Crisis States Research Centre

WORKING PAPERS

Working Paper no. 58
- Development as State-Making -

THE QUANDARIES OF CODING AND RANKING : EVALUATING POOR STATE PERFORMANCE INDEXES

Francisco Gutiérrez Sanín
Universidad Nacional de Colombia

November 2009

Crisis States Working Papers Series No.2

ISSN 1749-1797 (print)
ISSN 1749-1800 (online)

LSE DESTIN Development Studies Institute

Crisis States Research Centre

The Quandaries of Coding and Ranking: Evaluating Poor State Performance Indexes

Francisco Gutiérrez Sanín
Universidad Nacional de Colombia

'According to the Celestial Emporium of Benevolent Knowledge, animals are divided into: those that belong to the Emperor, embalmed ones, those that are trained, suckling pigs, mermaids, fabulous ones, stray dogs, those included in the present classification, those that tremble as if they were mad, innumerable ones, those drawn with a very fine camelhair brush, others, those that have just broken a flower vase, those that from a long way off look like flies.'
Jorge Luis Borges

Introduction

This paper discusses and evaluates 'poor state performance' (PSP) definitions and indexes. By PSP indexes I understand all those categorisations and their respective measurement tools that are developed to differentiate contemporary states according to a set of criteria that produces a performance mark in a scale that goes from 'better' to 'worse'. The main PSP-index result is a rank of countries per year; and according to the established policy perspective that inspires the majority of them, the most interesting cases are poor performers and this is why I have chosen the umbrella term PSP.

The indices are fed by an already huge literature about state collapse (Zartman 1995; Milliken and Krause 2002; Hale 2004), state fragility (Andersen 2006; Chataigner and Gaulme 2006; Marshall 2008) and state weakness (Patrick 2006; Eizenstat et al. 2004; Ignatieff 2002), which aspires to: a) differentiate unstable or non-viable states from others; b) understand, and/or describe, a specific set of state outcomes; and c) in some cases, predict – based on a series of symptoms (FFP 2005), for example through systems of early warning (Carment 2003) – these outcomes. The evaluation of PSPs implies a dialogue with this literature. The sheer size of the intellectual effort involved, and the number of studies it draws on, reveals the importance and the seriousness that is attributed to the issue of (extremely) poor state performance, both in the academic and the policy worlds. Thus state failure has been considered by the United States to be a key security issue, with state failure in poor countries triggering migrations, international economic turbulence and regional instability. Index crafters themselves stress this point:

> 'Since September 11, 2001, the United States and other governments have frequently asserted that threats to international peace and security often come from the world's weakest states. Such countries can fall prey to and spawn a host of transnational security threats, including terrorism, weapons proliferation, organized crime, infectious disease, environmental degradation, and civil conflicts that spill over borders.' (Rice and Stewart 2009)

Unfortunately, and against all evidence, the overwhelming majority of them – whether qualitative or quantitative – have failed to acknowledge that differentiating 'good' from 'bad' state performance can be a hard classificatory task.

That the PSP literature faces classificatory problems can be readily seen. It is very common to find lists offered by experts that appear haphazard and lacking in organising criteria. For example, Zartman (1995) – one of the most cited and respected authors in the field – provided an early list of African failures that included 'newly independent countries, others that have fallen into deep problems, some that have witnessed long periods of economic stagnation', and even South Africa as a state at risk of failure. This is not an isolated example, and there is a plethora of definitions of PSP that are rarely explicit about their theoretical underpinnings (Di John, 2007) and tend to be unaware of the minimal requirements for an effective classification. After a careful evaluation, Cammack et. al. (2006) came to the conclusion that presently both researchers and policy makers count with labels, but not with well-defined PSP concepts:

> 'Fragile states' is a label currently in use by the international community to identify a particular class of states. Actors conceptualize the FS agenda differently according to their concerns and goals. The word 'fragile' is often substituted without a precise change in meaning by 'failed', 'failing', 'crisis', 'weak', 'rogue', 'collapsed', 'poorly performing', 'ineffective', or 'shadow'; a fragile state may also be called a 'country at risk of instability' or 'under stress', or even a 'difficult partner'. In most cases, these labels do not have a meaning that is clearly understood far beyond the author who has used them. Moreover, many of the FS definitions mix up the meaning of the word fragility with propositions about correlates and causal relations. Further complicating the matter, donor definitions appear to fall into three general but overlapping types: where fragility is defined in terms of the *functionality* of states, of their *outputs* (including insecurity), or of their *relationship* with donors.

Why are these flaws so pervasive? Maybe they appear only in narrative accounts, but not in the indexes themselves. It is supposed, after all, that among the key advantages of quantification are rigor, and the capacity and need of making explicit definitions and assumptions. I will show here that in many respects PSP databases and indexes are possibly even worse classificatory tools than impressionistic descriptions. The main claim of this paper is that this does not happen by chance: coding and ranking state performance are difficult tasks. In this spirit, and following Gutiérrez and González (2009), throughout the paper I treat separately two types of problems:

A. Those exhibited by the majority of the PSPs, which are more or less solvable, and probably characteristic of quantitatively immature fields: poor conceptual definition, huge conceptual dispersion, inconsistencies, confusion between causes and definitions (Cammack et al. 2006), and an unresolved political economy of knowledge;

B. Those exhibited by PSPs, but shared with practically all cross-national contemporary political science databases where there is no benchmark solution: intrinsic ambiguity and lack of awareness of the problem of order.

Even a cursory study of PSP indexes reveals that they suffer many Type A vices, for which in some other fields scholars have already produced good solutions (Marshall and Jaggers 2009; Pearson and Baumann 1993). In this paper I claim that Type B concerns generally have not been unearthed, let alone tackled. I believe that they have been ignored because of a

mechanical exportation of standard assumptions and techniques used in economics to different conceptual worlds, without caring to consider their soundness in the new context. This operation generates deeply hidden assumptions which are comfortable to work with, but which can produce analytical disasters.

The discussion is organised around two types of object: *databases* and *indexes*. PSP indexes are generated from databases. Databases are rectangular arrays of data, where generally the rows are cases and the columns are variables. In this context, cases will always be countries/year. Thus, building a PSP index entails at least the following steps:

a. Establish a definition of state, and then a definition(s) for PSPs;

b. Identify the information that is needed to evaluate state performance, in the terms defined in a.;

c. Find, code and store the relevant information, or adequate substitutes (proxys);

d. Clean the data, eventually transforming it and reducing dimensionality;

e. Create a formula that uses all the relevant information to produces a single number for each case (aggregation);

f. Document the process.

I will suggest that PSP index builders have failed to identify, and thus to solve, the extremely serious difficulties that they face in each step of such a process. The paper is organised in the following fashion. In the first section, I focus on definitional issues, suggesting that definitions of poor state performance generally stem from prototyping, which in this context has clear downsides. Confusing, sometimes inconsistent, definitions generate haziness, endogeneity and ahistoricity. The second section is dedicated to intrinsic ambiguity (which riddles the whole process). I claim here that any precise (in the sense of admitting crisp coding) definition of PSP is spurious. The third section discusses order and aggregation, arguing that the deeply hidden assumption over which PSP databases operate is that there is a 'numeraire' that allows the agent to substitute units of one variable for units of the other. This assumption is at least moot, and has prevented a fine-grained discussion of the meaning of aggregation and multi-attribute decision making in the context of political and social databases. In particular, I claim that in a very specific, and important, sense, PSP indexes are conceptually vacuous. The conclusions put the pieces together, and suggest that at least some of the most severe problems I have highlighted are tractable.

The main examples are taken from the best known and most utilised databases – LICUS, Failed States Task Force, Fund for Peace (FFP), Brookings Institute, among others – not because I consider them particularly weak or open to criticism, but because they express very well mainstream practices and common assumptions.[1]

[1] Polity could be included here, because a focus of fragility research is the relation between poor state performance and lack of democracy. However, the high quality of this database, and the awareness of its administrators of potential pitfalls, make it an extraordinarily important 'laboratory' to investigate the problems of current quantitative cross-national research (Gutiérrez and González 2009). Thus, I only use it to illustrate potential general database and index problems, without engaging in a systematic analysis.

Conceptualisation

The reader will find in Annex 1 the definitions of poor performance of some major PSP databases. The FFP (2009) presents a list of negative attributes: loss of territorial control, loss of legitimacy, inability to provide reasonable services and inability to interact as a full member of the international community. The CIA-sponsored State Failure Task Force offers a 'narrow definition' consisting of the collapse of state authority; but since these events are rare they produce a broader one, which includes 'genocide', 'disruptive regime transitions' and 'revolutionary wars aimed at displacing the regime'. Carleton University (http://www.carleton.ca/cifp/ffs.htm) creates an explicit scale that ranges from fragility to failure and collapse. For USAID (2005), 'it is more important to understand how far and quickly a country is moving from or toward stability than it is to categorise a state as failed or not'. Though stability seems to be the most cherished value, the definition is organised around the notion of the direction of change (positive or negative). LICUS includes in its criteria weak institutions, poor governance, political instability and frequent violence (and its delayed effects) (World Bank 2006).

This reflects Cammack et al.'s (2006) observation, according to which there is a serious confusion in the literature between the concept and its consequences. Thus, numerous symptoms are included in the definition, without reflecting if they form a necessary and sufficient set. As King and Zeng (2000) say in their excellent criticism of the Task Force results:

> …in some sense the explanatory variables (infant mortality, partial democracy, legislative ineffectiveness, and so on) are really indirect indicators that the state has already failed, whereas their heterogeneous dependent variables (genocide, disruptive regime transitions, and revolutionary wars aimed at displacing the regime) are not really measures of state failure but instead are indicators of some of the disastrous *consequences* of state failure.

Furthermore, it is not very clear why a certain set of symptoms, and not another one, is plugged into the definition. The State Failure Task Force, as Di John (2007) notes, adopts a core definition, but then replaces it by another, quite different one, claiming that the latter – the one that is effectively used – could be operationalised, while the former not. But the difference between the true and the operational one is so big that it is not clear how the quantification is supposed to relate to the original idea. For some reason the FFP introduces environmental decay as a symptom of failure. The provision of services is also much trickier than what index creators and administrators seem to think, and it is a good example of the analytical issues at stake: first, it creates a severe problem of endogeneity for any model that wants to correlate level of development with failure; second, it does not differentiate between good and bad state performers at low levels of development; third, it needs to be questioned how the administrator and coders of the database differentiate between an 'adequate' and an 'inadequate' provision of services, who creates this yardstick, whether it remains unchanged or evolves, and if so how; and fourth, there is no metric to evaluate the intensity of the symptom. For example, weakness/fragility/collapse is related to situations when 'crime goes out of control', or to expressions like: 'failure to meet the needs of the citizens'. No clear instructions are to be found in the codebooks of the databases as to how this may be measured or coded.

Part of the maddening heterogeneity incorporated into PSP definitions is related to the fact that they generally are a product of prototyping. What several researchers have done is to find

typical instantiations of failure, and then have tried to extract their common characteristics. For example:

> State failure is a label that encompasses a range of severe political conflicts and regime crises exemplified by macro-societal events such as those that occurred in Somalia, Bosnia, Liberia, and Democratic Republic of Congo (Zaire) in the 1990s. …At its worst, political instability results in a total or near-total collapse of central political authority – in short, a failed state – that can produce humanitarian crisis and ungoverned zones. (Bates et al. 2003)

Prototyping 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 the others do not necessarily fit more moderate examples. This straightforward, but important, point will be returned to below.

Additionally, it does not take into account the possibility of the existence of different types of failure. If there are several ways of falling apart, then a single definition of failure (i.e. a sheer vacuum of organised activity from above at very low levels of development) will not necessarily do the job. This can be viewed from two perspectives. First, *cross national*. Over the recent years we have witnessed several events where *de facto* and *de jure* sovereignty have broken down, and the state – as a unitary entity – disappears. But they seem to be everything but homogeneous. Czechoslovakia, the Soviet Union and probably Albania collapsed in a very real tangible way (Rosenau 1969). Then there are cases like Congo and Somalia where 'the state authority has ceased to exist during long periods'. Iraq's recent breakdown, produced by an invasion, was different, and its strengthening or fragilisation continues to be a matter of heated debate. These examples reveal a much more complex landscape than a linear progression (or retrogression) from 'strength' (or any similar term) to 'weakness'. What we find instead is spectacular success punctuated by brutal shocks, with critical exogenous events playing a major role, and sudden and sharp recuperations. This picture matches much better both the historical evidence and state building theory (Skocpol 1984; Tilly 1992). PSP indexes put all these states of the world in the same basket. For example, the FFP rates very highly (which in its scale is negative) both North Korea and Congo (they get more or less the same marks, although the former appears systematically a bit worse). There is no doubt that for citizens of both countries life is bleak, but in radically different senses. In the first case they are smothered by the total presence of the state, in the second by its absence. Is it reasonable to assert that the North Korean state is as fragile or weak as the Congolese?

Second, *longitudinal*. Here, things appear even hazier. The basic problem is that – as, for example, Skocpol (1984) has shown very clearly in her works about revolutions – 'strength' or 'fragility' are relational, not absolute, terms. However, the point is not captured even by such weathered researchers as North, Wallis and Weingast who, in their most recent book on state performance (2009), systematically speak about England in the sixteenth century in terms of a 'fragile' state. It is true that they need a toolkit of terms that are placed high in the ladder of abstraction (Sartori, 1970) to fulfil their promise –no less than 'building a conceptual framework for interpreting recorded human history'-- but precisely the litmus test of that type of terminology is that it must preserve its basic desirable characteristics when applied to different periods (in this case, its relational nature).However, England may or may not have been fragile in some sense then, but certainly it was not fragile as Somalia or Afghanistan are today. At the other side of the quantitative-qualitative barricade we find

exactly the same issues. Polity, for example, gives Switzerland in the 1960s ten out of ten on the democratic scale, without taking into consideration that women were disenfranchised. I wonder if a country without female electoral participation would get the same generous mark today. The index administrators (users) do not have any tool to express that in the 1960s that restriction was already rare, but still acceptable for regimes classified as 'democratic', while today it is not.

Be this as it may, to be able to feed working indexes, prototypes have to be transformed into operational terms. Taking almost any clause of the definitions offered by the PSP databases, it becomes clear that they are really not operational. Practically no work related to the disambiguation of terms has been done. The result is that the *definiens* is many times more obscure and impenetrable than the *definiendum*. In reality, we confront here a double problem: analytical and empirical. Analytically, if we do not have the right to consider the expression 'state failure/collapse/fragility' obvious, then we cannot consider obvious the expressions 'good governance', 'legitimacy', 'adequate provision of vital services', among others. Empirically, it is not clear that state fragility is less observable than, say, legitimacy (which it should be, because we want to replace more abstract terms by others that capture part of their meaning and that are more directly observable). Using hazy notions as if they were genuine solutions to definitional problems interacts with a political economy of knowledge. You do not have to be an extreme relativist to understand that the authoritative definition of a state of the world can be part of that state of the world. In other words, codification by multilateral agencies can have direct political and economic consequences. For example, LICUS and CPIA ratings are used to approve credits and access to other resources that low development countries need desperately. Such terms as 'good governance' are not only very flexible; they have meant adopting policies similar to those favoured by the Bank (IEG 2006).

Finally, PSP indexes are plagued by inconsistencies (some of them pretty obvious). Take USAID's Fragile States Strategy. On the one hand, its assertion that what matters is the direction of change is rather obscure. On the other hand, it has four interrelated priorities: (i) enhanced stability; (ii) improved security; (iii) institutional and policy reform; and (iv) developing institutional capacities. It does not seem to have occurred to the strategy crafters that institutional and policy reform, if genuine, can get in the way of enhanced stability; and developing institutional capacity can also create huge problems (Huntington 1996). Expressed more generally, it may be the case that achieving good thing A prevents achieving good thing B, or that providing A for agent X prevents providing A for agent Y. Here the USAID appears to succumb to a typical composition fallacy: all good things and all bad things come together (Hirschman 1981; Putzel 1997).

In sum, since causes, consequences and correlates are conflated and inserted into the definitions, there is a huge problem of endogeneity (Woolridge 2002) sufficiently strong so as to make suspicious any regression that utilises PSP indexes[2].

Intrinsic ambiguity

In the previous section, it was seen that PSP databases tend to leave many loose ends – and Cammack et al. (2006) show many more examples of this. However, the gargantuan efforts by databases like Polity to specify the meaning of every single term reveal the limits of

[2] The endogeneity is increased with poor model specification and measurement error, which are typical of PSP exercises (see infra).

disambiguation in the social sciences. For example, Polity's codebook (Marshall and Jaggers 2009: 14) offers the following definition of the term autocracy: [3]

> Institutionalized Autocracy: 'Authoritarian regime' in Western political discourse is a pejorative term for some very diverse kinds of political systems whose common properties are a lack of *regularized* political competition and concern for political freedoms. We use the *more neutral* term Autocracy and define it operationally in terms of the presence of a distinctive set of political characteristics. In *mature form*, autocracies *sharply* restrict or suppress competitive political participation. Their chief executives are chosen in a regularized process of selection within the political elite, and once in office they exercise power *with few* institutional constraints. Most modern autocracies also exercise a *high degree* of directiveness over social and economic activity, but we regard this as a function of political ideology and choice, not a defining property of autocracy. Social democracies also exercise *relatively high degrees* of directiveness. We prefer to leave open for empirical investigation the question of how Autocracy, Democracy, and Directiveness (performance) have covaried over time.

Here can be seen not only the effort to be clear, but that nevertheless many of the terms can only be quantified through a purely subjective evaluation. This suggests that there is a ceiling to disambiguation. This is the phenomenon of 'intrinsic ambiguity'[4], which is an inevitable characteristic of the social sciences, since almost all social scientific definitions include clauses related to the degree to which the phenomenon is observed. From the linguistic point of view, these are heavily hedged expressions, full of modifiers such as 'almost all' or 'heavily'. In the definition above, 'authoritarian regime' is replaced by a more neutral terminology that depends on hedges like 'relatively high', 'few constraints' or 'high degree'. Similarly, in the FFP definition of state failure it is asserted that a state is in dire straits if it does not offer 'reasonable' public services, shows 'extensive' corruption, suffers 'sharp' decline and is exposed to 'severe' pressures (FFP 2009).

This of course is problematic for the coder, who has the difficult task of translating such vague indications as 'relatively high' and 'high' into a string of 0s and 1s, or numbers on a scale. The coder faces two distinct challenges: firstly, associate each term with a position on a scale; and secondly, establish the cut-off points. For example, if we want to make a regression to know whether a certain PSP index is associated with the presence of internal conflict in some previous period, then we go to a conflict database and capture the data about conflict presence or absence – which is almost always in binary form: is a country at war or not? The original criterion – at least 1,000 combat-related deaths, of which at least 5 percent are caused by the weakest side – has resulted too stringent, excluding some obvious situations of confrontation, so several agencies have made successive readjustments, the most radical one being the '25-casualties standard', above which there is war, and below peace. It is easy to see that solutions of this type do not solve the original problem. First of all, they simply replace 'type-α' (excluding genuine events) by 'type-β' errors (including false events)[5]. At the same time they do not address the cut-off point issue: why 25, and not 24 or 26? Can a

[3] I take the example from Polity, because by far it constitutes the most sophisticated effort to operationalise quantitatively a social scientific concept, and to walk the researcher transparently through each step of that operationalization.

[4] I do not give to this a negative or a positive connotation.

[5] More pedantically, this is a tradeoff between a probability of incurring in α 's and a probability of incurring in β 's

country that has had 24 casualties annually during ten straight years be considered any less at war than one that presented 26 in a single year?

But this takes us directly to the problem of the quality of the data: another big source of uncertainty in the social sciences, especially in the PSP field. It is in the nature of things that state-performance information is directly proportional to the level of development and the strength of the state. Indeed, there tends to be a fairly good correlation between the quantity of missing data in a database and GDP per capita[6]. This means that generally the data will be scarcer, and poorer, where there is most need of it. This is not a fully intractable problem, and there have been some excellent approaches and concrete solutions, such as that of Ball (1996), who set the gold standard for the evaluation of conflict-related deaths. But neither is it a trivial one. It is not very important when there are steep differentials of state capacity, because the fact that data is difficult to retrieve does not mean that anything goes. Afghanistan's GDP per capita is missing for several years, but nobody would claim that those missing values could be equal or higher than, say, Singapore's. On the other hand, if we want to compare Uganda with Rwanda, and we do not have reliable data about their GDP per capita, then things get uglier. Note that the presence of poor data increases the nastiness of the cut-off issue. If my criterion is 1000 casualties, or 25 casualties, and if I know that my sources are shaky, what should I do with a count of 24 or of 26?

One direct way of attacking this problem is by transforming numerical into categorical data[7]. This is a clear improvement with respect to spurious precision: for example, you know that you can put country A in the lower development category if its GNP falls within a certain band[8]. However, it does not eliminate ambiguity: on the one hand, the problem of cut-off points persists; and on the other, categorisation now combines numerical and linguistic labels. These are two quite different types of label, but nevertheless they tend to be treated in PSP databases –and more generally in cross-national research – as if they were identical. A person counting trade, or homicides, or transit accidents, and producing a numerical label, bases her result on some kind of written record, and operates under the assumption that if somebody else makes the same count he will (almost) always come to (nearly) the same result. However far can we be of this ideal in extreme cases that anybody can cite easily, this operational assumption holds[9]. Linguistic labels are essentially different in nature, since an observer – say, a coder[10]— produces a subjective evaluation based on an unspecified corpus of evidence. Once again, there are several situations in which nobody would seriously put in doubt which is the 'correct' label. Does North Korea have more than 'few' (formal) constraints on its executive? No. Are the constraints of Sweden high? Yes. But the majority of comparisons – and almost all the really interesting ones – are much less clear cut. For example, is political participation 'sharply' restricted in Singapore, the Philippines, Venezuela or Colombia? More

---

[6] The correlation is high but not perfect. For example, European centrally planned economies of the 1960's had high-intermediate levels of development and strong states, but produced almost no data for researchers.

[7] The other way is imputation of missing data. There are already very sophisticated imputation techniques (see for example Little and Rubin, 2002; Box et al,1994), which by the way are rarely used in PSP indexes (some simply delete cases with missing data, which is not correct and introduces serious distortions). I will wholly skip the issue here. Suffice it to say that above a percentage of missing data, imputation is limited.

[8] For example, the World Bank's GNI categorical data is extremely useful. Note that here we are dealing with intervals: we know that country A's GNI per capita is between 0 and 200.
http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/0,,contentMDK:20420458~menuPK:64133156~pagePK:64133150~piPK:64133175~theSitePK:239419,00.html

[9] Another wholly different problem is if the record is accurate.

[10] Not necessarily human. For example, the Fund For Peace uses machine classification (through CAST, an in-house developed software) to evaluate indicators of state performance through the press and specialized reports.

importantly for PSPs, because the ultimate product will be a rank, where is it more sharply restricted? We know that even in the most favourable scenarios (evaluations provided by experts) the answer is likely to diverge – sometimes sharply. It would be an error to assert that linguistic tags only reveal the preferences of the coder. It would be more correct to state that they are different from numerical ones, and that they cannot be treated on the same footing. A very substantial portion of the data offered by PSPs consists of linguistic, not of numerical, labels. However, in econometric and quantitative exercises – but also in policy making and lay discussions – they are treated as if they were *bona fide* hard data. This then becomes a circular fiction: we speak to produce the facts, and then let the facts speak.

All this simply underscores that social science quantification faces several types of uncertainty (not only probabilistic). Social conflicts, ambitions, aspirations and cravings are expressed in natural language; so too are social science concepts. Thus, as long as we are talking about history and macro interactions, total disambiguation is a failed project. I believe it can be comfortably claimed that any viable social scientific definition is heavily hedged. This, of course, vey much includes state performance. For example, North, Wallis and Weingast (2009: 41) assert that there are 'no sharp borders' between their ideal types of states, and continually stress the fluidity of the categories at the borders of the classificatory space. I read this in the following sense: disambiguation is impossible, but even if it were not it would nevertheless be incorrect, since it would introduce an un-called for assumption of crisp borders between the state types. Modellers should take this observation very seriously. A third source of uncertainty is the inverse relation between the quality of the observation and the interest of the event. For PSP and conflict researchers, the focus of attention is on those situations in which data are critically poor.[11]

PSP indexes have several idiosyncratic flaws that can and should be corrected. The real inquiry, though, starts after they have been overcome. Operationalising hedged concepts, establishing cut-off points, handling interval data and dealing with linguistic labels are not trivial tasks, and we can only criticise PSPs for wholly ignoring them: indeed, to date there is not a single reference to them. The problem is therefore not the use of linguistic labels and hedges, but the inability of our current practices and formalisms to take ambiguity on board and recognise and incorporate the resultant added layers of uncertainty into the models.

Aggregation and order

As was indicated in the introduction, the final necessary step to produce a PSP index – or more generally, any index useful for cross-national research – is aggregation. You have many variables, say N, with different numerical and possibly linguistic tags (see Figure 1), and a multitude of cases You want to put all your cases into a scale, perhaps allowing for ties. This is a typical multi-attribute decision-making operation (Kahraman 2008; Zanakis et al. 1998). There are many choices of this type in everyday life. Recurring to the typical example used in the multi-attribute decision-making literature: you want to purchase a car, and you take into consideration security, petrol efficiency, power and price (see for example Lootsma 1997). There can be cars that perform very well with respect to one or two variables, but poorly with respect to others (see Figure 2, where higher marks mean better). The operation appears relatively unproblematic (although, as we will see, it generally requires a more sophisticated solution than what several PSP indexes offer): you aggregate all the variables to get a single mark, or rank, on which to base your decision, and then choose the car that is rated higher.

[11] Polity has advanced in this regard, introducing a variable about the quality of data (Marshall and Jaggers 2009: 31).

Figure 1

	Variable 1	Variable 2	Variable 3	Variable 4
Case 1	1	5	6	8
Case 2	2	2	3	4
Case 3	5	3	2	1
Case N	4	3	2	1

Figure 2

	Security	Efficiency	Power	Price
Car 1	5	5	1	4
Car 2	3	3	3	5
Car 3	2	5	2	5
Car 4	5	3	5	2

Ranking state performance is heavily multi-attribute. Generally, state performance is evaluated by ten or more variables, thematically clustered in boxes (or baskets) of variables[12]. The index is generated from the boxes. Thus there are two sub-steps: aggregating variables and aggregating boxes. Here many problems appear. To start with the simplest possible scenario, suppose you have to make a final binary decision – the state is failed or not – and you have three boxes, A, B and C, all three also binary. How many, and which, 'ones' ('successes') will you require to classify the state as failed? All three? Two of three? One? Perhaps only the first two, or the third? This is the logical-conceptual dimension: will you use in the definition a logical AND, a logical OR, or a combination of both, to decide when are you in the presence of the phenomenon of interest? Taking an example from a related field, Rosenau's canonical definition of invasion included two variables, linked by an AND (Gutiérrez and González 2009). So you do not tag any event as an invasion unless you clearly observe *both* variables in action. As happens in many other fields, PSP indexes generally ignore the issue. They do not try to establish conceptually what combination of boxes, or of variables, will produce enough evidence to assert that we are in presence of the phenomenon of interest (for example, failure).

Instead of this, they directly proceed to the operational aggregation of both variables and boxes. Here there are two very fundamental, and interrelated, issues that deserve careful consideration.

The first is a typically 'deeply hidden assumption' about the nature of the variables. The aggregation proceeds as if there were a common abstract counting item, what economists call a'numeraire', that enables the decision maker to count how many units of variable A substitute one unit of variable B. In the example of the car buyer of Figure 2, he gives every variable exactly the same weight; he will tolerate the loss of quality in dimension 1 (security) if there is an identical improvement, say, in dimension 2 (price). The assumption of the existence of a numeraire is an essential clog in the machine of economic theory[13], and it is both powerful and reasonable for the type of problems for which it was crafted. The question,

[12] Given the complexity of the problem involved, this is a very good practice.

[13] And of course it need not be money. The underlying counting units are von Neumann utilities, but with a series of additional –and also very reasonable—assumptions, the observable numeraire becomes in effect money.

however, is if it holds in the new context of cross-national research and social and political variables.

Since all the databases take for granted the existence of the numeraire, they do not discuss explicitly their aggregation function; there is no theory behind the choice of one or another function. An *aggregation function* is a rule that attributes to every vector of data (case) a single number in the range (in this case, a mark, or a position in the rank). As seen in Annex 2, both LICUS and the FFP's aggregation functions are straightforward: they sum up the values of their variables. This means that according to FFP a loss in one variable – for example, an increase in atrocities – can be substituted by an improvement in another one – say reversing the brain drain: I compensate for a massacre if I am able to repatriate one scientist. I do not believe that anybody would be ready to say as much – and of course there is no theory to sustain such proposition – but this is the concrete meaning of that aggregation function. Please note that I am not suggesting that a weighted average is incorrect in general – this would be pure obscurantism. What I am saying is that the implicit assumption of weighted averages is that there is a substitution rate between the variables, and that as soon as one unpacks that assumption it becomes evident that it does not always hold in the context of social and political databses.

Table 1 - Order indicator h(x) for South Africa and Uganda according to Polity between 1991 and 2005
Source: Polity IV Project

Year	South Africa	Uganda
1980	0,926829	0,853659
1981	0,926829	0,853659
1982	0,926829	0,875
1983	0,925	0,926829
1984	0,97561	0,926829
1985	0,95122	Missing
2000	0,926829	0,341463
2001	0,921053	0,368421
2002	0,923077	0,333333
2003	0,921053	0,342105
2004	0,921053	0,342105
2005	0,923077	0,333333

h(x): Define a top and a bottom, for which the number of up-sets are respectively 0 and 1. Include in the count of up-sets the cases that are equal or better in the precedence relation established in the text. Divide this by the total number of cases. Then substract this from 1 and you get h(x). It is easy to demonstrate that it is a correct aggregation function.

How much monopoly of the state, then, would you be ready to sacrifice to obtain democracy, or vice versa? How much security for liberty? The immediate reaction of the reader will probably be to retort that, if there is an answer, and many times there isn't, it will depend: for whom? This is completely correct. When there is a numeraire, by a series of reductions and reasonable assumptions the existence of a homogeneous decision maker in a concrete domain

of activity can be reasonably hypothesised[14]. Of course, variations on the theme are allowed and fruitful[15], but the common theme is there. When there is no numeraire, multi-attribute choice typically relies heavily on the input of the decision maker (Kahraman 2008). Actually, one of the critical aspects of multi-attribute choice is its capacity for eliciting explicit criteria from those who decide and whose interests are being represented in the process of index creation: for example, the car buyer. However, in the PSP field there is no homogeneity – only the fiction of it, which produces serious anomalies. To offer but one example, suppose we are looking for a form for evaluating the relative position of any case X in my database. The simplest method would be to count the cases that are above (up-sets) or below (down-sets) X and then divide this number by the total of cases[16]. Let us use the former[17] to see how the relative democratic rankings of South Africa and Uganda evolved with respect to other African countries, as shown by Polity. After nnormalisations and some very simple tinkering (see Table 1), my index h(x) goes from 0 to 1,with 1 equivalent to 'placed at the top' and 0 'placed at the bottom'. The reader will see in the Table that, according to Polity: a) South Africa systematically ranked among the most democratic African countries in the 1980s, in the midst of apartheid; b) between 2000 and 2005 it slightly fell in the ranking, which means that in relative terms it lost democratic standing in Africa after the termination of apartheid; and c) Uganda was very near the top of African democracy between 1980 and 1985, under the murderous regime of Milton Obote, while between 2000 and 2005 it was near the bottom.

This has two explanations[18]: first, Polity favours the criterion of competitiveness of the executive over all those it does include[19]; and second, Polity does not include in its variable list either levels of repression or civic liberties.[20] It is legitimate to ask, though, if an index with such anomalies as the South African one can be utilised in a probabilistic model to arrive at credible conclusions. The mild answer is: at least with a lot of caveats. The Uganda rating is also moot, though in a different sense. Even if you restrain your definition of democracy to electoral competitiveness, the impact of very high levels of repression on this should be taken into account. Be it as it may, my hunch is that with high probability South African and Ugandan decision makers would find these ratings rather strange. Multi-attribute decision making without a discussion of the weights given to the different criteria by a decision maker

[14] 'The consumer', 'the producer', 'the politician', 'the bus customer'. We can suppose that, ceteris paribus, THE bus customer prefers a cheaper than a more expensive, and a shorter than a longer, trip; the genuine exceptions are few, and contrived. See the key comment of Przeworski (2004) on this: state and markets 'The theory ['rational choice' or 'strategic action'] works only if we can identify classes of individuals in some structure of conflict and plausibly attribute to them some objectives. To put it differently, the political economy approach works only when it is imbued with sociology. This is why it is hard to say anything about 'individuals' or even 'voters'. They are heterogeneous. Some want one thing, some want another…The more sociology we can build into theories, the greater the benefit of the economic approach' (Przeworski, 2004; 86)

[15] Degree of risk aversion, for example.

[16] Actually, this innocent looking example has more to it than what meets the eye. Counting the relative size of the up-sets or downsets of each case in a database is itself a well behaved aggregation function, which solves several of the problems I am discussing here, and which utilizes an important database order structure (Gutiérrez, in preparation).

[17] Using downsets we arrive at basically the same conclusion

[18] The counter-argument that the count of up-sets measures only the relative position is true, but has no implication for the present discussion. The African countries did not improve too much in absolute terms. The (absolute) mark of South Africa did not improve after the fall of the apartheid either, and the Ugandan one decreased.

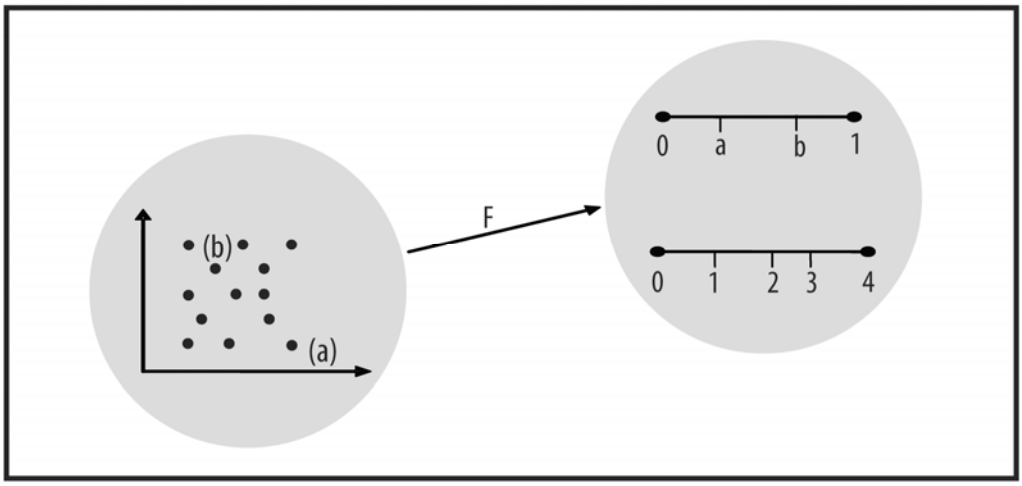
[19] Actually, Polity is an index of executive competitiveness, not of democracy, as is shown in Gutiérrez and González (in preparation).

[20] And explicitly says so in its codebook, which makes it so much better than several other databases. (Marshall and Jaggers 2009: 13).

with an explicit utility function is in reality an act of black magic –with the consequential results.

My claim is that they are analytically vacuous: in reality, they say nothing.

Figure 3
Source: Gutiérrez and González 2009.



This brings us to the second fundamental issue. In the absence of a numeraire, the aggregation function takes us from a partially ordered world to a totally ordered one. Figure 3 shows how a vector with many numbers is intrinsically different from a single (natural or real) number in the following regard: the latter can be perfectly ordered from bigger to lower (better to worse), while the former cannot. It might be expected that PSP indexes behave well at the borders – very well and very poorly performing states – because, as was seen above, they are prototyped to do so: by construction, they fit well at least some of these cases. But on the one hand, indexes promise full coverage (every country can be ranked and compared with any other one), and on the other you do not really need an index to conclude that, for example, Norway's state is stronger than Haiti's, or Germany's than Colombia's. The quality and utility of an index is a function of its discriminating capacity with respect to the intermediate cases. How do current PSP indexes fare in this regard?

I will present the argument very informally (anyway, it is quite straightforward[21]). First of all, transform and normalise all the variables and boxes so that they go in the same direction and have the same top and bottom – for example, the higher the mark the better. Then, create a relation of precedence ($\preceq$). Case A precedes case B, A$\preceq$ B, if all the numbers contained in A are less or equal than those contained in B[22]. For example, Haiti precedes Norway. Note that some countries cannot be compared by this precedence relation, because they are not better, or worse, in every aspect than the other, like cars 2 and 3 in Figure 2. Count all possible paired comparisons between cases[23]., then insert into one category (call it set C) those paired comparisons that can be made according to the precedence relation $\preceq$, and in another (say,

[21] The formal version is in Gutiérrez (in preparation).
[22] I.e., $A_i \leq B_i$ for all i=1,…,n. This is the last 'formula' I plug in; the rest of the discussion proceeds strictly verbally.
[23] There are Binomial(N,2) of these, N being the total number of cases.

set INC) those that cannot. Count both, and determine the proportion. This is, of course, a purely empirical question. For Polity, for example, the 'rate of non-comparability' (the relative size of INC) is rather low[24]. However, in PSP databases non-comparability is by far the most common situation, where INC is bigger than C (see Table 2). This produces a dilemma: if the rate of non-comparability is almost negligible, then one-dimensional decision making is being masked with fake multi-dimensionality; on the other hand, if it is high, then you have to decide what you will do with set INC.

Table 2 – Non comparability

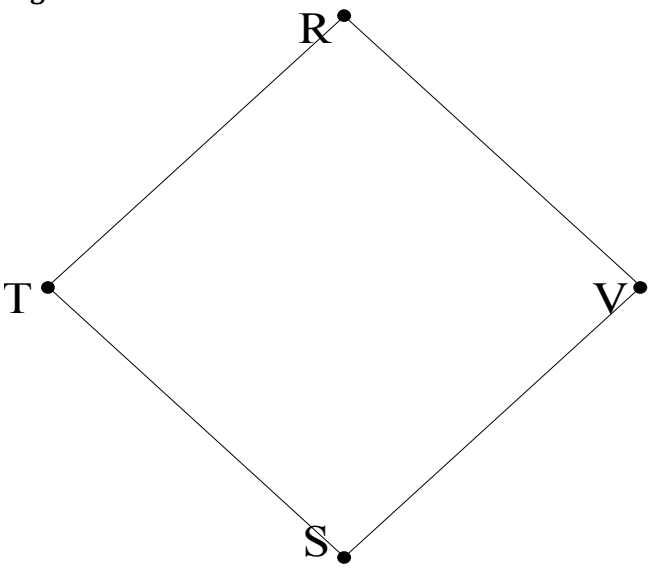
	Fund for Peace	LICUS
Number of pair wise comparisons	21316	5625
Members of C	8334	1861
Members of INC	12982	3764
% of INC	60.9	67

This boils down to the question of how to construct your aggregation function. 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 (if case A $\preceq$ B, then A is ranked lower than B); b) respects border conditions (if A has the higher marks in all its 1, 2, …, n characteristics, then it will be attributed the highest rank; if it gets the lower ones, it is bottom ranked) (Beliakov et al. 2007; Bustince et al. 2008).

If a correct aggregation function F is created, then F will rank all the cases. Suppose there are two cases, T and V, which are not comparable (they are members of INC). F will rank T above V, V above T, or will declare a tie. The claim is the following: for any T and V that belong to INC and for every correct function F, you can find at least one other correct aggregation function G that changes the relative positions of T and V while maintaining the relative positions of the members of C (the feasible comparisons). I will illustrate this with a very simple example. Suppose we have four countries, R, S, T and V. R is at the top, so S, T and V $\preceq$ R. Also, S is at the bottom, so T, V and R $\succeq$ S. However, T and V are not comparable (see Figure 4).

[24] Precisely because, as said above, *au fond* it is one dimensional.

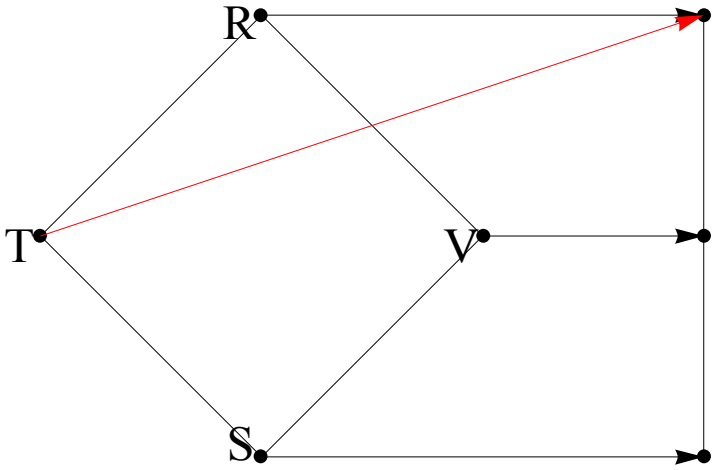
Figure 4



I want to rank the four countries, for example on a scale of 1 to 3. I build the correct aggregation function F (monotonous and well-behaved at the borders), as shown in Figure 5.

Figure 5 – The function F

It ranks T better than V. It is a correct aggregation function

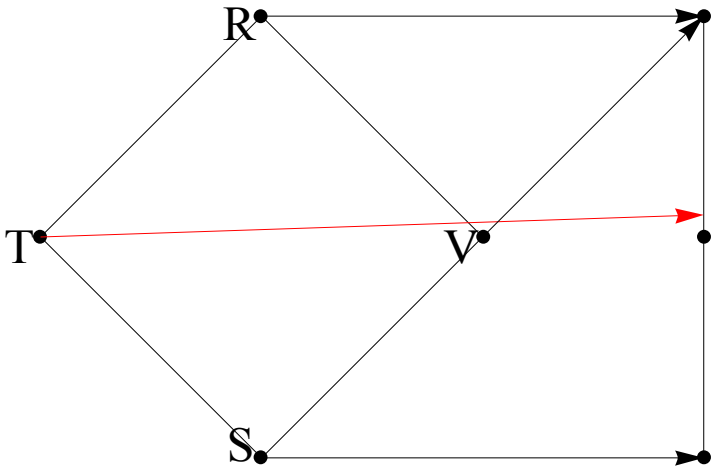


After applying the function F, T appears better ranked than V. Figure 6[25] shows that it is possible to build another correct aggregation function G stipulating a different ranking for the non-comparable cases and preserving order for the other rankings. This holds in the most general situations, at least for finite arrays of data[26]. Since we have no clear formal criterion to say that F is better than G, then the relative position of countries T and V is an artefact of our aggregation choice. In other words: a) the aggregation of comparable cases is trivial; and b) the aggregation of incomparable cases is arbitrary.

[25] *At least* one; in the example, there are more.
[26] The full demonstration is somewhat more involved, but the 'spirit' is identical. Gutiérrez, in preparation.

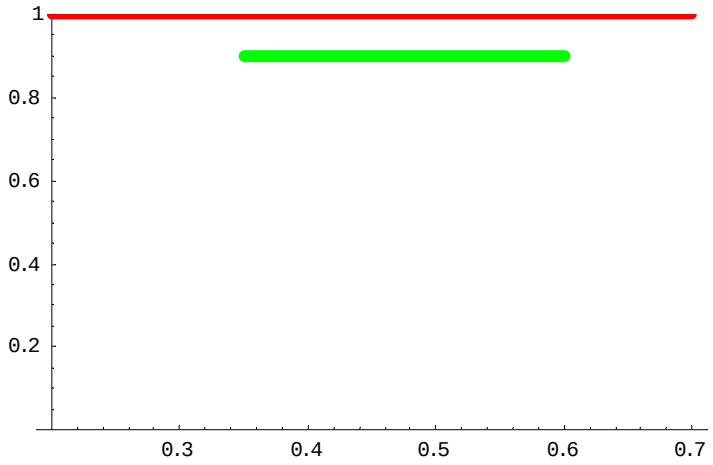
Figure 6 – The function G

It ranks V better than T. It is also a correct aggregation function



Let us ground the discussion on concrete examples. Suppose we want to know if Colombia is a stronger (less fragile, better performing) state than Bolivia. It is easy to understand why this comparison is so difficult, since it is a multi-attribute decision, without a numeraire or a homogeneous, abstract decision maker. You could put the quandary in Rawlsian terms: given a lottery, would you prefer to be born in one country or in another? Over a long period, Colombia has not had either a monopoly of legitimate violence or full control of its territory; on the other hand, it is a much better provider of basic services and has less ethnic fragmentation than Bolivia. In the FFP, Colombia is systematically rated higher than Bolivia, but the spread of the marks of the former is much broader than the spread of the latter. To visualise this, cases are not represented in Figure 7 as numbers, but as intervals, where the left and right ends are the minimum and maximum score respectively. This shows that there is no obvious way of comparing state strength (fragility) in Bolivia and Colombia in 2005, or more generally any pair belonging to set INC. Since there is a well formed aggregation function for every possible ranking of non-comparable cases, we are faced with a strong dilemma: either the overwhelming majority of the cases are comparable, in which case the ranking says nothing new; or there is a substantial portion of non-comparable cases, and therefore the ranking is arbitrary, predicating nothing about the relative position of the non-comparable cases, but rather about the rules of the game that have been established. Another, more positive way of wording the idea is that the relative positions of the paired comparisons pertaining to INC is wholly determined by the aggregation function. Note that this is not true for comparable cases (due to the restriction of monotonicity for correct functions). If this is the case, one would expect a mass of intellectual resources mobilised to support theoretically the specific aggregation function that is being used, and to explain why it is better than others, because if, for substantive reasons, we show that one aggregation function is preferable, then the arbitrariness of the exercise vanishes. This, however, has not happened. On the contrary: there is not a single discussion about the issue in the literature, nor in the codebooks – which Actually, some codebooks simply skip the theme. The deeply hidden assumption of the existence of a numeraire must have played a role in this lacuna.

Figure 7
The top line represents Colombia, the bottom Bolivia. This interval representation does not assume that there is a numeraire that makes dimensions interchangeable; the assumption, instead, is that the most relevant characteristics of each case are captured by its extreme values.



It may be argued that multidimensionality can be 'flattened out'. For example, factorial and principal component analyses, or multidimensional scaling, allow the researcher to see below an apparently baffling multi-dimensionality (Izenman 2008; Lebart Morineau and Warwick1984 Morineau and Piron 1995). Approaches such as structural equations permit the incorporation of multidimensionality directly into the model, assuming that a complex concept is a 'latent variable' regulated by other, explicit ones (Lebart 1995). The approach favoured by the Task Force is the following: reducing dimensionality through multi-variate statistics, and then using a neural network to make the final aggregation. It is easy to see that nothing of this can solve the problem of order and aggregation. For such a complex concept as state performance, the reduction of dimensionality is only a necessary, but not a sufficient, step in the process of aggregation; and after reduction, several dimensions will remain. On the other hand, the downside of neural networks – which in many respects are superior to the aggregation functions I have considered above – is that they are black boxes (King and Zheng 2000). In the context of the present discussion, this is a highly undesirable characteristic: for PSP indexes, the aggregation function *is* the theory.

To summarise this argument:

a. The existence of a numeraire that permits the reduction of the multiple dimensions of a PSP set of variables via a common accounting unit cannot be assumed;

b. It should be demonstrated, or discarded;

c. The theoretical construct of a homogeneous decision maker with a unique utility function, which in several ambits is an extremely powerful and correct simplification, in this context is uncalled for;

d. Not having at hand a numeraire nor a homogeneous decision maker, the core of any ordering – especially when multi-dimensionality is not trivial, and there is a high proportion of non-comparable cases – is the aggregation function, and this should be made explicit and supported theoretically in at least two senses: showing that it expresses a substantively defensible notion of order; and showing that it is not arbitrary (i.e. for the specific field it is better than other possible options);

e. Since this has not been done, PSP indexes predicate nothing about the relative position of incomparable cases.

Conclusions

PSP indexes succumb to several obvious errors. LICUS employs as a definitional criterion of state fragility 'bad policies'.[27] But who is to give the verdict?[28] These errors deserve to be flagged and corrected. More seriously, the treatment of poor and missing data, and of the severe endogeneity that plagues regressions based on PSP sources, can be improved.

But the real story starts when we focus on the Type B issues. They consist of the inability of current formalisms to take intrinsic ambiguity on board, and to unearth deeply hidden assumptions about the existence of a numeraire, which allow the establishment of substitution rates between variables, and the introduction of these substitution rates as if they were the utility function of a universal decision maker. The final result is necessarily contrived and full of anomalies. It is not idiosyncratic, or the product of poor solutions (though the discussion is muddled by their presence). Polity has taken the effort of clarity within the boundaries of current formalisms to its ultimate consequences, but all the lingering issues remain. There is a limit to disambiguation. Regarding order, the situation is even worse: purported rankings between cases that are not comparable in the domain are fully arbitrary, unless the black box of the aggregation function is opened, displayed and discussed. But the erroneous assumption of the existence of the numeraire, and thus the idea that total order can be taken for granted, has prevented this.

Note that I have not taken into the discussion 'second level' problems, derived from composition. For example: how to aggregate hedged variables? Or how much overall precision do you lose when you associate two variables that are poorly, or vaguely, defined?
One sort of reaction to all this would be to condemn the whole effort of formalisation altogether. However, technological conservatism does not offer such a brilliant future because of two reasons: first, it stops at the critical, destructive moment, without offering any alternative; and second, one of the characteristics of the contemporary world is the overabundance of unprocessed information. Both academics and policy makers are insatiable with respect to methods that allow them to search, order and structure information in a meaningful and cognitively economic way. It is inevitable that these tools exist, and one of the safest bets is that they will proliferate in the future.

Potent data mining, organising and clustering tools have been developed, in the context of technological developments to cope with huge datasets and information overflow, to deal with intrinsic ambiguities, linguistic labels and multiattribute-multiobjective decision making (Kahraman 2008; Shouhong and Archer 1994; Zanakis et al. 1998). Indeed, there is a trade-off here, because these tools incorporate higher levels of complexity, but they have the potential to produce tractable models that do not entail a complete deformation of the concept that supposedly is being operationalised. Better formalisms can and should be developed. The research programme should go in the following direction: incorporation of ambiguity and other forms of non-probabilistic uncertainty into the models; use of 'linguistic variables' (Zadeh 1975); overcoming crisp cut-offs and other instantiations of spurious precision;

[27] See also the extremely involved and subjective definitions of 'good government'. For example Grindle, M., 2007, 'Good Enough Governance Revisited', Development Policy Review, vol. 25 (5):554).
[28] As throughout this discussion, there are obvious situations in which anybody can tell if a policy is good or bad; but the interesting ones are essentially contestable, and may take time and historical perspective to answer.

careful discussion of the assumption of a numeraire; developing relational operationalisations that capture the historicity of certain key concepts; theoretically backing the aggregation function; acknowledgement of the problem of order, and creation of tools to deal with it; and management of objects different from numbers in a scale (for example, membership functions, intervals and fuzzy numbers). Movements in this direction should produce much improved models.

Naturally, no model will ever replace good qualitative and historical analysis (Elster 2007). What good models can do is to complement narratives, by allowing researchers to coordinate simultaneously many variables and dimensions that cannot be worked with only verbally. Models can fulfil this task, though, only if their assumptions are carefully discussed and if they are able to express reasonably well critical aspects of the concept they are supposed to operationalise.

Annex 1: Definitions of Poor Performance

Fund for Peace

Source: Fund for Peace 2009.

Concept	Definition	Examples of linguistic hedges
State failure	A state that is failing has several attributes. One of the most common is the loss of physical control of its territory or a monopoly on the legitimate use of force. Other attributes of state failure include the erosion of legitimate authority to make collective decisions, an inability to provide reasonable public services, and the inability to interact with other states as a full member of the international community. The twelve indicators cover a wide range of state-failure risk elements, such as extensive corruption and criminal behaviour, inability to collect taxes or otherwise draw on citizen support, large-scale involuntary dislocation of the population, sharp economic decline, group-based inequality, institutionalised persecution or discrimination, severe demographic pressures, brain drain and environmental decay. States can fail at varying rates through explosion, implosion, erosion or invasion over different time periods	Reasonable, full extensive, sharp

USAID

Source: USAID 2005.

Concept	Definition	Linguistic hedges
Fragile state	USAID uses the term fragile states to refer generally to a broad range of failing, failed and recovering states. However, the distinction among them is not always clear in practice, as fragile states rarely travel a predictable path of failure and recovery, and the labels may mask sub-state and regional conditions (insurgencies, factions etc.) that may be important factors in conflict and fragility. It is more important to understand how far and quickly a country is moving from or toward stability than it is to categorise a state as failed or not. Therefore, the strategy distinguishes between fragile states that are vulnerable from those that are already in crisis.	Generally, rarely, may be
Vulnerable	USAID uses 'vulnerable' to refer to those states unable or unwilling to adequately assure the provision of security and basic services to significant portions of their populations, and where the legitimacy of the government is in question. This includes states that are failing or recovering from crisis.	Adequately
Crisis	USAID uses 'crisis' to refer to those states where the central government does not exert effective control over its own territory or is unable or unwilling to assure the provision of vital services to significant parts of its territory, where legitimacy of the government is weak or nonexistent, and where violent conflict is a reality or a great risk.	

CIA Task Force
Source: Goldstone et al. 2000.

Concept	Definition	Linguistic hedges
State failure	Narrowly defined, state failures consist of instances in which central state authority collapses for several years. Fewer than 20 such episodes occurred globally between 1955 and 1998, however – too few for robust statistical analysis. Furthermore, events that fall beneath this total-collapse threshold often pose challenges to US foreign policy as well. For these reasons, the Task Force broadened its definition of state failure to include a wider range of civil conflicts, political crises and massive human-rights violations that are typically associated with state breakdown. For the purposes of this study, state failure was defined to include four categories of events.	

Carleton University Canada
Source: http://www.carleton.ca/cifp/app/serve.php/1138.pdf

Concept	Definition	Linguistic hedges
Fragile states	Fragile states lack the functional authority to provide basic security within their borders, the institutional capacity to provide basic social needs for their populations, and/or the political legitimacy to effectively represent their citizens at home and abroad.	Basic, effectively
Weak states	Weak states are susceptible to fragility or failure because of limited governance capacity, economic stagnation and/or an inability to ensure the security of their borders and sovereign domestic territory.	Limited
Failing states	Failing states exhibit key elements of fragility and are experiencing organised political violence. Peace processes are weak or non-existent.	Key
Failed states	Failed States are characterised by conflict, humanitarian crises and economic collapse. Government authority, legitimacy and capacity no longer extend throughout the state, but instead are limited either to specific regions or groups.	
Collapsed states	Collapsed states possess no meaningful central governments. These nations exist purely as geographical expressions, lacking any characteristics of state authority, legitimacy or capacity.	
Recovering states	Recovering states are states that exhibit key elements of fragility, but where substantial and at least partially successful 'nation-building' efforts are present.	Least partially

Annex 2 – Examples of aggregation functions

LICUS

$$f(x_t) = \frac{\frac{x_{1t} + x_{2t} + x_{3t}}{3} + \frac{x_{4t} + x_{5t} + x_{6t}}{3} + \frac{x_{7t} + x_{8t} + x_{9t} + x_{10\,t} + x_{11\,t}}{5} + \frac{x_{12\,t} + x_{13\,t} + x_{14\,t} + x_{15\,t} + x_{16\,t}}{5}}{4}$$

Fund For Peace

$$f(x_1, x_2, \ldots, x_{12}) = \sum_{i=1}^{12} x_1$$

After the sum has been done, then the following If-Then rules are used:

If $(\sum_{t=1}^{12} x_t) \geq 90$ then 'Alert'

If $60 \leq (\sum_{t=1}^{12} x) < 90$ then 'Danger'

If $30 \leq (\sum_{t=1}^{12} x_t) < 60$ then 'Moderate danger'

If $(\sum_{t=1}^{12} x_t) < 30$ then 'Sustainable state'

Table 6
Table – Order indicator h(x) for South Africa and Uganda according to Polity between 1991 and 2005

Year	South Africa	Uganda
1980	0,926829	0,853659
1981	0,926829	0,853659
1982	0,926829	0,875
1983	0,925	0,926829
1984	0,97561	0,926829
1985	0,95122	Missing
2000	0,926829	0,341463
2001	0,921053	0,368421
2002	0,923077	0,333333
2003	0,921053	0,342105
2004	0,921053	0,342105
2005	0,923077	0,333333

h(x): Define a top and a bottom, for which the number of up-sets are respectively 0 and 1. Include in the count of up-sets the cases that are equal or better in the precedence relation established in the text. Divide this by the total number of cases. Then substract this from 1, and you get h(x). It is easy to demonstrate that it is a correct aggregation function.

Table 7 – Non comparability

	Fund for Peace	LICUS
Number of pair wise comparisons	21316	5625
Members of C	8334	1861
Members of INC	12982	3764
% of INC	60.9	67

References

Andersen, L. 2006. *Security Sector Reform in Fragile States*. Copenhagen: Danish Institute for International Studies.

Ball, Patrick. 1996. *Who Did What to Whom? Planning and Implementing a Large Scale Human Rights Data Project*. American Association for the Advancement of Science,at http://shr.aaas.org/Ball/contents.html.

Bates, Robert H., Epstein, David L., Goldstone, Jack A., Gurr, Ted Robert, Harff, Barbara, Kahl, Colin H., Knight, Kristen, Levy, Marc A., Lustik, Michael, Marshall, Monty G., Parris, Thomas M., Ulfelder, Jay and Woodward, Mark R. 2003. *Political Instability Task Force Report: Phase IV Findings*. McLean, VA: Science Applications International Corporation.

Beliakov, Gleb, Pradera, Ana and Calvo, Tomasa. 2007. "Aggregation functions. A guide for practitioners" Springer, Heildelberg-New York

Bustince Humberto, Herrera Francisco, Montero Javier (Eds.) (2008): "Fuzzy sets and their extensions: representation, aggregation and models", Springer, Berlin-Heidelberg-New York

Cammack, Diana, McLeod, Dinah, Rocha Menocal, Alina and Christiansen, Karin. 2006 'Donors and the 'Fragile States' agenda: a survey of current thinking and practice', report submitted to the Japan International Cooperation Agenda', ODI-JICA.

Carment, D. 2003. 'Assessing state failure: implications for theory and policy', *Third World Quaterly* 24(3): 407-27.

Country Indicators for Foreign Policy (CIFP). 2006. *Failed and Fragile States, A concept Paper for the Canadian government*. Ottawa: CIFP, Carleton University.

Chataigner, J. M. and Gaulme, F. 2006. *Beyond the Fragile State: Taking Action to assist Fragile Actors and Societies*. Paris: Agence Française de Développement.

Chesterman, Simon, Ignatieff, Michael, Thakur, Ramesh. 2005. *Making states work. State failure and the crisis of governance*. New York: United Nations University Press.

Di John, Jonathan. 2007. 'Conceptualising the Causes and Consequences of Failed States: A Critical Review of the Literature', Background paper, Crisis States Research Centre.

Eizenstat, S., Porter, J. and Weinstein, J. 2004. *On the Brink: Weak States and US National Security*. Washington DC: Center for Global Development.

Elster, Jon. 2007. *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Ertman, Thomas. 1997. *Birth of the Leviathan. Building states and regimes in Medieval and Early Modern Europe*. Cambridge: Cambridge University Press.

Fund for Peace. 2009. *Failed States Index FAQ*. Washington DC: Fund for Peace, at http://www.fundforpeace.org/web/index.php?option=com_content&task=view&id=102&Itemid=327

Goldstone, Jack A., Gurr, Ted Robert, Harff, Barbara, Levy, Marc A., Marshall, Monty G., Bates, Robert H., Epstein, David L., Kahl, Colin H., Surko, Pamela T., Ulfelder, John C. and Unger Alan N. 2000. State Failure Task Force Report: Phase III Findings. McLean, VA: Science Applications International Corporation, at http://globalpolicy.gmu.edu/pitf/SFTF%20Phase%20III%20Report%20Final.pdf.

Gros, Jean Germain. 1996. 'Towards a taxonomy of failed states in the New World Order. Decaying Somalia, Liberia, Rwanda, and Haiti', *Third World Quarterly* 17(3): 455-71.

Gutiérrez, Francisco and González, Andrea. 2009. 'Force and ambiguity: Evaluating sources for cross-national research – the case of military intervention', *Crisis States Working Paper* No. 50, London: Crisis States Research Centre, London School of Economics.

Hale, H. E. 2004. 'Divided We Stand: Institutional Sources of Ethnofederal State Survival and Collapse', *World Politics* 56(2): 165-93.

Hermann, Margaret and Kegley, Charles. 2001. 'Democracies and Interventions: Is there a Danger Zone in the Democratic Peace?' *Journal of Peace Research* 38(2): 237-45.

Hirschman, Albert. 1977. *Salida, voz y lealtad*. Mexico DF: Fondo de Cultura Económica.

Hirschman, Albert. 1981. *Essays in trespassing: economics to politics and beyond*. Cambridge: Cambridge University Press.

Huntington, Samuel P. 1996. *The Clash of Civilizations and the Remaking of World Order*. New York: Simon and Schuster.

IEG: Independent Evalution Group (2006): "Engaging with fragile states. An IEG review of World Bank support to Low Income Countries Under Stress", World Bank, Washington

Ignatieff, M. 2002. 'Intervention and State Failure'. *Dissent Magazine*, at http://www.dissentmagazine.org/article/?article=641.

International Commission on Intervention and State Sovereignty (ICISS). 2001. *The responsibility to protect*. Ottawa: ICISS.

Izenman, Alan Julian (2008); *Modern Multivariate Statistical Techniques,* Springer

Kahraman, 2008: "Fuzzy multi-criteria decision making. Theory and applications with recent developments", Springer, Istanbul

King, Gary and Langche, Zheng. 2000. 'Improving quantitative studies of international conflict: a conjecture', *American Political Science Review* 94(1): 21-36.

Lebart, L., Morineau, A. & Warwick (1984), *Multivariate Descriptive Statistical Analysis,* Wiley, New York.

Lebart 1995. "Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices" John Wiley and Sons, NJ

Lootsma, 1997: Lootsma Freerk (1997): "Fuzzy logic for planning and decision making", Kluwer Academic Publishers, Dordrecht-Boston-London

Marshall, M. G. 2008. *Fragility, Instability, and the Failure of States*. New York: Council on Foreign Relations.

Marshall, Monty G. and Jaggers, Keith. 2009. Polity IV Project. Political Regime Characteristics and Transitions, 1800-2007. Dataset Users' Manual. Severn MD: Center for Systemic Peace, at http://www.systemicpeace.org/inscr/p4manual2007.pdf.

Milliken, J. and Krause, K. 2002. 'State Failure, State Collapse and State Reconstruction: Concepts Lessons and Strategies', *Development and Change* 33(5): 753-74.

Moore, Barrington. 1979. *Injustice: the social bases of obedience and revolt*. New York: Random House.

North, Douglass, Wallis, John and Weingast, Barry. 2009. *Violence and social orders. A conceptual framework for interpreting recorded human history*. Cambridge: Cambridge University Press.

Patrick, S. 2006. *Weak States and Global Threats: Assessing Evidence of 'Spillovers'*. Washington DC: Center for Global Development.

Pearson, Frederic S. and Baumann, Robert A. 1993. *International Military Intervention, 1946-1988*, Ann Arbor: Inter-University Consortium for Political and Social Research, at http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/06035..

Przeworski, Adam. 2004. *States and markets. A primer in political economy*. Cambridge: Cambridge University Press.

Putzel, James. 1997. 'Policy Arena: Accounting for the lsquodark sidersquo of social capital: reading Robert Putnam on democracy, *Journal of International Development* 9(7): 939-49.

Reno, William. 1998. *Warlord politics and African states*. London: Lynne Rienner.

Rice, Susan, y Patrick Stewar. *Index on state weekness in the developing world.* Washington: The brookings institution, 2009.

Rosenau, James. «Intervention as a scientific concept.» *Journal of conflict resolution* 2, nº 13 (1969): 149-171.

Sartori, Giovanni. 1970. 'Concept Misformation in Comparative Politics', *American Political Science Review* 64(4): 1033–53.

Shouhong Wang, Norman P. Archer, "A neural network technique in modeling multiple criteria multiple person decision making". *Computers & Operations Research,* Volume 21, Issue 2, February 1994, Pages 127-142

Skocpol, Theda. 1984. *Los Estados y las revoluciones sociales*. Mexico DF: Fondo de Cultura Económica.

Snyder, Jack. 2000. *From voting to violence: democratization and nationalist conflict*.New York: Norton.

Tilly, Charles. 1992. *Coerción, capital y los Estados europeos (990-1990)*. Madrid: Alianza.

USAID. 2005. *Fragile States Strategy*, at http://www.usaid.gov/policy/2005_fragile_states_strategy.pdf.

Wooldridge, Jeffrey. *Econometric analysis of cross section and panel data.* United States of America: Massachusetts Institute of Technology, 2002.

World Bank. 2006. *Engaging with fragile states. An IEG review of World Bank Support to Low Income Countries under Stress*. Washington DC: World Bank.

Zadeh L.A. (1975a) The Concept of a Linguistic Variable and its Application to Approximate Reasoning - II, *Information Sciences*, 8:4, 301-357

Zanakis, S., A. Solomon, N. Wishart, and S. Dublish (1998): Multi-attribute decision making: A simulation comparison of select methods. European Journal of Operational Research, 107(3):507-529

Zartman, W. I. 1995. *State Collapse: The Disintegration and Restoration of Legitimate Authority*. Boulder: Lynne Rienner.

Crisis States Working Papers Series 2

WP1 James Putzel, 'War, State Collapse and Reconstruction: phase 2 of the Crisis States Programme' (September 2005)
WP2 Simonetta Rossi and Antonio Giustozzi, 'Disarmament, Demobilisation and Reintegration of ex-combatants (DDR) in Afghanistan: constraints and limited capabilities', (June 2006)
WP3 Frederick Golooba-Mutebi, Gabi Hesselbein and James Putzel, 'Political and Economic Foundations of State making in Africa: understanding state reconstruction', (July 2006)
WP4 Antonio Giustozzi, 'Genesis of a Prince: the rise of Ismail Khan in western Afghanistan, 1979-1992' (September 2006)
WP5 Laurie Nathan, 'No Ownership, No Peace: the Darfur Peace Agreement', (September 2006)
WP6 Niamatullah Ibrahimi, 'The Failure of a Clerical Proto-State: Hazarajat, 1979-1984' (September 2006)
WP7 Antonio Giustozzi, "Tribes" and Warlords in Southern Afghanistan, 1980-2005' (September 2006)
WP8 Joe Hanlon, Sean Fox, 'Identifying Fraud in Democratic Elections: a case study of the 2004 Presidential election in Mozambique'
WP9 Jo Beall, 'Cities, Terrorism and Urban Wars of the 21st Century', (February 2007)
WP10 Dennis Rodgers, 'Slum Wars of the 21st Century: the new geography of conflict in Central America', (February 2007)
WP11 Antonio Giustozzi, 'The Missing Ingredient: non-ideological insurgency and state collapse in Western Afghanistan 1979-1992', (February 2007)
WP12 Suzette Heald, 'Making Law in Rural East Africa: SunguSungu in Kenya', (March 2007)
WP13 Anna Matveeva, 'The Regionalist Project in Central Asia: unwilling playmates', (March 2007)
WP14 Sarah Lister, 'Understanding State Building and Local Government in Afghanistan', (June 2007)
WP15 Pritha Venkatachalam, 'Municipal Finance Systems in Conflict Cities: case studies on Ahmedabad and Srinagar, India', (July 2007)
WP16 Jason Sumich, 'The Illegitimacy of Democracy? democratisation and alienation in Maputo, Mozambique', (September 2007)
WP17 Scott Bollens, 'Comparative Research on Contested Cities: lenses and scaffoldings', (October 2007)
WP18 Debby Potts, 'The State and the informal in sub-Saharan African economies: revisiting debates on dualism', (October 2007)
WP19 Francisco Gutiérrez Sanín, Tatiana Acevedo and Juan Manuel Viatela, 'Violent liberalism? State, conflict, and political regime in Colombia, 1930-2006: an analytical narrative on state-making', (November 2007)
WP20 Stephen Graham, 'RoboWar TM Dreams: Global South Urbanisation and the US Military's 'Revolution in Military Affairs'', (November 2007)
WP21 Gabi Hesselbein, 'The Rise and Decline of the Congolese State: an analytical narrative on state-making', (November 2007)
WP22 Diane Davis, 'Policing, Regime Change, and Democracy: Reflections from the Case of Mexico', (November 2007)
WP23 Jason Sumich, 'Strong Party, Weak State? Frelimo and State Survival Through the Mozambican Civil War: an analytical narrative on state-making', (December 2007)
WP24 Elliott Green, 'District Creation and Decentralisation in Uganda', (January 2008)
WP25 Jonathan DiJohn, ' Conceptualising the Causes and Consequences of Failed States: A Critical Review of the Literature', (January 2008)
WP26 James Putzel, Stefan Lindemann and Claire Schouten, 'Drivers of Change in the Democratic Republic of Congo: The Rise and Decline of the State and Challenges For Reconstruction - A Literature Review', (January 2008)
WP27 Frederick Golooba Mutebi, 'Collapse, war and reconstruction in Uganda: An analytical narrative on state-making', (January 2008)
WP28 Frederick Golooba Mutebi, 'Collapse, war and reconstruction in Rwanda: An analytical narrative on state-making', (February 2008)
WP29 Bjørn Møller, 'European Security: the role of the European Union', (February 2008)
WP30 Bjørn Møller, 'European Security: The Role of the Organisation for Security and Co-operation in Europe', (February 2008)
WP31 Laurie Nathan, 'Anti-imperialism Trumps Human Rights: South Africa's Approach to the Darfur Conflict', (February 2008)
WP32 Ben Moxham, 'State-Making and the Post-Conflict City: Integration in Dili, Disintegration in Timor-Leste', (February 2008)
WP33 Kripa Sridharan, 'Regional Organisations and Conflict Management: comparing ASEAN and SAARC', (March 2008)

WP34 Monica Herz, 'Does the Organisation of American States Matter?' (April 2008)
WP35 Deborah Fahy Bryceson, 'Creole and Tribal Designs: Dar es Salaam and Kampala as Ethnic Cities in Coalescing Nation States
WP36 Adam Branch, 'Gulu Town in War and Peace: displacement, humanitarianism and post-war crisis' (April 2008)
WP37 Dennis Rodgers, 'An Illness called Managua' (May 2008)
WP38 Rob Jenkins, 'The UN peacebuilding commission and the dissemination of international norms' (June 2008)
WP39 Antonio Giustozzi and Anna Matveeva, 'The SCO: a regional organisation in the making' (September 2008)
WP40 Antonio Giustozzi, 'Afghanistan: transition without end' (November 2008)
WP41 Niamatullah Ibrahimi, 'At the Sources of Factionalism and Civil War in Hazarajat' (January 2009)
WP42 Niamatullah Ibrahimi, 'Divide and Rule: state penetration in Hazarajat, from monarchy to the Taliban' (January 2009)
WP43 Daniel Esser, 'Who Governs Kabul? Explaining urban politics in a post-war capital city' (February 2009)
WP44 Francisco Gutierrez et al, 'Politics and Security in Three Colombian Cities' (March 2009)
WP45 Marco Pinfari, 'Nothing but Failure? The Arab League and the Gulf Cooperation Council as Mediators in Middle Eastern Conflicts' (March 2009)
WP46 Anna Matveeva, 'The Perils of Emerging Statehood: civil war and state reconstruction in Tajikistan' (March 2009)
WP47 Jennifer Giroux, David Lanz and Damiano Sguaitamatti, 'The Tormented Triangle: the regionalisation of conflict in Sudan, Chad and the Central African Republic' (April 2009)
WP48 Francisco Gutierrez-Sanin, 'Stupid and Expensive? A critique of the costs-of-violence literature' (May 2009)
WP49 Herbert Wulf and Tobias Debiel, 'Conflict Early Warming and Response Mechanisms: tools for enhancing the effectiveness of regional organsations? A comparative study of the AU, ECOWAS, IGAD, ASEAN/ARG and PIF' (May 2009)
WP50 Francisco Gutierrez Sanin and Andrea Gonzalez Pena, 'Force and Ambiguity: evaluating sources for cross-national research- the case of military interventions' (June 2009)
WP51 Niamatullah Ibrahimi, 'The Dissipation of Political Capital amongst Afghanistan's Hazaras: 2001-2009' (June 2009)
WP52 Juergen Haacke and Paul D. Williams, 'Regional Arrangements and Security Challenges: a comparative analysis' (July 2009)
WP53 Pascal Kapagama and Rachel Waterhouse, 'Portrait of Kinshasa: a city on (the) edge', (July 2009)
WP54 William Freund, 'The Congolese Elite and the Fragmented City', (July 2009)
WP55 Jo Beall and Mduduzi Ngonyama, 'Indigenous Institutions, Traditional Leaders and Elite Coalitions for Development: the case of Greater Durban, South Africa' (July 2009)
WP56 Bjorn Moller, 'Africa's Sub-Regional Organisations: seamless web or patchwork?' (August 2009)
WP57 Bjorn Moller, 'The African Union as Security Actor: African solutions to African problems?' (August 2009)

These can be downloaded from the Crisis States website (www.crisisstates.com), where an up-to-date list of all our publications including Discussion Papers, Occasional Papers and Series 1 Working Papers can be found.

LSE

Crisis States Research Centre

The Crisis States Research Centre aims to examine and provide an understanding of processes of war, state collapse and reconstruction in fragile states and to assess the long-term impact of international interventions in these processes. Through rigorous comparative analysis of a carefully selected set of states and of cities, and sustained analysis of global and regional axes of conflict, we aim to understand why some fragile states collapse while others do not, and the ways in which war affects future possibilities of state building. The lessons learned from past experiences of state reconstruction will be distilled to inform current policy thinking and planning.

Crisis States Partners

Ardhi University
Dar es Salaam, Tanzania

Collective for Social Science Research
Karachi, Pakistan

Developing Countries Research Centre (DCRC)
University of Delhi
Delhi, India

Department of Environmental and Geographical Sciences
University of Cape Town
Cape Town, South Africa

Instituto de Estudios Políticos y Relaciones Internacionales (IEPRI)
Universidad Nacional de Colombia
Bogotá, Colombia

Makerere Institute of Social Research
Makerere University
Kampala, Uganda

Research Components

Development as State-Making

Cities and Fragile States

Regional and Global Axes of Conflict

Development Studies Institute (DESTIN)
LSE, Houghton Street, London WC2A 2AE
Tel: +44 (0)20 7849 4631
Fax: +44 (0)20 7955 6844
Email: csp@lse.ac.uk
Web: www.crisisstates.com