lion's share of each dimension's scores. These variables are *homicides*, *child mortality*, *press freedom*, and *human rights*. Empirically, they exhibit sufficient amounts of exploitable variance in most countries in the world (unlike, e.g., battle deaths, which is often zero). Conceptually, these outcome variables proxy a deficiency of the state in its respective core function. The goal is thus not to normalize each indicator, but to give it a distribution that translates into dimension scores that correspond with their concept.

The chosen lower and upper bounds for truncation are listed in Table 1; the resulting impacts are listed in the last column of Table 2.⁸ After truncation, all variables are re-scaled to a zero-to-one scale. Some of the truncated and standardized indicator scores are strongly skewed, with very low frequencies at higher values. We assume that marginal effects decrease with higher values and thus take their logarithms (and bring them back to the zero-to-one scale). In a final step, we align all variables to range from their worst to their best extremes, inverting variables where necessary. Table 1 indicates how each indicator was treated in the transformation step.

A crucial question is now how to aggregate indicator scores within each dimension of fragility. The most widespread approach in index building is taking averages. This approach, however, has weak theoretical underpinnings. Why, for instance, should the absence of drinking water be made up for with higher enrollment rates? And if so, to what degree? Following Goertz (2006, pp. 128–131), we combine the transformed scores of our indicators with a “weakest link approach”: the score of each dimension per country-year is determined by the lowest value among the available indicators. Should less than two indicators be available, no dimension score is calculated.⁹

For example, a country with standardized scores of 1.0 for battle deaths, 0.5 for homicides, and 0.2 for the BTI assessment of the monopoly of violence will receive a violence control score of 0.2. The idea is that even if there is no civil war causing battle deaths, and even if reported homicide rates are rather average, there must be a reason for such a low expert assessment of the monopoly of violence that is not captured with the former two indicators. This reason could be severely under-reported homicide rates, or a latent threat to stability that does not yet translate into battles or violent crime. Our approach thus prevents the undesired effects of compensation

⁸We experimented with a large range of reasonable lower and upper bounds for most of our variables. The overwhelming majority of these alternative datasets produced clustering results that are very similar to our final result.

⁹Note that the more indicators are used in a dimension, or the more indicators are available for a particular country-year, the more likely is this dimension or this country-year to have a lower score than one drawing on less indicators. As more capable states are more likely to have complete data, however, this bias is unlikely to affect our results severely; it would also exist—to a lesser or larger degree—for other aggregation rules.

(Munck 2009, p. 32). When calculating dimension scores as averages, a country that experiences more severe civil war battles could set off this deterioration by achieving lower criminal murder rates. Such a trade-off is not a valid translation of our concept of violence control.¹⁰ The weakest-link approach is equivalent to considering each variable a necessary component of a functioning state in the respective dimension. Table 2 shows descriptive statistics of the imputed and transformed data and of our three dimension scores. For clarity, the [Supplementary File](#) describes the entire transformation procedure in mathematical notation.

Identifying Constellations of Fragility

Once we have generated the scores for each of the three dimensions of statehood in any given country year within our sample, we can turn to the task of identifying constellations of fragility across the dimensions. The clustering exercise shall answer how many distinctive constellations (or groups) exist and provide the properties of these groups. There is no need to attempt a statistical proof that the data has structure at all, i.e., that there is more than one latent cluster, since the clustering is externally motivated by the desire to bring order into the phenomenon of state fragility (cp. Everitt et al. 2011, p. 262). Our focus is on finding the *best possible clustering* that differentiates groups of states, given our data.

In order to increase the number of observations for the clustering exercise, we pool all country years in our sample. Pooling country years is equivalent to disregarding the temporal dependence between repeated observations for one country. We are thus asking: which constellations of fragility have ever existed over the entire period under investigation? Constellations of fragility (our groups) are hence assumed to be constant within our sample. For the short period of time we observe, we find this assumption defensible. Note that this does not mean that country classifications are fixed. Individual countries can move between groups if their characteristics change from one year to the next.

The number of groups we expect to obtain is also driven by our research question. Fewer than four groups would not provide sufficient variation for a substantially interesting interpretation. Three groups would simply provide one cluster of good performers, one of bad performers, and one in-between—similar to the anocracy category of the Polity typology (Marshall et al. 2016). More than ten groups could not be handled in any practical application. We thus aim to find the *best fitting clustering solution* which arranges the observations into *four to ten clusters*. As Grimmer and King (2011) argue, considerations of utility should prevail when selecting an optimal clustering (and may even trump statistical fit).

We employ finite mixture modeling (Fraley and Raftery 2002) to detect dominant constellations of fragility within our data, i.e., groups of countries exhibiting

¹⁰The same logic obviously also precludes employing factor analysis or other dimension-reducing techniques in this step. We are not interested in whether the indicators in one conceptual dimension load on one dimension empirically; they rather complement each other. This design also preempts the use of automatic variable-selection techniques.

similar combinations of strengths and weaknesses in the different dimensions. The underlying statistical assumption of this approach is that scores within the individual dimensions will be distributed normally within groups. The model is fitted simultaneously to all three dimensions, in an attempt to find the multivariate normal distributions that best describe the data for a given number of groups. By comparing measures of fit between solutions with differing numbers of groups, we can also determine the optimal number of clusters. In other words, we are asking the algorithm to help identify how many “clouds” of countries that are similar in terms of violence control, implementation capacity, and empirical legitimacy can be found in the data. And we are asking what location and shape these clouds have, i.e., the average scores and spread of the three dimension variables.

Mixture model clustering is still less common in political science than older methods such as hierarchical or k -means clustering. The latter is used, for example, by Tikuisis and Carment (2017). k -means, however, simply constitutes a special, restricted case of mixture modeling (Vermunt 2011). Employing a full mixture model provides various advantages. It allows to specify the shape that clusters can assume and to restrict parameters, preventing excessively flexible specifications. This is useful since we aim at obtaining compact clusters that do not spread widely over individual dimensions. Otherwise, it would be hard to derive meaningful conclusions on the interdependence of the three core state functions. Mixture models also allow us to calculate the probabilities of observations belonging to a particular cluster. This “soft classification” thus incorporates the uncertainty inherent in the process, instead of a “hard” classification that would simply assign binary class indicators (such as k -means). And finally, they allow us to draw conclusions about the number of clusters that best represents the variation on the data, using goodness-of-fit measures.

Following the notation of Scrucca et al. (2016: 291), the equation we optimize to find the best clustering solution for a given number of mixture components G is

$$f(x_i; \Psi) = \sum_{k=1}^G \pi_k f_k(x_i; \theta_k),$$

where $x = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ is a sample of n observations,¹¹ $\Psi = \{\pi_1, \dots, \pi_{G-1}, \theta_1, \dots, \theta_G\}$ are the parameters of the mixture model, $f_k(x_i; \theta_k)$ describes the k th component density for observation x_i with parameter vector θ_k , and $(\pi_1, \dots, \pi_{G-1})$ are the mixing probabilities that add to 1. The model is estimated by applying the expectation-maximization algorithm—a common maximum-likelihood estimator—to the corresponding log-likelihood function. As most model-based clustering approaches, we assume that the components follow a multivariate Gaussian distribution: $f_k(x; \theta_k) \approx N(\mu_k, \Sigma_k)$, where μ are the mean vectors and Σ the covariance matrices that determine the permissible shapes of the components.¹²

¹¹We obviously violate the assumption that our observations are independent and identically distributed by pooling the country years, but as we argue above, we find this decision reasonable for our application.

¹²Recall that our dimension indicators are bounded. However, hardly any of our observations reach the bounds, which is why the bounds do not interfere with the normality assumption.

We consider three model specifications to determine the shapes Σ of our fragility constellations. These variations are discussed in detail in the [Supplementary File](#). Here, we focus on our preferred shape $\Sigma_{EII} = \lambda I$, where λ is a scalar and determines the volume of the tri-axial ellipsoids representing the clouds of data points that constitute the groups. Variable I represents the identity matrix, restricting the multivariate normal distributions that constitute the ellipsoids to have identical spread in all directions, resulting—in our three-dimensional application—in spherical group properties, or “circular clouds.” Thus, all groups are equally shaped, and of approximately equal size in terms of their standard deviations across all dimensions (but not necessarily equal in terms of the number of countries). This prevents that groups either spread widely over particular dimensions or that individual countries with rare score combinations are identified as separate groups. As suggested by Scrucca et al. (2016), we refer to this specification as the “EII” specification.

Once we confront our models with data, we can calculate a statistical measure of model fit to assess solutions with varying numbers of groups.¹³ Our criterion of choice is the integrated complete-data likelihood criterion (ICL; Scrucca et al. 2016: 297). As the more commonly used Bayesian Information Criterion (BIC), the ICL penalizes models for the number of parameters. Other than the BIC, the ICL also penalizes cluster overlap. It thus helps the researcher select a specification that fits the data well and identifies groups that are clearly distinguishable, and thus more useful. Figure 1 shows the ICL scores for the substantively interesting range between four to ten groups. Since extreme outliers may interfere with our normality assumption, we remove them from our sample, leaving us with 1866 observations. We define outliers as the one percent of observations with the highest Mahalanobis distance from the dimensions’ means.¹⁴

The ICL reaches its maximum at ten within our range of desired solutions. More groups seem to better model the underlying data structure. Nonetheless, the statistic also reaches a local maximum at six clusters, suggesting a suitable solution that is also rather parsimonious. Extensive robustness checks in the [Supplementary File](#) show that while other finite mixture specifications and other clustering methods (k -means and hierarchical clustering) all tend to suggest that more clusters better represent the data, this inflation of clusters is a common phenomenon in large datasets. Many variations tested also suggest local maxima in the measures of fit for six groups, however. We thus find it justified to employ the parsimonious variation with six spherical groups as our main specification.

Among the mixture models, the EEI specification has the best fit for this number of clusters. When we compare how countries are classified in the competing mixture model solutions with similar numbers of groups (“EEE-5” and “EEI-7”; see the [Supplementary File](#) for details), we find that some clusters of the most parsimonious solution EII-6 are either joined two-to-one or split into two, while others remain

¹³All calculations were performed using the statistical environment R (R Core Team 2019). The [Supplementary File](#) lists the packages that were employed.

¹⁴The Mahalanobis distance is a multidimensional generalization of the standard deviation and thus appropriate for detecting outliers across our three continuous dimension scores.

Fig. 1 Model fit indicated by the Integrated complete-data likelihood criterion (ICL)

stable. Between 69 and 98% of countries are classified into the same group or into one of two groups that the original group has been split into (see Tables A4 and A5A in the [Supplementary File](#)). Hence, these specifications do not create entirely new constellations, which supports our preference for the most parsimonious solution. In substantive terms, this model provides sufficient disaggregation for our purpose.

The [Supplementary File](#) also provides a comparison of the six-group results of *k*-means and hierarchical clustering with our favored mixture model result. It shows that for both alternatives, 87% of all observations are classified in the most comparable clusters. Jointly, these robustness checks bolster our trust that the EII-6 solution is a good representation of latent fragility constellations.

A Typology of States

To classify countries based on our estimates, we assign each country-year to the cluster with the highest probability. Since we also provide the probabilities of belonging to the other clusters, it is possible to employ alternative assignment rules where necessary. For example, one could set a minimum probability of .5 (i.e., a higher probability than all other options combined; 97% of all observations pass this threshold), or even .9 (65% of all observations) for a country-year to be assigned to a cluster.

A convenient way of presenting the properties of the resulting fragility constellations are boxplots (Fig. 2).¹⁵ For ease of reference, the constellations are labelled A through F. Constellations A and F constitute the poles, displaying the worst and best performances across all dimensions. Constellations B through D perform particularly bad in one dimension each, on average: among these three, B has the worst score

¹⁵In addition, the [Supplementary File](#) contains the actual model parameters including bootstrapped standard errors (Table A8) and detailed descriptive statistics for all groups.

Fig. 2 Distribution of dimension scores within fragility constellations

in violence control, C the worst score in implementation capacity, and D the worst score in empirical legitimacy. Constellation E does not perform very badly in any dimension, but it does not reach the levels of constellation F. The ordering does not imply that constellations further to the right are necessarily “better” than those to the left. Only for constellations A, E, and F, there is a clear rank order across all dimensions: F is better than E is better than A. Constellations B through E, by contrast, rank differently in different dimensions; they are “unrankable.” This shows how our typology is able to disentangle the “messy middle” where one-dimensional aggregation procedures, which allow for reciprocal compensation of all indicators, project very different constellations onto the same scores (cp. Gutiérrez Sanín et al. 2013, pp. 312, 317).

While our result looks “neat” in the sense that severe deficiencies occur in either all or only one dimension, it is not trivial. Deductive approaches would most likely have arrived at different constellations, such as an eight-fold typology derived from a three-dimensional two-by-two-by-two table that allows all combinations of low and high performance across three dimensions. We show that some of these theoretically feasible constellations (such as a combination of low control, low capacity, and high legitimacy) do not form stable clusters.

We name the constellations such that the labels describe their most pronounced features:

- (A) *dysfunctional* in all dimensions;
- (B) *low-control* despite decent implementation capacity;
- (C) *low-capacity*, but rather decent control of violence;
- (D) *low-legitimacy*, despite decent violence control and implementation capacity;
- (E) *semi-functional* in all dimensions;
- (F) *well-functioning* in all dimensions.

The boxplots also show how many country years populate each constellation. Dysfunctional and low-control states constitute the smallest groups (with 5 and 10

percent of all country years). Low-capacity states constitute the largest group (with 30% of all country-years).

Presenting these “fragility constellations” does not imply that we propose to stretch the meaning of the attribute “fragile” to all groups. This is especially evident for the well-functioning states. Instead, this group provides a useful benchmark for assessing the performance of the other fragility constellations. We suggest that countries belonging to the other constellations face particular challenges related to their statehood and that the respective extent and configuration of these challenges differ substantially between the groups. While detailed policy implications at the country level will require additional analyses, the types offer valuable intermediate level information between case studies and one-size-fits-all approaches by giving policy makers an instant idea of the directions that change to the better should take (Grävingholt et al. 2015).

Typical examples of countries that are classified as dysfunctional in 2015 are the Central African Republic, Libya, and Somalia. Low-control states include Guatemala, Jamaica, and South Africa; low-capacity states, Haiti, Togo, and Zimbabwe; and low-legitimacy states, Belarus, Saudi-Arabia, and Turkey. The group of semi-functional states comprises Cape Verde, Panama, and Peru. Examples of well-functioning states are Austria, Japan, and Slovenia. Countries that were particularly uncertain to belong to any one group include Bolivia (semi-functional or low-capacity), Singapore (well-functioning or low-legitimacy), and the USA (semi-functional or well-functioning).

The advantage of a more disaggregated picture “in the middle” of the fragility syndrome becomes clearer when our constellations are compared to The Fund for Peace’s (2014) “Fragile States Index” (FSI; formerly “Failed States Index”) for all country-years over the time period 2005 to 2015. Figure 3 shows that the FSI considers low-control, low-legitimacy, and semi-functional states to be equally fragile—their boxes overlap. This is due to the fact that the FSI collapses its 12 dimensions of state fragility into an aggregate index, allowing mathematical compensation between issues that can hardly be put on the same scale from a theoretical point of view (cp. Munck 2009, pp. 30–35).

A better understanding of the intricacies of state fragility and its dominant manifestations should help improve policy responses. Knowledge of real-world types of

Fig. 3 Distribution of the Fragile States Index (FSI) over fragility constellations, 2005–2015

Fig. 4 World map of fragility constellations, 2015 (see online copy for color version)

fragility will of course not solve all challenges. Individual country data show that some states in the middle range of fragility face gaps in more dimensions than the one that dominates their type. And in any case each country still needs to be analyzed in its own right. But fragility types can serve to start the analysis from a better basis than the sweeping general assumption that a state is “fragile.”

The map in Fig. 4 gives an overview of the regional distributions of fragility constellations in the year 2015. Tables A9 through A18 in the [Supplementary File](#) provide detailed descriptions of group dynamics. One interesting development when looking at group sizes over time is that the number of countries in low-capacity constellations has been clearly declining, while the number of countries in low-legitimacy and well-functioning constellations has been increasing slightly. This is mainly due to the positive trend in implementation capacity scores over the past 10 years (see Fig. A5 in the [Supplementary File](#)).

The transition plot in Fig. 5 shows which constellations countries have transitioned from and to—if they have—in any year between 2005 and 2015.¹⁶ In the transition plot, thicker and darker lines represent more transitions. Three dominant pairs of mutual interchange emerge:

1. dysfunctional (A) and low-capacity states (C) (e.g., Haiti moves back and forth);
2. low-control (B) and low-legitimacy states (D) (e.g., Libya);
3. and low-control (B) and semi-functional (E) (e.g., Georgia).

The dominant exit for low-capacity states is to low-legitimacy states, but they seem to fall back less frequently than in the pairs listed above. Only low-legitimacy and semi-functional states ever manage to transition directly to well-functioning

¹⁶Tables A17 and A18 in the [Supplementary File](#) specify the exact number of transitions within each directed pair and list all transitions that have occurred.

Fig. 5 Transitions between constellations, 2005–2015

states (e.g., low-legitimacy Bosnia-Herzegovina and semi-functional Estonia to well-functioning). Well-functioning states could—for the sample period—be considered an endpoint, since only few countries ever exit this group (Estonia, North Macedonia, and Montenegro), and none of these classified as well-functioning with high certainty.

The transition plot also implies that increases in implementation capacity only occur when violence control is high. Dysfunctional and low-control states rarely if ever improve their implementation capacity. At first sight it may seem that capacity increases even in low-legitimacy states—an argument common to proponents of authoritarian development. A closer look at the countries that do transition from low-legitimacy to well-functioning, however, attenuates the argument: they include Estonia, Romania, and former Yugoslav republics, all of which are under strong influence of the European Union—or have even become members during the examined time period.

Conclusion

This article argues that fragility constellations are better suited to investigate state fragility than one-dimensional indices or classifications derived from such indices. Our empirical typology shows that fragile states come in different types that also perform in very different ways, despite being given similar fragility scores in popular rankings such as the Fragile States Index (FSI). The inherent multidimensionality of state fragility that should preclude one-dimensional aggregation has previously been discussed in academia (Call 2011)s and has also been taken up in development cooperation (e.g., Organisation for Economic Co-operation and Development 2015) but it was largely ignored by index builders (see Tikuisis et al. 2015 for an exception).

Our empirical clustering contributes to the measurement literature by providing an alternative perspective to one-dimensional aggregations while nonetheless providing

manageable aggregate information. It is based on rigorous methods to determine the number of constellations and the thresholds between them. Despite this non-trivial methodological approach, the resulting typology is intuitive to understand and thus open to usage by a wider audience of both practitioners and policy makers. In contrast to ideal-type concepts, policy makers have the option of delving deeper into the specific case by looking at dimension scores, thus getting an impression of the depth of the gap.

A word of caution to potential users of our typology is opportune. Considering the data limitations we face and the necessity to impute data points, one should not be overconfident that our results will remain unchanged with future data updates. For now, we consider our typology the best possible model of fragility constellations, and a useful one. But when addressing substantial questions, it is often crucial to also consider the scores that countries receive in individual dimensions, in a “dashboard” style, avoiding the excessive “mashup” that aggregate indices of development tend to create (Ravallion 2012). It may be that an individual state has deficiencies in two core functions although it is classified as, for instance, “low-legitimacy,” the group that is defined by particular weakness only in the empirical legitimacy dimension. Our model tells us, however, that this is an exception and that most states are captured by the constellations described above. The uncertainty score attached to every country-year classification provides a useful indicator for detecting exceptions. Untypical cases tend to have lower probabilities of belonging to a group.

One important extension to the typology that could be aimed at in future iterations of this approach is the improved attribution of policy outcomes. We employ certain outcomes such as under-5 mortality as proxy variable for measuring a state’s implementation capacity, but advances in reducing mortality may also originate from other actors, such as donors or non-governmental organizations. At the moment, we rely on the assumption that in many situations, provisions made by other actors are at least partly credited to the state. A field study from Afghanistan confirms this assumption for the dimension of legitimacy (Böhnke and Aid 2013). However, better distinguishing a state’s endogenous capacity from that of competing or complementary actors would improve the validity of country classifications. It would be preferable to measure state functions directly, but this is hard to do and can only be approximated with expert assessments, such as the BTI employed here. Nonetheless, with the adoption of the 2030 Agenda at the United Nations level and the ensuing quest for suitable indicators of progress, one may soon expect advances in looking into the black box of state capacity. This will enable scholars to better measure fragility.

Another step needed to better analyze state fragility is overcoming methodological nationalism. We provide information on state fragility on the country level in this article. This does not mean that we assume state fragility to be homogeneous within countries. Various scholars working on extractive, administrative, or coercive state capacity have made promising suggestions how to better measure subnational variation in state fragility (e.g., Gingerich 2013; Harbers 2015; Lee and Zhang 2017; Stollenwerk 2018). None of these approaches, however, is currently scalable to global coverage. Achieving this goal will require substantial funding and a cooperative effort of the academic community. For now, it holds the promise of rich insights into how states may overcome fragility.

Acknowledgements We are grateful for constructive feedback and/or support with data processing to Julian Bergmann, Julia Leininger, Constantin Ruhe, and Christopher Wingens. We also thank the editors as well as three anonymous reviewers for their detailed comments and suggestions.

Funding Information Work for this project was supported by research funding from the Federal Ministry for Economic Cooperation and Development.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ahlquist JS, Breunig C. Model-based clustering and typologies in the social sciences. *Political Anal.* 2012;20(1):92–112.
- Anderson B. *Imagined communities: reflections on the origin and spread of nationalism*. London: Revised edition Verso; 1991.
- Bertelsmann Transformation Index. BTI 2016 Codebook for country assessments Bertelsmann Stiftung. 2016.
- Bertocchi G, Guerzoni A. Growth, history, or institutions: What explains state fragility in sub-Saharan Africa? *J Peace Res.* 2012;49(6):769–83.
- Besley T, Persson T. Fragile states and development policy. *J Eur Econ Assoc.* 2011;9(3):371–98.
- Böhnke JR, Aid ZC. Minds and hearts: the impact of aid in conflict zones. *Conflict Management and Peace Science.* 2013;30(5):411–32.
- Brinkerhoff DW. State fragility and governance: conflict mitigation and subnational perspectives. *Dev Policy Rev.* 2011;29(2):131–53.
- Buchanan JM. *The limits of liberty: between anarchy and Leviathan*. 1975: University of Chicago Press, Chicago.
- Bøås M, Jennings KM. Insecurity and development: the rhetoric of the ‘failed state’. *Eur J Dev Res.* 2005;17(3):385–95.
- Call CT. Beyond the ‘failed state’: toward conceptual alternatives. *Eur J Int Rel.* 2011;17(2):303–26.
- Carment D, Prest S, Samy Y. *Security, development, and the fragile state: bridging the gap between theory and policy*. Abingdon: Routledge; 2010.
- Collier D, LaPorte J, Seawright J. Putting typologies to work: concept formation, measurement, and analytic rigor. *Polit Res Q.* 2012;65(1):217–32.
- Dogan M. Conceptions of legitimacy. *Encyclopedia of government and politics*. In: Hawkesworth M and Kogan M, editors. Routledge; 1992. p. 116–26.
- Draude A. How to capture non-Western forms of governance: in favour of an equivalence functionalist observation of governance in areas of limited statehood. SFB-Governance Working Paper Series, (2). 2007.
- Eizenstat SE, Porter JE, Weinstein JM. Rebuilding weak states. *Foreign Aff.* 2005;84(1):134–46.
- European Union. Shared vision, common action: a stronger Europe – a global strategy for the European Union’s foreign and security policy http://eeas.europa.eu/archives/docs/top_stories/pdf/eugs_review_web.pdf. 2016.
- Evans PB, Rueschemeyer D, Skocpol T. *Bringing the state back in*. Cambridge: Cambridge University Press; 1985.
- Everitt BS, Landau S, Leese M, Stahl D. *Cluster analysis*, 5th. New York: Wiley; 2011.
- Fariss CJ. Respect for human rights has improved over time: modeling the changing standard of accountability. *Am Polit Sci Rev.* 2014;108(2):297–318.

- Faust J, Grävingholt J, Ziaja S. Foreign aid and the fragile consensus on state fragility. *J Int Relat Dev*. 2015;18(4):407–27.
- Fraley C, Raftery AE. Model-based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc*. 2002;97(458):611–31.
- Fukuyama F. *State-building: governance and world order in the 21st century*. Ithaca: Cornell University Press; 2004.
- Gilley B. The meaning and measure of state legitimacy: results for 72 countries. *Eur J Polit Res*. 2006;45(3):499–525.
- Gingerich DW. Governance indicators and the level of analysis problem: empirical findings from south america. *Br J Polit Sci*. 2013;43(3):505–40.
- Gisselquist RM. Developing and evaluating governance indexes: 10 questions. *Policy Studies*. 2014;35(5):513–31.
- Gisselquist RM. Varieties of fragility: implications for aid. *Third World Q*. 2015;36(7):1269–80.
- Goertz G. *Social science concepts: a user's guide*. Princeton: Princeton University Press; 2006.
- Grävingholt J, Ziaja S, Kreibbaum M. Disaggregating state fragility: a method to establish a multidimensional empirical typology. *Third World Q*. 2015;36(7):1281–98.
- Grimmer J, King G. General purpose computer-assisted clustering and conceptualization. *Proc Natl Acad Sci*. 2011;108(7):2643–50.
- Gros J-G. Towards a taxonomy of failed states in the new world order: decaying Somalia, Liberia, Rwanda and Haiti. *Third World Q*. 1996;17(3):455–72.
- Gutiérrez Sanín F, Buitrago D, González A. Aggregating political dimensions: of the feasibility of political indicators. *Soc Indic Res*. 2013;110(1):305–26.
- Harbers I. Taxation and the unequal reach of the state: mapping state capacity in Ecuador. *Governance*. 2015;28(3):373–91.
- Hausdorf B, Hennig C. Species delimitation using dominant and codominant multilocus markers. *Syst Biol*. 2010;59(5):491–503.
- Helman GB, Ratner SR. Saving failed states. *Foreign Policy*. 1992;89:3–20.
- Hirschman AO. *Exit, voice, and loyalty: responses to decline in firms, organizations, and states*. Cambridge: Harvard University Press; 1970.
- Holland AC. Forbearance. *Am Polit Sci Rev*. 2016;110(2):232–46.
- Kaplan S. Identifying truly fragile states. *Wash Q*. 2014;37(1):49–63.
- Krasner SD, Risse T. External actors, state-building, and service provision in areas of limited statehood: introduction. *Governance*. 2014;27(4):545–67.
- Lee MM, Zhang N. Legibility and the informational foundations of state capacity. *J Polit*. 2017;79(1):118–32.
- Levi M, Sacks A, Tyler T. Conceptualizing legitimacy, measuring legitimating beliefs. *Am Behav Sci*. 2009;53(3):354–75.
- Marshall MG, Cole BR. *Global report 2014: conflict, governance, and state fragility*. Center for Systemic Peace. 2014.
- Marshall MG, Gurr TR, Jagers K. *Polity IV project: political regime characteristics and transitions, 1800–2015: Dataset users' manual*. Center for Systemic Peace. 2016.
- Martis RJ, Chakraborty C, Ray AK. A two-stage mechanism for registration and classification of ECG using gaussian mixture model. *Pattern Recogn*. 2009;42(11):2979–88.
- McLoughlin C. When does service delivery improve the legitimacy of a fragile or conflict-affected state? *Governance*. 2015;28(3):341–56.
- Munck GL. *Measuring democracy: a bridge between scholarship and politics*. Baltimore: Johns Hopkins University Press; 2009.
- Olson M. Dictatorship, democracy, and development. *Am Polit Sci Rev*. 1993;87(3):567–76.
- Organisation for Economic Co-operation and Development. *Concepts and dilemmas of state building in fragile situations: from fragility to resilience*. Paris: OECD Publishing; 2008.
- Organisation for Economic Co-operation and Development. *The state's legitimacy in fragile situations: unpacking complexity*. Paris: OECD Publishing; 2010.
- Organisation for Economic Co-operation and Development. *Supporting statebuilding in situations of conflict and fragility: policy guidance*. Paris: OECD Publishing; 2011.
- Organisation for Economic Co-operation and Development. *State of fragility 2015: meeting post-2015 ambitions*. Paris: OECD Publishing; 2015.
- Pospisil J, Kühn FP. The resilient state: New regulatory modes in international approaches to state building? *Third World Q*. 2015;37(1):1–16.

- Quinn KM, Monroe BL, Colaresi M, Crespin MH, Radev DR. How to analyze political attention with minimal assumptions and costs. *Am J Polit Sci.* 2010;54(1):209–28.
- R Core Team. R: A language and environment for statistical computing R Foundation for Statistical Computing. 2019.
- Ravallion M. Mashup indices of development. *World Bank Res Obs.* 2012;27(1):1–32.
- Rawls J. A theory of justice. Cambridge: Belknap; 1971.
- Rice SE, Patrick S. Index of state weakness in the developing world. *Brookings.* 2008.
- Risse T, Stollenwerk E. Legitimacy in areas of limited statehood. *Annu Rev Polit Sci.* 2018;21:403–18.
- Saylor R. Concepts, measures, and measuring well: an alternative outlook. *Sociol Methods Res.* 2013;42(3):354–91.
- Scrucca L, Fop M, Murphy TB, Raftery AE. Mclust 5: clustering, classification and density estimation using gaussian finite mixture models. *The R Journal.* 2016;8(1):289–317.
- Stollenwerk E. Measuring governance and limited statehood. *Oxford handbook of governance and limited statehood.* In: Risse T, Börzel TA, and Draude A, editors. Oxford University Press; 2018. p. 106–27.
- The Fund for Peace. The fragile states index. *Foreign Policy.* 2014;207(4):72–7.
- The White House. The national security strategy of the United States of America. 2002.
- The World Bank. World development report 2011: conflict, security, and development. The World Bank. 2011.
- Tikuisis P, Carment D. Categorization of states beyond strong and weak. *Stability: International Journal of Security and Development.* 2017;6(1):1–23.
- Tikuisis P, Carment D, Samy Y, Landry J. Typology of state types: persistence and transition. *International Interactions.* 2015;41(3):565–82.
- Tilly C. War making and state making as organized crime. Bringing the state back in. In: Evans P, Rueschemeyer D, and Skocpol T, editors. Cambridge University Press; 1985. p. 169–91.
- United Nations Office on Drugs and Crime. Global study on homicide 2013: trends, contexts, data. UNODC. 2013.
- Vermunt JK. K-means may perform as well as mixture model clustering but may also be much worse: comment on Steinley and Brusco (2011). *Psychol Methods.* 2011;16(1):82–8.
- von Haldenwang C. The relevance of legitimation – a new framework for analysis. *Contemp Polit.* 2017;23(3):269–86.
- Wade R. Governing the market: economic theory and the role of government in East Asian industrialization. Princeton: Princeton University Press; 1990.
- Weatherford MS. Measuring political legitimacy. *Am Polit Sci Rev.* 1992;86(1):149–66.
- Weber M. Politik als Beruf. Berlin: Duncker & Humblot; [1919] 2010.
- Zartman IW. Collapsed states: the disintegration and restoration of legitimate authority. Boulder: Lynne Rienner Publishers; 1995.
- Ziaja S. What do fragility indices measure? *Zeitschrift für Vergleichende Politikwissenschaft.* 2012;6(1):39–64.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Sebastian Ziaja is a Senior Researcher at the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE). Currently he studies the measurement of political regime transitions and the effect of foreign aid on subnational variation in state capacity.

Jörn Grävingholt is a Senior Researcher and Project Lead at the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE). His areas of research include state fragility, drivers of forced displacement, peacebuilding and governance support, and the dynamics of democratic and non-democratic regimes.

Merle Kreibaum was a Researcher at the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE). She is a development economist. Her research focusses on the causes and consequences of political violence, in particular at the micro level.