



EUI Working Papers

LAW 2012/32

DEPARTMENT OF LAW

MEASURING STATENESS, RANKING POLITICAL ORDERS:
INDEXES OF STATE FRAGILITY AND STATE FAILURE

Nehal Bhuta

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
DEPARTMENT OF LAW

*Measuring Stateness, Ranking Political Orders:
Indexes of State Fragility and State Failure*

NEHAL BHUTA

This text may be downloaded for personal research purposes only. Any additional reproduction for other purposes, whether in hard copy or electronically, requires the consent of the author(s), editor(s). If cited or quoted, reference should be made to the full name of the author(s), editor(s), the title, the working paper or other series, the year, and the publisher.

ISSN 1725-6739

© 2012 Nehal Bhuta

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Abstract

This paper examines two indexes of state failure and state fragility. It considers the wider historical context for the emergence of interest in state failure and state fragility, and examines attempts to define the concepts. After reflecting on the limitations and difficulties of various definitions, the paper scrutinizes two attempts to “measure” stateness by creating indexes of state failure and fragility. It argues that both indexes are flawed methodologically, and poses the question of why there nonetheless continues to be interest in quantifying the concepts despite the insurmountable problems of measurement.

Keywords

state failure, state fragility, indexes, development, international law, international relations

Author Contact Details

Nehal Bhuta,
Professor of Public International Law,
European University Institute,
Florence, Italy

Email: Nehal.Bhuta@eui.eu

MEASURING STATENESS, RANKING POLITICAL ORDERS: INDEXES OF STATE FRAGILITY AND STATE FAILURE

Nehal Bhuta

Introduction

The concept of “failed,” “failing” or “fragile” states has become ubiquitous. The failing or fragile state is referred to as a source of grave security threats,¹ as a particularly challenging context for development assistance,² and as an impediment to the achievement of human development goals. The term “failed state” appears to have emerged in the early 1990s, and was used in reference to dramatic cases of state collapse, generally occasioned by severe internal conflict. Indeed, one of the earliest attempts to measure the incidence of state failure – George Mason University’s State Failure Task Force – took events such as revolutionary war, regime change and genocide as instances of state failure. A paradigmatic failed state in this understanding was the former Yugoslavia, Rwanda, Somalia or Afghanistan (1992-96), where severe conflict meant that no governing authority had effective control over the territory. Obviously, such circumstances are associated with a variety of crises which would be of concern to the international community: forced displacement and refugee flows, violations of humanitarian law and international criminal law, massive destruction of human and physical capital, and possible “ungoverned spaces” which might become operational homes to terrorist organizations or conduits for transborder flows of people, drugs and weapons.

The externalities associated with situations of complete state collapse appear to have triggered an interest in understanding the correlates and preconditions for such situations. It is unclear when notions such as “weak” or “fragile” states emerged as significant concepts, but in September 2002 the United States’ National Security Strategy referred to “failing” states as a threat to U.S. security, connoting a set of states which have not yet failed, but are at risk of doing so. Robert Rotberg’s influential article in *Foreign Affairs* in the same year,³ the author differentiates between “strong,” “weak” and “failed” states and argues that there are observable pathways by which states move from one of these conditions to the other. Similarly, Krasner and Pascual refer to “weak and failed” states, while the UK’s National Security Strategy discusses “failed and fragile” states.⁴

In the context of development assistance, DFID, USAID and the OECD’s Development Assistance Committee (DAC) began referring to “fragile states” from 2004. From 2002, the World Bank labeled certain countries as “low income countries under stress” (LICUS), based on a poor rating in the annual Country Policy and Institutional Assessment (CPIA). The LICUS category has been replaced by “fragile states,” which includes a subset of “Post-Conflict Countries.” In the development assistance context, the classification of states as “fragile” was directly or indirectly connected with judgments

¹ National Security Council, *National Security Strategy of the United States*, Washington DC, September 2002, p.1; *National Security Strategy of the United States*, March 2006, pp. 37, 44; *National Security Strategy of the United States*, May 2010, 8, 11, 13.

² European Report on Development (EU/EUI), *Overcoming Fragility in Africa*, San Domenico di Fiesole, 2009, p.12; UK Department for International Development, “Why We Need to Work More Effectively in Fragile States.” London: DFID, 2005; United States Agency for International Development, *Failed States Strategy*, Washington DC: USAID, January 2005; OECD Development Assistance Committee, *Concepts and Dilemmas of State Building in Fragile Situations*, OECD Discussion Paper, Paris: OECD, 2008.

³ Robert Rotberg, “Failed States in a World of Terror,” *Foreign Affairs*, July/August 2002.

⁴ Stephen Krasner and Carlos Pascual, “Addressing State Failure,” *Foreign Affairs*, July/August 2005; Cabinet Office, *National Security Strategy of the United Kingdom*, London: Cabinet Office, March 2008, pp.13, 14.

concerning aid allocation. Where allocation was performance based, certain countries facing severe internal conflict or its immediate aftermath were unlikely to meet performance-based goals for aid eligibility, and yet were in desperate need of development assistance. Thus, in order to accommodate these countries' specific needs for assistance, "fragile states were first identified ... as countries where Performance Based Allocation should not apply ... the answer was then to give [these countries] a specific treatment for effectiveness or security reasons."⁵ A state deemed fragile may also be designated for particular kinds of assistance and aid programming, such as access to a special World Bank Trust Fund and the provision of incentives to Bank staff to work in these countries.

These 2 different contexts for the assessment of whether a state is "failing," "failed" or "fragile" have engendered two somewhat distinct imperatives for mechanisms to identify and measure such concepts. One purpose of measurement is to identify that group of countries eligible for special kinds of development assistance, or which should be considered for exceptions to performance-based allocation. The other purpose of measurement is to identify countries at risk of generating security threats, developing zones of ungovernability, or suffering from severe internal conflict – and thus warranting special scrutiny, multilateral diplomatic action or other responses.

⁵ Patrick Guillaumont and Sylviane Guillaumont Jeanneney, "State Fragility and Economic Vulnerability: What is Measured and Why?" CERDI-CNRS/University d' Auvergne, July 2009, p.4.

Definitions of State Failure and Fragility

Despite the ubiquity of the terms “failed,” “failing” and “fragile,” there is no consensus on the meaning of these concepts, or how to measure the extent to which a state is fragile or failing. Indeed, under these conditions, the process of constructing the measure also becomes an exercise in defining the concept to be measured. There is some degree of convergence, at a high level of generality, concerning what the notion of state failure or state fragility is *believed to be associated with*. That is, most explications of the concepts involve descriptions of the *kinds of symptoms* characteristic of a state which is deemed fragile or failing. Definitions are “prototypical”.

In the various definitions, the kinds of symptoms which are said to be characteristic of state failure and state fragility include: illegitimate or ineffective government; the absence of the rule of law; lack of political will or capacity to deliver basic public goods such as border control, crime prevention and essential services; lack of will or capacity of the government to provide functions needed for development, poverty reduction and human rights; a “broken social contract” between state and society, or a poor “state society relationship.”

The UK’s National Security Strategy defines a “failed state” as one whose “government is not effective or legitimate enough to maintain the rule of law, protect itself, its citizens and its borders, or provide the most basic services, and a “fragile state” as “one in which those problems are likely to arise.”⁶ The *Fragile States Strategy* of USAID describes fragile states as “the product of ineffective and illegitimate governance. ... Effectiveness refers to the capability of the government to work with society to assure the provision of order and public goods and services. Legitimacy refers to the perception by important segments of society that the government is exercising state power in ways that are reasonably fair and in the interests of the nation as a whole. Where both effectiveness and legitimacy are weak, conflict or state failure is likely to occur.”⁷ The OECD DAC defines states as fragile “when state structures lack political will and/or capacity to provide the basic functions needed for poverty reduction, development and to safeguard the security and human rights of their populations.”⁸ The Council of the European Union defines fragility as referring to “weak or failing structures and to situations where the social contract is broken due to the state’s incapacity or unwillingness to deal with its basic functions, meet its obligations and responsibilities regarding the rule of law, protection of human rights and fundamental freedoms, security and safety of its population, poverty reduction, service delivery, the transparent and equitable management of resources and access to power.”⁹

Among academic writers and think tanks, the definition of failed, failing and fragile states tends to be more expansive, and involves (sometimes quite lengthy) lists of qualities said to correspond to state failure. Rotberg¹⁰ describes failed states in contrast to strong states. Strong states are “places of peace and order,” which “control their territories and deliver a high order of political goods to their citizens ... Strong states offer high levels of security from political and criminal violence, ensure political freedom and civil liberties, and create environments conducive to the growth of economic opportunity. Failed [and presumably, fragile] states are tense, conflicted and dangerous. They have some or all of the following characteristics, depending on where they lie on the spectrum between failed and fragile (or failed, weak and strong):

⁶ *National Security Strategy of the United Kingdom*, page 13.

⁷ USAID, *Fragile State Strategy*, p. 3.

⁸ OECD DAC, *Principles for Good International Engagement in Fragile States and Situations*, Paris: OECD, 2007, p.2

⁹ Council of the European Union, Council Conclusions on an EU Response to Situations of Fragility.” 2831st External Relations Council Meeting, 19-20 November, Brussels.

¹⁰ Rotberg, “Failed States in a World of Terror.”

- A rise in criminal and political violence
- A loss of control over their borders
- Rising ethnic, religious, linguistic, and cultural hostilities
- Civil war
- Use of terror against their own citizens
- Weak institutions
- A deteriorated infrastructure
- Inability to collect taxes without undue coercion
- High levels of corruption
- A collapsed health system
- Rising levels of infant mortality and declining life expectancy
- The end of regular schooling opportunities
- Declining levels of GDP per capita
- Escalating inflation
- Widespread preference for non-national currencies
- Basic food shortages.

RAND describes failed states as those which “typically suffer from cycles of violence, economic breakdown, and unfit governments that render them unable to relieve their people’s suffering, much less empower them.”¹¹ Brookings (as a prelude to their attempt to measure state weakness) defines weak states as “countries that lack the essential capacity and or will to fulfill four sets of critical government responsibilities: fostering an environment conducive to sustainable and equitable growth; establishing and maintaining legitimate, transparent and accountable political institutions; securing their populations from violent conflict and controlling their territory; and meeting the basic human needs of their population.”¹² The US Fund for Peace (USFFP) (which authors the Failed State Index (FSI)) describes weak and failing states as having “common attributes. These include loss of physical control over territory, lack of monopoly on the use of force, declining legitimacy to make authoritative decisions for the majority of the community, an inability to provide security or social services to its people, and, frequently, a lack of capacity to act as a full member of the international community.”¹³

¹¹ Marla Haims, David Gompert, Gregory Treverton, Brooke Stearns, *Breaking the Failed State Cycle*, RAND Occasional Paper, Santa Monica, 2008 p.xi.

¹² Susan Rice and Stewart Patrick, *Index of State Weakness in the Developing World*, Brookings Institution: Washington DC 2006, p.3.

¹³ Pauline Baker, “Fixing Failed States: The New Security Agenda,” *The Whitehead Journal of Diplomacy and International Relations*, Winter/Spring 2007, p.90. (Baker is President of the US Fund for Peace).

The Limits of Symptomatic Definitions

A common feature of all of these definitions is that they derive the meaning of state failure and fragility from its symptoms, but are vague in the specification of relationships between symptoms and causes (or between symptoms and underlying illness). Does poor governance cause conflict or does conflict cause poor governance? Under most of the above concepts of state weakness, we would answer “both,” but as a result the concept of state weakness is not adding much to our understanding of either parameter (conflict or governance). Is a poorly governed state weak even if it is not prone to internal conflict? The above definitions occasionally allude to or mention causal pathways from fragility to failure (such as USAID’s definition), but most do not. As such, it is unclear whether the concept of state fragility and failure could ever avoid significant over-determination. It is equally unclear whether the concepts can be analytically useful except as a short-hand for describing a collection of associated phenomena, without being able to assign weight to the relative significance of a given symptom in hastening or restraining failure. If the primary function of the terms is as a categorical short-hand, their utility will be greatest for identifying and labeling members of the set of fragile and failing states (defined stipulatively as exhibiting the symptoms set out above) in need of special attention or development assistance.

But the terms’ utility as a means of differentiating between different degrees of weakness (which presupposes some capacity to differentiate the relative role and impact of specific symptoms in driving failure within the set of weak states) seems limited. One related question which arises is whether the concept over-aggregates distinct social and political problems, and so obscures our understanding of specific relationships between symptoms and causes (and the possible tradeoffs in addressing them). Some scholars make this claim, questioning the utility of the concept for understanding the drivers of conflict and how to address them.¹⁴

A second important commonality is that each of these definitions includes notions of the legitimacy, effectiveness and capacity of the state as critical dimensions of the state’s strength or weakness/fragility. Several definitions also refer to the “will” of the state to provide certain goods and services to its population. These terms are latent or unobservable concepts. No consensus exists on how to measure them or which proxies reliably capture these phenomena, if any. Similarly, the kinds of legitimacy, effectiveness and capacity which augment or diminish the fragility of the state are themselves subject to argument.¹⁵ The exact relationship between a particular kind of legitimacy and the strength and weakness of a state is not specified in most of the definitions, although some refer to democratic legitimacy.

Among the more foundational critiques of the concept of fragile and failing states is that the concept assimilates many problems of security with problems of development, and too broadly associates security risks with a lack of development.¹⁶ The critique here is twofold. First, there is the claim that the “securitization of development” shifts the focus towards development initiatives which are believed to contain or manage security risks, and thus may skew development funding priorities. There is also the danger that developmental problems within a territory are artificially placed in the framework of “security” in order to attract more funding. Second, the bundling of development,

¹⁴ See Charles T. Call, “Beyond the ‘failed state’: Toward Conceptual alternatives.” (2010) 20 *European Journal of International Relations* 1-24; Edward Newman, “Failed States and the International Order: Constructing a Post-Westphalian World.” (2009) 30 *Contemporary Security Policy* 421-443.

¹⁵ See, for example, OECD, *The State’s Legitimacy in Fragile Situations: Unpacking Complexity*, Paris: 2010.

¹⁶ See Eka Ikpe, “Challenging the Discourse of Fragile States,” (2007) 7 *Conflict, Security and Development* 85; Clement Eme Adibe, “Weak States and the Emerging Taxonomy of Security in World Politics,” (1994) 26 *Futures* 490; Mark Duffield, *Global Governance and the New Wars: The Merging of Development and Security* (London: Zed, 2001); Mark Duffield, *Development, Security and Unending War: Governing the World of Peoples* (Cambridge: Polity, 2008).

governance and security under the “metaconcept” of state fragility incentivizes deeper and more intensive forms of intervention in the developing world, and thereby promotes a “civilizing mission” that may be both unrealistic and detrimental in the long term.¹⁷ It also encourages a hierarchy of sovereignty, in which states classified as failing or failed are deemed incapable or unwilling of exercising the prerogatives of sovereign equality and thus are (or should be) less entitled to international legal protections against the use of force and intervention.¹⁸ Other academic writings also contest the extent to which state weakness and so-called “ungoverned spaces” are empirically more likely to foster or become home to terrorist groups.¹⁹

The concept of the state found in the fragility/failure literature is predicated on a highly particular image of the state, an ideal-type of the state form which is sometimes termed ‘Weberian’. In this ideal-type, the state is a unified and coherent actor, set apart from other social organizations and sources of social power. Through its officials, the state has its own preferences and can act on those preferences in order to change the behaviour of others. The state enacts its domination through a uniform set of rules, backed by a credible threat of violence.²⁰ At least for some Western European states, this ideal type approximates an historical reality – although even among this relatively small group, variation is substantial throughout recent history.²¹ Importantly, for Weber, this ideal-type was not to be understood as a normative or normalizing typification of what qualifies as a ‘true’ or ‘proper’ state; rather, an ideal-type is an exaggerated abstraction from a particular historical reality, used only to better grasp the features of the phenomenon under analysis.²² This ideal typical concept of the state – including the much-cited definition of it as an organization which exercises a monopoly of violence over a territory – was an abstraction that sought to identify a common form of political organization that emerged from a specific European experience. As Migdal and Schlichte observe, ‘an ideal-type is not itself a hypothesis, but it allows one to build hypotheses on deviations, variations and totally different forms. Once these differences are noticed, the need for a vocabulary of description and explanation becomes obvious.’²³ To the extent that the ideal type of the state is taken as the measure of state-ness, it becomes a norm rather than a heuristic for the development of explanatory or interpretive accounts of how a given political organization works (or does not work). The crypto-normative quality with which the Weberian ideal type of the state has been imbued in the fragility discourse has meant that deviations from the type are taken as evidence of weaknesses or failures.

¹⁷ See Duffield, *ibid* in particular.

¹⁸ For example of a recent paper which argues that the challenge of failing and failed states requires the revision of norms of sovereign equality, see John Yoo, “Fixing Failed States,” http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1552395.

¹⁹ See: E. Newman, “Weak States, State Failure and Terrorism,” *Terrorism and Political Violence* (2007) 19: 463; Charles Call, “The Fallacy of the Fragile State,” *Third World Quarterly* (2008) 29.

²⁰ Weber in fact never completed a state theory. His concept of the state is scattered throughout his sociology of law and his omnibus work, *Economy and Society*.

²¹ See, for example, Thomas Ertman, *The Birth of the Leviathan: Building States and Regimes in Medieval and Early Modern Europe* (Cambridge: Cambridge University Press, 1997); and Samuel Finer, ‘State and Nation-Building in Europe: The Role of the Military’, in Charles Tilly (ed), *The Formation of National States in Europe* (Princeton: Princeton University Press, 1975) 84-163.

²² Max Weber, *Economy and Society, Volume One* (Berkeley: University of California Press, 1978), 9, 20-21.

²³ Migdal and Schlichte, ‘Rethinking the State’, 3.

Efforts at Measurement and Indexes: The Fund for Peace and USAID

In the early 1990s, the CIA established and funded the State Failure Task Force (later renamed the Political Instability Task Force) to explore methods of risk assessment and early warning systems for state failure.²⁴ The Polity IV dataset generated by this project, which spanned more than a decade, would become an important source for current fragility/failure indexes, but it is reported that the Task Force also produced classified models and algorithms which are in use by the CIA to identify lists of countries at risk of instability.²⁵ After 2001, there was a proliferation of attempts to measure failure and fragility, and to produce ordered lists.²⁶

The claim to be able to quantify degrees of state failure or fragility is seductive. It holds out a promise of greater objectivity and transparency in making judgments about risks of conflict and instability in a given state. It also promises comparability across states in terms of the extent of failure or fragility, notionally allowing judgments to be made about which states should be prioritized for policy attention. But as noted above, a key challenge of measurement is definitional. Definitions of failure or fragility are prototypical in the sense that they identify typical instantiations of strength or failure and derive common characteristics.²⁷ The fundamental difficulty with prototyping is that ‘the characteristic traits of the phenomenon are collapsed with putative causes and consequences ... This is a bit like defining cancer as a consisting of smoking, uncontrolled growth of cells, and family crisis.’²⁸ Such prototypical definitions do not specify causal pathways between observed characteristics and fragility or strength, and provide no means to understand the relative role of each characteristic or how they interact with each other temporally, and do not propose how each dimension can be logically assembled into a single notion of strength or weakness. But as Munck and Verkuilen point out, specifying the meaning of a concept and its components ‘affects the entire process’ of measurement, from data generation to aggregation, and conceptualization is intimately linked with theoretical claims about the nature of the phenomenon one is seeking to measure.²⁹

In this part of the chapter, I examine two significant indexes of state failure or fragility. Among the most visible and widely reported indexes of state failure is that produced annually by USFFP, and published in *Foreign Policy*. The Index ranks 177 countries based on an aggregate of their scores on 12 indicators. The higher the aggregate, the greater the purported degree of state failure (and thus the higher the rank of the state on the index). In the 2010 version of the FSI, the top five “failed states” were Somalia, Chad, Sudan, Zimbabwe and the Democratic Republic of the Congo. According to the USFFP, their website (which hosts the FSI) receives hundreds of thousands of page views a month,³⁰ and there is some evidence that the FSI is becoming regularly cited in academic literature as a relevant measure of state failure.³¹ Much less known, but potentially more influential is USAID’s “Alert List” for Conflict and Instability, produced annually from 2006, which ranks 160 countries in order of fragility. The USAID index is not public, but is distributed widely within the US Government. In the 2010 version of the USAID Index, the top 5 fragile states were Chad, Somalia, the Democratic

²⁴ Robert Adler, ‘The Crystal Ball of Chaos’, *Nature*, vol 29 (2001) 480-81.

²⁵ Interview with USAID official, July 23, 2010.

²⁶ I have identified ten lists or rankings of state failure/fragility/weakness in the course of this research.

²⁷ Francisco Gutierrez Sanin, ‘Evaluating State Performance: A Critical View of State Failure and Fragility Indexes’, *European Journal of Development Research*, vol 23 (2011) 24.

²⁸ *Ibid.*

²⁹ G. Munck and J. Verkuilen, ‘Measuring Democracy: Evaluating Alternative Indices’, *Comparative Political Studies* vol 35 (2002) 7.

³⁰ Email correspondence with Mark Loucas, USFFP Program Officer, July 8, 2010.

³¹ Based on preliminary Scholar Google search

Republic of the Congo, Guinea and Sudan. It is noteworthy that there is substantial convergence between these two indexes at the “top 5.” It might be taken as evidence of their robustness, but it is equally indicative of the extent to which prototypical concepts “allows the creation of definitions that fit well cases on the margins of the definitional space – that is, the extreme ones – but not necessarily elsewhere. In other words, it can produce an over-fit to a certain set of situations, but the variables that separate well the extreme cases from others do not necessarily fit more moderate examples.”³² In fact, the 2 Indexes follow completely different and non-comparable methodologies, suggesting that any convergence in rankings at the extreme ends of the definitional space may well be a consequence of “over-fit” rather than rigorous conceptualization and measurement.

³² Francisco Gutierrez Sanin, “The Quandries of Coding and Ranking: Evaluating Poor State Performance Indexes.” Crisis States Working Papers Series No.2, November 2009 (LSE Development Studies Institute), no page numbers.

USFfP's Failed States Index

Unlike all of the other indexes referred to above, the FSI's twelve indicators are not based on pre-existing measures. Rather, each country is scored between 1 (best) and 10 (worst) for each of the twelve indicators based on an assessment conducted internally by USFfP. The 12 indicators for which each state receives a score are as follows:

1. Mounting Demographic Pressures
2. Massive Movement of Refugees or Internally Displaced Persons
3. Legacy of Vengeance-Seeking Group Grievance or Group Paranoia
4. Chronic and Sustained Human Flight
5. Uneven Economic Development Along Group Lines
6. Sharp and or Severe Economic Decline
7. Criminalization and or Delegitimization of the State
8. Progressive Deterioration of Public Services
9. Suspension of the Rule of Law and Widespread Violation of Human Rights
10. Security Apparatus Operates as a State within a State
11. Rise of Factionalized Elites
12. Intervention of Other States or External Political Actors.

The meaning of each of these indicators is subjective. The allocation of a score may be subject to high degrees of variation if not carefully controlled. In other Indexes which rely on a method of intensity scoring (such the Cingranelli and Richards dataset and the Polity IV data set), detailed coding manuals (often over 100 pages in length) are created and coders are trained in the assignment of scores based on conformity of the data source with specific formulae and phrases in the code book. Underlying data sources are kept as homogeneous as possible and more than one coder is assigned to same source in order ensure consistency and check for divergences or disputable interpretations. Publication of coding manuals ensures both transparency and replicability.

By contrast, the method by which a state is scored between 0 and 10 on each of the FSI indicators is complex. It is not transparent and it is not replicable. No detailed instructions for the interpretation or application of the descriptive terms contained in the indicators are provided publicly.³³ The score is allocated by an incompletely explained combination of qualitative and quantitative methods.

The process as a whole is referred to by USFfP as the "Conflict Assessment System Tool" (CAST), and the software used in the process has been successfully patented.³⁴ The process begins with the CAST software conducting specified Boolean searches of a mixed data source combining news reports, policy documents and other materials collected for USFfP. The material is provided to USFfP in English, although USFfP contends that the sources include material "originating from 110 countries in 50 languages."³⁵ The number and variation of Boolean searches conducted for each indicator is

³³ A "ratings guide" and a document entitled "Indicators and their Measures" can be found on the USFfP website, but these do not amount to code books. They provide more detailed descriptions of what is meant by some of the terms in the 12 indicators, but these descriptions add new layers of interpretative uncertainty to how the scores are derived from the underlying data.

³⁴ Jan 2, 2003; US 2003/0004954A1.

³⁵ "The CAST Methodology: Its Use in Producing the Failed State Index." No date. The Fund for Peace.

unclear, although the USFP explains that for each of the 12 indicators there are a subset of “measures,” each of which has a specific search string.³⁶ The reliability of the sources available for each country in the index is evaluated by the software, through a “computation of the difference in information from the source compared to the same category of information reported from a core of five world sources that include the CIA, NY Times, CNN, BBC and NPR.”³⁷

The output of the search strings is combined with existing statistical measures which are deemed relevant to the level of the indicator.³⁸ It is unclear how these numbers are aggregated with numbers of generated hits by the Boolean search process. It appears that the CAST software produces a score from the content analysis and aggregates this with baseline quantitative measures associated with the indicator. The software also produces an estimate of how much this indicator has improved or deteriorated since the previous year.

Parameters of a country’s scores are set by taking the number of hits generated through Boolean searches and dividing that number by the sample size (documents per time period per country). The number thus generated is known as the “salience”. A “10” for that indicator is set by reference to the worst salience for that country since 1960. For example, in the case of Rwanda, 1994 would be a “10” for the indicator labeled “Legacy of Vengeance-Seeking Group.” The intensity of content generated by the relevant Boolean searches for Rwanda for 1994 would thus represent a “10”, and all others results could be scaled against this maximum.

The indicator value which is generated in this manner is then reviewed by an analyst, who has formed his own view about whether the indicator has improved or deteriorated for the country in question since the last year. The analyst is an in-house analyst, who is trained in the CAST software.³⁹ However, it is not clear whether the analysts are substantively country experts. According to USFP, there are 15 analysts for 177 countries. Initially, outside country experts were used to check country scores, but this practice was discontinued because the experts were not “socialized” into the methods of the index.⁴⁰

Where the analyst’s judgment concerning the score and the degree of change of the score is at odds with the software-generated number, there is a process of review and reconciliation. The steps in this process, and the criteria for adjusting a score upward or downward are unknown, but seem to rely heavily on the analyst’s “feel” for the country situation based on several years of maintaining the index. The results are said to be “peer-reviewed,” but it appears that this means that results are reviewed by other in-house analysts, not submitted to academic experts for external review.

A further adjustment in the score may be brought about by an events-sensitive assessment of recent developments. This assessment is supposed to capture “surprises, triggers, idiosyncrasies, national temperaments and spoilers.” The software is pre-programmed with incidences that comprises these kinds of events for a particular country, and the events will be allocated a score between 0 and 5 by “country experts” (who also determine which indicators are affected by its occurrence). If the event occurs, the software will adjust the score of the affected indicators.

³⁶ Ibid.

³⁷ Patent Application Publication, p.3.

³⁸ By way of example: for Indicator 1, the following quantitative sources are also consulted in forming a view: Proportion of population under 15, Population undernourished (UNDP), Life expectancy at birth, Population annual growth rate, Infant mortality (WHO). Email correspondence with Mark Loucas, July 21, 2010.

³⁹ Correspondence with Nate Haken, USFP, July 2010.

⁴⁰ Interview with Nate Haken, US Fund for Peace, July 20, 2010.

The indicator score may also be affected by an additional measure of the representativeness, legitimacy and professionalism of institutions deemed to constitute the “immutable core” of a state, namely:

1. A competent domestic police force and corrections system.
2. An efficient and functioning civil service.
3. An independent judicial system.
4. A professional and disciplined military accountable to civilian authority.
5. A capable leadership.

How the evaluation of the legitimacy, representativeness and professionalism of these five institutions is undertaken is not explained. The evaluation does not have a direct impact on indicator scores, but becomes part of the “country profile” which helps set a baseline for the allocation of scores.

The final score of a country, and the determinant of its overall rank in the index, is the aggregate of the 12 indicator scores. The higher the number, the more severe the degree of state failure. However, it should be observed that, due to the numerous points in the process at which the basis for aggregation of different kinds of data is not verifiable, it is doubtful whether the indicator levels for each country are truly comparable. Or at least, the basis for comparability is uncertain. Comparability is further complicated by what seems to be an inherently relative scoring standard, in which a “10” for a particular state may be determined by reference to specific historical examples representing a nadir for state failure for that state. As such, the ordinal ranking – which generates so much attention for the index – may be of little or no validity as a measure of degrees of state failure.

Another implication is that the number generated by the scoring and ranking is a poor signal for the nature of the problems facing a particular country; without considerable additional context and country-specific knowledge understanding the difference between an “8” and a “6” on a given indicator is impossible. Indeed, in light of the questions arising about the transparency and reliability of the scoring, it is far from clear that anything is actually being “measured.”

The producers of the FSI and the CAST software emphasize that of greater interest is the generation of trend lines concerning changes in the level of stability of a state (as measured by the values assigned to the 12 indicators). These trend lines, according to the authors, allow some measure of early warning of states at risk of intensifying conflict and deteriorating stability.⁴¹ FSI and CAST do not purport to predict state failure or state collapse. It does claim to be a tool “for diagnosing the risk of violence in weak and failing states.”⁴²

Apart from the FSI, USFfP generates a variety of knowledge products which it makes available on a commercial basis. It licenses its CAST software and provides training to use it. It provides client specific country profiles and conflict assessments, but does not disclose its clients. It also provides “real time” updates in the form of Alerts and notifications for developments in specific countries. It seems that these are also provided to clients for a price. Country Profiles are available for free on their website. These short documents provide an explanation of the country’s FSI score and shed some light on how the score changed from the previous year.

⁴¹ Pauline Baker, “The Conflict Assessment System Tool (CAST). An Analytical Model for Early Warning and Risk Assessment of Weak and Failing States.” 2006. Page v.

⁴² Ibid.

USAID's Fragility and Risk for Instability Rankings

USAID's definition of fragility emphasized 'legitimacy' and 'effectiveness' of the government of a state as the essential dimensions of its strength or fragility. Measuring fragility therefore required creators of the index to render countable qualities of a political order that are inherently unobservable and uncountable. Proxies had to be determined for each of these terms, based on theories about how a certain countable phenomenon (such as infant mortality or economic growth) was to be understood as representing a dimension of an uncountable property, such as legitimacy. Indeed, the very definitions of legitimacy and effectiveness required theoretical foundations, in order to articulate their claimed relationship with another non-observable term, 'state strength' (or state weakness).

In a Strategy Paper written for USAID as part of the creation of the Index,⁴³ the strength or weakness of a state is defined in terms of the legitimacy and effectiveness of its institutions, especially those institutions charged with 'the management of conflict.' A strong state is one which successfully creates and maintains 'the institutional mechanisms of 'neutral ground,' which is composed on 'those elements of the social life of the country that are non-partisan in ways relevant to the social conditions of that country, and which contain a range of neutral legitimating factors such as legality, erudition and technical or professional competence.'⁴⁴ The 'neutral ground' that characterizes state strength is thus composed of – identified with – the presence of institutions possessing certain qualities of effectiveness and legitimacy, which are in turn associated with features such as legality and technocratic competence.

These are clearly associative or correlative definitions, and causalities could run in either direction.⁴⁵ They are also prototypical definitions, as noted above. In their attempt to 'operationalize' the concepts of 'effectiveness' and 'legitimacy' in order to render them useful for the assessment of state fragility, the strategy paper authors develop a four row by two column accounting matrix, in which rows represent 'four dimensions of state-society relations: political ...; economic ...; social ...; and security,' while the columns represent effectiveness and legitimacy for each of these dimensions. An assessment of a state's fragility entails gathering evidence of the effectiveness and legitimacy of the state's political, economic, social and security functioning, and evaluating this evidence against a benchmark of state strength.

The accounting matrix, with its four by two enumeration of dimensions of effectiveness and legitimacy would become a template for developing measures of fragility that would serve as inputs into the Index. A way had to be found to quantify each of 8 possible dimensions of effectiveness and legitimacy,⁴⁶ rendering them countable and so susceptible to aggregation in a one-dimensional rank ordering. Countability requires that data be found or created which successfully measures the phenomena sought to be counted. Where the phenomena (however well specified) are unobservable, it is necessary to resort to proxies which can be observed and counted, implying a theory of the relationship between the proxy and the attributes of the underlying concept. Counting is always reasoning.

⁴³ Goldstone et al., 'Strategy Framework'.

⁴⁴ Jack Goldstone, Jonathon Haughton, Karol Soltan and Clifford Zinnes, 'Strategy Framework for the Assessment and Treatment of Fragile States', Executive Summary, June 2004, 5-6. (My emphasis).

⁴⁵ Ibid, 8. Additional glosses on the meaning of effectiveness and legitimacy are provided in the document. Effectiveness is defined as "the degree to which a state has the administrative capability and resources to carry out the tasks of governance," where governance includes "a disciplined military and bureaucracy" and "intelligence and administrative capability. Legitimacy is further described as "rulers being judged – by both elites and popular groups – as being reasonably fair and just in their exercise of power."

⁴⁶ USAID, 'Measuring Fragility: Indicators and Methods for Rating State Performance.'

Analytically, concepts such as political legitimacy or social legitimacy are not well specified in the sense that their stipulated attributes are themselves unobservable. For example, political legitimacy is defined as ‘political institutions and processes that are transparent, respect societal values, and do not favor particular groups.’⁴⁷ Each of these terms is non-self-evident, and what might qualify as a measure of each of these attributes is unclear, requiring further disaggregation and conceptualization. Moreover, each of the attributes of political legitimacy is difficult to logically distinguish from attributes of other dimensions of legitimacy, such as social legitimacy or security legitimacy. The former is defined as ‘tolerance for diversity, including opportunities for groups to practices ... cultures and beliefs,’ while the latter is defined as ‘military and police services that are provided equitably and without violation of civil rights.’ The problem with logically non-exclusive definitions of attributes of unobservable variables such as political legitimacy, is redundancy and conflation.⁴⁸ How do we know that a measure of tolerance is not also a measure of respect for societal values? If we do not know, how can we avoid problems of double counting, or inaccurate differentiation between 2 measures of the same underlying phenomena? The analytical difficulties of conceptualization in fact point to the deep difficulty of measurement of multi-dimensional characteristics such as legitimacy, a difficulty that is at once compounded and made less visible where measures are aggregated into a one-dimensional number, as in an index.

Conceptual problems of this kind are aggravated in the attempt to identify proxies which can be used as observable data representing unobservable phenomena. Given that definitional problems remain severe, it is a puzzle how proxies could be determined: how could one differentiate between good and bad proxies for a concept, when the concept itself is not well specified? The creators of the USAID Index sidestepped this puzzle by substituting the proxy for the concept. Instead of identifying measures that successfully capture attributes of the concepts sought to be measured, the creators identified pre-existing indicators which they contended measured political, social, economic and security outcomes or perceptions, and maintained that these outcomes or perceptions amounted to measurements of the 8 dimensions of effectiveness and legitimacy decomposed by the accounting matrix.

Definitions of effectiveness and legitimacy developed by the strategy paper were thus decoupled from the identification of measures, with an emphasis instead on where ‘good data’ could be found that could conceivably be related to the measurement of effectiveness and legitimacy. The result is a bricolage of data sets chosen because of their coverage, frequency of updating, assumed credibility of their source and some notion of robustness. But the critical question of how and why the outcome or perception data is in fact a measure of the unobservable variable is answered only with the introduction of further, diverse, theoretical and factual claims. For example, among the proxies chosen to measure ‘political effectiveness’ is the government effectiveness indicator generated by the World Bank’s Governance Matters dataset, based on an aggregation of surveys of perceptions of the quality of public service. The rationale for choosing this as a component measure of political effectiveness is that ‘the quality of public service provision is a good, directly observable outcome of effective governments.’⁴⁹ Yet, the quality of public service provision is not at all directly observable, which is why the Governance Matters data set relies on subjective perception surveys to develop a measure.⁵⁰ The relationship between perception and reality is taken as linear and direct by the USAID Index, whereas it has been commonly observed that it is difficult to know what underlying quality is being measured by such surveys: ‘Are they really making judgments about the quality of governance and particular institutions, or are they implicitly answering the question, ‘how do you think things are

⁴⁷ Ibid, 2.

⁴⁸ See Munck and Verkuilen, ‘Measuring Democracy: Evaluating Alternative Indices’, 12.

⁴⁹ USAID, ‘Measuring Fragility’, 6.

⁵⁰ See Fearon, ‘Do Governance Indicators Predict Anything?’, 27.

going today in country X (perhaps compared implicitly to countries in the region)?”⁵¹ Because the concept of ‘political effectiveness’ remains underspecified for the USAID Index, we cannot begin to determine how this indicator could be theorized as a component measure of this dimension or what weight to give it.

The USAID Index identified 33 outcome indicators claimed to measure effectiveness and legitimacy across political, security, economic and social dimensions. Each of these indicators suffers from one of the two difficulties described above: it is a subjective perception index, with an erroneous assumption that perception stands in a linear relationship to reality, rather than at best a very “noisy” approximation of complex trends, or; it equates a policy measure with an outcome, or assumes that outcome data reflects effective policy.⁵² But in all cases, the relationship between the indicator and higher order variable of effectiveness or legitimacy is dubious, both because the concepts of effectiveness and legitimacy are poorly specified, and because the arguments made for why the indicator adequately measures these terms are unsubstantiated. The result is a collection of indicators drawn from a large variety of datasets with widely disparate methodologies, which stand in for rather than measure concepts of effectiveness and legitimacy. The numbers become the concepts rather than represent or approximate them, creating a misleading impression that analytical problems of specification, commensuration and aggregation have been resolved because all inputs now appear as equivalent quantities.

Of the datasets relied upon by USAID to generate index inputs, some are regarded as well-settled, already widely in use and ‘hard’ enough to stand in as real measurements of the phenomena they purport to measure, and not something else. Others remain much more controversial, and the relationship between measure and referent is regularly contested. By summarizing data from these datasets into a one-number per country manner, the USAID Index works to absorb and erase the uncertainty surrounding the data and what it measures. This is perhaps most clear in the reliance on datasets that purport to measure the quality of governance, extent of discrimination, human rights abuse and regime type. The governance indicators are already aggregations of a large number of subjective perception surveys, and as noted above, it is difficult to determine whether each underlying survey measures institutional quality or some more vague set of perceptions.⁵³ The Index also uses indicators derived from the Minorities at Risk, Polity IV and Political Terror Scale to measure discrimination, regime-type and human rights abuses, respectively. Like the USAID Index into which they are aggregated, each of these data sets attempts to count inherently uncountable, qualitative phenomena. In order to do so, each relies on elaborate coding processes whereby a coder allocates a number (0 or 1, 0-2, 0-10, 0-5, depending on the dataset) based on his or her interpretation of qualitative information (‘content analysis’) about the political and social circumstances in the relevant state, or based on whether a given event has taken place (eg. a military coup) in a given year. Vast swathes of history and politics across numerous, heterogeneous, political orders are thus commensurated and transformed into a number rating of the country situation.

Reliability, in the sense of complete identity between information and ratings across different coders, is never achieved. But even where very high levels of consistency (above 90 percent) are achieved, fundamental uncertainties remain as to what a given rating actually represents in terms of a state’s political and social reality. One obstacle is endogeneity, as described by Fearon in relation to datasets coding discrimination: ‘these measures of political exclusion and discrimination are based on the

⁵¹ Ibid, 28. See also Andrew Williams and Abu Siddique, ‘The Use (and Abuse) of Governance Indicators in Economics: A Review’, *Economics of Governance* vol 9 (2008) 131-175; Svend-Erik Skaaning, ‘Measuring the Rule of Law’, *Political Research Quarterly*, vol 63 (2010) 449-460.

⁵² The latter assumption may be more plausible in the case of infant mortality or literacy, which the Index counts as a measure of “social effectiveness” of a state: see, for discussion, Gary King and Langche Zeng, ‘Improving Forecasts of State Failure’, *World Politics*, vol 53 (2001) 652.

⁵³ Fearon, ‘Do Governance Indicators Predict Anything?’, 28.

subjective judgments of diverse coders, trying to code somewhat impressionistic things. Countries where there has been no ethnic conflict and where ethnic relations have been calm are for that reason judged to have [little] exclusion ... one can reasonably worry that a coder's knowledge that there was an ethnic conflict in a country increases the probability that he or she judges that ... groups were discriminated against or politically excluded.'⁵⁴ McCormick and Mitchell note that even where a good level of reliability is achieved in coding human rights violations, there is a basic problem of the meaning of a given rating: 'human rights violations differ in type not just amount, such that they cannot be clearly represented on a single scale. Imprisonment and torture are different types of government activity [involving] differing uses of governmental resources and capabilities, and differing costs for the government ... Regimes use different mixes of methods of political control, a variation missed by a one-dimensional scale.'⁵⁵ Cingranelli, one of the authors of a major human rights scale, concludes that, due to inadequacies in the consistency of underlying data quality and availability, a human rights scale is not reliable beyond 3 or 4 categories, and ought not be used to derive an ordinal ranking.⁵⁶

The key point here is not that these various attempts to develop cross-national datasets of political qualia are necessarily meaningless; rather it is that we cannot readily reach conclusions about what they mean, or what a higher or lower number within a dataset represents in terms of the "effectiveness and legitimacy" of a state. Yet global scale indexes such as the USAID Index perform a further aggregation of aggregations: the interpretive ambiguity, noisiness, uncertainty and measurement error which are an inherent part of each of these datasets are simply erased through addition into a single score. Espeland and Stevens note March and Simon's description of the way in which uncertainty is absorbed as information travels upward within an organization:

'Raw' information typically is collected and compiled by workers near the bottom of organizational hierarchies; but as it is manipulated, parsed and moved upward, it is transformed so as to make it accessible and amenable for those near the top, who make the big decisions. This "editing" removes assumptions, discretion and ambiguity, a process that results in "uncertainty absorption": information appears more robust than it actually is. As March and Simon put it: "Uncertainty absorption takes place when inferences are drawn from a body of evidence, and the inferences instead of the evidence itself, are then communicated."⁵⁷

This insight has particular application to the new generation of global scale indexes. What is moved upward to a higher level of abstraction and aggregation is not 'raw' information, but already highly artificial composites and aggregates, the result of laborious construction, parsing, concept-making, judgment-calling and contestation. Transmitted upwards are sets of inferences and claims bundled into a one-dimensional figure, which the higher-level index then bundles further.

USAID's Index ranking for each country is obtained by averaging the 'effectiveness score' and 'legitimacy score' for each country into a 'fragility score.' The effectiveness and legitimacy scores are the aggregations of component indicators selected as proxies for each of the four dimensions of effectiveness and legitimacy. The step of aggregation is indispensable for deriving a single score for each state, allowing a seemingly transitive rank ordering to be obtained. Rank ordering is one of the

⁵⁴ Fearon, 'Do Governance Indicators Predict Anything?', 16.

⁵⁵ James M. McCormick and Neil Mitchell, 'Human Rights Violations, Umbrella Concepts and Empirical Analysis', *World Politics* vol 49 (1997) 513-14. In relation to the Polity IV dataset, Gleditsch and Ward note, "Categorical data encompass a small number of subdimensions that interact in non-obvious ways ... Data are multipath – there is a wide variety of ways in which politics can receive a single scale value. Vastly different temporal, spatial and social contexts support the same democracy and autocracy scale values." Kristian S. Gleditsch and Michael D. Ward, 'Double Take: A Reexamination of Democracy and Autocracy in Modern Politics', *Journal of Conflict Resolution*, vol 41 (1997) 380.

⁵⁶ Cingranelli, 'The Cingranelli and Richards (CIRI) Human Rights Data Project', 406, 408. The Political Terror Scale, used by the USAID Index, has 5 categories and purports to measure human rights "conditions" rather than the narrower "practices."

⁵⁷ Espeland and Stevens, 'A Sociology of Quantification', 422. Inline reference omitted.

most seductive features of unidimensional indexes: it creates an impression of precise differentiation between objects (in this case, states) in respect of a complex property (stateness) that seems to have been successfully reduced to a unidimensional measure. As Merry, Kingsbury and Davis note, indicators such as the Human Development Index gain authority in part because they become a ‘shorthand’ for a country’s circumstances and rankings imply standards to be met or shortcomings made legible (and perhaps actionable).⁵⁸

For rank ordering to fulfill these ambitions of diagnosis and prediction, the differences between country ranks must in fact capture some real distinctions and interactions in the dimension of fragility – and these differences must be amenable to transitive ordering in the sense that if Sierra Leone is ranked above Cambodia, and Cambodia is ranked above Cameroon, it must be true that that Sierra Leone is more fragile than Cameroon.⁵⁹ For fragility and its underlying dimensions (and the indicators of those dimensions) to be susceptible to this treatment, the problems of conceptual specification and proxy identification pointed out earlier must have been suitably resolved:

First, the analyst must make explicit the theory concerning the relationship between attributes [dimensions]. Second, the analyst must ensure that there is a correspondence between this theory and the selected aggregation rule, that is, that the aggregation rule is actually the equivalent formal expression of the posited relationship. For example, if the aggregation of two attributes is at issue and one’s theory indicates that they both have the same weight, one would simply add the scores of both attributes. If one’s theory indicates that both attributes are necessary features, one could multiply both scores, and if one’s theory indicates that both attributes are sufficient features, one could take the score of the highest attribute. In this regard, then, it is crucial that researchers be sensitive to the multitude of ways in which attributes might be linked and avoid the tendency to limit themselves by adherence to defaults, such as additivity.⁶⁰

To put it another way, an aggregation function for a multidimensional phenomenon like ‘fragility’ requires the solving of 2 kinds of problem: determining direction of causalities between fragility and its dimensions, and between dimensions,⁶¹ and determining a *numeraire* which allows one to calculate how many units of dimension A substitute one unit of dimension B. Without resolving these 2 puzzles, rank ordering beyond pair-wise comparison is arbitrary.⁶²

USAID’s background paper on Measuring Fragility contains no discussion of the difficulties of aggregation, and the Index states that Principal Component Analysis is used to derive a weighted average for the component indicators for effectiveness and legitimacy. The average is then standardized to make the mean for all countries 0, and a country’s deviation from the mean becomes its fragility score; a positive score connotes higher fragility; a negative score reflects lesser fragility. This generates the appearance of unidimensionality, but only by placing the statistical cart before the theoretical horse, sidestepping the validity problems raised by poor specification, doubtful proxies and no theory of aggregation.⁶³

⁵⁸ See generally Kevin Davis, Benedict Kingsbury and Sally Merry, ‘Indicators as a Technology of Global Governance’.

⁵⁹ A correct aggregation function is a set of rules that attributes to a vector of characteristics a single element in the range and additionally: (a) behaves monotonically, and (b) respects boundary conditions: G. Beliakov, A. Pradera and T. Calvo, *Aggregation Functions: A Guide for Practitioners* (New York: Springer Berlin Heidelberg, 2007) cited in Gutierrez Sanin, ‘Evaluating State Performance’, 32.

⁶⁰ Munck and Verkuilen, ‘Measuring Democracy: Evaluating Alternative Indices’, 24.

⁶¹ Recall that the USAID Index asserts 8 dimensions of fragility (4x2). King and Langche note that establishing interactions for a six dimensional space (assuming, plausibly linear and non-linear interactions) is “almost incomprehensibly immense.” King and Langche, ‘Improving Forecasts of State Failure’, 639.

⁶² See Gutierrez Sanin, ‘Evaluating State Performance’, 30-33 for detailed discussion.

⁶³ Gutierrez Sanin rightly points out that for indexes of state fragility, “the aggregation function is a substantial part of the theory” of the fragility concept itself.

Conclusion

Within the style of reasoning of contemporary statistical method, the numbers generated by USFFP's Failed States Index and USAID's Index do not appear robust or easily defensible. They are not very good numbers. Rather than consider how one might more successfully measure stateness, it seems to me that a pressing puzzle is the existence and growth of demand for rankings based on relatively poor concepts, data and methods.⁶⁴ As Lampland suggests, it may well be a mistake to assume that the effective uses of a number depends upon their veracity or some defensible argument for veracity.⁶⁵ Quantification is a formalizing practice that can serve a variety of purposes, even if a number's claim to objectivity and scientificity cannot be readily cashed out under the prevailing exchange rate for scientific method: 'Numbers are instruments, not simply transparent signs.'⁶⁶ But what are they instruments for, in this case?

Synoptic, attention-grabbing, provocative: these terms are used to describe the effects of ordinal ranking state-strength, even as they acknowledge that the methodological bases for ranking may not be persuasive to many.⁶⁷ The reduction of a complex, multidimensional reality (nothing less than the political and social order of a territory) to a single number is admitted by most to be susceptible to criticism. But as argued above,⁶⁸ the category of fragility seems to demand a means of classifying and comparing across vast scales of people, space and history. Closely intertwined with policy demands to globally forecast security risks and target sources of disorder within states, it is a problematic and heuristic which elicits panoramic techniques – a tendency evidenced by the numerous other attempts to quantify and index state fragility. Panoramas, as Latour argues, are artifacts that aspire to see the whole and show it in an ordered and coherent way: 'What is so powerful in those contraptions is that they nicely solve the question of staging the totality, of ordering the ups and downs, of nesting 'micro,' 'meso' and 'macro' into one another ... They collect, they frame, they rank, they order, they organize.'⁶⁹

A panorama of stateness, such as that presented by the FSI and the USAID Index, integrates a theory of political order (founded in institutions, achieved through technical efficiency and legitimate conflict resolution) with a raft of measures and numbers to generate a macro-level snapshot of degrees of effectiveness and legitimacy of 161 states across the world. Rendering degrees of political order calculable in this manner also changes the way in which it can be made visible and represented: unlike an expert opinion, a human intelligence report, or a historical monograph, the judgments and narratives embedded in a unidimensional number are easy to transpose into two-dimensional space, whether tabular (a list color coded by degrees of fragility) or pictorial (a map color coded by degrees of fragility or instability risk). In the color code used by USAID and USFFP, highly fragile is represented by red or orange, and low fragility by green or blue. The map creates a regionalized picture of fragility, in which there are zones of the world suffering lesser or greater degrees. Unsurprisingly, those zones in which red and orange predominate overlap significantly with underdeveloped or lesser developed countries. Those zones where green and blue predominate are first world/Northern countries and regions, with some exceptions. The effect of this aesthetic is striking: zones associated

⁶⁴ See Christopher Hood, Ruth Dixon and Craig Beeston, 'Rating the Rankings: Assessing International Rankings of Public Service Performance', *International Public Management Journal*, vol 11 (2008) 298-398 for similar puzzlement in relation a range of rankings.

⁶⁵ Martha Lampland, 'False Numbers as Formalizing Practices', *Social Studies of Science*, vol 40 (2010) 377-404.

⁶⁶ *Ibid*, 383.

⁶⁷ Similar terms were used by officials involved in the production of other Indexes of state fragility, to describe the value of an ordinal ranking: Author interviews, Washington DC, July 20 and 22, 2010.

⁶⁸ See above page 5.

⁶⁹ Latour, *Reassembling the Social*, 187-189.

with poor governance and instability are marked with a color connoting danger/threat/heat (red, orange), and zones associated with good governance and stability are marked with a color connoting safety/ease/calm or cool (green, blue). Indeed, the aesthetics of the fragility map resemble a ‘heat map,’ sometimes used to represent risks of contagion and spillover of epidemics or (more recently) systemic financial crises.⁷⁰

Panoramas of this kind, then, effect distributed cognition, by tying together different scales of information and different genre systems (theories of politics, models of political organization, measures of economic growth, ratings of government policy, subjective perception surveys) in order to make ‘an otherwise amorphous composite ... into a thing that holds together in the imagination of government officials and the general public.’⁷¹ Quantification of inputs eases the way to this knitting together, especially where scales are large, and obscures the nature of the genres being linked together. The (messy, unpersuasive) world-making that lies beneath this discrete numerical knowledge-object is obscured, and its seemingly solid and transparent qualities engenders what Sally Merry calls ‘knowledge-effects’: ‘Numerical measures produce a world knowable without the detailed particulars of context and history.’⁷² A striking example is the way in which the FSI is presented annually in *Foreign Policy* magazine: the Index is always matched with narrative stories and opinion pieces which make strong claims about the causes of state failure and how to address them – claims which are not in any way tested by the FSI. Thus, in the July 2010 *Foreign Policy* issue publishing the FSI for that year, the Index was published alongside an opinion piece by economist Paul Collier, entitled “Bad Guys Matter” in which Collier rehearses his well-known argument that the quality of governance and natural resource endowments are causally related to the onset of civil war and state failure.

The numbers are associated with global judgments about the quality and nature of political order in the territory. The numbers stand in for a judgment about a complex social reality (e.g., the ‘relationship between government and society’) but also tie in with beliefs, common sense notions and normative claims about what characterizes good and bad political orders and outcomes. The normalizing tendency of the concept of fragility and failure is reinforced but also rendered less transparent by the process of quantification; it is hidden behind claims of methodological authority and elaborate numerical artifacts.

⁷⁰ My thanks to Bruce Carruthers for bringing this similarity with heat maps to my attention.

⁷¹ Espeland and Stevens, ‘A Sociology of Quantification’, 412. See also Bowker and Leigh Starr, *Sorting Things Out*, 136.

⁷² Sally Merry, ‘Measuring the World: Indicators, Human Rights and Global Governance’, *Current Anthropology*, vol 52 (2011) S84.

