Letter from the President

Space and Comparative Politics

Sidney Tarrow
Cornell University
sgt2@cornell.edu

A few years ago, a group of colleagues and I put together a collective book called *Silence and Voice in the Study of Contentious Politics* (Aminzade et al. 2001). One of our number, Bill Sewell, found an empty silence in the study of contentious politics – the concept of space itself (Sewell 2001). Oddly, the same lacuna exists in comparative politics more generally.

Of course, our subdiscipline deals implicitly with space but for the most part, we have specified the spatial dimensions of politics as what Sewell called *abstract space* – “continuous, unbounded, unlimited extension in every direction, regarded as void of matter or without reference to this” (p. 53). But space also means *concrete space* – “a definite location of a particular size and shape.” Used in this way, Sewell continues, “space is defined not by objective quantifiable characteristics,” but is “defined in relation to human occupation, use, or gaze” (ibid.).

Here is an example: Jim Fearon and David Laitin map countries in which civil wars are likely to break out; such countries are more likely to be mountainous than more pacific places (Fearon and Laitin 2003:97). That is what I mean by “abstract space.” But when we ask why mountains should be correlated with civil unrest, we have to turn to “concrete space.” It is concrete mechanisms – like the isolation of mountain residents from the centers of power, their knowledge of the terrain, and the interpersonal trust that grows up in small communities – that enhance “the likelihood that mountain dwellers will be able to engage successfully in illegal or subversive activities like smuggling or guerrilla warfare” (Sewell, p. 55; also see Tilly 2000, 2003).

Consider the long survival of the Waldensian sect in the French Alps above the town of Embrun with which Charles Tilly begins his *Trust and Rule* (2005: pp. 1ff.). Their occupation of “abstract space” high above the valley predicted the likelihood that they could escape the intrusion of the monarchy. And they did. But only by specifying the properties of the “concrete space” they inhabited can we understand how they survived for as long as they did.

Adjacent disciplines have done a better job than we have in integrating spatial variables within their models and methods. Geographers like John Agnew (1987), Sallee Marston (2000), Doreen Massey (1994), and Byron Miller (1994) have specified ways in which space structures politics. Sociologists like Herbert Gans (2002), Tom Gieryn (2000), Edward (Continued on page 2)
Section Officers

President
Sidney Tarrow
Cornell University  
sgt2@cornell.edu

Vice President, President-Elect
Peter A. Gourevitch
University of California, San Diego  
pgourevitch@ucsd.edu

Secretary-Treasurer
Karen Remmer
Duke University  
remmer@duke.edu

2006 Program Chair
Deborah J. Yashar
Princeton University  
dyashar@princeton.edu

Executive Council

Elizabeth Perry
Harvard University  
eperry@gov.harvard.edu

Mathew Shugart
University of California, San Diego  
mshugart@ucsd.edu

Susan Stokes
Yale University  
susan.stokes@yale.edu

Anna Grzymala-Busse
University of Michigan  
abusse@umich.edu

(Continued from page 1)

Soja (1989), and Dingxin Zhao (1999; 2001) have invested space and place with concrete meanings. Anthropologists and historians with an anthropological vision have taught us that space must be constructed into “place” (Boone 2002; Edelman 2001; Rugasa and Torre 2003; Wood 2001).

“Soja (1989), and Dingxin Zhao (1999; 2001) have invested space and place with concrete meanings. Anthropologists and historians with an anthropological vision have taught us that space must be constructed into “place” (Boone 2002; Edelman 2001; Rugasa and Torre 2003; Wood 2001).

“Adjacent disciplines have done a better job than we have in integrating spatial variables within their models and methods.”

When we look inside “concrete space,” we can identify numerous processes in which proximity matters that can enrich our subdiscipline’s contributions. From among many possible ones, I choose to illustrate three: space and co-presence, spatial diffusion, and scale shift.

Spatial Co-Presence

As a Berkeley graduate student in the 1960s, I went off to Italy trying to understand how the different structural conditions of North and South impacted the organizational, ideological, and the electoral behavior of the Italian Communist Party (PCI) and its relations with the southern peasantry. Serendipitously, I discovered that in Southern Italy, both the big cities and the villages voted for the Right while communes of 30,000 to 100,000 voted for the Left. Why? The properties of concrete space inspired a hypothesis: the medium-sized cities were large agrarian borghi with little infrastructure, in which small artisanal workers existed alongside peasant workers (Schneider and Schneider 1976). It was not the size of these towns but the relations between newly-industrialized workers and peasants that produced the higher-than-average vote for the Left. Like Fearon’s and Laitin’s mountains, size was the abstract spatial category that could only be understood in terms of the relations within concrete space.

Concrete space also helped Dingxin Zhao understand the dynamics the Tienanmen Square rebellion. As is well known, Chinese communist leaders had established 67 universities in the Haidan district, isolated from the surrounding neighborhoods (Zhao 1998: 1502; I quote Sewell’s summary from p. 72 of his article). A good part of the explosion of contention in 1989 can be explained by traditional social network theory (McAdam 1988; Osa 2003; Gould 1995). But it was the co-presence of students from many universities in public space that produced the massing of student groups marching on Tienanmen Square (Zhao 1998: 1508-09). For that, Zhao turns to the brokerage role of the “liaison men” who linked students from different campuses to one another (Zhao 1998: 1514-15; Sewell: 74). It was these relations within “concrete space” that made “abstract space” work to produce a near-revolution.

Spatial Diffusion

A second process — diffusion — is an important spatial factor in contentious politics. By diffusion, I mean the spread of a form of contention, an issue, or a way of framing it from one site to another. The diffusion of social movements, revolutions, and other forms of contentious politics is similar in many ways to the diffusion of innovations (Soule and Zylan 1997; Strang and Soule 1998). For example, in their examination of the Swing rebellion in southern England, Hobbsbawm and Rudé (1975) found...
that the incidents of Swing violence followed the main lines of road communication between London and the coast. Rudé’s work on rural riots in 18th century France similarly showed that contentious episodes followed the river valleys (Rudé 1981:24).

“The diffusion of social movements, revolutions, and other forms of contentious politics is similar in many ways to the diffusion of innovations.”

Brokerage and diffusion combine in many episodes of contention, producing new coordination. Brokerage reinforces diffusion, allowing individuals and groups that have had no previous contact to coordinate their actions through third parties with contacts with both. When Slobodan Milosevic tried to steal local elections in Serbia in 1996, Otpor militants used a variety of performances to oppose the regime (Tarrow 2005: ch. 6). Between 2000 and 2004, the Otpor model diffused from Serbia to Georgia. Three weeks of peaceful street protests culminated in a “March of the Angry Voters,” led by Mikhail Saakashvili, leader of the opposition coalition and the country’s president after Shevardnadze was forced to step down.

Where had the opposition learned such tactics? In the months before the election, Saakashvili had traveled to Serbia to contact the former organizers of the anti-Milosevic movement. He returned with a plan for nonviolent action modeled closely on the success of Otpor. Ex-Otpor activists traveled to Georgia, where they led training sessions for Georgian reformers. The Georgians, in turn, trained a cadre of grassroots activists. Diffusion and brokerage combined to produce the spread of anti-election fraud campaigns in former state socialist regimes in Eastern Europe.

Scale Shift

The Internet’s importance is critical to a special form of diffusion – scale shift (Tarrow and McAdam 2005). Scale shift is an essential element of all contentious politics, without which contention that arises locally would remain at that level. We can define it as a change in the number and level of coordinated contentious actions to a different focal point, involving a new range of actors, different objects, and broadened claims (McAdam, Tarrow, and Tilly 2001: 331). It can also generate a change in the meaning and scope of the object of a claim as a function of its transfer.

Students of American politics are familiar with the process of scale shift in the areas of welfare and regulatory policy. Until the 1930s, most policy innovations began at the state level; only after long experimentation at that level did they percolate up to the federal realm. The same is true of contentious politics. For example, the 19th century temperance movement shifted back and forth from the state and local to the federal level several times in response to opening and closing political opportunities, before settling on the federal strategy of a constitutional amendment (Szymanski 2003: ch.3).

In his article, Sewell shows how a new national scale for politics was established in France during that country’s great revolution. This development began soon after the taking of the Bastille, when local uprisings put new “patriot” administrations in power all over the country (Sewell, 2003: 84-5). But the nationalization of politics was not a natural or an automatic phenomenon; it was the Jacobin club, with its network of provincial affiliates, that nationalized that revolution. When they seized power in 1792, the Jacobins’ provincial societies became instruments of the Paris-organized emergency government. As Sewell concludes, “the country-wide network of concerted opinion and civic activism constituted by the Jacobin societies was probably as central to the creation of a national scale of politics as was the organization of national elections” (p. 85).

Scale shift can operate in two directions: upward – in which case local action spreads outward from its origins; or downward, when coordination at a higher level generates coordination among lower-level units. Today’s international system offers a special opportunity for activists because it opens conduits for upward shift of contention and brings new “global” practices back to the local level (Tarrow 2005). The “world” social forums in Brazil and India were the result of upward scale shift from European counter-summits in the 1990s; many who attended these summits returned to their own countries to establish versions of the social forum that soon became engaged with local alignments (Tarrow 2005: ch. 7).

Some observers see in the global justice protest events of the last decade signs of the formation of a “global civil society” (Clark 2003; Fiorini 2003). There is much anecdotal evidence to support their view: the diffusion of an international support network for the Zapatista rebels in Mexico (Olesen 2005); the formation of groups like People’s Social Action following that episode (Wood 2005); and the drama of contested international summits like the WTO summit in Seattle, the G-8 in Genoa, and the more recent
Cancun and Hong Kong WTO summits (Andretta and Mosca 2004). Do these campaigns demolish national space – at least as far as contentious politics is concerned?

The death of the nation-state has been announced before, but its passing has always been found to be premature. Not even the activists who engage in “global” social movements escape the spatial constraints of their national regimes, if only because they need to appeal to constituencies whose vistas are far less global when they return home.

Consider the massive number of European activists who joined local “social forums” in Europe after taking part in the Porto Alegre and Mumbai World Social Forums. When Donatella della Porta and her collaborators interviewed many of them, they identified strongly with the struggle against globalization, but the majority were deeply embedded in domestic forms of activism (della Porta, et al. 2006). Rather than demolishing space constraints, they were taking the lessons of transnational activism back to the contested spaces of domestic politics.

**Conclusions**

In this article, I have made some scattered observations about space and the need to enter into concrete space to understand the processes and mechanisms in which proximity and distance matters. Had I had more space (and more time!), I could have raised a number of research questions, answers to which would need to be built on a better operationalized notion of space. About each of the dimensions I have sketched, I might have asked:

- **About spatial co-presence:** what was it about the co-presence of workers and peasants in the small cities of Southern Italy that produced disproportionately high votes for the Communist party? And were the dynamics of interaction there anything like other class interactions in our field – for example, the interaction among workers and intellectuals that Robert Fishman has found in Spain (2004)?

- **About spatial and non-spatial diffusion:** are the patterns of participation in the Tienanmen Square rebellion typical of local forms of diffusion, while diffusion to higher levels depends more on impersonal channels like the Internet? Or do the three forms of diffusion – through interpersonal networks, third-party brokerage, and impersonal media channels – complement one another, as they seemed to do in the formation of the Zapatista solidarity network (Olesen 2005)?

- **About scale shift:** are globalization and internationalization producing increasingly transnational patterns of contentious politics? Or do the inducements and constraints of domestic politics continue to structure contention?

As yet, I see little systematic evidence that domestic social movements are “spacing out” into the uncharted terrain of global social movements. But even if they were, I would bet that the spatial boundaries of national politics constrain their activities – much as the boundary of the Mason-Dixon line constrained civil rights activism in the US in the 1960s.

And that is because boundaries – and the spaces they enclose – are not objective factors that can be erased by technology, travel, or mass communication. Boundaries are the relational markers of identities, and identities are the spaces within which activists and other people mobilize. But that is another story.

**Notes**

1 The contentious politics project was sponsored by the Mellon Foundation and sited at the Center for Advanced Study in the Behavioral Sciences, and owed much to the support of Phil Converse and Bob Scott at the Center and to Harriet Zuckerman at the Foundation. Members of the group to whom I owe thanks for the ideas in this letter are Ron Aminzade, Jack Goldstone, Doug McAdam, Liz Perry, Chuck Tilly and especially Bill Sewell, on whose work 2001 article this letter is heavily dependent. Other products of the CASBS/Mellon contentious politics gang will be found in Goldstone, et al. 2003, in McAdam, Tarrow and Tilly 2001, and in the PhD dissertations of the fourteen graduate fellows associated with it.

2 Social control of contention also diffuses across space. For examples, see della Porta and Reiter 1998. For an account of how the two streams of diffusion can intersect, see McAdam 1983.

3 We prefer this more neutral term to Sewell’s “scale-jumping,” which implies that scale shift is only an upward movement. Downward scale shift seems to us an important part of contentious politics as well.

4 As often happens to me, it was Chuck Tilly who found a more felicitous definition for a concept I was using loosely — “downward scale shift.”
Innovative Methods in Comparative Politics

Introduction

After more than a decade of intense debate about methodological principles, some new ideas about how to practice comparative politics are beginning to emerge. This symposium provides short descriptions of several of the most recent innovative methods for evaluating theoretical propositions in comparative politics. Methodological innovation is not a completely new phenomenon in our discipline, especially in statistical methods, where it is a constant. However, it seems to us that the pace of innovation has accelerated recently. Whereas over the past few decades comparativists could do a case study, write a comparative history, or run a regression, we now have alternative options such as designing computational simulations, embedding experiments in surveys, or subjecting fuzzy sets to Boolean logic. Although few scholars currently employ these methods, all have the potential to transform the way we study politics. As a result, we thought it would be helpful to introduce you, our readers, to some of them.

Toward this end this symposium brings together concise descriptions of three non-traditional methods by scholars who have pioneered their use. Each of these brief pieces describes the method and the problem it addresses, defends its superiority over more traditional methods, specifies the phenomena that it can and cannot address, offers an example from published research, and advises us on how we can become proficient in using it.

Kenneth Benoit and Michael Laver lead off with an introduction to automated content analysis using a software package called Wordscores, which they developed in collaboration with Will Lowe. Wordscores, they claim, enables researchers to mine texts for useful information without the taxing burden of actually reading them. Although their claim is a bold one, the authors offer examples showing that this method has reliably located parliamentary speeches, party manifestos, and amicus briefs on relevant ideological dimensions. In a field that relies as heavily on the content of communications as ours, the potential of such an approach seems boundless ... so long as the authors' claims withstand careful scrutiny. Benoit and Laver helpfully provide references for further reading for those who wish to give automated content analysis more careful consideration.


Few methods are more traditional in comparative politics than writing narratives that reconstruct causal processes. In recent years, however, many scholars have suggested ways to improve upon this method, usually by putting greater emphasis on theory. In our second contribution, Tulia Falleti surveys the family of methods that fall under the umbrella of “theory-guided process-tracing.” They include methods proposed by Alexander George and Timothy McKeown, Alexander George and Andrew Bennett, Ronald Aminzade, Tim Büthe, Peter Hall, Robert Bates, James Mahoney, David and Ruth Collier, and others, which have been labeled “process-tracing,” “analytic narratives,” “historical narratives,” and “systematic process analysis,” among other terms. In her contribution Falleti defines and contrasts the most prominent examples of these related methods, clarifying their common elements, how they differ, and what they do uniquely well.

The final contribution to the symposium is that of Simon Hug, who explains selected innovations in (or imports into) quantitative comparative analysis. Comparativists are increasingly aware that many of our common explanatory variables are endogenous, i.e., caused by other explanatory variables. This is a problem when the unspecified determinants of the explanatory variable are correlated with the unspecified determinants of the outcome. Hug shows that another oft-cited problem, selection bias, can be understood as a special case of endogeneity: it is a situation in which our ability to observe a variable, not merely its value, is a function of some other variables. Hug argues that these problems are just as relevant for qualitative analysis as for quantitative methods, but that quantitative methods provide various ways of solving them. For the benefit of readers who wish to learn to these methods, he cites a variety of sources that explain more fully how to address endogeneity and selection bias.
The Richness of Text as Political Data

Of the many sources of information we can collect to describe human activity, the words that humans write and speak offer perhaps the most immense, yet also most underused, data resource. In the course of almost any form of political activity, political actors tend to generate text in the form of spoken or written activity, and this text typically contains valuable information about the political orientations and motivations of the actors. While not every type of political activity gives rise to easily observable texts, it is certainly the case that political texts such as campaign speeches, party manifestos, legislative deliberations, judicial decisions and submissions, executive addresses, government statements, and treaty negotiations are increasing in number and availability. The problem with political text as a data source, indeed, stems not from its lack of easy availability, but rather from precisely the opposite problem: there is so very much text that it becomes almost impossible to analyze systematically. To separate and analyze the handful of needles from the truckloads of hay, we need efficient and proven methodologies for sorting through huge volumes of text.

Textual analysis offers numerous advantages as a source of data in political analysis, provided proper methodologies can be developed to analyze it. Once texts have been written or speeches recorded, they become part of the historical record. Quite unlike techniques such as survey research or other forms of interviewing, for example, the analysis of texts does not require that a measurement instrument be constructed in advance with some particular purpose in mind, and does not by the very act of measurement disturb the environment under investigation.

“The words that humans write and speak offer perhaps the most immense, yet also most underused, data resource.”

The systematic analysis of text can be deeply retrospective and perfectly non-intrusive, while the same corpus of text can be analyzed for many different reasons that were quite unanticipated at the time the text was generated. These are prodigious advantages, especially in comparative politics where other methods of gathering information about political actors – such as interviews, surveys, or directly observing non-textual behavior – are costly, difficult, or even dangerous. While it would be silly to claim that text analysis is a panacea for all research problems, there should be no doubt that it has immense potential as a source of information about political actors.

Text Analysis as Statistical Analysis

The analysis of the text written or spoken by humans has traditionally taken one particular form: being “read” for its “meaning,” in the same way that an APSA-CP reader would read an article about Wordscores in a symposium. This form of essentially qualitative reading allows humans to communicate with each other using words that users believe to have some mutually understood meaning. Various attempts by researchers to be more systematic about this have evolved into the measurement technique of “content analysis,” the vast bulk of which still involves the search for systematic and replicable methods of reading texts for meaning. Content analysis may be performed by human readers, who code either whole texts or units of text using a predefined coding scheme; or, conversely, it may be performed with the assistance of computers that use “dictionaries,” designed by human analysts with the particular purpose of assigning individual words, or groups of words, into predefined and meaningful coding categories. Text coding schemes that aim to uncover the meaning of texts, however, are typically difficult to implement, invariably rely upon subjective judgment calls by researchers, and often tend to be very costly in terms of both time and intellectual effort. And, because nothing comes close to a functioning human brain when it comes to interpreting meaning, automated text coding schemes that attempt to determine the meaning of texts are inherently unreliable.

The Wordscores method, by contrast, takes a fundamentally different approach to extracting information from a corpus of text. Rather than being concerned with the meaning of
text, we view a text as containing behavioral information about its author that can be analyzed statistically, based on systematic patterns of word usage. More specifically, the relative frequencies of the use of particular words by political actors provide observable manifestations of underlying political orientations, beliefs, and attitudes. The words contained in texts, therefore, may be analyzed statistically, without requiring the analyst to interpret, synthesize, or even read the texts being examined, in much the same way that a researcher concerned with public opinion would not attempt to read each respondent’s coded survey responses prior to analyzing a survey dataset. Indeed, because Wordscores works in any language by analyzing patterns of word usage, it is not necessary that the analyst be capable of reading the texts being analyzed.

Wordscores Explained

Wordscores is a form of uni-dimensional scaling technique that allows researchers to extract single-dimensional information from texts about which nothing is known a priori. For example, it can extract information about left-right positions from election manifestos, about attitudes towards a new “constitution” for Europe from treaty negotiations, or attitudes towards the government expressed in impeachment or no-confidence debates. No method — short of divine revelation — can transform total ignorance into valid new knowledge, however, and Wordscores certainly makes no such claim. With Wordscores, what happens is that the researcher starts with a policy dimension of substantive interest and a set of “reference” texts for which the investigator has valid independent information, or is confident of being able to make assumptions, about their positions on the dimension under investigation. By analyzing the reference texts in terms of relative word frequencies both within and across texts, Wordscores gives each word in the set of reference texts a score, representing the most likely policy position of a user of the word in question. The technique is now “primed” for an analysis of “virgin” texts, about which the analyst has no a priori information, but would like to locate on the dimension under investigation. This list of scored words, derived from analyzing the reference texts, is used to produce estimated positions for each of the virgin texts based on their patterns of word usage. The core of the method is very simple and can be performed easily using only a spreadsheet. (For details, see Laver, Benoit, and Garry 2003).

Rather than being concerned with the meaning of text, we view a text as containing behavioral information about its author that can be analyzed statistically, based on systematic patterns of word usage.”

As a fundamentally statistical method, Wordscores has three great advantages over more traditional methods of text analysis. It requires no substantive judgment call on the part of the investigator during the process of data analysis and is thus perfectly replicable; it operates in any language, since it does not involve the creation or maintenance of a coding dictionary; and because it is a statistical procedure, it generates an estimate of the uncertainty associated with any estimated policy position. The last feature in particular distinguishes Wordscores from nearly every previous form of quantitative content analysis, redressing a critical omission with preexisting quantitative measures.

“The words contained in texts [...] may be analyzed statistically, without requiring the analyst to interpret, synthesize, or even read the texts being examined...”

Because Wordscores maps a priori knowledge of reference texts and scores onto the virgin texts under investigation, the analyst’s role in identifying and scoring appropriate reference texts is crucial. If the reference texts are inappropriate for the dimension under investigation, or if their estimated or assumed policy positions are misleading, then this will produce misleading estimates of the positions of the virgin texts. The key, therefore, is to build on a solid foundation of well-chosen reference texts, with solid estimates or assumptions about their policy positions.

Applications to Comparative Research

Wordscores assumes that texts contain information about the positions of their authors on separable dimensions on which they may be individually scaled (although nothing constrains a text to contain information on only a single dimension). The prototypical case, and the one on which Laver, Benoit, and Garry (2003) successfully demonstrated the use of Wordscores to extract party position-
Wordscores has also been shown to work well when two opposing extreme positions are known, and the object is then to place a number of unknown texts along a “pro- versus anti-” dimension. Laver and Benoit (2002), for example, successfully demonstrated the ability of Wordscores to place 58 different speakers in a 1991 Irish confidence debate, using the speeches of two opposition party leaders (with a reference scores of -1.0 each) and the incumbent Irish prime minister’s speech (reference score 1.0). A similar research design was used by Benoit et al. (2004) to place 2,493 written submissions from participants in the Convention on the Future of Europe along a dimension of being for or against the proposed “constitution” of Europe. In another example, speeches made during the Clinton impeachment trial were used to show the different positions of 73 senators, taking the speeches of Jesse Helms (+1.0) and Edward Kennedy (-1.0) as reference texts, correctly predicting two thirds of the impeachment votes on the perjury charge (Bertelli and Grose 2006). In a first application to judicial decisions, McGuire and Vanberg (2005) have applied the method to extract ideological positions from U.S. Supreme Court decisions.

The method also has been reliably demonstrated to work in non-English languages, correctly classifying German parties based on their manifestos (Laver, Benoit, and Garry 2003), French presidential speeches (Laver, Benoit, and Sauger forthcoming), and Italian ministerial positions (Giannetti and Laver 2005a).

Unpublished trials using European Union political documents have also demonstrated Wordscores to work in a statistically indistinguishable identical manner on both French and English versions of the same texts.

Thus far Wordscores has mainly been used to estimate the positions on some known dimension of political actors whose texts, but not positions, are observable directly. Wordscores is thus a useful tool for extracting positional information from the texts as estimates of the unobservable positions of their authors. We see a huge potential for Wordscores in political analysis, however, in two fields that remain as yet largely untapped.

The first of these fields lies in the area of document classification, and the closely associated classification of actors who generate the documents. Bertelli and Grose (2006), for instance, use Senatorial speeches to classify votes based on legal-constitutional grounds, contrasting this to a spatial-ideological prediction from other data, concluding that the impeachment vote was predominantly explained by spatial positions.

Evans et al. (2005) show that, of 104 amicus curiae briefs submitted to affirmative action cases before the Supreme Court, Wordscores accurately classified 89% of the amici for petitioners and 75% for respondents, using no additional knowledge of any kind besides reference text based on four principal briefs in one of the cases. As a whimsical yet also successful example of document classification and as part of the third and final round of revision of the original Wordscores paper for the American Political Science Review, we used the technique to show the movement towards acceptance of the three referees between their first and second reports (a brief report on this is available from the Wordscores web site, below).

The second tantalizing application for automated content analysis tools such as Wordscores relates to intra-party politics. Spatial models of party competition that go beyond the unitary actor assumption must model decision-making within parties when they select actions that serve as input to inter-party competition. Empirical implementations will involve plotting maps of, for example, the positions of all party legislators. Particularly in parliamentary democracies, where highly disciplined parties mean that legislative roll calls are uninformative about the policy positions of individual legislators, Wordscores analysis of texts generated by members of parliament offers the prospect of significant progress. We have already noted a version of this in the work of Bertelli and Grose (1996). A recent paper by Giannetti and Laver (2005b) used a Wordscores analysis of delegate speeches at party congresses to map the internal factional structure of the Italian DS party and to predict party
splits on key foreign policy roll calls. The next empirical phase of our own work will involve mapping the factional structure of each of the main British parties, using every word spoken in the House of Commons between 1992 and 2001.

Although Wordscores was originally designed for the analysis of political texts, it has the potential to work well in other fields where text analysis is performed. This includes disciplines related to political science, such as history, law, sociology, and anthropology, but also more applied fields such as media studies, business and marketing, and even comparative literature. Indeed, much emerging work using Wordscores since the publication of our 2003 original paper has taken place in fields not directly related to political analysis. In any field where researchers need to extract information from the texts generated by the actors they study, but who previously have lacked a feasible means to retrieve such information, Wordscores may be a useful research tool.

How to Use Wordscores

Since the publication of Laver, Benoit, and Garry (2003), a suite of programs known (unsurprisingly) as Wordscores is available as a library for the Stata statistical package, and is available from www.wordscores.com

That website also contains a beta version of a standalone Java implementation of the software. Over the next year and a half we plan to make many improvements in both the methodology as well as the software available for implementing it, including a book-length project on Wordscores and related methods of computerized content analysis.

References


Theory-Guided Process-Tracing in Comparative Politics: Something Old, Something New

Tulia G. Falleti
University of Pennsylvania
falleti@sas.upenn.edu

The method of process-tracing was originally proposed to incorporate historical narratives within highly abstract theories and explanations in the social sciences. As such, the method is not new. Although it has been unearthed and polished recently, the debates on the relationship between historical narratives and theoretical explanations, and between the specificity of historical events and the generalizations of law-like propositions span several decades and various disciplines.

In political science, one of the earliest explicit definitions of process-tracing was provided by Alexander George
and Timothy McKeown (1985), who defined it as a method of within-case analysis to evaluate causal processes. According to George and McKeown (1985: 35), this method does not solely rely on the comparison of variations across variables in each case, but also “investigate[s] and explain[s] the decision process by which various initial conditions are translated into outcomes” (emphasis added). More concretely, the authors (1985: 35) argue that “[t]he process-tracing approach attempts to uncover what stimuli the actors attend to; the decision process that makes use of these stimuli to arrive at decisions; the actual behavior that then occurs; the effect of various institutional arrangements on attention, processing, and behavior; and the effect of other variables of interest on attention, processing, and behavior.” Thus defined, the method of process-tracing is deeply rooted in the tradition of methodological individualism. It attempts to uncover the microfoundations of individual behavior that connect hypothesized causes and outcomes and to reduce the difficulties associated with unobserved contextual variables. Moreover, the authors conceptualize process-tracing as a method that not only permits the testing of hypotheses but also, unlike the statistical method, the generation of them.

More recently, Alexander George and Andrew Bennett (2005: 206) defined process-tracing as the “method [that] attempts to identify the intervening causal process - the causal chain and causal mechanism - between an independent variable (or variables) and the outcome of the dependent variable.” According to these authors, the main ontological difference between the statistical method and the method of process-tracing is that while the former attempts to define causal effects (i.e., the expected value of the change in outcome when - in theory - only one independent variable changes), the latter identifies the causal mechanisms that connect causes and effects. They define causal mechanisms as “ultimately unobservable physical, social, or psychological processes through which agents with causal capacities operate, but only in specific contexts or conditions, to transfer energy, information, or matter to other entities” (2005: 137). To the extent that these agents do not need to be individuals, this more recent definition of process-tracing allows for the identification of causal mechanisms that do not have to be rooted at the individual level. Moreover, following Mahoney’s (2001) conceptualization of causal mechanisms, George and Bennett grant that these may ultimately be unobservable.

**Recent Approaches**

Other approaches have been applied to what we can label the “theory-guided process-tracing” (TGPT) method. There are important commonalities among them but also significant differences, particularly regarding the utility of the TGPT method to generate and test theories.

Sociologist Ronald Aminzade (1993) provides one of the most succinct and illuminating definitions of TGPT. According to Aminzade (1993: 108), the researcher has to provide “theoretically explicit narratives that carefully trace and compare the sequences of events constituting the process” of interest. Narratives, he writes (1993: 108), “allow us to capture the unfolding of social action over time in a manner sensitive to the order in which events occur. By making the theories that underpin our narratives more explicit, we avoid the danger of burying our explanatory principles in engaging stories. By comparing sequences, we can determine whether there are typical sequences across [cases] … and can explore the causes and consequences of different sequence patterns.”

The celebrated book by Ruth Berins Collier and David Collier (1991) on the different patterns of working class incorporation in eight Latin American countries and their effects on the political systems and regimes of the region constitutes an excellent example within the comparative politics literature of the type of research design and method described by Aminzade. In their book, Collier and Collier analyze the political developments of eight countries along four paired comparisons. Countries with similar types of incorporation of institutionalized labor are paired together. At the same time, major socioeconomic differences between the countries of each pair allow the authors to control for one of the alternative explanations (the impact of socioeconomic change). The case narratives and structured comparisons are organized around the conceptualization of reform, incorporation, aftermath, and heritage periods, spanning most of the twentieth century. TGPT shows that varying patterns of institutionalization of the conflicts among dominant and middle classes, organized labor, and the state in the early stages of the incorporation period had significant effects on the features of the political parties, their links to unions, and the political regimes of later periods.

Other methodological approaches to TGPT include those advanced by Robert Bates et al. (1998), Tim Büthe (2002), and Peter Hall (2003). The “analytic narrative” approach proposed by Bates and his coauthors combines the theoretical tools of rational choice and game theory with the narrative method. They “seek to account for outcomes by identifying and exploring the mechanisms that
generate them” (Bates et al. 1998: 12). Similar to George and McKeown’s uncovering of the decision-making process, the five collaborators (Bates et al. 1998: 11) in this book “seek to understand the actors’ preferences, their perceptions, their evaluation of alternatives, the information they possess, the expectations they form, the strategies they adopt, and the constraints that limit their actions.” Analytic narratives serve to illustrate the workings of a previously specified and deductively derived model. The ultimate epistemological goal of analytic narratives is to test models that yield unique equilibria. This is, however, the most limiting aspect of this approach. When narratives seek to incorporate relatively long periods of time, changes in variables that are exogenous (or endogenous) to the model are likely to occur and can significantly change the actors’ preferences, options, and strategies. The idea of stable equilibria is severely confining when applied to historical narratives. Hence, analytic narratives are not well suited to incorporating changes that reshape the preferences of actors or the effects of events that may render previously available options no longer viable.

In an excellent methodological article, Tim Büthe (2002) proposes the concept of “historical narratives.” These narratives seek to provide support for a theoretical argument and can be viewed as simplifications of reality that depict a deductively sound, systematic, regular relationship between variables (2002: 482). Unlike the authors of Analytic Narratives, Büthe allows for an analysis of endogeneity in his proposed approach. He recognizes the importance of feedback loops in social phenomena. These are cases in which changes in the dependent variable at one point in time may lead to changes in the independent variables at a later point in time, which in turn will lead to further changes in the dependent variable.

As other scholars have also noted (Pierson 2004), “[t]ime itself thus becomes an element of the causal explanation, a factor in the model” (Büthe 2002: 486). Historical narratives are useful for contextualizing the different steps of the process rather than leaving them fragmented into analytical stages. Büthe (2002: 489), however, remains skeptical about the extent to which historical narratives can assess alternative explanations. He argues that historical narratives are, at best, “plausibility probes” of models.

Peter Hall (2003), instead, gives a higher epistemological status to his proposed method of “systematic process analysis.” He observes that theories of strategic interaction and path dependence have recently brought to light complex ontological aspects of social phenomena. Issues of timing, sequencing, complex interaction effects, and multiple causalities are rendering traditional methods in comparative politics inadequate and wanting. The method he proposes begins by specifying the set of theories that identify the relevant causal factors and how they operate. From each theory, the investigator then derives predictions about the patterns that should be observed if the causal theory is valid and if it is false. Other relevant observations, as many and diverse as possible, are then made. The patterns present in these observations are then inspected for consistency with the predictions of each of the relevant theories to determine which causal theory is superior to the others (Hall 2003: 391-2).

Hall argues that in order to unfold the process that connects causes and outcomes, small-N comparisons provide much more explanatory leverage than has conventionally been recognized by the comparative method. Small-N comparisons allow scholars to assess more complex causal processes in a much richer set of observations. According to Hall, systematic process analysis is the most promising and fruitful venue to achieve an understanding of causal complexity.

The main difference among these recent approaches to TGPT is their conceptualization of the relationship between theory and method. In the Analytic Narratives project, TGPT provides an edge to understanding actors’ preferences, expectations, and strategies, which could not be elucidated through the application of other methodologies. Yet, the function of the method is to illustrate how the theory or formal model works in the real world rather than to generate theories. According to Büthe (2002), TGPT is more appropriate than other methods in the study of phenomena characterized by complex causality. However, he argues that, as a method, TGPT can only present us with plausible explanations, and cannot rule out alternative theories. In other words, it can generate but cannot test theories. In contrast, Hall (2003) maintains that TGPT is an epistemologically superior method in that it can map the ontological complexity of the social world and also rule out competing theories. It can generate and test theories.

Causal Mechanisms, Periodization, and Proficiency

What do these different approaches to TGPT have in common? In its different variants, the method of TGPT explains the outcomes of interest by going back in time and identifying the key events, processes, or decisions that link the hypothesized cause or causes with the outcomes. Note that this procedure entails that the researcher start with a set of hypothe-
“The method of [theory-guided process-tracing] explains the outcomes of interest by going back in time and identifying the key events, processes, or decisions that link the hypothesized cause or causes with the outcomes.”

Unfortunately, there is astonishingly little consensus about what the “causal mechanisms” are. James Mahoney (2001: 579-80) identified twenty-four definitions of causal mechanisms in the literature. In many cases, causal mechanisms become synonymous with “intervening variables.” However, if causal mechanisms were to be reduced to intervening variables, which could be operationalized and measured using other methods, then TPGT as a method would lose most of its comparative advantage. Ideally, causal mechanisms should be portable concepts. They should be of a higher level of aggregation than intervening variables. Causal mechanisms should identify relationships between conditions and outcomes that can be applied to other contexts (which should pertain to the theoretical domain of the theory proposed; see Geddes 2003: 152). Examples of these mechanisms are “learning,” “competition” (Pierson 2004: 40-1, 124-9), “institutional conversion,” and “institutional layering” (Thelen 2004: 35-7). These mechanisms are not simply intervening variables; instead, they refer to processes that can be applied to empirical situations different from those in which they were first conceived.

The utilization of the TPGT method poses two difficult questions to the researcher: when the narrative should start and when it should end. How far back the researcher goes in tracing the process is totally dependent on her theoretical framework and the competing theories under consideration, which define the theoretically relevant cause or causes to be studied. This, however, is not always as conclusive a criterion as we might like. The issue of starting points may be highly contentious. Collier and Collier (1991) start their TPGT in moments of critical junctures, whereas Mahoney (2000) argues that only contingent events (and not all critical junctures are contingent events) can trigger path-dependent processes. Whether or not we adopt a more stringent definition of starting point, we should always justify the choice of period under study. Whereas the justification of the starting point may be contested, the end point is easier to establish; there is less disagreement on this issue since the end point is largely determined by the presence of the outcome of interest.

In linking the start and the end of the sequence, the researcher who applies the TPGT method has to reconstruct an explicit chronology of the sequence of events that constitute the process of interest. This is not a simple task; it requires a precise conceptualization of the types of events that are constitutive of the process as well as those that are not - even if they pertain to related, albeit different, processes. Therefore, an explicit theory underpinning our process of interest as well as a profound familiarity with our cases are both necessary.

Successful use of the TPGT method requires the researcher to tap into her sociological or political imagination in order to identify the theories relevant to the problems and puzzles she seeks to explain, and to be able to derive feasible causal mechanisms. Moreover, familiarity with the case or cases under study is crucial. Familiarity with the history, historiography, and politics of the cases of study makes it possible to avoid problems of selection bias (Lustick 1996) and to improve the validity and reliability of our proposed narratives (Vitalis 2006). This requirement, in turn, poses two other practical questions: Is fieldwork always necessary? How many cases can a single author study using the TPGT method?

“How far back the researcher goes in tracing the process is totally dependent on her theoretical framework and the competing theories under consideration, which define the theoretically relevant cause or causes to be studied.”

Lots of ink has been spilled fueling the debate between the advocates of fieldwork and archival research, on the one hand, and the advocates of the sufficiency of secondary sources, on the other. Exemplary books and articles have been written using both research strategies, thus making it
difficult to establish a priori the supremacy of either approach.

However, in the early stages of our academic careers as comparativists, it would be hard to compensate for the many insights we get through in-depth interviews, the analysis of primary documents and archives, and the overall experience of living in the society whose politics we seek to explain. Independent of the research strategy we choose, the TGPT method requires in-depth knowledge of our cases. For this reason, works that apply TGPT will be closer to the small-N end than to the large-N end of the sampling spectrum. If, as Hall (2003) says, this is the price we pay to understand complex causality, the trade-off is worth it.

“...the TGPT method requires in-depth knowledge of our cases. For this reason, works that apply TGPT will be closer to the small-N end than to the large-N end of the sampling spectrum.”

The increasing salience of TGPT as a qualitative method can be appreciated in the sessions offered by the Consortium on Qualitative Research Methods (CQRM). In January of this year, the CQRM (for the fifth consecutive year) offered a ten day-course on Qualitative Research Methods (CQRM). In January of this year, the CQRM (for the fifth consecutive year) offered a ten day-course on Qualitative Research Methods (CQRM). In January of this year, the CQRM (for the fifth consecutive year) offered a ten day-course on Qualitative Research Methods (CQRM). In January of this year, the CQRM (for the fifth consecutive year) offered a ten day-course on Qualitative Research Methods (CQRM).

The TGPT method can provide a cognitive map for the type of ontological world described by Hall (2003). In this regard, the TGPT method has several advantages compared with traditional methods. First, as emphasized by George and Bennett (2005) and Hall (2003), it permits the study of complex causal relationships such as those characterized by multiple causality, feedback loops, path dependencies, tipping points, and complex interaction effects. Second, it can lead to the formulation of new theories or hypotheses on the causal mechanisms that connect correlated phenomena. Third, TGPT in structured, focused comparisons (i.e., measuring the same causal mechanisms and outcomes in the same way across each case) permits the testing of hypotheses and theories. Finally, the TGPT method can reveal how endogenous changes affect the evolution of our variables of interest. For all these reasons, the careful application of TGPT to generate and test hypotheses will continue to advance our knowledge about complex causality phenomena in comparative politics.

References


1 Introduction

Scholars in comparative politics almost by definition can only resort to quasi-experimental studies to test their theories. Few are the situations where a comparativist finds her/himself in the lucky situation of having control over an explanatory variable. This becomes most obvious when we turn to the question of whether institutions matter (e.g., Przeworski 2004). While comparativists have become more self-conscious in making theoretically derived causal claims, they have also been alerted to two related problems which make testing their theories difficult, or, perhaps even impossible. These two intertwined problems relate to the quasi-experimental nature of most empirical research in comparative politics and are described as selection biases and the problem of endogeneity.

In this short note, I wish to demonstrate that these two problems are of tantamount importance, independent of the research strategy, i.e., quantitative or qualitative, employed. Given the more solidly developed tools to address these problems in quantitative research, I will discuss these problems in this context and highlight ways in which they can be addressed, albeit sometimes only imperfectly. I start by presenting the problem of endogeneity bias, showing also how it relates closely to selection biases.1 I examine this second problem in section three, before concluding in section four.

2 Models of endogeneity

A frequent approach in comparative politics is the attempt to assess why particular outcomes differ among countries. An important research strand attributes considerable weight to the role of institutions:

\[ \text{outcome} = f(\text{institutions}) \]  

Underlying this simple equation is the assumption that if in a particular country different institutions, e.g., those of a democracy instead of those of an autocracy, were present, the outcome would be different. To tie things down more precisely, let’s assume that \( x_i \) indicates the presence of a particular institution and \( x_i = 0 \) its absence. Then our statement would involve a comparison of the outcome \( y \) under two scenarios. Let’s assume that we observe in our dataset the outcome \( y_i \) for a particular observation \( i \) for which \( x_{i,1} = 1 \). To assess the effect of \( x_{i,1} = 1 \) we would need to know the outcome \( E(y_i \mid x_{i,1} = 0) \). To circumvent this problem we normally assume that the effect

\[ E(y_i \mid x_{i,1} = 1) - E(y_i \mid x_{i,1} = 0) \]

can be assessed by comparing cases with 0 and 1 as values of \( x_i \). More precisely we resort (in a linear regression framework) to the following model:

\[ y_i = \beta_0 + \beta_x x_i + \epsilon \]  

Estimating this model under the well-known assumptions of the classical linear regression (CLR) would yield an estimate of our quantity of interest, \( \beta_x \). Obviously, this model is very unlikely to satisfy the assumptions of the CLR model. Most glaringly, it is rather unlikely that no omitted variables influencing our outcome are unrelated to our included variable.

There are at least two reasons why this might occur. First, several variables might affect our outcome variable and be related to \( x \), for instance, because the chosen institution affects this other variable. Controlling for these variables by including them as a set of variables \( z \) would address this issue. Second, variables affecting \( y \) might also be related to the presence of \( x \). Any results in the problem of endogeneity (of \( x \) to be precise). Again, including these additional variables as controls may help address this issue. In order to assess the effect of \( x \) precisely, these variables included in \( z \) which explain the presence of \( x \) must either explain the presence of \( x \) perfectly, or whatever is not explained is unrelated to our outcome \( y \). Both of these eventualities are unfortunately quite unlikely, especially in comparative politics. Our models which try to explain, for instance, the presence of a particular institution, are often entailed with a considerable error:

\[ x_i = \delta_0 + \sum_j \delta_j z_{i,j} + \theta_i \]
In this equation, the variables $z_1,...,z_l$ correspond to additional explanatory variables which we think contribute to the presence of institution $x$. Equation 2 clearly shows under what circumstances adding variables $z$ in equation 3 is an adequate strategy. As long as $\theta$ in equation 2 is uncorrelated with $\varepsilon$ in equation 1, adding these controls solves our problem of endogeneity. In most cases the assumption that $\Sigma_{z\theta} = 0$ is unwarranted, which leads to biased estimates of the effect of a particular institution as determined by $\beta_x$.

Equation 2 also shows a way to address the problem of endogeneity. By construction

$$\hat{x}_i = \delta_0 + \sum_{j=1}^{l} \delta_j z_{ji}$$

obtained by estimating equation 2 and using the estimated coefficients to obtain predicted values for $x_i$ will be uncorrelated with $\varepsilon$, provided that the $z_j$'s are uncorrelated with any control variables we might wish to include in equation 2. Inserting $\hat{x}$ instead of $x$ will solve our problem. This solution, based on instrumental variables (IVs), has become common practice. It is not, however, without its pitfalls. First, as Bartels (1991) and Bound, Jaeger, and Baker (1995) have authoritatively shown, if the instruments $z$ for $x$ are weak, i.e., the fit of equation 2 is low, our estimated coefficient for $\beta_x$ may be very imprecise. Second, finding instruments uncorrelated with other included variables is often difficult. An illustrative example comes from the ongoing debate about whether good institutions enhance economic growth. Awareness of the endogeneity of good institutions Acemoglu, Johnson, and Robinson (2001) employ as instrumental variables $z$ settler mortality and indigenous population density in 1500. Glaeser, Porta, López de Silanes, and Shleifer (2004) argue, however, that these two variables fail to be good instruments, since they are correlated with another variable, namely human capital, which also explains economic growth.

Another approach to the problem of endogeneity is matching. Intuitively we need to find for observations $i$ with $x_i = 1$ cases which are (almost) identical in all other variables to observation $j$, except that $x_j = 0$.

These matching methods, as nicely illustrated in Persson and Tabellini’s (2003) work on the economic effects of political institutions, involves estimating equation 2. The predicted probability gives us an indication of how close a particular observation is to another one with a different value on $x$. We can compare, for instance, the value of $y$ for pairs of observations that are similar in their estimated probabilities of having the institution but different with respect to whether the institution is actually present. The average of the differences in the values of $y$ for all the pairs is our estimate of the effect of $x_i$.

While I discussed the problem of endogenous regressors in the context of a classical linear regression model, similar solutions exist for limited dependent variables (Rivers and Vuong, 1988; Bollen, Guilkey, and Mroz, 1995). Interestingly studies employing this approach also allows researchers to address the problem of endogeneity.

To derive these biases, I will detail in the next section the basic elements of selection bias models.

### 3 Models of selection bias

At least since the publications of Achen (1986) and Geddes (1991), scholars in comparative politics have been aware of the problem of selection biases. Increasingly interesting applications of models addressing this problem appear in journals and books.

The problem arises when (at least) our dependent variable $y$ is observed for only a subset of the observations of sample we wish to study. Thus, in the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (4)$$

we have

$$y = \begin{cases} y_{obs} \text{ if } t = 1 \\ \cdot \text{ if } t = 0 \end{cases} \quad (5)$$

Under the assumption that $t$ is related to some independent variables $z$, we have
\[ t_i^* = \delta_0 + \sum_{j=1}^t \delta z_{ji} + \theta_i \]

and \[ t_i = \begin{cases} 1 & \text{if } t_i^* \geq 0 \\ 0 & \text{if } t_i^* < 0 \end{cases} \quad (6) \]

If \( \sigma_{e,b} = 0 \), then estimating equation 1 on the subset of cases for which \( y \) is observed yields unbiased estimates. If \( \sigma_{e,b} \neq 0 \), however, then biases are to be expected. This is all the more likely if the set of variables \( z \) contains elements appearing in equation 3. If these are imperfect measures, omitted variables are likely to be similar across the equations 3 and 5, inducing covariation between the two error terms. Solutions to this problem exist but differ, however, according to the availability of data. In general, the setup consists of equations 3 and 5.

The question now is what remains unobserved if \( t \leq 0 \). The simplest case is if only \( y \) is unobserved (as in equation 4), while all the remaining variables are observed. Heckman (1976) derived a simple estimator discussed in detail in Achen (1986). In essence, it consists of estimating in a first step the equation \( t_i \) as a probit model and then use the inverse of the Mills ratio as additional regressor in the outcome equation.\(^{10}\)

In the more complicated case, neither \( y \) nor \( x \) are observed when \( t \leq 0 \). In that case a truncated regression results. For obvious reasons, a first stage selection equation cannot be separately estimated, since for all observed cases \( t_i \geq 0 \), hence no variation appears in the probit model. Outcome and selection equation can, however, be estimated jointly, since the two equations together induce a log-likelihood function for the observed cases (e.g., Maddala 1983, 176ff). This function reflects the likelihood that, conditional on being observed, observations take on particular values for the dependent variable.

Hug (2003), building up on Muthen and Jöreskog (1983), provides Monte Carlo simulations for this estimator. He demonstrates that with sufficient correlation between the error terms in the equations 3 and 5, a sufficiently large sample, and a minimum of selectivity, the estimator clearly outperforms a simple OLS estimation of the outcome equation. He also shows in three empirical examples dealing with the success of new political parties, the degree of violence in demonstrations, and the rebellious behavior of “minorities at risk,” that selection effects are quite real.

For the simpler models, Achen (1986) based on Heckman’s (1976) work, offers a very accessible introduction, as well as Breen (1996). While most of the issues discussed here are presented in a concise fashion in Maddala (1983), Greene (2003, 780-790) discusses these issues in great detail. Dubin and Rivers (1990) offer a general treatment of selection models when dealing with limited dependent variables.

More recently, scholars have become interested in addressing selection biases in more complex situations. Both Brehm (2000) and Boehmke (2003) suggest ways to use external information to arrive at an estimate of the Mills ratio necessary in the second stage of a Heckman model, even if a first-stage selection equation cannot be estimated separately.

Sartori (2003) derived, under the assumption of identical errors across the two equations, an estimator for a binary selection model without exclusion restrictions. Finally, Boehmke, Morey, and Shannon (forthcoming, 2006) present an elegant way to address selection bias in duration models.

Nice illustrations of the relevance of models of selection bias appear in Przeworski and Vreeland (2000), who highlight the effect of selection biases in assessing the role of IMF loans. While other authors find very diverging effects of IMF loans on economic growth, Przeworski and Vreeland (2000) report very subtle differences. During the loan program, countries see lower economic growth than they would have if they had not requested a loan. After the end of the program, these same countries grow, however, faster than comparable countries that refrained from entering into it. The decision to join a loan program appears to be influenced by the governments’ willingness to adopt politically costly adjustment policies.

Similarly, Nooruddin (2002) shows selection effects in the results of economic sanctions. When one does not consider why the US imposed sanctions against a country, the success of the sanctions does not appear to be related to the country being allied to the US; but when the selection mechanism is taken into account, US ties contribute to the explanation. Specifically, the success of sanctions depends on whether the target was involved in a military dispute and whether there was cooperation with the US.

4 Conclusion

Comparative research, whether qualitative or quantitative, is almost by definition confronted with problems of endogeneity and selection bias. To test our theories, we would ideally need experimental control, but in practice we can only rely on quasi-experimental data. This should make us all more sensitive to the problems I discussed in this note. While some authors (e.g., Przeworski, 2004) come close to suggesting that studying the effect of institutions is impossible because of these problems, I
would argue that adequate empirical models may alleviate some of the most egregious biases.

“Comparative research, whether qualitative or quantitative, is almost by definition confronted with problems of endogeneity and selection bias [...] Adequate empirical models may alleviate some of the most egregious biases.”

Hence, even if our explanatory variables are endogenous or our sample is only a subset of the cases we wish to study, ways exist not only to assess whether this results in biases but also to correct them. Required for this, however, are theoretically well-grounded empirical models explaining the presence of institutions and their effects. Given that the problems of endogeneity and selection biases are almost a mainstay of comparative politics, comparativists should always be aware of these potential problems.

Notes

1 This close link is discussed in detail by Przeworski, Alvarez, Cheibub, and Limongi (2000, 279-289) and Persson and Tabellini (2003, 113-153).

2 In technical terms this is often referred to as “selection on observables.”

3 Here I gloss over the problem that a linear model is normally hardly appropriate if \( x \) is dichotomous, even though Angrist and Krueger (2001) suggest using a linear probability model in such situations.

4 If control variables are likely to be correlated with the instruments, Acemoglu (2006) pleads for the inclusion of the control variables in the first stage regression as additional regressors.

5 Accessible introductions and surveys of the relevant literature on IV estimation appear in Angrist and Krueger (1999) and Greene (2003), while Hausman (1978) discusses the possible specification tests.

6 Persson and Tabellini (2003, 113-153) provide an accessible survey of the problems discussed in this note, and present the basic logic of matching methods (138-142). Acemoglu (2006 forthcoming) criticizes, however, their implementation of the IV-estimator and the usefulness of some of the instruments used to predict the presence of particular institutions.

7 This form of matching is only one to derive estimates for the quantity of interest to us (see Persson and Tabellini 2003, 138-142).

8 These authors rely both on instrumental variables and selection models to address the endogeneity of political institutions.

9 Given that \( x \) is a constant in this subset of observations, the effect of this variable is absorbed into the constant.

10 A constraint has to be respected, namely that in the set of variables \( z \) at least one variable does not appear in the set of variables \( x \). A full-information likelihood estimator exists as well for this model.

References for this article are on-line at www.nd.edu/~apsacp

In 2005, the World Bank made public the latest update of its Governance Indicators, a comprehensive panel dataset that covers six dimensions of governance for practically all countries in the world, plus several associated territories. Data are available for alternating years during the period 1996-2004. This update not only adds the 2004 estimates to the series, but also revises estimates for previous years on the basis of newly available data. Clear and complete methodological and substantive information about this data source is found in Kaufmann, Kraay, and Mastruzzi (2004) and, for the last update, in Kaufmann, Kraay, and Mastruzzi (2005). The World Bank does not plan to extend the database back into the past (because many of the data sources are not available before 1996), but will continue updating it in the future. The next wave of indicators for 2006 will be available in early 2007.

The six dimensions of governance identified and measured by the World Bank are: 1) Voice and Accountability (political, civil, and human rights); 2) Political Stability and Violence (the likelihood of violent threats to governments, such as terrorism); 3) Government Effectiveness (the quality
Datasets

of the bureaucracy and public services); 4) Regulatory Quality (the incidence of market-unfriendly policies); 5) Rule of Law (the quality of contract enforcement, the police and the judiciary, and the likelihood of crime); and 6) Control of Corruption (the exercise of public power for private gain). Political scientists in general, and comparativists in particular, will recognize in these labels some of their most interesting dependent variables and theoretically crucial independent variables.

Each of the five waves from 1996 to 2004 has incorporated more countries and data sources. The average number of countries covered in each of the six dimensions was 173 in 1996 (with a minimum of 151 in the “Control of Corruption” dimension), monotonically increasing since then to the current 207 (with a minimum of 204 in the “Regulatory Quality” and “Control of Corruption” dimensions). Likewise, the number of sources per country used in the construction of the measurement for each dimension grew from a median of 4 in 1996 to a median of 9 in the 2004 wave. The percentage of countries with only one data source declined from 15% to 7% in the same period.

The array of data sources used to construct each indicator is impressive. For the 2004 wave, the World Bank used 352 variables drawn from 37 data sources (12 of them used for the first time in 2004), produced by 31 different organizations, such as the Economist Intelligence Unit, Freedom House, Gallup International, the Heritage Foundation, Political Risk Services, State Department/Amnesty International, and the World Bank. All the data sources are subjective (or perceptions-based) measures. They can be broadly divided into two types: 1) surveys of individuals and domestic firms with first-hand knowledge of the governance situation within a country; and 2) assessments by specialized agencies such as risk rating companies, universities, NGOs, and multilateral aid organizations whose staffs include knowledgeable country analysts or in-country correspondents. The data available for each country from this diverse array of sources is then combined through an unobserved components model, which allows calculating the precision (i.e., the standard error) of the scores.

Each of the six indicators is a standard normal variable (by definition mean=0 and standard deviation=1), so that almost all scores lie in the -2.5 to 2.5 range (higher scores correspond to better performance). The fact that each wave includes more sources implies that the estimates are more precise in the most recent years: the average standard error for all countries in each of the six dimensions ranges from 0.26 to 0.36 in the 1996 wave, while it ranges from 0.18 to 0.27 in the last series. However, the accompanying methodological paper emphasizes that, even in 2004, the magnitude of standard errors is not trivial, and therefore small differences among countries are typically not significant.

Although relying exclusively on subjective measures may be seen as a weakness, especially by researchers accustomed to working with allegedly more accurate ones, the authors make a strong case for the former. They not only argue that key governance dimensions, such as corruption, are basically impossible to measure objectively, but also point out that perceptions-based measures often capture information missed by objective ones, especially when the latter assess de jure situations that are often poor descriptions of de facto realities (Kaufmann, Kraay, and Mastruzzi 2005, 27-31).

The database, downloadable as an Excel spreadsheet (from http://www.worldbank.org/wbi/governance/govdata/), does not have the typical “long” format but the “wide” format for panel (or longitudinal) data. Each of the 209 rows is a country and each of the 90 substantive columns corresponds to the dimension score, its standard error, and the number of sources on which the estimate is based (that is, 3 columns*6 dimensions*5 years=90).

Despite the obvious usefulness of the six indicators on their own, they are in fact the product of the operationalization of the concept of governance, understood as having three dimensions: 1) “the process by which governments are selected, monitored and replaced”; 2) “the capacity of the government to effectively formulate and implement sound policies”; and 3) “the respect of citizens and the state for the institutions that govern economic and social interactions among them” (Kaufmann, Kraay, and Mastruzzi 2005, 130). Voice and Accountability and Political Instability and Violence measure the first dimension, Government Effectiveness and Regulatory Quality the second, and Rule of Law and Control of Corruption the third one.

Given this conceptualization, it is important to assess whether these indicators are measuring a single underlying “governance” variable, the three independent dimensions mentioned above, or maybe other dimensions not foreseen by the authors of the study. My factor analysis of the six indicators in 2004 provides strong evidence in favor of the single dimension hypothesis: the first factor explains 87.3% of the total variance and correlates very highly with all indicators (from a maximum of .984 for Rule of Law to a minimum of .856 for Voice and Accountability). This finding suggests that an aggregate
“Index of Governance” can and should be constructed. Although each dimension taps a conceptually (and for some countries empirically) different trait, it is apparent that the high inter-dimension correlations points to the existence of a latent variable that strongly influences all of them.

Governance is generally assumed to be a relatively enduring feature of political systems. The World Bank indicators confirm this idea: changes from wave to wave are not very large. The correlations between the scores in 1996 and the scores in the subsequent waves decline over time but remain always very high. For Voice and Accountability, the Pearson correlation coefficient between 1996 and 2004 scores is .927, for Government Effectiveness .925, for Rule of Law .920, for Control of Corruption .913, for Regulatory Quality .861, and for Political Stability .810. In sum, the stability of scores is similar and very high for the first four indicators, somewhat lower for Regulatory Quality, and still lower for Political Stability. But even in this case, a 1996 country value is a rather good predictor of its 2004 score: the adjusted $R^2$ is .655 and the mean residual is .58 (a relatively low figure for a variable that has a range of approximately 5).

The website for the database has a built-in utility for obtaining data and graphs for specific countries and regions, which permits an easy and quick access to snapshots of the data. More complex graphs and statistics require downloading the full Excel database and making minor changes to it in order to make it readable by statistical, graphing, or file conversion packages.

Figure 1 summarizes the data by region and year. In order to avoid cluttering the graph, I collapsed the six series of data into a single line per region that represents the unweighted average of the six governance indicators. Because of the relative nature of the indicators, the aggregate statistics are zero-sum: if a country or region goes up another one has to go down. Therefore, the trends in the graph have to be interpreted in relative rather than absolute terms.

For example, the slight decline of the OECD countries in the period 1998-2004 is probably more a function of absolute improvement in Eastern...
Europe and East Asia than of its own decline in absolute terms.

The authors of the accompanying paper are modest about the accuracy and usefulness of this important and constantly improving database: “While these aggregate governance indicators have been useful in providing a general snapshot of the countries of the world for various broad components of governance, now for 8 years, and while the margins of error have declined over time, they remain a rather blunt instrument for specific policy advice at the country level. As we have argued in the past, these aggregate indicators need to be complemented with in-depth in-country governance diagnostics, based on micro-surveys of households, firms and public officials within the countries” (Kaufmann, 2005, 41). Notice that this caveat does not necessarily apply to aggregate analysis: even if they carry relatively large standard errors, the governance scores can still yield unbiased (if inefficient) statistical estimates when they are used as dependent variables in regression analysis. As independent variables, however, these not-too-high levels of precision will result in important attenuation biases.

The World Bank’s Governance Indicators are a comprehensive, source-rich, dynamic, and methodologically sophisticated source of data for some of the most important variables that comparativists and other political scientists study. The project’s on-line information and data access functions are clear, useful, and easily accessible, as are the accompanying methodological and substantive papers. The methods and sources are transparent, and the authors are cautious in highlighting the limitations of the data.

Given the human, organizational, and financial resources of the World Bank, and its intention to keep producing and improving the database in the future, the uses, analyses, and criticisms that political scientists make of these data will likely lead to a virtuous feedback circle. This will, a few years from now, allow the profession to count on a temporally longer and qualitatively better Governance Indicators database.

References


Notes

1 Several of the polities included in the database, such as the Cook Islands, the Virgin Islands, Reunion, and the West Bank, are not strictly sovereign states. In the rest of the review I use the word “countries” to refer to all the polities included in the database.

2 This list of sources may raise questions about possible ideological biases, in the sense that some of them may assess the governance situation more favorably in countries with, for example, conservative administrations. The authors conducted a test to detect such bias, and concluded that there is not evidence of it for any source except the Heritage Foundation, which is the “only source that appears to have a consistent ideological bias,” although “fairly modest in magnitude” (Kaufmann, Kraay, and Mastruzzi 2005, 22-3).

3 Details on the statistical procedure are provided in the appendix of Kaufmann, Kraay, and Mastruzzi (2005) and in Kaufmann, Kraay, and Zoido-Lobaton (1999). The Unobserved Components Model is basically a weighted average of the observed scores from each source for each country (re-scaled to common units) in which weights are proportional to the precision of each data source. Precision, in turn, depends on the strength of the correlation among the sources.

4 It is important to note that the imprecision of the estimates does not arise from their being based on subjective measures but from the fact that alternative measures provide different estimates of the underlying dimension. The more sources an estimate is based on and the more precise these sources are, the better the estimate’s precision.
Recent writings concerning the measurement of political democracy offer sophisticated discussions of the problems of conceptualization, operationalization, and aggregation. Yet they have less to say about the error that derives from the use of inaccurate, partial, or misleading data sources. Drawing on evidence from five Central American countries, we show that this data-induced measurement error compromises the validity of the principal, long-term cross-national scales of democracy (*Comparative Political Studies*, October 2005).

In response, we call for an approach to index construction that relies on case expertise and the use of a wide range of data sources. We employ this approach in developing a new index of political democracy for the Central American countries during the 20th century. The index draws on a comprehensive set of secondary and primary sources even as it rigorously pursues standards of conceptualization, operationalization, and aggregation. The raw data for each country, the BLM democracy index, and the scores for each of the 5 dimensions of democracy are available for download at [http://blmdemocracy.gatech.edu/](http://blmdemocracy.gatech.edu/).

Our index is dynamic in the sense that new data or sources may refine our coding of individual dimensions for individual years, particularly in the early part of the 20th century.

If you believe that our current raw dataset has errors or shortcomings, please contact the authors.
International Political Science Association Prize

The International Political Science Association has bestowed the first Foundation Mattei Dogan Prize for Lifetime Achievement in Political Science on Guillermo O’Donnell of the University of Notre Dame. O’Donnell will deliver a prize lecture at the IPSA World Congress in Fukuoka, Japan, in July. The text of his lecture will be published in the summer 2006 issue of APSA-CP.

Sidney Tarrow on the 2006-2007 APSA programs

Thanks to Program Chair Deborah Yashar, it looks as if the Comparative Politics section will make a large and interesting contribution to the 2006 APSA meeting in Philadelphia. The Program Chairs for the 2007 meeting (Deborah Gerber and David Lake) are already at work planning for that meeting and have asked me to get to work soon to name a chair or chairs for our section. Here is a copy of their message to the Section heads:

We write at this time to ask for your nominations for your section’s panel organizers. These individuals will be responsible for organizing your section’s panels and poster sessions, and collectively will make up the meeting program committee. Duties include attending a program committee luncheon at the 2006 Annual Meeting, writing the section’s call for papers, reviewing and selecting electronically submitted paper and panel proposals, collaborating via phone and email with other program committee members to create co-sponsored panels, securing discussants and panels chairs, keeping track of acceptances and rejections, and assisting APSA staff in finalizing the program listings. Each organized section is asked to nominate one or two individuals to organize their section’s panels. Nominees should be people with excellent professional judgment who are well connected in their subfield, and who are reliable and well organized. We strongly encourage nominees who are members of under-represented groups or are from under-represented institutions. I would be most happy to get your suggestions for this important (and challenging!) assignment. Sid Tarrow.

Michael Wallerstein, 1951-2006

Political Science and Comparative Politics lost one of its most distinguished and beloved theorists and practitioners with the passing away of Michael Wallerstein on January 7, 2006 from brain cancer. Wallerstein served as president of the Comparative Politics section from 1999 to 2001.

A political economist with enormous empirical range and methodological sophistication, Michael Wallerstein focused his attention on understanding relationships among political-economic institutions, electoral politics, and market outcomes. Some of his best-known work explores the social and economic impact of wage bargaining institutions, with an eye toward understanding the contributions that centralized wage bargaining and social democratic politics can make both to economic efficiency and social equality.

Among other things, Wallerstein’s work showed that powerful labor unions can reduce wage inequality without increasing unemployment, and that centralized unions play an important role in reducing inequality. Wallerstein also investigated how social democracy, such as that found in the Scandinavian nations, can counteract the inefficiencies and inequities of a market society. In addition to influential single-authored works, Wallerstein was also known for fruitful long-term collaborations, among others with Adam Przeworski and Karl Ove Moene. He recently completed two articles, one with Miriam Golden on the impact of globalization and occupational change on wage inequality, and a second with
David Austen-Smith, on the impact of racial divisions in society on political support for welfare spending.

Wallerstein earned his undergraduate degree in political science from Stanford University in 1974, and his Ph.D. in political science from the University of Chicago in 1985. He began his teaching career in 1984 at the University of California at Los Angeles and spent ten years there before moving to Northwestern University in 1994. Wallerstein spent the next decade at Northwestern, where he served as chair from 1997 to 2000. In 2004, he moved to Yale University, where he assumed the Charlotte Marion Saden Professorship of Political Science. In 2005, Wallerstein was elected to the American Academy of Arts and Sciences. Other honors that Michael Wallerstein earned over the course of his career include the Franklin L. Burdette/Pi Sigma Alpha Award for the best paper presented at the annual political science meetings (1985) and runner-up for the Gregory Luebbert Award in comparative politics (1999).

Over the course of his career, Wallerstein’s work received support from the National Science Foundation, the Norwegian Research Council and the MacArthur Foundation. He was a scientific advisor for the Foundation for Research in Economic and Business Administration in Norway, where he had been a visiting scholar at both the University of Oslo and at the Institute for Social Research during the academic year 1989-1990. Since 2001 he served as a member of the Scientific Committee of the Center for the Advanced Study in the Social Sciences at the Juan March Institute in Madrid, Spain, where he visited and taught regularly. In the American Political Science Association, he had served on the Executive Council and was also a member of the Task Force on Graduate Education.

Wallerstein was also an accomplished competitive sailor. While at Northwestern, he spent many hours on Lake Michigan pursuing excellence in sailing as in everything else. A guitar player, he had an abiding love of music, especially jazz, and especially John Coltrane.

He was a loving husband and father, survived by his wife, Elizabeth Atlas, an artist, whom he met as a student on an archaeological dig in Mexico; his son Jonah, a recently graduated software engineer; and his daughter Hannah, a student at Wesleyan College. He leaves behind a rich intellectual legacy in the many graduate students whose scholarship he helped to shape. Many colleagues were, in a sense, also his grateful students. On a personal level, he will be remembered across the country and abroad as an extremely generous scholar who gave freely of his time. He will also be remembered as a scholar whose work was imbued with a deep passion for politics and an abiding desire to use his scholarship to advance the cause of justice and equality.

Kathleen Thelen, Miriam Golden, and Peter Swenson

New Master's in Democracy and Democratization, University College London

The Department of Political Science and School of Public Policy at University College London have launched a master’s program in Democracy and Democratization, currently in its first year. The program focuses on the design and operation of democratic institutions in old and new democracies. More information about the program is available at http://www.ucl.ac.uk/spp/teaching/msc-democracy-democratization/ or by contacting Dr. Sherrill Stroschein at s.stroschein@ucl.ac.uk.
In this Newsletter:  
Winter 2006

How to Subscribe

Subscription to the APSA-CP Newsletter is a benefit to members of the Organized Section in Comparative Politics of the American Political Science Association. To join the section, check the appropriate box when joining the APSA or renewing your Association membership. Section dues are currently $8 annually, with a $2 surcharge for foreign addresses. The printing and mailing of the Newsletter is paid for out of member dues. To join APSA, contact:

American Political Science Association  
1527 New Hampshire Ave., NW  
Washington, DC 20036, USA  

Telephone: (202) 483-2512  
Facsimile: (202) 483-2657  
e.mail: membership@apsanet.org

Changes of address for the Newsletter take place automatically when members change their address with the APSA. Please do not send change-of-address information to the Newsletter.

Copyright 2006 American Political Science Association. Published with financial assistance from the members of the Organized Section in Comparative Politics of the American Political Association and the College of Arts and Letters, the Helen Kellogg Institute for International Studies, the Nanovic Institute for European Studies, and the Graduate School of Arts and Sciences of the University of Notre Dame du Lac.

Letter from the President  
Space and Comparative Politics, Sidney Tarrow

Symposium: Innovative Methods in Comparative Politics

Automated Content Analysis of Political Texts using Wordscores,  
Kenneth Benoit and Michael Laver

Theory-Guided Process-Tracing in Comparative Politics: Something Old, Something New, Tulia G. Falleti

Endogeneity and Selection Bias in Comparative Research,  
Simon Hug

Datasets

Data Set Review: The World Bank’s Governance Indicators (1996-2004),  
Carlos Gervasoni

Other datasets

News & Notes

Awards

Announcements

American Political Science Association  
University of Notre Dame  
Decio Hall, Box “D”  
Notre Dame, IN 46556  
USA