

Comparative Political Studies (CPS), an interdisciplinary journal, publishes theoretical and empirical research articles by scholars engaged in cross-national comparative study. **CPS** will not consider for publication manuscripts that have been simultaneously submitted to other journals or books unless such manuscripts are accompanied by an explanatory letter giving strong reasons for considering an exception to this rule.

Manuscripts should be sent electronically to the Comparative Political Studies Editorial Office as Word files: cps@u.washington.edu. Submit an anonymous manuscript file (all identifying information removed), plus a separate page with the title, authors' names and affiliations, keywords, and complete contact information for the corresponding author (mail and e-mail addresses as well as telephone and fax numbers). If the manuscript is submitted in PDF format, provide a separate 100- to 150-word abstract as a Word file for use in querying reviewers. The abstract must state precisely the intellectual problem under consideration, the method or argument addressed to the problem, and the conclusions reached by the author(s). Articles should be double spaced, prepared with a 12-point font (Times Roman preferred), and 25 to 35 pages long. Longer submissions may not be considered. Count all pages and begin the abstract, reference section, footnotes, tables, figures, and appendices (if any) on separate pages. Footnotes and bibliographic citations should follow the *Publication Manual of the American Psychological Association* (5th edition). A style sheet may be obtained upon request. A brief bibliographical paragraph describing each author's current affiliation, research interests, and recent publications should accompany the manuscript. Submission of a manuscript implies commitment to publish in the journal. Please state in your cover letter that the material has not been published elsewhere in substantially similar form or with similar content. Authors in doubt about what constitutes prior publication should consult the editor. E-mailed submissions will be assigned a manuscript number and acknowledged. Unless requested otherwise, decision letters will be sent by e-mail to the corresponding author.

Comparative Political Studies (ISSN 0010-4140) (J203) is published monthly by SAGE Publications, Thousand Oaks, CA 91320. Copyright © 2008 by SAGE Publications. All rights reserved. No portion of the contents may be reproduced in any form without written permission of the publisher. Periodicals postage paid at Thousand Oaks, California, and at additional mailing offices. POSTMASTER: Send address changes to Comparative Political Studies, c/o 2455 Teller Road, Thousand Oaks, CA 91320.

Subscription Information: Annual subscription rates for institutions and individuals are based on the current frequency. Prices quoted are in U.S. dollars and are subject to change without notice. Canadian subscribers add 7% GST (and HST as appropriate). Outside U.S. subscription rates include shipping via air-speeded delivery. **Pricing:** Institutions: \$1,053 / single issue: \$95. Individuals: \$139 / single issue: \$15. Noninstitutional orders must be paid by personal check, VISA, or MasterCard. **Claims:** Claims for undelivered copies must be made no later than six months following month of publication. The publisher will supply replacement copies when losses have been sustained in transit and when the reserve stock will permit.

All subscription inquiries, orders, and renewals with ship-to addresses in North America, South America, and Canada should be addressed to SAGE Publications, 2455 Teller Road, Thousand Oaks, CA 91320, U.S.A.; telephone (800) 818-SAGE (7243) and (805) 499-0721; fax (805) 371-1700; e-mail journals@sagepub.com; <http://www.sagepub.com>. Inquiries and subscriptions with ship-to addresses in the U.K., Europe, the Middle East, and Africa should be sent to Sage Publications Ltd, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP, England; telephone +44 (0)20 7324 8500; fax +44 (0)20 7324 8600. For ship-to addresses in India and South Asia, please write to Sage Publications India Private Ltd, B-42 Panchsheel Enclave, Post Box 4109, New Delhi 110-017, India; telephone (91-11) 2649 1290; fax (91-11) 2649 2117. For additional ordering instructions to customers outside the Americas, please visit <http://www.sagepub.co.uk/customerCare.nav> for information.

Copyright Permission: Permission requests to photocopy or otherwise reproduce material published in this journal should be submitted by accessing the article online on the journal's Web site at <http://cps.sagepub.com> and selecting the "Request Permission" link. Permission may also be requested by contacting the Copyright Clearance Center via their Web site at <http://www.copyright.com>, or via e-mail at info@copyright.com.

Abstracting and Indexing: This journal is abstracted or indexed in Academic Abstracts, Academic Search, America: History and Life, Asia Pacific Database, Central Asia: Abstracts & Index, Corporate ResourceNET, Current Citations Express, Current Contents: Social & Behavioral Sciences, Expanded Academic Index, Historical Abstracts, International Bibliography of the Social Sciences, International Political Science Abstracts, ISI Basic Social Sciences Index, MasterFILE FullTEXT, Middle East: Abstracts & Index, National Information Services Corporation (NISC), North Africa: Abstracts & Index, PAIS International, Peace Research Abstracts Journal, Periodical Abstracts, Political Science Abstracts, Public Library FullTEXT, Scopus, Social Science Source, Social Sciences Citation Index, Social Services Abstracts, Sociological Abstracts, Southeast Asia: Abstracts & Index, Standard Periodical Directory (SPD), TOPICsearch, Wilson OmniFile V, and Wilson Social Sciences Index/Abstracts and is available on microfilm from ProQuest, Ann Arbor, Michigan.

Advertising, Reprints, and Back Issues: Current advertising rates and specifications may be obtained by contacting the advertising coordinator in the Thousand Oaks office at (805) 410-7160 or by sending an e-mail to advertising@sagepub.com. To order reprints, please call (805) 410-7763 or e-mail reprint@sagepub.com. Information about availability and prices of back issues may be obtained from the publisher's order department (address above). Single-issue orders for 5 or more copies will receive a special adoption discount. Contact the order department for details; write to the London office for sterling prices.

Change of Address: Six weeks' advance notice must be given when notifying of change of address. Please send the old address label along with the new address to the SAGE office address above to ensure proper identification. Please specify name of journal.

Social Cleavages and the Number of Parties

How the Measures You Choose Affect the Answers You Get

Heather Stoll

University of California, Santa Barbara

This article explores the issues surrounding the operationalization and measurement of social cleavages in a large- N context from the perspective of the party systems literature. It first highlights the difficulty of measuring the concept, which it argues has impeded hypothesis testing, and suggests strategies for improving upon existing measures. It then demonstrates that the conclusions we draw about well-known hypotheses linking social cleavages to the number of electoral parties in legislative elections depend upon the measure we employ. Hence, the article makes the case that how we measure social cleavages matters.

Keywords: *social cleavages; social heterogeneity; electoral systems; party systems; measurement*

Which factors explain the cross-national and cross-temporal variance in the number of electoral parties contesting legislative elections? In the past two decades, scholars have attempted to integrate the institutional and sociological approaches to this research question, which had previously "coexisted uneasily" (Amorim Neto & Cox, 1997, p. 150) since the publication of Duverger's (1963) classic text. Contemporary theories such

Author's Note: Early versions of this article were presented as parts of larger papers at the 2004 Annual Meeting of the American Political Science Association, Chicago, September 2-5, and at the 2004 Society for Comparative Research (SCR) Graduate Student Retreat, University of California, San Diego, La Jolla, May 14-15. An even earlier version is found as parts of chapters 2 and 3 of my dissertation. I thank David Laitin; conference participants such as Gary Cox; colleagues at the University of California, Santa Barbara (particularly Chris Parker, Rob Rauchhaus, Lorelei Moosbrugger, and Garrett Glasgow); Kevin Deegan-Krause; and several anonymous referees for helpful comments on earlier drafts. Any remaining errors of course remain my own. The replication data set and supplemental materials are available from my Web site.

as Cox's (1997) accordingly ground their explanations in the institutional variable of the electoral system and the sociological variable of social cleavages. Empirically, a large body of quantitative research has attempted to test hypotheses derived from this theoretical literature. Most studies argue that the empirical evidence supports the general proposition that the more socially heterogeneous a country is, the more electoral parties it will have, either controlling for or conditional upon the strength of the electoral system.

However, hypothesis testing in previous studies has been hampered by a variety of methodological issues, which scholars have recently begun to address. For example, both Brambor, Clark, and Golder (2005) and Golder and Clark (2006) dealt with the proper testing of conditional hypotheses, and Golder and Clark with the consequences of pooling across time as well as space. Moreover, measures of political institutional variables have gradually been refined. What has thus far received relatively little attention is the operationalization and measurement of the social cleavages variable. Although this comparative neglect is understandable in light of the other advances that have been made, it nevertheless raises two important questions. Are our measures of this important variable valid representations of the concept ultimately of interest? And what are the consequences for our hypotheses if they are not?

To answer these questions, this article reviews the empirical literature that explores the relationship of social cleavages to the number of electoral parties, with a focus on recent (post-1980) quantitative studies. Its first contribution is to identify the issues that confront us when operationalizing and measuring social cleavages in a large-*N* context. This exercise reveals the difficulty of developing valid measures of the concept in such a setting, which the article argues has effectively precluded meaningful tests of our hypotheses. On a more positive note, it also suggests strategies for improvement. The article then demonstrates that even when we limit consideration to existing measures instead of dreaming about what might ideally be, these issues have consequences for our hypotheses, its other major contribution. Specifically, although the predominant measure in use supports the arguments advanced by the literature, alternative measures do not. In other words, this article both argues and demonstrates that how we measure social cleavages matters. We accordingly hope that it will spark scholarly interest in this variable, the poor relation of political institutions.

The article initially summarizes the quantitative empirical literature. It then discusses the issues surrounding large-*N* measures of social cleavages. As a follow-up, it shows that conclusions about the literature's hypotheses

are sensitive to the measure employed. The article concludes by sketching the implications of the inquiry for future research.

Theory, Hypotheses, and the Empirical Literature

A large literature in comparative politics has sought explanations for the variance in the number of electoral parties that compete in legislative elections.¹ Scholars have primarily focused upon two political institutional variables as explanatory factors: the electoral system and the regime type. The independent variable that the literature has most often called "social cleavages" and sometimes "social heterogeneity" enters as a counterpoint to these institutional variables. Social cleavages are usually defined as large-scale divisions within a society that are *exogenous* to the political system (see, e.g., Ordeshook & Shvetsova, 1994, p. 107). Many scholars from Duverger (1963) onwards have hypothesized that the more social cleavages a country has, which the article will henceforth call its *latent diversity*, the more political parties that country will have, *ceteris paribus*.² Specifically, the latent diversity of a country determines its "natural number" of parties (Cox, 1997, p. 140): the number that would compete if there were no institutional incentives for coordination. However, more recent studies have refined this simple additive story, instead painting the relationship as conditional upon institutional, and specifically electoral system, incentives (e.g., Amorim Neto & Cox, 1997; Cox, 1997; Golder & Clark, 2006; Ordeshook & Shvetsova, 1994). In other words, an increase in latent diversity is hypothesized to only lead to an increase in the number of electoral parties if the electoral system is sufficiently permissive.

Over the past 25 years, prominent quantitative studies have attempted to test these theoretical claims (e.g., Amorim Neto & Cox, 1997; Chhibber & Kollman, 2004; Cox, 1997; Filippov, Ordeshook, & Shvetsova, 1999; Golder, 2006; Golder & Clark, 2006; Jones, 1997; Ordeshook & Shvetsova, 1994; Powell, 1982), focusing on four testable hypotheses:

Hypothesis 1: Latent diversity is statistically related to the number of electoral parties. In other words, a model including latent diversity is superior to a purely institutional specification.

Hypothesis 2: The effect of an increase in latent diversity on the number of electoral parties is conditional upon the strength of the electoral system. That is, an interactive model specification is superior to an additive one.

Hypothesis 3: Increasing latent diversity will lead to an increase in the number of electoral parties, although only for sufficiently permissive

electoral systems when an interactive model specification is employed. In other words, the marginal effect of latent diversity is either unconditionally positive and significant, or positive and significant for sufficiently permissive electoral systems (and insignificant otherwise).

Hypothesis 4: The conditional marginal effect of latent diversity on the number of electoral parties increases as electoral system permissiveness increases. That is, when employing an interactive model specification, the coefficient on the interaction term between latent diversity and electoral system permissiveness is positive.

The significance of these hypotheses for the theoretical literature varies. Hypothesis 1 is the most basic but least interesting of the four because it merely tests for the existence of a statistically significant relationship between latent diversity and the number of electoral parties. Hypothesis 2, first introduced by Ordeshook and Shvetsova (1994), asks a simple question about the nature of this relationship: whether it is conditional upon electoral system permissiveness. Hypothesis 3 is implied but not formally part of many scholars' theoretical accounts; it is fleshed out by Golder and Clark (2006), who also introduce the more specific Hypothesis 4. The latter two hypotheses capture important additional, yet theoretically distinct, claims about the nature of the relationship, such as its positive direction.

Testing these hypotheses of course requires the development of measures of latent diversity. An early convergence upon two closely related operationalizations, either the effective number of ethnic groups or ethnic fractionalization, has for the most part not been challenged. However, different measurements of ethnic groups have sometimes been employed, such as Golder and Clark's (2006) use of data compiled by Fearon (2003) instead of the more conventional Soviet geographers' data from the 1960s. Only Powell (1982), Ordeshook and Shvetsova (1994), and Jones (1997) have reported the results from employing different operationalizations, two of which (the effective number of linguistic groups in Ordeshook and Shvetsova and the effective number of racial groups in Jones) are nevertheless closely related to the dominant approach. The more novel alternatives are the effective number of religious groups in Ordeshook and Shvetsova; an index based on the Catholic proportion of the population in Powell; and an index based on the agricultural proportion of the population in Powell, the latter two of which were employed simultaneously with ethnic fractionalization.

Most scholars have used these measures to argue that the evidence supports the hypotheses. However, similar conclusions have not always been drawn, which can sometimes be directly attributed to the measure of latent diversity employed. For example, Ordeshook and Shvetsova (1994) found

little support for the unconditional version of Hypothesis 3 when using the effective number of religious groups to operationalize latent diversity, a finding that is buried in a footnote.³ More fundamentally, we might ask about the validity of these measures: Have they allowed for meaningful tests of the hypotheses?

Operationalization and Measurement Issues

The two processes of operationalization and measurement are obviously critical steps in a social science that seeks to "make descriptive and causal inferences about the world" (King, Keohane, & Verba, 1994, p. 7). In his classic text, Blalock (1982) defines operationalization as the "theoretical process by which we move from ideas or constructs to suggesting appropriate research operations," whereas measurement is the "general process through which numbers are assigned to objects" (p. 11). Although we plow no new ground by claiming that substantial error can be introduced by either operationalizing or measuring without explicit reference to theoretical structure (King et al., 1994, p. 153), this section of the article offers what we believe is the most sustained argument to date that existing measures of latent diversity have not served our theories well. To be clear, this article is not the first to reflect upon the operationalization and measurement of this variable. Jones (1997, 1999, 2004) in particular has repeatedly voiced concerns about the validity of the dominant operationalization. There has been more but still only limited discussion of the measurement of the dominant operationalization (see, e.g., Golder, 2006, p. 37). Yet, surely latent diversity's time has come. With this belief, we attempt to shine a critical light upon our measures of this important variable. First addressed is the issue of variable operationalization, which is followed by a discussion of measurement itself.

Operationalization

Several issues, which are often closely tied to our definition of the concept, confront us when operationalizing latent diversity.

Diversity in Social Cleavages or Groups?

The first of these issues is where the diversity in latent diversity resides. Is our interest in the diversity of the *groups* generated by the country's set of social cleavages or in the diversity of the *social cleavages* themselves?

Depending upon which conceptualization we adopt, we may paint a very different picture of the latent diversity of a country. A simple example will, it is hoped, illustrate the point. Take Country A, in which there is one social cleavage (say, ethnicity) and four groups along the cleavage. Now take Country B, in which there are two social cleavages (say, ethnicity and religion) but only three groups because of the way in which the two cleavages jointly partition individuals into groups. Which country is more diverse? By conceptualizing diversity in terms of cleavages, it is Country B, but by conceptualizing diversity in terms of groups, it is Country A.

The empirical literature has vacillated between these two conceptualizations, although it has always employed a group-based operationalization, with the possible exception of Powell (1982). For example, contrast Amorim Neto and Cox's (1997) conceptual focus on cleavages (in spite of their group-based operationalization) to Ordeshook and Shvetsova's (1994) conceptual focus upon groups. The more theoretically compelling story linking diversity to the number of electoral parties can be told using the group-based conceptualization, as Golder and Clark's (2006, p. 680) "rehabilitation" of Duverger's (1963) theory makes clear. This suggests that future work should continue moving in this direction. However, measuring the diversity in groups across many countries and time periods is not straightforward, a topic to which we will return. A cleavage-based operationalization, potentially easier to measure, may accordingly be an expedient choice. This is an avenue ripe for exploration.

One Social Cleavage or Many?

The second issue that we must confront is how many social cleavages produce the diversity of interest to us. In other words, do we conceptualize and operationalize diversity with respect to a single cleavage or with respect to multiple cleavages? It is not surprising that this choice has important ramifications for our measure. To see this, consider the case of Country B introduced above. Say that it has two ethnic and three religious groups, which together partition the citizenry into three ethnoreligious groups. If we choose to conceptualize and operationalize diversity in terms of a single cleavage, say ethnicity, then the country has two groups. However, if we instead consider both cleavages, it then has three groups and accordingly appears more diverse.

The empirical literature has always opted for a single-cleavage operationalization, with the one exception of Powell (1982); however, at the same time, it has arguably adopted the multiple-cleavage conceptualization.

Given this overarching interest in diversity writ large, not narrowly defined, a single-cleavage operationalization does not truly test the hypotheses. To elaborate, bedeviling the latter are two facts: First, different cleavages are politicized in different countries and time periods; second, several cleavages are often simultaneously salient. For example, although ethnicity is politicized in some countries, such as Belgium, other cleavages are salient elsewhere, such as the largely orthogonal socioeconomic and religious cleavages that pillarized Dutch society through the early postwar years (Lijphart, 1968). In fact, we can identify *seven* oft-politicized cleavages: the socioeconomic, cultural-ethnic, religious, foreign policy, urban-rural, post-materialist, and regime support (see, e.g., Lijphart, 1999). Accordingly, we should ideally operationalize diversity in terms of *multiple* cleavages such as these seven, particularly if our conceptualization is cleavage based.⁴ Fearon (2003) optimistically opines that the task of measuring groups with respect to multiple cleavages cross-nationally, which includes taking into account how the various cleavages relate, is "relatively straightforward" (p. 215). We are aware of no studies that have collected such data, however, which suggests that it is actually an exceedingly difficult task. Nevertheless, developing such measures should be a priority in light of their importance to any empirical evaluation of the literature's hypotheses.

Three strategies exist for coping in the interim. The first, from another literature, is to combine the indicators of group diversity for each separate cleavage, forming an index of overall group diversity. One example is Annett (2001), who averages a measure of ethno-linguistic group diversity and a measure of religious group diversity. This strategy obviously does not account for how the cleavages relate and thus introduces measurement error.⁵ However, the exact nature of the measurement error produced will depend upon the method of index construction chosen. For example, Annett's measure usually underestimates the true latent diversity.⁶ A second, related strategy is to simultaneously make use of separate multiple indicators. The only example in the literature is Powell (1982), as noted above. We note in passing that the American politics literature has employed both of these strategies (e.g., Aistrup, 2004; Sullivan, 1973). The third strategy, taken by the remainder of the empirical literature, is the previously mentioned single-cleavage operationalization, such as ethnic diversity: that is, to take as our indicator a part of the theoretical construct. This strategy will also generate measurement error, arguably of an even more serious type.⁷ For example, Jacobson and Lulu (1974) found that the single-indicator strategy usually performed the worst of the three. Accordingly, it is instead advisable to adopt one of the first two interim strategies.

If One Social Cleavage, Which?

If we decide upon a single-cleavage operationalization (and hence a group-based conceptualization) of diversity, we face a third issue, one that is more operational than conceptual in nature: Around which of the seven social cleavages mentioned above do we base our measure? Again, take the example of Country B. If we choose to operationalize diversity in terms of ethnicity, then this country has two groups; however, if we instead operationalize diversity in terms of religion, it then has three groups. Country B accordingly appears more diverse when operationalizing diversity in terms of religion than ethnicity.

As noted earlier, the empirical literature has always operationalized diversity with respect to ethnicity or its close cousins, with the two exceptions of Ordeshook and Shvetsova (1994) and Powell (1982). But how can this be justified, and are there valid alternative choices? Two theoretically motivated strategies present themselves for picking one cleavage as an overall proxy for the cleavage structure. The first strategy, originally advocated by Rose and Urwin (1969, p. 9), is to choose the cleavage that has been politicized for the longest period of time in the largest number of countries: in other words, the cleavage that has had the greatest chance of shaping the party system. Of the seven cleavages identified above, the primary contender in the advanced industrial world, which until recently has been at the core of the literature, is socioeconomic (class).⁸ The second strategy is to choose the cleavage that has most often led to differentiation in the politicized cleavage structure. As long as this cleavage crosscuts at least one existing group and the electoral system is sufficiently permissive, its politicization might be expected to increase the number of electoral parties. Caramani's (2004) study makes an equal case for the religious and cultural-ethnic cleavages as the most differentiating cleavage, although we might also recognize the foreign policy cleavage as a contender. Hence, either way, there are good reasons to broaden our operational focus beyond the cultural-ethnic cleavage. Moreover, this discussion raises an as-yet-unexplored theoretical issue that also recommends a broader empirical focus: that different types of diversity might have different effects upon the party system.

Diversity as "How Many?"

Regardless of how we weigh in on the above matters, a fourth issue deserves consideration. Thus far, we have followed the empirical literature in conceptualizing diversity as the *number* of either groups or cleavages, with larger numbers being equated with greater diversity. In other words, we have counted *how many* groups or cleavages a country possesses. But

there are other ways of conceptualizing diversity, at least when we are dealing with a group-based conceptualization, that might better match the measure to the mechanism, to paraphrase Posner (2004, p. 852). There are also various operational indicators of these concepts.

One such alternative conceptualization of diversity is *polarization*: that is, the intensity of the differences between groups (Posner, 2004, pp. 851-852), with intense or large differences equated with greater diversity, perhaps even regardless of how many groups there are. Other literatures offer some guidance as to how we might operationalize polarization. An example is Montalvo and Reynal-Querol (2001), who derive an index of polarization, P , that takes its maximum value when there are two groups of equal size.⁹ This operationalization is appropriate if the likelihood of conflict (and hence politicization) is hypothesized to be the greatest when groups have similar sizes. It is thus not tapping what we more conventionally think of as polarization: the above intensity of differences. The only existing large- N operationalization along these lines, however, is Fearon's (2003) index of cultural fractionalization, C .¹⁰ We might view this operationalization as integrating the "how many" and polarization concepts in that it weights groups by *both* their sizes and differences when counting.

Moreover, the how many concept has usually been operationalized for groups using the size-weighted number of items to be counted, and specifically using the effective number, N .¹¹ However, other methods of counting may have both practical and theoretical advantages. One example is the index of fractionalization, F , which carries equivalent information to N .¹² Some comparativists have preferred N to F because of its more intuitive interpretation, but the price is usually a positively skewed empirical distribution that is problematic for many statistical models (Fearon, 2003, p. 209). On these grounds, both F and the log of N might be better choices than N . Another example is Molinar's (1991) modification of N , NP , which corrects for N 's tendency to overcount the largest item. There are also operationalizations that make no reference to size, such as the raw (unweighted) number, although this particular approach has many obvious drawbacks. Other possibilities from other literatures include various indicators of group dominance such as a dummy variable for a group that composes between 45% and 90% of a country's population (see, e.g., Fearon & Laitin, 2003); a dummy variable for the presence of a sizable minority group; and Lieberman's A_w , which was popularized in the American politics literature by Sullivan (1973) and has a similar interpretation to F . The causal mechanism underlying both of the former two operationalizations links groups of particular sizes to the politicization of the cleavage, much as Montalvo and Reynal-Querol's (2001) index of polarization does.

Diversity in the Aggregate or in the Districts?

Another issue that we face, also regardless of the choices that we make with respect to the previous issues, is the level at which latent diversity should be measured. Are we interested in conceptualizing and operationalizing diversity at the *aggregate* (national) level, or is our interest in diversity at the *district* level? Again, the approach that we choose has clear ramifications for our measure. For example, consider a country with two comparably sized groups. Using the aggregate approach, this country has two groups. If each electoral district's population mirrors that of the country as a whole, then the typical district will also have two groups, and the two approaches will paint identical pictures of the country's diversity. However, these aggregate data are also consistent with a very different district-level situation. For example, if there are only two equally populous electoral districts, the population of the first district might belong exclusively to the first group, and the population of the second district might belong exclusively to the second group. In this case, the typical district has one group, and the district and aggregate pictures diverge.

All existing empirical research adopts the aggregate conceptualization and operationalization of diversity, with the exception of Jones (1997). However, to again paraphrase Posner (2004), we match the measure to the mechanism of current theorizing by employing the district-level approach, because electoral system incentives—with which latent diversity is hypothesized to interact—operate at the district level. Ideally, we should operationalize latent diversity as the mean or median latent diversity in the districts, analogous to the current operationalization of electoral system permissiveness (e.g., Amorim Neto & Cox, 1997, pp. 156-158). A broader advantage of the district approach is that it may allow us to disentangle the effect that cross-district diversity has upon cross-district coordination, that is, upon party aggregation (Chhibber & Kollman, 2004), from the effect that intradistrict diversity has upon intradistrict coordination—the two processes that together produce the national party system. Unfortunately, but not surprisingly, no comparativist has yet taken on the daunting task of measuring diversity at the district level within a single country, though, let alone across many countries and time periods.¹³ Without district-level data, we obviously cannot measure district-level operationalizations. In the absence of such measures, we may obtain better tests of our hypotheses by shifting the unit of analysis from legislative to presidential elections, where district- and aggregate-level measures are usually commensurate.

Exogenous or Endogenous Diversity?

Finally, we turn to the least tractable—but arguably the most critical—issue that we face. Recall that the “latent” in latent diversity indicates that the diversity of interest is exogenous to the political process. The problem with instead conceptualizing diversity as *politicized*, following Jones (1999, 2004) in a related literature and Mozaffar, Scarritt, and Galaich (2003) in this literature, is that politicized diversity is endogenous by definition. Specifically, at minimum, one can plausibly argue that political institutions influence the politicization of groups along a cleavage (see, e.g., Posner, 2005). Accordingly, we have a simultaneous equation model with two structural equations, one of which has politicized diversity as the dependent variable and the other of which has the electoral party system as the dependent variable. Using ordinary least squares (OLS) to estimate such a model, and specifically an equation with the electoral party system on the left-hand side and politicized diversity on the right-hand side, is likely to produce biased and inconsistent coefficient estimates.¹⁴ Conversely, OLS can be used to estimate the reduced form equations, which include a model with the electoral party system on the left-hand side and latent diversity on the right-hand side.

These practical considerations almost certainly explain why the empirical literature has employed an exogenous conceptualization and operationalization of diversity, with the few exceptions (mostly in related literatures) noted above. The problem, however, is that constructivists have imploded the very concept of exogenous or “eternal” groups (Laitin & Posner, 2001, p. 17). It is consequently difficult to develop a valid operationalization of such a variable. For example, we cannot even define the population of all potential latent cleavages, let alone the population of the potential groups generated by such cleavages (see Fearon, 2003, regarding ethnicity). Laitin and Posner's (p. 15) solution is to focus on the groups that are actually “doing the competing” relevant to the outcome of interest. This approach allows us to match the measure to the mechanism, but at the same time undercuts exogeneity. Another problem is that the groups actually measured will often be politicized, not latent. For example, religious groups can only be measured independently of the individual's identification with the group, and hence of its politicization, with great difficulty. Even cleavages such as the ethnic that are defined by relatively more objective criteria such as skin color and language contain an element of subjectivity. Hence, although it may be easier to construct measures of some types of latent diversity, validity will be difficult to attain across the board. Fearon suggests one very advisable solution to the problem: Use past (lagged) measurements. Another is to employ a

cleavage- instead of a group-based conceptualization and operationalization of diversity, because the constructivist literature arguably suggests that the groups with which individuals identify are more likely to be endogenous to political institutions than are the cleavages themselves.

Measurement

We now turn from operationalization to measurement. It is initially worth reiterating that many of the operationalizations of latent diversity discussed above unfortunately remain unmeasured, including those that would best test the literature's hypotheses. But what measurements are available to us, and how do they compare? All operationalizations currently in use require us to identify the latent groups along individual cleavages and at minimum to also compile population share data for these groups. Measurements of this sort have only been gathered for two of the seven potential latent cleavages identified in the prior section: the religious and the cultural-ethnic cleavages. The strengths and weaknesses of the existing sets of measurements are assessed below. An important caveat about *all* such measurements, though, is that there are usually multiple ways to specify the set of groups in a country. Accordingly, we must take to heart Fearon's (2003) admonition "[to] not, in effect, choose the coding that best supports our theory, after the fact" (p. 198).

Measuring Cultural-Ethnic Groups

The earliest and most well-known set of measurements of cultural-ethnic groups is an atlas of ethnolinguistic groups and population figures compiled by Soviet geographers in the early 1960s (Taylor & Hudson, 1972). The popular ethnolinguistic fractionalization index measure (ELF) results when these data are summarized using the index of fractionalization, *F*. Ordeshook and Shvetsova (1994) used an updated version of the atlas from the mid-1980s to add a time-series dimension to the measure. Roeder (2001) has helpfully created a data set based on the Soviet atlases that contains several versions of ELF at these two time points.

Recently, four alternative sets of measurements have been developed by political scientists and economists working in other literatures.¹⁵ First, Annett (2001) relied on Barrett (1982) to construct a list of the most detailed possible breakdown of ethnolinguistic groups and their population shares for the early 1980s. Second, employing a broader definition of ethnicity, Fearon (2003) developed a list of the ethnic groups comprising at least 1% of a country's population. He gathered data on the population

shares of these groups for the 1990s using a variety of sources. Note that he also collected data on their linguistic similarity, a proxy for their cultural similarity, in order to create measurements of his index of cultural diversity, *C*. Third, Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003) independently developed a list of ethnic groups and their population shares for the early 2000s, also using a variety of sources. They too employed a broader definition of ethnicity that encompassed racial as well as linguistic characteristics. However, they did not exclude groups comprising less than 1% of a country's population and aimed for the most disaggregated listing possible. Fourth, Alesina et al. in addition constructed a data set of linguistic groups and their population shares for the early 2000s, basing their data exclusively on the *Encyclopedia Britannica*.¹⁶ In so doing, they returned to the more minimalist, language-centered definition of ethnicity.

Despite the different sources, time periods, definitions of ethnicity, and coders involved in the creation of these sets of measurements, reasonable correlations are reported between the indexes of ethnic fractionalization derived from them (Fearon, 2003, p. 210; Posner, 2004, p. 857). Nevertheless, there are reasons to prefer some of the measurements to others. One very prominent such reason is the exogeneity of the measurements. In this light, a drawback to Fearon's set of measurements is that his data for Africa as well as his overall approach are based on that of Scarritt and Mozaffar (1999). We should therefore be concerned that he is measuring politicized, not latent, diversity. Conversely, as discussed above, an advantage of the Soviet geographers' data is precisely that it is "more than 40 years out of date" (Posner, 2004, p. 850).¹⁷ Both Alesina et al.'s (2003) and Annett's (2001) measurements seem to fall somewhere in between the prior two with respect to exogeneity, although the latter may be slightly preferable when accounting for contemporary outcomes due to their greater age. Moreover, favoring both the Soviet geographers' measurements of ethnic groups and Alesina et al.'s measurements of linguistic groups is the relatively objective nature of an individual's primary linguistic affiliation. In the short to medium run, an individual's mother tongue is for all intents and purposes something with which he or she is born. Measurements of linguistic groups should accordingly be more exogenous to political institutions than measurements of ethnic groups writ large. To be sure, we pay a price for this exogeneity because the cultural-ethnic cleavage is not fully reducible to language: One need only think of African Americans in the United States. A final consideration is reliance upon multiple sources, which should improve accuracy. This criterion favors the data sets compiled by the Soviet geographers, Alesina et al., and Fearon.

Measuring Religious Groups

Turning to the religious cleavage, Alesina et al. (2003), Fearon and Laitin (2003), and Annett (2001) have all recently constructed cross-national lists of religious groups and population figures.¹⁸ Annett again relied upon Barrett (1982) to collect data for the early 1980s. He defined the population of religious groups to be 15 major world religions or denominations, of which only Christianity is disaggregated. In other words, only the population shares of the religious groups, not the groups themselves, were allowed to vary from country to country. In contrast, Alesina et al. and Fearon and Laitin allowed the groups to vary across countries as appropriate, as with all lists of ethnic groups. However, although their lists almost always disaggregated Christianity, other religions are less frequently disaggregated. Their population share data is from the 1990s and again is based upon a variety of sources.

The pairwise correlations between the fractionalization indices derived from the three sets of religious group measurements are substantial; however, those between the religious and ethnic fractionalization indices are low, as one might expect (Alesina et al., 2003, p. 162). A source of concern with all of these sets of measurements is the bias toward disaggregating Christianity relative to other religions. For example, none disaggregate Judaism, the predominant religion of Israel, despite always disaggregating Christianity in predominantly Christian countries. For an analysis confined primarily to the advanced industrial world, this bias does not have much bite: Only two non-Christian cases—Japan and Israel—are conventionally included in such analyses. However, for analyses moving beyond the advanced industrial world into the non-Christian regions of Asia, Africa, and the Middle East, it is more problematic. Because Alesina et al. and Fearon and Laitin (2003) allow for at least some disaggregation of non-Christian religions, unlike Annett (2001), their measurements seem preferable. Furthermore, an argument in favor of choosing an ethnic over a religious operationalization is that measurements of ethnic groups are not plagued by this problem.

Other considerations follow. First, not surprising, is the exogeneity of the measurements. Ordeshook and Shvetsova (1994) criticize measurements of latent religious diversity because “religion is subject to the inherent ambiguity of how people choose to report weak or nonexistent affiliations” (p. 114). The broader critique is that *all* religious affiliations are self-chosen and self-reported. However, the same is true of ethnicity (see, e.g., Fearon, 2003, pp. 197-199, 203), particularly because ethnic group membership is usually obtained from surveys, such as the U.S. census. As

argued above, though, because ethnicity is frequently grounded in the relatively objective characteristic of language (and to a lesser degree, race), measurements of religious groups may be less exogenous than measurements of ethnic groups. Second, Annett’s (2001) relatively rigid approach, which allows no variation in group aggregation, seems less preferred than the more nuanced approaches of the others.

Do Measures Matter? A Sensitivity Analysis

These issues surrounding the operationalization and measurement of latent diversity beg the question of whether our conclusions are in fact sensitive to the measure of latent diversity that we employ. Our earlier review of the empirical literature suggested that they might be. In this section of the article, we investigate. To do so, we replicate one of the most well-known quantitative empirical studies in the literature, that of Amorim Neto and Cox (1997), itself replicated in Cox (1997). We stress that our goal is neither to evaluate Amorim Neto and Cox’s study on its own merits, nor to draw general conclusions about the causal relationships between variables on the basis of the analyses performed here. Rather, our goal is merely to assess the sensitivity of conclusions about the literature’s four key testable hypotheses to some of the measures of latent diversity currently available to us, however flawed those measures might ultimately be.

We focus on Model 4 from Table 2 in Amorim Neto and Cox (1997, p. 164), which provides the most appropriate test of the authors’ hypotheses about the relationship between what we have called latent diversity, electoral system permissiveness, and the number of electoral parties.¹⁹ This model is as follows (notation generally as in the original):

$$\text{ENEP}_i = \beta_0 + \beta_1 \text{LML}_i + \beta_2 \text{LATDIV}_i + \beta_3 \text{LML}_i \times \text{LATDIV}_i + \beta_4 \text{UPPER}_i + \beta_5 \text{PROXIMITY}_i + \beta_6 \text{PROXIMITY}_i \times \text{ENPRES}_i + \epsilon_i \quad (1)$$

In this equation, ENEP is the effective number of electoral parties, LML is the logged median district magnitude, LATDIV is the latent diversity (generalizing from the original, which used ENETH to represent Amorim Neto and Cox’s measure of latent diversity), UPPER is the proportion of seats allocated in an upper tier, PROXIMITY is the time to the closest presidential election, and ENPRES is the effective number of candidates in the presidential election. We estimate 16 versions of this model (labeled Models 1-16) via OLS, as in the original. The first of these models is a direct replication of Amorim

Neto and Cox. In other words, it uses their data, which consist of measures of the above variables for a cross-section of 51 country elections in the mid-1980s, as published in their appendix.²⁰ The measure of latent diversity employed in this replication, as in the original, is the effective number of ethnic groups based on the Soviet geographers' measurements from the 1960s, that is, the effective number transformation of ELF.

The 15 additional models vary the measure of latent diversity, but otherwise continue to use Amorim Neto and Cox's (1997) data. These measures of latent diversity encompass the most prominent existing operationalizations and all existing, publicly available measurements in various combinations. To elaborate, the 2nd model employs the effective number of ethnic groups based on Annett's (2001) measurements. The 3rd through 7th models use the same operationalization, the effective number of ethnic groups, but different measurements: Alesina et al.'s (2003) data on ethnic groups; the Soviet geographers' measurements of ethnic groups from the 1960s, but with groupings defined by Roeder (2001);²¹ the updated (1980s) version of the latter; Alesina et al.'s data on linguistic groups;²² and Fearon's (2003) measurements of ethnic groups. The 8th and 9th models use Fearon's measurements to instead calculate different operationalizations, the ethnic fractionalization and ethnic polarization as defined by Montalvo and Reynal-Querol (2001), respectively. The 10th uses Fearon's index of cultural fractionalization. Models 11 through 13 employ the effective number of religious groups based on data from Annett, Fearon and Laitin (2003), and Alesina et al., respectively. The 14th uses religious fractionalization based on data from Alesina et al., whereas the 15th uses the same data to calculate the religious polarization, again as defined by Montalvo and Reynal-Querol. Finally, the 16th model uses Annett's index, the average of his values of religious and ethnic fractionalization. We employed simple listwise deletion as our missing data strategy throughout, which left between 39 and 51 observations per model.²³

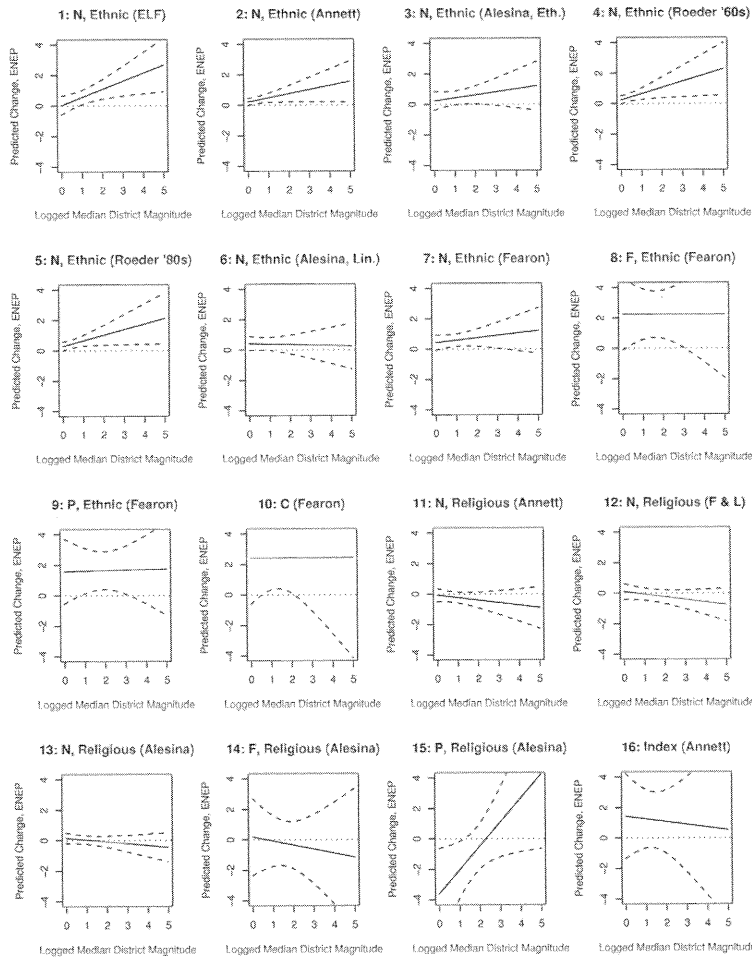
Perhaps it is not surprising that the empirical results do in fact vary, often dramatically. Table 1 contains the relevant estimated coefficients and standard errors for the 16 models, the latter of which are shown in parentheses. More germane to Hypothesis 3 is the estimated marginal effect of latent diversity, which is shown in Figure 1 for each of the 16 models. This is the partial derivative of Equation 1 with respect to latent diversity (see Brambor et al., 2005), evaluated over the range of the logged median district magnitude. Confidence intervals (95%) for the marginal effects are shown as dashed lines. Table 2 draws upon the above to summarize the support for Hypotheses 1-4. For each measure and hypothesis, a "Yes" denotes that the

Table 1
Selected Estimated Coefficients (and Standard Errors) for the 16 Replication Models (Models 1-16)

Model	1	2	3	4	5	6	7	8
Measure	<i>N</i> , Ethnic (ELF)	<i>N</i> , Ethnic (Annett)	<i>N</i> , Ethnic (Alesina, Ethnic)	<i>N</i> , Ethnic (Roeder, 1960s)	<i>N</i> , Ethnic (Roeder, 1980s)	<i>N</i> , Ethnic (Alesina, Linguistic)	<i>N</i> , Ethnic (Fearon)	<i>F</i> , Ethnic (Fearon)
LATDIV	0.018 (0.30)	0.21 (0.12)	0.21 (0.33)	0.23 (0.14)	0.29 (0.15)	0.42 (0.25)	0.43 (0.27)	2.2 (1.3)
LML × LATDIV	0.53 (0.21)	0.27 (0.14)	0.20 (0.21)	0.41 (0.19)	0.37 (0.18)	-0.033 (0.18)	0.16 (0.19)	-0.0085 (0.62)
<i>n</i>	51	50	51	43	44	51	39	39
Model	9	10	11	12	13	14	15	16
Measure	<i>P</i> , Ethnic (Fearon)	<i>C</i> (Fearon)	<i>N</i> , Religious (Annett)	<i>N</i> , Religious (Fearon & Laitin)	<i>N</i> , Religious (Alesina)	<i>F</i> , Religious (Alesina)	<i>P</i> , Religious (Alesina)	Index (Annett)
LATDIV	1.6 (1.2)	2.4 (1.7)	-0.069 (0.23)	0.11 (0.30)	0.15 (0.19)	0.18 (1.4)	-3.6 (1.6)	1.4 (1.5)
LML × LATDIV	0.029 (0.50)	0.0047 (0.96)	-0.16 (0.17)	-0.17 (0.16)	-0.12 (0.12)	-0.27 (0.71)	1.6 (0.76)	-0.18 (0.83)
<i>n</i>	39	39	49	39	51	51	51	49

Source: Annett (2001); Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003); Roeder (2001); Fearon (2003); Fearon and Laitin (2003). Note: *N* = effective number; ELF = ethnolinguistic fractionalization index; *F* = index of fractionalization; *P* = index of polarization; *C* = index of cultural fractionalization; LATDIV = latent diversity; LML = logged median district magnitude. All entries rounded to two significant digits. The dependent variable is the effective number of electoral parties. Each model employs a different measure of latent diversity.

Figure 1
Estimated Marginal Effects of Latent Diversity on the Effective Number of Electoral Parties (ENEP) for the 16 Replication Models (Models 1-16), All Calculated Over the Range of Logged Median District Magnitude



Source: Annett (2001); Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003); Roeder (2001); Fearon (2003); Fearon and Laitin (2003).

Note: ELF = ethnolinguistic fractionalization index measure; Eth. = ethnic; Lin. = linguistic; F & L = Fearon and Laitin. Each model employs a different measure of latent diversity.

Table 2
Summary of Conclusions Drawn About Hypotheses 1-4 From the 16 Replication Models (Models 1-16)

Model	Measure	Hypothesis 1	Hypothesis 2	Hypothesis 3	Hypothesis 4
1	N, ethnic (ELF)	Yes ($p = .0064$)	Yes ($p = .014$)		Yes
2	N, ethnic (Annett)	Yes ($p = .0082$)	No ($p = .060$)		Yes
3	N, ethnic (Alesina, Eth.)	No ($p = .15$)	No ($p = .35$)	Maybe (insignificant for medium to large M)	Yes
4	N, ethnic (Roeder, 1960s)	Yes ($p = .0071$)	Yes ($p = .033$)		Yes
5	N, ethnic (Roeder, 1980s)	Yes ($p = .0036$)	Yes ($p = .047$)		Yes
6	N, ethnic (Alesina, linguistic)	No ($p = .22$)	No ($p = .85$)	No (insignificant)	No
7	N, ethnic (Fearon)	Yes ($p = .040$)	No ($p = .42$)	Maybe (insignificant for large M)	Yes
8	F, ethnic (Fearon)	Yes ($p = .039$)	No ($p = .98$)	Maybe (insignificant for large M)	No
9	P, ethnic (Fearon)	No ($p = .075$)	No ($p = .95$)	Maybe (insignificant for large M)	Yes
10	C (Fearon)	No ($p = .12$)	No ($p = .99$)	Maybe (insignificant for medium to large M)	Yes
11	N, religious (Annett)	No ($p = .37$)	No ($p = .34$)	No (insignificant; negative for medium to large M)	No
12	N, religious (Fearon & Laitin)	No ($p = .51$)	No ($p = .28$)	No (insignificant; negative for medium to large M)	No
13	N, religious (Alesina)	No ($p = .61$)	No ($p = .33$)	No (insignificant; negative for medium to large M)	No
14	F, religious (Alesina)	No ($p = .89$)	No ($p = .71$)	No (insignificant; negative for medium to large M)	No
15	P, religious (Alesina)	No ($p = .082$)	Yes ($p = .046$)	No (significant and negative for small to medium M; insignificant for large M)	Yes
16	Index (Annett)	No ($p = .48$)	No ($p = .83$)	No (insignificant)	No

Source: Annett (2001); Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003); Roeder (2001); Fearon (2003); Fearon and Laitin (2003). Note: N = effective number; ELF = ethnolinguistic fractionalization index; F = index of fractionalization; P = index of polarization; C = index of cultural fractionalization; M = magnitude. Each model employs a different measure of latent diversity.

model employing the measure supports the hypothesis, a “No” denotes that it does not support it, and a “Maybe” denotes that it partially supports it. A quick glance over this table reveals little consensus across measures, particularly with respect to the key Hypothesis 3.

First, just fewer than half of the measures provide support for the most basic of the four hypotheses, Hypothesis 1. This can be seen by examining column 3 of Table 2. For each of the 16 measures, it lists the *p* value for the *F* test that compares the model employing that measure to the simpler, purely institutional model. Conventional levels of statistical significance are obtained for only 6 of the measures (Models 1, 2, 4, 5, 7, and 8). Of these 6 measures, 5 are operationalized as the effective number of ethnic groups; the remaining supporting measure is operationalized as ethnic fractionalization. However, not all measures of the effective number of ethnic groups support the hypothesis. Furthermore, the support of some measurements of ethnic groups is contingent upon their operationalization.

Second, fewer than one third of the measures provide support for Hypothesis 2. For each of the 16 measures, column 4 of Table 2 lists the two-sided *p* value for the coefficient on the interaction term between latent diversity and logged median district magnitude. This interaction term is statistically significant at conventional levels, suggesting that the relationship between the effective number of electoral parties and latent diversity is conditional upon the district magnitude, for only 4 of the measures (Models 1, 4, 5, and 15). Three of these measures are again operationalized as the effective number of ethnic groups. However, as before, other measures of this operationalization fail to support the hypothesis, and the support of some measurements depends upon their operationalization.

Third, only one quarter of the measures (Models 1, 2, 4, and 5) provide unequivocal support for Hypothesis 3. This can be seen by examining Figure 1. Any contradictory evidence revealed by this figure is summarized in column 5 of Table 2. For example, because the confidence bands always encompass 0 for Model 11, this measure of latent diversity—the effective number of religious groups as measured by Annett (2001)—is never predicted to have a statistically significant effect on the effective number of electoral parties. Moreover, its effect is always predicted to be negative. An additional five measures (Models 3, 7, 8, 9, and 10) provide partial support for the hypothesis. For example, the effective number of ethnic groups calculated using Fearon’s (2003) measurements (Model 7) is only predicted to have a positive and statistically significant effect on the effective number of electoral parties for moderately small to medium district magnitudes. Hence, just greater than half of the measures provide at

least some support for the hypothesis, with support again coming most strongly (but not universally) from the effective number of ethnic groups operationalization.

Finally, just over half of the measures provide support for Hypothesis 4 (Models 1-5, 7, 9, 10, and 15), that the marginal effect of latent diversity is increasing in electoral system permissiveness. This can be seen by examining the sign of the coefficient on the interaction term between latent diversity and logged median district magnitude in Table 1, the results of which are summarized in column 6 of Table 2. As before, the strongest support comes from the measures that tap the cultural-ethnic cleavage, but again as before, this support is not universal.

Accordingly, the conclusions that we draw about the literature’s four testable hypotheses depend greatly upon how we measure latent diversity. The hypotheses fare best when we operationalize latent diversity as the effective number of ethnic groups and use the measurements compiled by the Soviet geographers. Together, these two choices yield the measure that many quantitative studies have employed. However, other measures produce results that are contradictory to at least some, if not all, of the hypotheses. Exploring the reasons for these contradictory results is beyond the scope of this article, although we note the hypothesis posed earlier that different types of diversity might have different effects. Here, we simply seek to draw attention to the difficulty of our current position: Which conclusions should we draw about the relationship of latent diversity to the number of parties? Only greater consideration of the issues of operationalization and measurement raised in the prior section will allow us to privilege some measures, and hence some conclusions, over others.

Quo Vadis?

This article has argued that measurement matters when we try to relate what it has called latent diversity, and what the literature has called either social cleavages or social heterogeneity, to the number of electoral parties in legislative elections. It is worth noting at the outset how far we have come. From early work that leaned toward either institutional or sociological determinism, recent work has moved to integrate the two, as exemplified by Cox’s (1997) magisterial study. However, although we have gradually accumulated evidence that supports our theories about political institutions such as the electoral system, this article has argued that the empirical book is still open with respect to social cleavages. Although

scholars are aware that existing measures of this variable leave much to be desired, the issues of its operationalization and measurement have received relatively little sustained attention, for any number of good reasons. This article's goal was to begin redressing the balance.

In this light, the article's first contribution was to take a rigorous look at these operationalization and measurement issues. For example, should diversity be conceptualized and operationalized at the aggregate or the district level? With respect to cleavages or groups? And with respect to one cleavage or many? We identified the key issues, illuminated the consequences of particular choices, and where possible, weighed in on the debate, in the process identifying the weaknesses of existing measures. Most fundamentally, though, the article's take-away message is that latent diversity is an exceedingly difficult concept to measure. As a result, the measures that would best allow us to test the literature's hypotheses, such as measures of diversity at the district level that take into account the relationships between multiple cleavages, unfortunately have yet to be developed. Looking solely at existing measures of social cleavages, the article's second contribution was to demonstrate that conclusions about the literature's four key testable hypotheses are sensitive to the measure employed. Replicating the well-known study of Amorim Neto and Cox (1997), we showed that some measures, such as the effective number of ethnic groups based upon ELF, support at least some of the hypotheses; however, other measures, such as the effective number of religious groups and the effective number of ethnic groups based upon Alesina et al.'s (2003) data, do not. In other words, to paraphrase Geddes (1990), the measures that you choose affect the answers that you get to this research question. Taken together, these points suggest that future research must place the operationalization and measurement of social cleavages on center stage, instead of allowing them to languish in the wings.

Hence, we arrive at the million-dollar question. How can future research move beyond the "measurement by fiat" Blalock (1982, p. 19) approach that has thus far shaped the empirical literature's response to the difficulty of measuring the social cleavages variable? Our options seem threefold. First, we can admit defeat and characterize the research question as "one that cannot be resolved by empirical means" (p. 19). This is not an attractive approach for obvious reasons. Second, a more constructive but still closely related strategy is to redefine the research question so as to retain only those hypotheses that are actually testable with the data at hand: for example, to forgo framing our hypotheses in terms of social cleavages and to instead frame them in terms of, say, ethnic diversity. This stance, with

which some scholars have flirted, is attractive for both its honesty and simplicity but still leaves much to be desired. We can do more. This brings us to the third option: returning to the drawing board. To an optimist, the second option is a half-full (instead of a half-empty) glass. Why not take the opportunity provided by this reflection upon our current limitations to make progress where it is most needed? Having gone as far as we have with political institutions, the obvious challenge for future research is to tackle social cleavages, which should entail greater attention to both theory and measures.

With respect to the latter, the focus of this article, we pointed out many immediately actionable ways to improve upon existing measures, such as tapping nonethnic sources of diversity, taking into account group polarization instead of simply the number of groups, and employing the lagged values of measures. In the longer term, exciting contemporary developments in comparative politics might enable the construction of the currently elusive ideal measures. One example is new cross-national surveys such as the Comparative Study of Electoral Systems, which might be used to generate measures inspired by, but moving beyond, those employed in the American politics literature (e.g., Sullivan, 1973). Another example is the recent explosion in district-level analyses (e.g., Caramani, 2004; Chhibber & Kollman, 2004), which encourages us to believe that the generation of district-level data, at least for a few countries, is not a pipe dream. There are also ways to better put even existing measures to work. As suggested, one such strategy is to work with presidential instead of legislative elections. Finally, we might consider supplementing quantitative studies with qualitative ones. For example, we might look for cases where groups have either been added to or removed from a polity due to, say, boundary changes and explore the impact that this has had on the party system. Qualitative research designs like these may ultimately enable the best tests of our hypotheses, given the difficulty of constructing large-*N* measures.

Notes

1. Our arguments easily extend from legislative to presidential elections, although the focus here is on the former. Examples of quantitative studies addressing the latter include Amorim Neto and Cox (1997), Cox (1997), Jones (1999, 2004), and Golder (2006). They also extend to the more distantly related literature exploring the relationship between diversity and party competition in the American states (e.g., Sullivan, 1973), if less straightforwardly.

2. We introduce the term *latent diversity* instead of utilizing existing terms for two reasons. First, we prefer *latent* to *social* because it draws attention to the focus on exogenous

(nonpoliticized or latent) cleavages. Second, we prefer the term *diversity* to *cleavages* because it signals what is really of interest: the diversity of either the cleavages or groups generated by the cleavages, not the cleavages or groups themselves. *Heterogeneity* would be equally acceptable.

3. Results in other literatures have also been sensitive to the use of this alternative operationalization. For one example, see Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003) regarding the relationship between religious, ethnic, and linguistic heterogeneity on one hand and both economic growth and government quality on the other. For another, see Aistrup (2004) regarding constituency diversity and party competition in the United States.

4. For example, it seems difficult to justify the operationalization of both Amorim Neto and Cox (1997) and Cox (1997), the number of groups along a *single* cleavage, given their conceptualization of diversity as the number of cleavages.

5. To elaborate, for j cleavages, with n_j groups generated by the j th cleavage, the number of joint groups formed ranges from a minimum of $\max(n_1, \dots, n_j)$ to a maximum of $n_1 \times \dots \times n_j$, depending on the relationship between the cleavages. For example, with two groups along each of two cleavages, the number of joint groups ranges from two to four: Two if the cleavages perfectly overlap, three if they partially crosscut, and four if they perfectly crosscut.

6. It is always less than or equal to the theoretical minimum. Specifically, it is equal to the theoretical minimum if and only if the numbers of groups generated by the individual cleavages are all equal (i.e., if $n_1 = \dots = n_j$). An operationalization that will always have less measurement error than Annett's (2001) takes the theoretical minimum as the index value.

7. Specifically, any such measure will usually underestimate the true quantity of interest, because it ignores all types of diversity except for the chosen type, although all countries are not affected by the same constant amount. Systematic, nonconstant measurement error of this sort gives rise to biased and inconsistent coefficient estimates; however, it is difficult to determine the direction of the biases in a multivariate context.

8. Religion takes a distant but still unchallenged second place. Lijphart (1999, pp. 80-81), for example, found the socioeconomic cleavage salient in all 36 democracies studied, the religious in 22, the cultural-ethnic in 12, the foreign policy in 12, the urban-rural in 8, and the postmaterialist in 4. Many other scholars such as Caramani (2004) also note the preeminence of the socioeconomic cleavage. In fact, Jones (2004) tapped socioeconomic for his measure of politicized diversity for just this reason.

9. Their polarization index is calculated as follows:

$$P = 1 - \sum_{i=1}^n \left(\frac{0.5 - p_i}{0.5} \right)^2 p_i$$
, where p_i is the population share of the i th group and n is the number of groups.

10. Cultural fractionalization is calculated using the following equation:

$$C = 1 - \sum_{i=1}^n \sum_{j=1}^n p_i p_j r_{ij}$$
, where p_i is the population share of the i th ethnic group, p_j is the population share of the j th ethnic group, n is the number of ethnic groups, and r_{ij} is a resemblance factor capturing the cultural similarity between the two ethnic groups.

11. This well-known statistic was popularized in comparative politics by Laakso and Taagepera (1979) and is calculated as follows:

$$N = \frac{1}{\sum_{i=1}^n p_i^2}$$
, where p_i is the population share of the i th group and n is the number of groups.

12. As is also well known, $F = 1 - 1/N$. N is closely related to a member of the family of Rényi entropies, which are used in information theory to measure the amount of uncertainty

in a probability mass function (Cover & Thomas, 1991). This member is the quadratic Rényi entropy. Other members lead to different measures: For example, alternatives are $N' = n$, the number of groups, or $N'' = 1/\max(p_i)$, the reciprocal of the largest population share. These three measures quantify different properties of the sequence p_1, p_2, \dots, p_n . In comparison, F , a transformation of N , and N quantify the same properties and thus can be viewed as equivalent measures, to be contrasted with alternative measures such as N' and N'' . Rényi entropies are defined by a parameter $\alpha \in [0, +\infty]$. N' has parameter $\alpha = 0$, N has $\alpha = 2$, and N'' has $\alpha = +\infty$. Increasing α emphasizes larger groups and provides greater robustness to incomplete information about the smaller groups. In a way, N is a compromise between the two extremes of N' and N'' .

13. The closest is Jones (1997), who collected data for a single U.S. state. We note that the related American politics literature (e.g., Aistrup, 2004) has collected district-level data.

14. Because this simultaneous equation model is triangular, ordinary least squares can be employed if the model is recursive, that is, if the disturbance terms are contemporaneously uncorrelated. However, the assumption of recursiveness does not seem tenable in this case.

15. Explicitly omitted from consideration here are two well-known data sets that are measurements of *politicized* cultural-ethnic groups: the ethnopolitical groups of Scarritt and Mozaffar (1999) as well as Posner's (2004) politically relevant ethnic groups (PREG).

16. Ordeshook and Shvetsova (1994) also constructed a list of linguistic groups and their population shares based on the same source (for the year 1980), but this data set is not publicly available.

17. Posner's objection to using measurements of ethnic groups from the 1960s to explain contemporary economic growth does not seem applicable here. Although not as frozen as Lipset and Rokkan (1967) portrayed them to be, contemporary party systems are commonly thought to be "sticky." Hence, it is plausible that they, unlike economic growth, might be shaped by past latent diversity.

18. Others such as Montalvo and Reynal-Querol (2001) have also constructed cross-national lists of religious groups and their population shares, none of which are publicly available. All based their work on Barrett (1982), except for Ordeshook and Shvetsova (1994), who turned to the *Encyclopedia Britannica*.

19. This model is nevertheless flawed in that it omits a constitutive term (see Brambor, Clark, & Golder, 2005), ENPRES (the effective number of candidates in the presidential election). In the interests of replication, we report the results from comparably flawed models. However, including the omitted term yields effectively identical results, which are presented in a supplemental paper available from the author's Web site.

20. The data set described in both Amorim Neto and Cox (1997) and Cox (1997) contains 54 countries. However, both studies eliminated the 3 countries of Bolivia, Honduras, and Uruguay when the dependent variable was the electoral party system, as it is here. See Amorim Neto and Cox (1997, pp. 163-164).

21. This version of the Soviet geographers' original groupings, found in column 6 of Roeder's (2001) data set, is relatively disaggregated. For example, Roeder treats racially distinct subgroups within an ethnolinguistic group, such as White and African Americans, as separate groups.

22. Note that we correct a mistake in Alesina et al.'s (2003) score of linguistic fractionalization for the Netherlands, replacing their value of 0.514 with our value of 0.263. Details are found in the supplemental paper referenced in Note 19.

23. We also estimated the models on the same maximally listwise deleted set of 39 cases to ensure that any differences in results were due to differences in measures, not to differences

in cases. The only real change of note regards Hypothesis 2, which is now almost universally rejected. These results are available in the supplemental paper (see Note 19).

References

- Aistrup, J. (2004). Constituency diversity and party competition: A county and state level analysis. *Political Research Quarterly*, 57(2), 267-281.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2), 155-194.
- Amorim Neto, O., & Cox, G. (1997). Electoral institutions, cleavage structures, and the number of parties. *American Journal of Political Science*, 41(1), 149-174.
- Annett, A. (2001). Social fractionalization, political instability, and the size of government. *International Monetary Fund Staff Paper*, 48(3).
- Barrett, D. (Ed.) (1982). *World Christian encyclopedia*. New York: Oxford University Press.
- Blalock, H., Jr. (1982). *Conceptualization and measurement in the social sciences*. Beverly Hills, CA: Sage.
- Brambor, T., Clark, W., & Golder, M. (2005). Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14(1), 63-82.
- Caramani, D. (2004). *The nationalization of politics: The formation of national electorates and party systems in Western Europe*. New York: Cambridge University Press.
- Chhibber, P., & Kollman, K. (2004). *The formation of national party systems: Federalism and party competition in Britain, Canada, India, and the United States*. Princeton, NJ: Princeton University Press.
- Cover, T. & Thomas, J. (1991). *Elements of information theory*. New York: John Wiley.
- Cox, G. (1997). *Making votes count: Strategic coordination in the world's electoral systems*. New York: Cambridge University Press.
- Duverger, M. (1963). *Political parties: Their organization and activity in the modern state*. New York: John Wiley.
- Fearon, J. (2003). Ethnic structure and cultural diversity by country. *Journal of Economic Growth*, 8(2), 195-222.
- Fearon, J., & Laitin, D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review*, 97(1), 75-90.
- Filippov, M. G., Ordeshook, P., & Shvetsova, O. (1999). Party fragmentation and presidential elections in post-communist democracies. *Constitutional Political Economy*, 10(1), 3-26.
- Geddes, B. (1990). How the cases you choose affect the answers you get: Selection bias in comparative politics. *Political Analysis*, 2(1), 131-150.
- Golder, M. (2006). Presidential coattails and legislative fragmentation. *American Journal of Political Science*, 50(1), 34-48.
- Golder, M., & Clark, W. (2006). Rehabilitating Duverger's theory: Testing the mechanical and strategic modifying effect of electoral laws. *Comparative Political Studies*, 39(6), 679-708.
- Jacobson, A., & Lulu, N. M. (1974). An empirical and algebraic analysis of alternative techniques for measuring unobserved variables. In H. Blalock, Jr. (Ed.), *Measurement in the social sciences* (pp. 215-242). Chicago: Aldine.
- Jones, M. (1997). Racial heterogeneity and the effective number of candidates in majority runoff elections: Evidence from Louisiana. *Electoral Studies*, 16(3), 349-358.
- Jones, M. (1999). Electoral laws and the effective number of candidates in presidential elections. *Journal of Politics*, 61(1), 171-184.
- Jones, M. (2004). Electoral institutions, social cleavages, and candidate competition in presidential elections. *Electoral Studies*, 23(1), 73-106.
- King, G., Keohane, R., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- Laakso, M., & Taagepera, R. (1979). Effective number of parties: A measure with application to Western Europe. *Comparative Political Studies*, 12(1), 3-27.
- Laitin, D., & Posner, D. (2001, Winter). The implications of constructivism for constructing ethnic fractionalization indices. *APSA-CP Newsletter*, 12, 13-17.
- Lijphart, A. (1968). *The politics of accommodation: Pluralism and democracy in the Netherlands*. Berkeley: University of California Press.
- Lijphart, A. (1999). *Patterns of democracy*. New Haven, CT: Yale University Press.
- Lipset, S. M., & Rokkan, S. (1967). Cleavage structures, party systems and voter alignments. In S. M. Lipset & S. Rokkan (Eds.), *Party systems and voter alignments: Cross-national perspectives* (pp. 1-64). New York: Free Press.
- Molinar, J. (1991). Counting the number of parties: An alternative index. *American Political Science Review*, 85(4), 383-391.
- Montalvo, J., & Reynal-Querol, M. (2001). *The effect of ethnic and religious conflict on growth* (Project on Religion, Political Economy, and Society Working Paper No. 4). Cambridge, MA: Harvard University, Weatherhead Center for International Affairs. Available from <http://www.wcfiareligionproject.org>
- Mozaffar, S., Scarritt, J., & Galaich, G. (2003). Electoral institutions, ethnopolitical cleavages, and party systems in Africa. *American Political Science Review*, 97(3), 379-390.
- Ordeshook, P., & Shvetsova, O. (1994). Ethnic heterogeneity, district magnitude, and the number of parties. *American Journal of Political Science*, 38(1), 100-123.
- Posner, D. (2004). Measuring ethnic fractionalization in Africa. *American Journal of Political Science*, 48(4), 849-863.
- Posner, D. (2005). *Institutions and ethnic politics in Africa*. New York: Cambridge University Press.
- Powell, G. B. (1982). *Comparative democracies: Participation, stability and violence*. Cambridge, MA: Harvard University Press.
- Roeder, P. (2001). *Ethnolinguistic fractionalization (ELF) indices, 1961 and 1985*. Retrieved June 2004 from <http://weber.ucsd.edu/~proeder/elf.htm>
- Rose, R., & Urwin, D. (1969). Social cohesion, political parties and strains in regimes. *Comparative Political Studies*, 2(1), 7-67.
- Scarritt, J., & Mozaffar, S. (1999). The specification of ethnic cleavages and ethnopolitical groups for the analysis of democratic competition in contemporary Africa. *Nationalism and Ethnic Politics*, 5(1), 82-117.
- Sullivan, J. (1973). Political correlates of social, economic and religious diversity in the American states. *Journal of Politics*, 35(1), 70-84.
- Taylor, C., & Hudson, M. (1972). *World handbook of political and social indicators* (2nd ed.). New Haven, CT: Yale University Press.

Heather Stoll is an assistant professor in political science at the University of California, Santa Barbara. Her research interests range from comparative politics to political methodology.