Less Is Not More: The Insufficiency of Current Data for Understanding the Relationship between Social Diversity and Party System Development†

Rob Moser
University of Texas, Austin

Ethan Scheiner
University of California, Davis

Heather Stoll
University of California, Santa Barbara

Abstract

Scholars commonly argue that in democratic societies, the size or fragmentation of party systems — as well as other important features of party systems, such as their nationalization or aggregation — is a function of social heterogeneity, in interaction with political institutions such as the electoral system. Although societies exhibit different types of heterogeneity, from their religious diversity to how socioeconomically equal they are, and although heterogeneity often varies at different political levels, from the national level to the electoral district, political scientists typically characterize countries’ heterogeneity almost exclusively according to measures of national level ethnic diversity. We draw attention to the importance of both different types of heterogeneity and the level at which social heterogeneity is measured. Using original census data for a number of advanced industrial and developing democracies, we show that characterizing heterogeneity solely according to nationwide ethnic diversity often paints a misleading portrait, particularly where it matters the most: at the level of the electoral district. We conclude briefly with the implications for theories that seek to relate social heterogeneity to key aspects of democratic party systems.

† Paper presented at the 2014 Annual Meeting of the American Political Science Association, 28-31 August, Washington, D.C. Contact information is as follows: Moser, rmoser@austin.utexas.edu; Scheiner, escheiner@ucdavis.edu; Stoll, hstoll@polsci.ucsb.edu. We thank Neil Fasching for excellent research assistance.
Social diversity has been one of the most utilized and important independent variables in all of social science. Scholars have examined the impact of various types of social diversity—from ethnic heterogeneity to income inequality—on a vast range of political and economic outcomes, including conflict, democratic electoral competition, party systems, collective goods provision, and regime stability and change (e.g., Fearon and Laitin 2003; Sambanis and Shayo 2013; Cox 1997; Clark and Golder 2006; Moser and Scheiner 2012; Stoll 2013; Easterly and Levine 1997; Alesina and Glaeser 2004; Crisp, Olivella and Potter 2013; Horowitz 1985; Rabushka and Shepsle 2008).

Yet, despite the significance of social diversity in innumerable social science analyses, data collection efforts to measure social diversity remain remarkably underdeveloped. Relatively recent initiatives have significantly upgraded our empirical measures of social diversity at the national level (e.g., Alesina et al. 2003; Fearon 2003; Posner 2004; Stoll 2013), but available data lag far behind what is needed to address social science theories and hypotheses properly. In particular, social science generally lacks cross-national data on social heterogeneity in two vital arenas: 1) measures of social diversity at *the appropriate geographical level of analysis*, such as the electoral district in studies of party systems, and 2) measures of *different types* of social diversity that capture the *variety* of salient cleavages in democratic societies. On 2), many studies include measures of ethnicity but far fewer pay attention to other types of diversity, such as religion, rural/urban divisions, and socio-economic divisions rooted in income, occupation, and levels of education.

In this paper, we highlight the inadequacy of social science data on social diversity. We argue that political scientists still tend to rely too heavily on measures of *national-level ethnic heterogeneity* as their sole estimate of social diversity, and that this is a case where less is not more. Specifically, we show how relying on a single, national-level indicator of diversity fails to capture the variance of diversity found *within* countries across politically salient subnational units such as federal regions or electoral districts. We also show how this misses the vast differences in the types of conflicts that divide societies. Countries may be similar on some dimensions (e.g., low levels of ethnic diversity), but differ wildly on others (e.g., their degree of religious or socio-economic heterogeneity).

We address these issues surrounding social diversity data collection through a first-cut descriptive analysis of social heterogeneity in a number of countries. Focusing in particular on Ireland, New Zealand, Spain, and the United States, we construct data sets that allow us to illustrate differences in the degree of social diversity—as measured by age, race/ethnicity, religion, education, occupation, and degree of ruralness or urbanness—both across countries, as well as within them. Following Stoll (2008), we use these data to highlight how empirical questions over social diversity measurement may have important implications for researchers’ findings, with a particular focus upon the party system.

**Social Diversity and Party Systems: The Existing Literature**

For more than half a century, scholars have looked to social diversity as the foundation of party system size—that is, the number of salient parties within a democratic polity. Although better known for his “law” about the effects of electoral systems on party systems, Maurice Duverger
declared that political parties were a reflection of the “spiritual families” within society (as cited in Clark and Golder 2006, 682). Lipset and Rokkan’s (1967) classic study further solidified the central role that social divisions, which emerged from fundamental societal transformations such as industrialization and modernization, had on party systems. Not only did these divisions determine the size of party systems, they also “froze” party systems in place until the next social upheaval created significant enough new social forces to create new parties (1967: 52). If social divisions are the raw material for political parties, producing what Cox calls a society’s “natural” number of parties (1997, 140), it stands to reason that more diverse societies would have more parties. As Clark and Golder (2006) succinctly argue, “Duverger’s theory . . . implies that the number of parties should be an increasing function of the number of politically salient spiritual families in a polity” (2006: 682).

Although political scientists — especially since the advent of the new institutionalism — have tended to give greater attention to the effect of electoral systems on party system size, the impact of electoral rules is arguably secondary to social forces, manifesting only as a filter that allows or prevents each social division from forming its own set of political parties (e.g., Amorim Neto and Cox 1997; Clark and Golder 2006; Stoll 2013). Permissive electoral systems (such as proportional representation formulae coupled with large district magnitudes) provide few obstacles to the process of discrete social groups becoming manifest as individual political parties, while restrictive electoral systems (such as plurality formulae coupled with single member districts) produce powerful dynamics that largely prevent each politically salient social group from successfully forming its own party. Work by Powell (1982), Lijphart (1984), Ordeshook and Shvetsova (1994), Amorim Neto and Cox (1997), Cox (1997), Jones (1997, 2004), Mozaffar, Scarritt, and Galaich (2003), Clark and Golder (2006), Golder (2006), Stoll (2008, 2013), Singer and Stephenson (2009), Singer (2012), Moser and Scheiner (2012), and Potter (forthcoming), among others, firmly establishes the foundational influence of social divisions on the size of party systems, while also acknowledging the strong role that electoral systems play in mitigating the connection between social divisions and parties.

More recently, scholars have argued that social divisions have other effects on party systems as well. Most notably, while social diversity within electoral districts is expected to drive the number of parties at the level of the electoral district, diversity between districts and/or at the national level is thought to affect party system aggregation — that is, the establishment of large, national parties that can successfully compete in many districts (e.g., Cox and Knoll 2003; Stoll 2008, 2013; Hicken and Stoll 2011; Potter forthcoming). As Cox (1997:181-202) notes, it is party aggregation (which he calls “linkage”) that is central to the projection of district level party systems to the national level for the many countries with more than one electoral district, and hence to determining the size of the national party system. For example, social diversity across districts (i.e., the geographic concentration of social groups in different regions) provides incentives for the formation of regional parties; this in turn decreases the aggregation of the party system and drives up the national number of parties (e.g., Brancati 2008). Moreover, other political institutions such as the regime type may condition this effect (Stoll 2013).

In short, over the past thirty-plus years, the field has progressed, developing a clearer theoretical sense of how social heterogeneity and political institutions do (and do not) combine to shape the party system. Nevertheless, as we next discuss, despite these advances, the field
has typically lacked the data necessary to provide fully convincing tests to confirm that real-world accuracy of these theories

**Data Problem #1: Lack of District-Level Data on Social Heterogeneity**

The first problem with existing empirical approaches to the study of social diversity and party systems is that while the theoretical and empirical focus has rightly shifted to the level of the electoral district, measures of social heterogeneity have not followed suit.

More specifically, as noted in numerous studies (e.g., Riker 1986; Cox 1997), the number of parties at the district level is a function of social forces *within the district* and the electoral rules used there. As the foundation for the number of parties in a district, the variable of social heterogeneity is meaningful only insofar as it represents diversity within that district. And the logic behind Duverger’s Law is founded on the idea that political actors seek to affect the outcome of their own district’s race.

Nevertheless, prominent studies conducted at the level of the electoral district such as Singer and Stephenson (2009) and Singer (2012) all employ national level measures of social heterogeneity. Given the lack of available district-level measures of diversity, the use of national measures of course makes great sense, but doing so forces the researcher to make the unrealistic assumption that all electoral districts are perfect mirrors of the nation as a whole. And the preliminary evidence suggests that the failure of existing studies to match the measure to the mechanism (to borrow Posner’s 2004 phrase) has consequences: the few studies that have recently have been conducted using district level measures have come to markedly different conclusions from the standing wisdom: they have found non-linear instead of linear relationships between social diversity and party system size; relationships that vary with the type of diversity considered; and relationships that exist even under restrictive electoral rules, contrary to Duverger’s Law (e.g., Moser, Scheiner, and Milazzo 2011; Moser and Scheiner 2012; Stoll 2013).

To date, Jones (1997), Moser and Scheiner (2012), and Potter (forthcoming) are the only studies of legislative elections to employ data on social heterogeneity at the level of the electoral district.¹ Jones (1997) and Moser and Scheiner (2012) both draw upon census data; however, the latter study is unique in how it uses cross-level statistical models that account for the unobserved heterogeneity that characterizes legislative districts.

¹ Some scholars have gotten around this data problem to some extent by examining presidential elections, which allow for the use of national level data at the level of the electoral district because in almost all presidential elections, the electoral district is the country as a whole (e.g., Jones 1999, 2004; Golder 2006; Dickson and Scheve 2010; Stoll 2013). However, it is unclear whether the relationship between social diversity and the number of presidential candidates generalizes to legislative elections. Among other things, the stakes involved in presidential elections are often so much greater that voters and elites may have stronger incentives to behave strategically in such races. Moreover, the greater availability of information regarding presidential contests – as well as the fact that presidential candidates tend to be more well-known than candidates running for legislative office – makes it easier for voters to behave strategically. Indeed, Dickson and Scheve (2010: 365) argue in a footnote that they are not optimistic that their findings regarding ethnic diversity’s effects on presidential party system size would extend to the legislative party system. Further, presidential electoral systems are all extremely restrictive, which poses obvious challenges for testing hypotheses about the interaction between electoral system restrictiveness and social diversity (Stoll 2013).
however, they are limited in their comparative scope. For example, Jones studies only a single state (Louisiana) in a single country (United States), while Moser and Scheiner’s recent study is confined to a handful of countries that use single-member districts within a mixed-member electoral system setting. Both of these studies, accordingly, hold the electoral system constant by design, confining their analyses to relatively restrictive electoral systems. Potter (forthcoming) offers a very different approach, using Bayesian analysis of survey data (drawn from the Comparative Study of Electoral Systems [CSES] project) to estimate district- and cross-district diversity across a range social divisions and countries. By measuring numerous forms of diversity at both of these levels for a number of different countries, this approach is especially promising, but also has one significant limitation: researchers can only use the approach for years and countries in which the survey is conducted. CSES appears committed to conducting its surveys regularly for many countries for the foreseeable future, but researchers will be hard pressed to use Potter’s approach for years prior to the first CSES module (1996-2001) and for countries not covered by CSES.

Moreover, the use of national level measures makes it impossible to differentiate between the effect of social diversity on electoral coordination within electoral districts and the effect of social diversity on electoral coordination across electoral districts. As Stoll (2008, 2013) argues, diversity at different levels may have different effects: for example, diversity within districts might be expected to have a powerful effect upon the size of the district level party system, whereas diversity across districts and at the national level might be expected to powerfully shape party system aggregation. Yet only a handful of studies, such as Cox and Knoll’s (2003), have studied how national level ethnic diversity shapes party system aggregation. Potter’s (forthcoming) recent study is the first to empirically take up the issue of cross-district diversity using a quantitative design, although he confines his focus to its effect on the number of parties in the districts. However, a comprehensive, multi-level analysis — which requires measures of diversity at all of these different levels — has yet to be conducted, and those seeking to build upon Potter’s approach will face empirical limitations, as discussed both above and below.

Data Problem #2: Lack of Data on Different Types of Social Heterogeneity

A second problem with existing approaches is that to date, almost all quantitative studies seeking to link social diversity to party system size in legislative elections focus solely upon ethnic diversity, despite the (more qualitative) literature on party system development having long identified a number of consequential social divisions besides the ethnic one (e.g., Lipset and Rokkan 1967; Rae and Taylor 1970; Powell 1982; von Beyme 1985; Mainwaring and Scully 1995; Lijphart 1999; Caramani 2004). More specifically, most of the literature on the interactive impact that social diversity and electoral systems have on party systems focuses solely or primarily on ethnic heterogeneity to the exclusion of other social divisions (e.g.,

---

2 Crisp, Olivella and Potter (2013) also draw upon CSES data to create a related measure of cross-district diversity: what they call the cross-district constituency similarity, which is the diversity across districts in some demographic characteristics (which, interestingly, do not include ethnicity) of governing parties’ supporters.
Amorim Neto and Cox 1997; Cox 1997; Jones 1997, 1999; Mozaffar, Scarritt and Galaich 2003; Chhibber and Kollman 2004; Clark and Golder 2006; Golder 2006; Hicken and Stoll 2008; Singer and Stephenson 2009; and Singer 2013).

Of course, few would deny that ethnic divisions have a particularly powerful effect upon the party system, especially in ethnically divided societies where ethnicity is the primary political fault line, but we have good reason to believe that other kinds of social diversity matter too. For example, Lipset and Rokkan’s (1967) famous study identified a number of social divisions (“cleavages”) related to the national and industrial revolutions, such as class and urban-rural differences, that have had a major impact upon party system development. And once again, the preliminary evidence suggests that failing to match the measure to the mechanism has consequences: that different types of diversity do have different effects upon the party system (e.g., Stoll 2008, 2013). Ultimately, the relative impact of different types of social divisions on party systems is an empirical question, one that deserves systematic examination based on the available data aggregated at the appropriate level of analysis.

There are some notable exceptions to the emphasis on ethnic diversity in most work on social heterogeneity, but, even still, most studies that use other (i.e., non-ethnic) indices of diversity tend to look at a relatively narrow range of measures. For example, Powell (1982) includes measures of religious and urban-rural diversity; Ordeshook and Shvetsova (1994) use religious diversity; Jones studies (2004) ideological (left-right) diversity; and Moser and Scheiner (2012) examine urban-rural diversity. Stoll (2004, 2008, 2013) is the most prominent exception to date: in her most recent work, Stoll constructs measures of social diversity along six historically important divisions (ethnic, religious, urban-rural, socioeconomic, foreign policy, and post-materialist), as well as measures of three historically prominent and exogenous sources of changes in a democracy’s citizenry (changes in the franchise, changes in territory, and immigration). However, harkening back to our critique in the previous section on the lack of district-level indices, Stoll’s measures are all at the national level. Potter’s (forthcoming) survey-based measures of diversity are other recent exceptions that hold great promise, but a number of drawbacks remain. Most importantly from our perspective, his use of probabilistic topic modeling, a relatively new measurement strategy that returns only the size-weighted number of groups at the district and cross-district levels, does not straightforwardly allow for the testing of hypotheses about the impact of different types of diversity. Moreover, it is not clear that the survey approach is a marked improvement over Stoll’s, which was based on data from governmental sources, such as countries’ national censuses.

**But Less Is Not More**

In the remainder of this paper, we explore these two fundamental problems of data collection on social diversity. In the following section, for several illustrative cases, we construct an original data set of both district- and national-level measures of a variety of different forms of social diversity, as measured by age, race/ethnicity, religion, education, occupation, and degree of ruralness or urbanness. We use this data to argue that measuring social heterogeneity exclusively at the national level – and primarily with respect to ethnicity – paints a misleading portrait of a country’s social structure, and likely promotes a misreading of the relationship between that social structure and party system development.
The Data

The source of our data is national censuses. Collected periodically with great care (and often at great cost) by most democratic states, national censuses ask questions about the most relevant demographic, economic, and social characteristics of the people living within a country’s borders. The resulting data directly shapes party systems in the many democracies that have electoral districts whose boundaries are periodically re-drawn based on population. Further, a number of countries that employ list proportional representation electoral systems (the countries that are the least likely to re-district) still rely upon census data to determine how many seats are assigned to each electoral district whose boundaries are delimited by law. Moreover, even where census data is not directly implicated in the electoral system, it frequently serves as a critical input into democratic policy-making. Finally, census data, in contrast to public opinion data, dates back to the dawn of modern democratic politics in the oldest democracies, enabling an unparalleled longitudinal analysis of the relationship between social structure and party systems. For these reasons, the most accurate and authoritative portrait of a country’s social heterogeneity comes from census data.

In the remainder of this paper, we take a detailed look at different types of social diversity in four democracies at the level of the electoral district. Collectively, these countries employ some of the most common types of electoral systems currently in existence and provide us with substantial variation on the variable posited to condition the relationship between social diversity and the party system: the restrictiveness of the electoral system. Specifically, we examine data from Ireland (small magnitude, proportional representation via the single transferable vote); New Zealand (mixed-member proportional); Spain (medium magnitude list proportional representation); and the United States (plurality with single member districts, or first-past-the-post). To aid in comparability, we draw our data for all four countries from censuses conducted in the 2000s to early 2010s.

We cast a wide net and present data on a number of potentially politically important demographic, economic, and social divisions. These divisions include age, ethnicity, religion, occupation, education, and the degree of ruralness or urbanness. However, data on all of these divisions were not available for all countries. An example is the United States, where the Census Bureau is prohibited by law from collecting data on religious affiliation in either its demographic surveys or decennial censuses. Accordingly, for each of the divisions for which data are available, we calculate for each district the effective number of groups (denoted “N” in

---

3 For example, while the process of drawing constituency boundaries, usually called delimitation, is a relatively understood component of electoral systems, the most comprehensive comparative survey to date (of 87 countries) determined that 69% delimit electoral districts; for all of these countries, equality of population is one of, if not the most important, criterion (Handley 2008).

4 Following the approach taken by the countries themselves in their national censuses, our understanding of ethnicity is eclectic, embracing different ways of defining ethnic groups. For example, some countries such as the United States primarily define ethnicity racially, whereas for others such as Spain, it has a heavily linguistic component. And while some countries collect data on different dimensions of ethnicity, from race to language, others do not. Accordingly, for ease of discussion, we refer to all of these social and cultural characteristics as “ethnic”.
tables and figures), which serves as our primary measure of a district’s diversity with respect to that division. For example, for each district, we use the census measures of the share of the population that is associated with each ethnic group to calculate the effective number of ethnic groups. Sometimes we present a second measure of diversity for a given division: for age, we also present the percentage of the population that is 65 years of age or older, and, for urbanization, we indicate the percentage of the population that is engaged in agriculture as an occupation.

We take two strategies with respect to the types of groups we measure. First, salient ethnic and religious groups naturally differ from country to country, making it difficult, if not impossible, to impose a common categorical structure upon all countries. Therefore, for ethnicity and religion, we use the highest-level categories defined by each country’s census. Second, for the remaining divisions, we are able to create a (relatively) consistent list of groupings that we can use across countries. For age, we create three distinct categories: under-18, 18-64, and over-64. For education, we create five categories: less than a secondary education; a completed secondary education; some tertiary education, but less than a bachelor’s degree; a bachelor’s degree; and at least some post-graduate education (such as a master’s degree). For occupation, we employ a three-category schema: managerial and professional; service (including sales, office positions, and unskilled labor); and working class (skilled and semi-skilled labor, e.g., in manufacturing and trades). For ruralness, we use a two-category schema that divides the population into those who are engaged in agriculture and related occupations (such as forestry, fishing, and, in some cases such as New Zealand, mining) and those who are not. In all cases, we eliminated individuals placed in residual “other” categories as well as those who did not state a category membership.

Figures 1-4 present these data. Boxplots summarize the distribution of each measure of diversity for the country’s electoral districts. In addition to the measures already described, some country-specific data is presented, such as the black and Hispanic/Latino (henceforth, Latino) share of the population in the United States. Figure 1 contains data for the 43 Irish constituencies in 2011 from the 2011 census. Figure 2 contains data for the 63 New Zealand constituencies in 2008 from the 2006 census. Figure 3 presents data for the 52 Spanish constituencies in 2011, drawing from the 2011 census. Figure 4a presents data for the 436 constituencies for the 110th Congress (elected in 2006) of the United States, drawing upon the

---

5 This is identical to the Laakso-Taagepera (1979) equation for the effective number of parties, the most common operationalization of party system size, with a group’s share of the population replacing a party’s seat or vote share.
6 It is for this reason that all existing national level measures of ethnic and religious diversity employ country-specific lists of groups (Stoll 2008).
7 We exclude from this calculation those working in agriculture and related occupations such as forestry and fishing.
8 Data is for the 2007 constituencies, which are the same constituencies that were used for the 2011 election, given that constituency boundaries were only redrawn in 2005 and 2013.
9 Provinces serve as electoral districts in Spain, so this is also data for the Spanish provinces. Note that unusually, seat apportionment in Spain is not based upon this census data, but rather upon the electoral census, an electoral roll that is maintained (and updated monthly since 1995) by the Electoral Census Office.
2000 census, and Figure 4b presents data on the black share of electoral districts for this Congress as well as a historical one (the 42nd, elected in 1870). The Appendix lists the specific census questions and categories, as well as how they were aggregated, for each country and division.

Given our goal of exploring the differences between national and district level measures of diversity, we also present the comparable national level measures of diversity for these social divisions for the four countries in Table 1. That is, for the same censuses, we gather the data provided by the country’s census or statistics bureau for the country as a whole. We take this approach because it mirrors the conventional measurement strategy, providing insurance against there being differences between the reported national level data and the alternative, an aggregation of the district level data. The exception is New Zealand, for which we aggregate the district level totals.

Finally, in Table 2, we provide basic descriptive statistics – drawn from national censuses – on the most studied social division, ethnicity, at different sub-national levels for a number of countries. These sub-national units range from the municipality to the province. Some, but not all, of these sub-national boundary lines coincide with the electoral districts.

**The National versus the District Level**

One of the most important conclusions that can be drawn from comparing the different measures of diversity at the national and district levels for our four countries is that the electoral districts are not perfect mirrors of their entire body politics. For each country, there are many districts that are more homogenous than the country as a whole, and many that are more heterogeneous.

Take, for example, ethnic diversity in the United States. In 2000, the national effective number of ethnic (racial) groups was about 1.5 – a fairly, but not completely, homogenous picture. Yet Figure 4a shows that in U.S. electoral districts in 2006, the effective number ranges from a perfectly homogenous 1.0 to a much more heterogeneous 2.8. In fact, almost ten percent of electoral districts have an effective number of ethnic groups that is greater than 2.0, whereas more than a third have an effective number of ethnic groups that is less than 1.2. At an even more fine-grained level, Latinos make up about 13% of the United States population as a whole. At the same time, Latinos make up more than one-third of the population in roughly ten percent of all districts, but, at the same time, compose less than three percent of the population in nearly one-third of all districts. With respect to blacks, as Figure 4b shows, a substantial majority of districts (about 70%) have only a small minority black contingent (less than 10% of the electorate), but, simultaneously, blacks make up more than 30% of the electorate in a large number (11-18%) of districts.

An even more extreme example on the ethnic front is provided by New Zealand. In the nation as a whole in 2006, the effective number of ethnic groups is equal to 2.0. By way of contrast, in the districts, the effective number ranges from a homogenous 1.2 to a heterogeneous 3.9. The nature of some of this heterogeneity is made clear by examining the

---

10 This includes the delegate district of the District of Columbia. Districts for the 110th Congress, elected in 2006 and serving from 2007-09, were the first districts drawn using the 2000 census data.
percentage of the district that is Maori: this percentage ranges from a minimum of about four percent to a mean of fourteen percent to a maximum of forty-three percent.11

As another example, take the urban-rural division. In all four countries at the national level, the effective number of urban/rural groups is about 1.0 — an essentially homogenous portrait. Whether these countries are homogenously rural or urban seems to be clarified by the percent of the population engaged in agriculture and other conventionally rural occupations such as forestry: this statistics ranges from a miniscule less than one percent in the United States to a more significant but still small seven percent in New Zealand. Accordingly, one would conclude from the national level data that these are effectively urban countries in the early twenty-first century. However, the picture changes in important ways when we look at the district level. In Ireland and the United States, the average effective number of urban/rural groups in the districts increases slightly to about 1.4, although it remains at just over 1.0 in New Zealand and Spain. Yet in all countries, the value of the effective number ranges from a minimum of 1.0 (totally homogeneous) to a more heterogeneous maximum of 1.3 in Spain, 1.4 in the United States, and 1.6 in both Ireland and New Zealand. The even more telling picture is painted by the share of the district populations engaged in agriculture. This ranges from a minimum of effectively zero percent in some districts in all four countries, to a mean of five percent and a maximum of 13% in Spain; a mean of less than one percent and a maximum of 16% in the United States; a mean of seven percent and a maximum of 17% in Ireland; and a mean of seven percent and a maximum of 25% in New Zealand. Some electoral districts, in other words, are substantially rural in character, in contrast to the portrait painted by the national level data.

For a final example, consider patterns of age groups. About twelve percent of the national population in Ireland, New Zealand, and the United States is 65 or older, and about seventeen percent in Spain. Yet at the level of the electoral district, some districts possess an age structure tilted towards the young, whereas in others, the age structure tilts towards the elderly. Of the more youthful districts, the minimal percentages of individuals aged 65 or older are four percent in Ireland, five percent in the United States, six percent in New Zealand, and ten percent in Spain. Of the more elderly districts, the maximal percentages of individuals aged 65 or older are eighteen percent in Ireland, twenty-three percent in New Zealand, and a

---

11 We should note a significant problem with our district-level measures of diversity for New Zealand: in New Zealand, voters have the option of either voting in their geographical district or casting their ballot for a seat reserved for the Maori. Not surprisingly, a very large proportion of Maori choose to place themselves on the Maori rolls, rather than voting in their more geographically-focused single-member district (SMD). As a result, while the measures that we report here may accurately represent the residents of each SMD, they may not be a fair representation of the groups that vote in the district. One solution to this problem is, for our measures of ethnic diversity, to subtract out the number of Maori who reside in the SMD but vote on the Maori rolls (Moser and Scheiner 2012, Ch. 7). However, unfortunately, such a strategy does not address the problem for the other measures of diversity, because we do not know how many Maori fall into each category of the other measures. In many SMDs, this problem is not significant: the median Maori population is only 11%, with a low of under 4%. However, in other districts, it is more problematic, as Maori make up more than 24% of the population in seven districts. We welcome suggestions on approaches to address this issue.
whopping thirty percent in both Spain and the United States. Again, substantial variation exists at the level of the electoral district, as anyone who has visited the sunny American state of Florida would know.

Skimming over Table 2, which provides a broader comparative perspective, also drives home the point that districts can vary substantially in how homogenous or heterogeneous they are. With rare exceptions such as Uruguay, ethnic heterogeneity at the sub-national level typically ranges from the extremely homogenous (an effective number of about 1.0) to the fairly heterogeneous (usually an effective number of at least 2.0, and often ranging into the 3.0s). Over all of the countries and years for which we have data, the effective number of ethnic groups in the sub-national units on average deviates from the mean by almost 0.5 – a non-trivial amount.

A second important conclusion to be drawn from our data is that, while countries that are homogenous at the national level will also be fairly homogenous at the district level, countries that are heterogeneous at the national level may be homogenous at the district level. Canada is a case in point here. Data for 2006, shown in Table 2, reveal that at the level of the electoral district, most districts are ethnically (defined in linguistic terms in the Canadian case) homogenous: the effective number of ethnic groups ranges from 1.0 to 2.0, but the mean and the third quartile are both only 1.1, and the standard deviation is a mere 0.21. The corresponding national level figure, however, is a much more heterogeneous 1.7. In this case, the difference between the more homogenous districts and the more heterogeneous nation can be explained by the geographic segregation of English and French speakers, with French speakers concentrated in the province (and hence the districts) of Quebec.

An even more extreme example of national heterogeneity coupled with district homogeneity is provided by Spain, an example so extreme as to seem almost contrived. Data for 2011 shows that while not all Spanish districts fall on the homogenous side of the ethnic spectrum, most do: the minimum effective number of ethnic groups in the districts is 1.1, the mean is 1.4, the third quartile is 1.5, and the maximum is 2.4, which means that seventy-five percent of Spanish districts have an effective number of ethnic groups less than 1.5. This stands in marked contrast to the effective number of ethnic groups in the country as a whole: a whopping 9.9. Here, as in the Canadian case, ethnic groups are geographically concentrated in either single provinces or groups of neighboring provinces, and hence in electoral districts, given that the provinces serve as the electoral districts in Spain.

**Ethnic versus Other Types of Diversity**

This brings us to the second major take-away point from our exploration of social diversity and its implications for the party system: that assessing diversity solely with respect to ethnicity often paints a misleading portrait of a country.

A prime example is provided by Ireland, the least ethnically diverse of our four countries. Nationally, with an effective number of ethnic groups equal to 1.3 in 2011, Ireland falls on the homogenous side of the spectrum for minimally democratic countries. And while Figure 1 shows that there is more heterogeneity at the level of the electoral district, Irish districts in 2011 are still fairly ethnically homogenous: seventy-five percent of districts have an effective number of ethnic groups that ranges between 1.1 and 1.3, and the effective number
of ethnic groups in ninety-one percent of districts is less than 1.5. This would lead us to predict an extremely small party system size for Ireland. Yet in the 2011 election, the effective number of electoral parties at the national level in Ireland was about four, and over the entire post-war period, the average effective number of electoral parties has been almost 3.5 nationally and about three in the electoral districts, with higher numbers approaching four at both levels in more recent decades. While there are many countries with larger party systems, the Irish system clearly does not qualify as small. Accordingly, if social diversity is the demand-side explanation for the “natural” number of parties in a country (Cox 1997), one might suppose that there is not sufficient demand in Ireland for so many candidates and parties. The relatively large number of parties in Ireland becomes even more puzzling when one considers that the “true” natural number of parties is likely to be even greater, given that the Irish electoral system’s relatively small district magnitude (an average of almost four over the post-war period) is probably serving as a constraint.

The answer is to be found by looking beyond ethnic diversity. Ireland is much more diverse with respect to other historically important social divisions, most notably socio-economic ones (see Table 1 and Figure 1). For example, Table 1 shows that at the national level, the effective number of educational groups in Ireland in 2011 is 4.0 and the effective number of occupational groups is 3.1. Unlike the measures of ethnic diversity, these socio-economic figures place Ireland on the heterogeneous side of the socio-economic spectrum for our four countries, with only the United States arguably edging it out. Moreover, Figure 1 shows that there is also great diversity with respect to occupation and, especially, education in the Irish electoral districts. Similarly, as already discussed, Ireland also exhibits more variation with respect to ruralness at both the national level and across its districts, with a significant minority of districts maintaining a fairly rural character. Turning to age, one of the most basic demographic fault lines, Ireland also looks more diverse at both the national and district levels. Ireland is largely homogeneous with respect to religion at the national level, but exhibits more diversity at the level of the electoral districts.

To a lesser extent, a similar story can be told for the United States and New Zealand, as can be seen in Table 1 and Figures 2 and 4a. Of particular note, the United States exhibits even greater educational heterogeneity at both the national and district levels than Ireland does, as well as comparable occupational heterogeneity. Even in Spain, where other types of diversity do not rival ethnicity to the same degree, we can see that heterogeneity is present with respect to other divisions such as socio-economics.

**Conclusion: Implications for Party System Development**

In this paper, we have identified a significant disjuncture between the theories social scientists have developed regarding the relationship between social diversity and party systems and the data the scholarly community has compiled to empirically investigate them. Scholars have developed strong expectations about how social divisions serve as the foundation for the size and aggregation of democratic party systems; yet, we have not assembled data sets at the appropriate level of analysis and in the full spectrum of types of social diversity to conduct valid empirical tests of some of our most fundamental research questions. We argue that this “data
problem” has serious implications for the accuracy and validity of our understanding of party systems and, potentially, other key political phenomena.

Our goals here have been limited. We have attempted to demonstrate some of the ways that national level ethnic heterogeneity falls short if — as is often the case in contemporary social science research — it is used as the sole indicator of a state’s social diversity. We have shown that national level ethnic diversity differs sharply from what is experienced in a country’s electoral districts. Substantial numbers of electoral districts are significantly more homogenous, and others more diverse, than the nation as a whole. Moreover, we have shown that the extent of diversity within a country varies, depending on the dimension on which diversity is measured: a country can be simultaneously diverse on one dimension (e.g., ethnic diversity) but much more homogenous on others (such as religion or socioeconomic equality). Put differently, social diversity is more complicated than a single indicator such as national level ethnic diversity suggests.

This discussion raises important implications for our understanding of how social diversity affects party systems. To point out one example, take the divergence between national level and district level ethnic heterogeneity in the United States. This discrepancy potentially has important implications for the American party system. If many black voters voted for political parties designed explicitly to represent their racial group, they would be strong competitors in some districts, driving up the size of these districts’ party systems and, in turn, increasing the number of parties in the national party system. At the same time, however, because this increase in the number of parties would not occur in all districts, it would reduce the aggregation of the party system — that is, it would lead to less of a match between the district level and national level party systems. The same would be true with respect to Latinos, although to a lesser extent. The importance of the district-level measures here should be clear: these measures suggest the possibility of racial/ethnic representation, where national-level measures never would, particularly given the United States’ restrictive electoral system.

As a second example, we showed that in Canada and Spain, electoral districts tend to be more homogenous than the nation as a whole due to the regional segregation of ethnic groups. We expect this to have very different implications for the party system. We would expect that while the politicization of the ethnic division in these two countries should not drive up the size of the party system in the electoral districts, it might nevertheless do so at the national level, with regionally-based ethnic parties contributing to poorer party system aggregation and hence to a larger national party system. These divergent predictions across levels of analysis in turn have broader implications for a host of downstream political outcomes, such as the provision of public goods — predictions that cannot be arrived at when solely relying upon the national level ethnic data.

For a third and final example, our analysis of different types of diversity (other than ethnicity) in Ireland has important implications for the party system there. As we discussed above, Ireland is a relatively homogenous country in terms of its ethnic structure, but it is much more heterogeneous with respect to the socio-economic characteristics of its people. In short, we painted a very different picture of Ireland’s social diversity when we looked beyond ethnicity. This picture of greater diversity may help provide a partial explanation — one that
we otherwise would have missed if we had focused solely on ethnic diversity — for the relatively large number of parties found in Ireland.

Surely more detailed knowledge of social divisions, both at the national and the district levels, would enable us to better predict characteristics of countries’ party systems than would knowledge of their national ethnic heterogeneity alone. Less is not always more. The primary purpose of this paper is to make a call for, and begin the process of developing, the necessary measures to improve our data, which in turn will enable us to improve our analyses of these important questions.
### Appendix

<table>
<thead>
<tr>
<th></th>
<th>Ethnicity</th>
<th>Religion</th>
<th>Education</th>
<th>Occupation</th>
<th>Ruralness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ireland</strong></td>
<td>5 groups: White Irish, White Irish Traveller, Other White, Black, Asian</td>
<td>3 groups: Catholics, Other (not available at a lower level of aggregation; the exception to exclusion of residual &quot;others&quot; category), None</td>
<td>5 groups: Less than Secondary (No Formal Education; Primary Education; Lower Secondary), Completed Secondary (Upper Secondary; Technical), Some College (Advanced Certificate; Higher Certificate), College (Bachelors; Honors Bachelors); Post-College (Postgraduate; Doctorate)</td>
<td>3 groups: Managers and Professionals (Employers; Higher Professionals; Lower Professionals), Service and Sales (Non-Manual; Own Account; Unskilled), Working Class (Manual Skilled; Semi-Skilled)</td>
<td>2 groups: Agriculture (Agriculture; Farming) and Non-Agriculture (all others)</td>
</tr>
<tr>
<td><strong>New Zealand</strong></td>
<td>5 groups: European, Maori, Pacific Islander, Asian, Middle Eastern/Latin American/African</td>
<td>11 groups: None, Buddhist, Catholic, Protestant, Jehovah's Witness, Mormon, Hindu, Islam, Jewish, Maori, New Age</td>
<td>5 groups: Less than Secondary (No Qualification; Level 1 Certificate; Level 2 Certificate; Level 1, 2 or 3 Certificates Gained Post-School), Completed Secondary (Level 3 or 4 Certificate Gained at School);</td>
<td>3 groups: Managers and Professionals (Managers; Professionals), Service and Sales (Community and Personal Service Workers; Clerical and Administrative Workers; Sales Workers), Working Class (Technicians &amp; Trade Workers;</td>
<td>2 groups: Agriculture (Agriculture; Forestry; Fisheries; Mining), and Non-Agriculture (all others)</td>
</tr>
</tbody>
</table>
The table shows the educational qualifications and occupational groups in Spain by autonomous communities. The educational qualifications include Overseas Secondary School Qualification, Some College (Level 4 Certificate Gained Post-School; Level 5 Diploma; Level 6 Diploma), College (Bachelor Degree and Level 5 Diplomas), Post-College (Post-Graduate and Honours Degree; Masters Degree; Doctorate Degree).

In terms of occupational groups, there are 3 groups: Less than Secondary (Illiterates; Uneducated; Primary education; Compulsory Secondary), Completed Secondary (High School; Medium Grade Compulsory), Some College (Superior Grade Formative Education; Three Year University Certificate), College (Degree; Five

Spain | 19 groups (autonomous community of birth): Andalucia, Aragon, Principado de Asturias, Illeas Balears, Canarias, Cantabria, Castilla y Leon, Castilla - La Mancha, Cataluna, Comunitat Valenciana, Extremadura, Galicia, Comunidad de Madrid, Region de Murcia, Comunidad Foral de Navarra, Pais Vasco, La Rioja, | N/A | 5 groups: Less than Secondary (Illiterates; Uneducated; Primary education; Compulsory Secondary), Completed Secondary (High School; Medium Grade Compulsory), Some College (Superior Grade Formative Education; Three Year University Certificate), College (Degree; Five | 3 groups: Managers and Professionals (Managers; Professionals; Technicians and Associate Professionals), Service and Sales (Clerical Support Workers; Service and Sales Workers; Armed Forces; Elementary Occupations), Working Class (Craft and Related Trades; Plant and Machine Operators and Assemblers) | 2 groups: Agriculture (Skilled agricultural, forestry, and fishing workers) and Non-Agriculture (all others) |
<table>
<thead>
<tr>
<th>Ceuta, Melilla</th>
<th>Year University Certificate), Post-College (Master Degree; Doctorate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>5 groups (race; those identifying one race): White, Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander</td>
</tr>
<tr>
<td></td>
<td>5 groups: Less than Secondary (Less Than 9th Grade; 9th-12th Grades, No Diploma), Completed Secondary (High School Graduate, Includes Equivalency), Some College (Some College, No Degree; Associates Degree), College (Bachelors Degree); Post-College (Postgraduate or Professional Degree)</td>
</tr>
</tbody>
</table>
Bibliography


<table>
<thead>
<tr>
<th></th>
<th>N, Age</th>
<th>% 65 and Over</th>
<th>N, Race/Ethnicity</th>
<th>N, Religion</th>
<th>N, Education</th>
<th>N, Occupation</th>
<th>N, Urban-Rural</th>
<th>% Agriculture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ireland</td>
<td>2.1</td>
<td>12</td>
<td>1.3</td>
<td>1.3</td>
<td>4.0</td>
<td>3.1</td>
<td>1.1</td>
<td>5.0</td>
</tr>
<tr>
<td>New Zealand</td>
<td>2.2</td>
<td>12</td>
<td>2.0</td>
<td>3.2</td>
<td>2.9</td>
<td>3.0</td>
<td>1.2</td>
<td>7.0</td>
</tr>
<tr>
<td>Spain</td>
<td>2.1</td>
<td>17</td>
<td>9.9</td>
<td>N/A</td>
<td>2.7</td>
<td>2.7</td>
<td>1.1</td>
<td>2.9</td>
</tr>
<tr>
<td>United States</td>
<td>2.2</td>
<td>12</td>
<td>1.5</td>
<td>N/A</td>
<td>4.4</td>
<td>2.9</td>
<td>1.0</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Table 1. National level measures of diversity for several important social divisions for Ireland (2011), New Zealand (2006), Spain (2011), and the United States (2000). Note that the percentage of the population engaged in agriculture includes occupations such as forestry, fishing, and (where applicable) mining.
Figure 1. Electoral district level measures of diversity for several important social divisions for Ireland (2011).
Figure 2. Electoral district level measures of diversity for several important social divisions for New Zealand (2006).
Figure 3. Electoral district level measures of diversity for several important social divisions for Spain (2011).
Figure 4a. Electoral district level measures of diversity for several important social divisions for the United States (2000).
Figure 4b. Variation in the racial demography of Congressional districts in the United States over time: the percentage of Congressional districts for the 42\textsuperscript{nd} (1870) and 100\textsuperscript{th} (2006) Congresses with electorates less than ten percent black; between ten and thirty percent black; between thirty and fifty percent black; and at least fifty percent black. Source: Stoll (2013).
<table>
<thead>
<tr>
<th>Country and Census Year</th>
<th>Level and Number of Units</th>
<th>Corresponding Election Year(s)</th>
<th>Minimum</th>
<th>Median</th>
<th>Mean</th>
<th>Maximum</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bolivia (2001)</td>
<td>Department (9)</td>
<td>2009</td>
<td>1.4</td>
<td>1.9</td>
<td>1.9</td>
<td>3.0</td>
<td>0.47</td>
</tr>
<tr>
<td>Bolivia (2002)</td>
<td>Municipality (314)</td>
<td>2005</td>
<td>1.0</td>
<td>1.3</td>
<td>1.6</td>
<td>3.6</td>
<td>0.52</td>
</tr>
<tr>
<td>Brazil (2009)</td>
<td>States (27)</td>
<td>2010</td>
<td>1.3</td>
<td>2.0</td>
<td>2.0</td>
<td>2.4</td>
<td>0.29</td>
</tr>
<tr>
<td>Canada (2006)</td>
<td>Constituency (308)</td>
<td>2006</td>
<td>1.0</td>
<td>1.0</td>
<td>1.1</td>
<td>2.0</td>
<td>0.21</td>
</tr>
<tr>
<td>Colombia (2005)</td>
<td>Department (33)</td>
<td>2014</td>
<td>1.0</td>
<td>1.3</td>
<td>1.5</td>
<td>2.6</td>
<td>0.44</td>
</tr>
<tr>
<td>Costa Rica (2011)</td>
<td>Province (7)</td>
<td>2014</td>
<td>1.1</td>
<td>1.2</td>
<td>1.2</td>
<td>1.4</td>
<td>0.079</td>
</tr>
<tr>
<td>Ireland (2006)</td>
<td>Constituency (43)</td>
<td>2007</td>
<td>1.1</td>
<td>1.2</td>
<td>1.2</td>
<td>1.6</td>
<td>0.12</td>
</tr>
<tr>
<td>Ireland (2011)</td>
<td>Constituency (43)</td>
<td>2011</td>
<td>1.1</td>
<td>1.3</td>
<td>1.3</td>
<td>1.9</td>
<td>0.15</td>
</tr>
<tr>
<td>Ireland (2011)</td>
<td>Constituency (40)¹</td>
<td>2016</td>
<td>1.2</td>
<td>1.3</td>
<td>1.3</td>
<td>2.0</td>
<td>0.16</td>
</tr>
<tr>
<td>New Zealand (1986)</td>
<td>Constituency (93)</td>
<td>1987</td>
<td>1.0</td>
<td>1.3</td>
<td>1.4</td>
<td>3.1</td>
<td>0.33</td>
</tr>
<tr>
<td>New Zealand (1986)</td>
<td>Constituency (93)</td>
<td>1990</td>
<td>1.0</td>
<td>1.3</td>
<td>1.4</td>
<td>3.1</td>
<td>0.33</td>
</tr>
<tr>
<td>New Zealand (1991)</td>
<td>Constituency (95)</td>
<td>1993</td>
<td>1.1</td>
<td>1.4</td>
<td>1.6</td>
<td>3.7</td>
<td>0.54</td>
</tr>
<tr>
<td>New Zealand (1996)</td>
<td>Constituency (60)</td>
<td>1996</td>
<td>1.2</td>
<td>1.7</td>
<td>1.9</td>
<td>3.9</td>
<td>0.59</td>
</tr>
<tr>
<td>New Zealand (1996)</td>
<td>Constituency (61)</td>
<td>1999</td>
<td>1.2</td>
<td>1.7</td>
<td>1.9</td>
<td>4.0</td>
<td>0.62</td>
</tr>
<tr>
<td>New Zealand (2001)</td>
<td>Constituency (62)</td>
<td>2002</td>
<td>1.2</td>
<td>1.6</td>
<td>1.8</td>
<td>3.5</td>
<td>0.57</td>
</tr>
<tr>
<td>New Zealand (2001)</td>
<td>Constituency (62)</td>
<td>2005</td>
<td>1.2</td>
<td>1.6</td>
<td>1.8</td>
<td>3.5</td>
<td>0.57</td>
</tr>
<tr>
<td>New Zealand (2006)</td>
<td>Constituency (63)</td>
<td>2008</td>
<td>1.2</td>
<td>1.8</td>
<td>1.9</td>
<td>3.9</td>
<td>0.60</td>
</tr>
<tr>
<td>Peru (2006)</td>
<td>Department (26)</td>
<td>2011</td>
<td>1.1</td>
<td>1.8</td>
<td>1.8</td>
<td>2.8</td>
<td>0.48</td>
</tr>
<tr>
<td>Russia (2001)</td>
<td>Raion (62)</td>
<td>2003</td>
<td>1.0</td>
<td>2.1</td>
<td>2.0</td>
<td>3.5</td>
<td>0.50</td>
</tr>
<tr>
<td>Spain (2001)</td>
<td>Province/ Constituency (52)</td>
<td>2004, 2008</td>
<td>1.1</td>
<td>1.3</td>
<td>1.4</td>
<td>2.3</td>
<td>0.29</td>
</tr>
<tr>
<td>Spain (2011)</td>
<td>Province/ Constituency (52)</td>
<td>2011</td>
<td>1.1</td>
<td>1.3</td>
<td>1.4</td>
<td>2.4</td>
<td>0.26</td>
</tr>
<tr>
<td>Ukraine (2001)</td>
<td>Raion (1036)</td>
<td>2002, 2012</td>
<td>1.2</td>
<td>1.6</td>
<td>1.8</td>
<td>3.5</td>
<td>0.51</td>
</tr>
<tr>
<td>Ukraine (2010)</td>
<td>District (225)</td>
<td>2012</td>
<td>1.0</td>
<td>1.2</td>
<td>1.4</td>
<td>2.0</td>
<td>0.32</td>
</tr>
<tr>
<td>United States (2000)</td>
<td>Constituency (436)²</td>
<td>2006 (110th)</td>
<td>1.0</td>
<td>1.3</td>
<td>1.4</td>
<td>2.8</td>
<td>0.36</td>
</tr>
<tr>
<td>Uruguay (2011)</td>
<td>Department (19)</td>
<td>2009</td>
<td>1.1</td>
<td>1.1</td>
<td>1.2</td>
<td>1.4</td>
<td>0.11</td>
</tr>
<tr>
<td>Wales (2003)</td>
<td>Constituency (40)</td>
<td>2003</td>
<td>1.2</td>
<td>1.4</td>
<td>1.5</td>
<td>2.0</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Table 2. Descriptive statistics for the effective number of racial/ethnic groups at different sub-national levels, particularly the level of the electoral district. Source for Bolivia 2002, New Zealand 1987-2005, Russia 1999, Ukraine 2002, and Wales 2003 is Moser and Scheiner (2012); all other sources are national censuses.

¹ Data for the constituencies as redrawn in 2013, which will apply to the next election (scheduled for 2016, unless an earlier election is called).
² Includes the Delegate District for the District of Columbia.