Parsimony: Limited Diversity, Counterfactual Cases and Comparative Analysis." In *Vergleichen in der Politikwissenschaft*, eds. Sabine Kropp and Michael Minkenberg. Wiesbaden: Verlag für Sozialwissenschaften, 180–197.

Rihoux, Benoît. 2006. "Qualitative Comparative Analysis (QCA) and Related Systematic Comparative Methods: Recent Advances and Remaining Challenges for Social Science Research." *International Sociology* 21 (5): 679–706.

Rihoux, Benoît and Charles C. Ragin. 2009. *Configurational Comparative Methods*. Thousand Oaks: Sage Publications.

Rosenbaum, Paul R. 2002. *Observational Studies*, 2nd edn. New York: Springer-Verlag.

Schneider, Carsten Q. and Claudius Wagemann. 2006. "Reducing Complexity in Qualitative Comparative Analysis (QCA): Remote and Proximate Factors and the Consolidation of Democracy." *European Journal of Political Research* 45 (5): 751–786.

Schneider, Carsten Q. and Claudius Wagemann. 2010. "Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-Sets." *Comparative Sociology* 9 (3): 397–418.

Schneider, Carsten Q. and Claudius Wagemann. 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. New York: Cambridge University Press.

Thiem, Alrik. 2013. "Clearly Crisp, and Not Fuzzy: A Reassessment of the (Putative) Pitfalls of Multi-Value QVA." *Field Methods* 25 (2): 197–207.

Thiem, Alrik and Adrian Dusa. 2013. "QCA: A Package for Qualitative Comparative Analysis." *The R Journal* 5 (1): 87–97.

Measuring Partial Membership in Categories: Alternative Tools

Zachary Elkins
University of Texas, Austin
zelkins@austin.utexas.edu

Almost any attempt at classification runs into a boundary problem. Some cases fit neatly into one category, some fit one category only partially, and some fit multiple categories. This is a well-understood issue among both cognitive psychologists, who have documented how the brain's hard-wiring classifies stimuli, and taxonomists,[1] who seek to "soft-wire" additional sorting schemes. My focus here is mostly on the soft wiring. How, exactly, *should* researchers build classification systems—referred to here as taxonomies—that account for partial membership in categories, if at all? An important reference point is fuzzy sets, an intriguing concept that has gained some traction in sociology and political science. I explore a set of measurement strategies for assigning partial membership scores in the context of executive-legislative relations, a research domain overdue for innovation in conceptualization and measurement.

Measuring Partial Membership in Categories?

I define a "partial membership score" as a measure of the extent to which a particular case belongs to a specific category. While scholars disagree on the necessity and utility of calculating such a score, one common rationale for doing so is that this calculation fits nicely with how the human mind works. Are we, perhaps, hard-wired to classify information according to a partial membership process? Some important insights from cognitive psychology on how we process and classify phenomena lead us to think so.

In the "classical" view of categorization (Murphy 2004), categories are defined by necessary and jointly sufficient conditions for membership. For example, parliamentary democracies may be defined by "assembly confidence," wherein the executive is both selected and removed by the legislature. This view of concepts admits no borderline cases and treats each member of the category as a full instance of the concept, with no significant distinctions among members.

The modern view, associated closely with Ludwig Wittgenstein and Eleanor Rosch, shifted toward a more graded view of concepts, thereby challenging the idea of well-defined membership and non-membership. Wittgenstein's (1953) concept of "family resemblance" undermines the idea that there is any common (much less necessary) attribute of category members. In Wittgenstein's view, parliamentary systems might be a family of systems whose members share—in varying combinations—a substantial number of characteristics, such as executive decree, minimal legislative oversight of the executive, and a figure-head for head of state. Rosch's (e.g, 1975) large body of experimental work advances the idea that people differentiate with respect to the *degree* of belonging to a prototype. For instance, Rosch showed that—in the framework of protoypes—a chair is a highly typical instance of furniture, a bookcase less typical, and a piano even less so. This focus on degree of belonging shifted the understanding of classification processes away from the idea of sharply defined category membership based on the conception of necessary and sufficient conditions.

In political science, David Collier's (Collier and Mahon 1993; Collier and Levitsky 1997) work on classical versus radial subtyping highlighted the necessity of using graded approaches to categorization, particularly with central yet contested concepts like "democracy." Collier's work left political scientists with a stronger appreciation for partial membership in categories, though he stopped short of recommending particular measurement instruments with which to assign scores.

I should note a parallel set of studies in cognitive psychology that reveal a certain "categoriness" to the mind. That is, for some concepts at least, we tend to lump phenomena into classes and to minimize the conceptual distance between co-classified items and exaggerate the distance between cross-classified items. This phenomenon, *categorical perception*, is evident with phenomena such as color, sounds, and—I suspect—a fair number of learned categories such as those in social science (Harnad 1990).

In sum, there appears to be a strong basis in cognitive psychology for the idea that partial membership is central to our neurological hard-wiring and "natural" categorization. It also seems likely that the continuum underlying many of these

Thanks to David Collier and Pam Paxton for helpful comments on earlier drafts.

[1] In this context, taxonomist refers broadly to scholars concerned with classification, and not narrowly to specialists in biological taxonomy.

classification schemes is subject to perceptual discontinuities that lead to a natural clustering of items. Thus, membership, but membership by degree. The question remains, however, of how an analyst can adapt existing measurement practices to reflect these ideas, in particular with respect to the learned categories of social science.

A Point of Departure: Fuzzy Sets

These insights explain the appeal of fuzzy sets, which extend the logic of set theory to graded membership (Zadeh 1965, Smithson and Verkuilen 2006, Ragin 2008). Charles Ragin has taken the lead in introducing fuzzy sets to the social sciences, in connection with an analytical method known as Qualitative Comparative Analysis (QCA—the fuzzy set version is fsQCA). For our purposes, it is important to separate QCA, the inferential method, from the concept of fuzzy sets. While the utility of QCA is the subject of debate (including contributions to this newsletter), fuzzy sets, at least as a descriptive device, are considerably less contested. Still, much of the measurement technology for fuzzy sets (at least for social science) has developed in the context of fs/QCA, and so it makes sense to start there.

The appeal of fuzzy sets, as summarized by its proponents, is clear:

> With fuzzy sets, it is possible to have the best of both worlds, namely the precision that is prized by quantitative researchers and the use of substantive knowledge to calibrate measures that is central to qualitative research. (Ragin 2008: 82)

Assigning fuzzy set scores to cases is challenging, no matter how one does it. Charles Ragin has offered a transparent approach, which seems to have some currency among fs/QCA scholars. Ragin identifies two related methods, which he labels "direct" and "indirect" (Ragin 2008). Both, at least in his examples, build on continuous measures of an underlying concept. In his classification of countries into the set of "developed" countries, for instance, he uses a continuous base measure of GDP/capita.

For the direct method, researchers "calibrate" the measurement by identifying three "anchor points" in the base continuous measure: the points at which a case reaches (1) full membership, (2) full non-membership, and (3) the crossover point between membership and non-membership. Researchers use these values to sort cases into one set or another and compute scores between 0 and 1 by transforming deviations in GDP/capita from the cross-over point with a log-odds function.[2]

The indirect method is similar to the direct method, except that the analyst codes each case with one of the following six membership scores, all of which reflect the level of membership in a target set—for example, developed countries (Ragin 2008: 84). The six scores would be (1) full membership, (2) mostly in, but not fully, (3) more in than out, (4) more out than in, (5) mostly out, but not fully, or (6) full non-membership. The analyst assigns each of these categories an equally spaced number between zero and one (1.0, 0.8, 0.6,...) and then regresses the scores on the base measure (GDP per capita) using a fractional logit model. The predicted scores thereby become the fuzzy set scores. In some fsQCA applications, the number of scores may be greater or less than six.

As is probably clear, both the direct and indirect approaches to creating fuzzy-set membership require some strong theoretical assumptions regarding the location of the calibration points. (All measurement approaches, of course, lean on theory to some degree in order to build the ship at sea, as it were). It seems likely that the location of these calibration points will vary significantly across researchers (descriptive heterogeneity) and, relatedly, will vary with respect to the relationship of the measure with other constructs (causal heterogeneity). fsQCA seeks to take context into account, but the relevant features of context can readily be well beyond the reach of any standard approach to contextualization.

Consider an everyday example of descriptive heterogeneity. For any individual, there is some noticeable and abrupt cross-over point between cold and hot. But this cross-over point will depend on whether one is a Texan or a Minnesotan, young or old, playing soccer or watching it from the stands, and a never-ending list of other factors. If asked, each observer would identify a different cross-over point based on their own perception of temperature. In such a case, is it helpful to have a fuzzy-set score that indicates to which category a certain temperature belongs? Does it make sense to say that the temperature has a fuzzy-set membership of 0.43? Perhaps, but it will depend crucially on an inter-subjective and inter-contextual agreement about the location of crossover points.

It seems implausible that scholars can agree on crossover points for a great many political variables of importance, such as democracy or economic development. Ragin is quite candid on this point:

> The collective knowledge base of social scientists should provide a basis for the specification of precise calibrations. For example, armed with an adequate knowledge of development, social scientists should be able to specify the per capita income level that signals full membership in the set of developed countries. However, the social sciences are still in their infancy and this knowledge base does not exist. (Ragin 2008: 86)

This acknowledgment of the difficult theoretical exercise of assigning calibration points for economic development and GDP per capita is telling. These two make for a well-known concept/indicator pair and uncertainty in this domain suggests that these decisions will be even more fraught in other domains.

With respect to *causal* heterogeneity, it is quite possible that base measures like GDP/capita are related to outcomes along different functional forms. Imagine, for example, that two outcomes interest a researcher: democracy and happiness. To the extent that there are discontinuities in either of the two GDP/capita-outcome relationships, it is likely that democracy and happiness "kick in" at different levels of GDP/capita.

[2] See Ragin (2008: Chapter 5) for more detail.

If this is the case, the problem of causal heterogeneity highlights the risky rescaling process inherent in both the direct and indirect approaches to fuzzy-set membership. If one agrees with Ragin that full membership is hard to establish, even for familiar, widely-studied phenomena like development and per capita income, then the complex gradations of full membership and non-membership for many other phenomena may be illusive indeed. Fuzzy-set measures may well add a layer of complexity to the continuous measure, without a corresponding gain in meaning. Without a useful calibration point, the fuzzy-set measure rescales the base measure into units that are no longer directly observable or meaningful. Compare a GDP/capita of $4500 to a fuzzy-set membership in the set of developed countries equal to 0.43. While 0.43 is some function of GDP/capita, it is no longer directly observable, not particularly meaningful, and unclear whether two researchers with different outcomes of interest will interpret the measure in the same manner.

To be fair, fuzzy set scores are not alone in their intangibility. Many measurement strategies involve rescaling observable scores, either by constraining their quantities between endpoints or constraining their distributional parameters (e.g., normalizing, with a mean of zero and a standard deviation of one). Researchers must therefore ask two questions. First, how much does the rescaling procedure reduce interpretability? Second, does the gain of rescaling outweigh the cost regarding interpretation?

In fs/QCA, the benefit is presumably that the anchor points have real meaning, indicating full membership, full non-membership, and the point between the two. And, in fact, Ragin sees this calibration procedure as comparable to the creation of a Celsius scale, in which zero and 100 degrees mean something real with respect to the effects of temperature on water. To the extent that these calibration points *do* have real intersubjective meaning, then perhaps moving beyond concrete units is a large benefit of rescaling. But, as discussed above, the tenuous nature of the assignment process, even for merely establishing full membership and non-membership, makes it difficult to believe that the calibration points are *actually* interpretable in any consistent manner.

Fuzzy set measurements also seem to have some difficulty, at least as they are specified, in fully representing the meaning of a systematized concept. Let's think of some examples of categories and cases that exhibit "boundary" problems: Olives (fruit), Poker (sport), and Duckbill Platypi (mammals). These cases induce categorical head-scratching because they share attributes with both co-classified cases and cross-classified cases. In each case, they have been categorized as such because researchers have preferred to privilege one dimension (respectively, seeds, competition, and mammary glands) over another. However, other secondary characteristics are associated with cases in each category (again, respectively: sweetness, athleticism, and internal gestation).

So, variance within categories derives in part from multiple, semi-related dimensions of the concept. A satisfying measurement strategy would be one that could represent and test the dimensionality of the category using multiple measures. Combining multiple indicators both to represent the concept more fully and to improve reliability is a virtue of most measurement models, and something on which fuzzy sets—as conventionally measured—fail to capitalize. Conventional fuzzy-set measures typically identify, and measure membership for, each set/dimension separately.

A Framework for Evaluation

The limitations of current fuzzy-set measurement practices cast something of a shadow on the use of such measures in analysis. As Ragin (2008: 71) himself notes, "the key to useful fuzzy-set analysis is well-constructed fuzzy sets." Unfortunately, the measurement challenges leave the method—by its own criteria—limited in its applications to the social sciences. One wonders whether more attractive solutions are available for measuring partial membership scores. But more attractive in what way? Here, the foregoing examination of fuzzy-set measurement practices can be helpful. Not only do these practices serve as a focused reference point for comparison, but the comparison suggests some useful criteria for evaluating such measures.

Of course, researchers will have different analytic and descriptive uses for partial membership scores. However, some basic concerns seem relevant to anyone who builds or uses partial membership—concerns that I express here in terms of three points of inquiry. Whatever else they do, helpful measurement strategies should be able to shed some light on one or more of these points.

1. **Homogeneity within Categories**. How much diversity is evident in the categories in question? That is, to what degree can an analyst make the claim that categories are sufficiently uniform?

2. **Conceptual Architecture of Categories**. Which attributes are responsible for potential heterogeneity within a category? That is, can we identify the multiple dimensions, or components, that structure a category and produce its diversity?

3. **Degree of Membership in Categories**. To what degree (with what probability) does a particular case "belong" to a given category? That is, can we assign membership meaningfully and generate useful partial membership scores?

Alternative Measurement Approaches and an Application

This exploration of partial membership methods is not an abstract exercise for me. I have devoted many years to the Comparative Constitutions Project (CCP), with the goal of describing the world's written constitutions, historic and contemporary, and testing theories regarding the origins and consequences of constitutional choices (Elkins, Ginsburg, Melton 2013). By reading and re-reading 800 constitutions, I have become acutely aware that the constitutional landscape of the world's states is not particularly well-conceptualized. Specifically, we lack a well-developed sense of how constitutions express rights, duties, powers, and principles and how these properties co-vary. Moreover, we do not seem to have provided those who are in the business of *writing* constitutions

with a helpful conceptual framework for understanding their choices.

Executive-Legislative Relations

One of—if not *the*—most important decisions that constitutional designers must make is how to structure the roles of the executive and the legislature. For well over a century, political scientists (and constitutional drafters) have conceptualized this choice predominantly as one between two basic types: presidentialism and parliamentarism. At least since de Gaulle, a mixed type—sometimes called semi-presidentialism—has entered the academic and political dialogue. Though other scholars have suggested additional intermediate categories (e.g., Shugart and Carey 1992), for the purpose of this analysis we will consider these three basic categories.

Perhaps *because* of its familiarity, this typology is highly relevant to our purposes. The definitional criteria are rarely in dispute. Presidential and parliamentary systems are usually distinguished by two related attributes—the procedures for the *election* and *survival* of the head of government. In a parliamentary system, the assembly selects the head of government, who serves at their pleasure; in a presidential system, citizens select the president, who serves for a fixed term.

In conjunction with these defining attributes, strong stereotypes have developed about what these types look like across a set of other secondary properties, which are often denoted by three prototypic systems: the U.S., the British, and the French. Mounting evidence, however, suggests that these three types mask great diversity in executive-legislative systems (Shugart and Carey 1992, Tsebelis 2002, Cheibub et al. 2013); many presidential systems take on parliamentary attributes and many parliamentary systems take on presidential attributes. How can we take these partial memberships into account?

Strategy One: A Continuous Scale Instead of Classification

A first potential strategy is to sidestep classification entirely—thereby deliberately setting aside these three questions. After all, the baseline conditions for measuring partial membership require that concepts exhibit some combination of "categoriness" and partial membership in such categories. Many concepts, it seems, exhibit one but not the other. The potential objection to measuring partial membership, therefore, is that a researcher either does not recognize any discontinuity in a variable or does not know where to draw the boundary lines.

In such a case, the dominant strategy is to build continuous scales that tap gradations on a dimension (or dimensions) of the particular concept. It is a familiar task for many scholars. Indeed, the goal in most measurement models is not to classify units, but rather to assign a score to the units across a continuous measure. Ideally, one would assemble a set of multiple measurement items that both represent the systematized concept adequately and improve the reliability of the overall score. A long tradition of measurement strategies, based principally on covariance structure modeling, allow for the construction and testing of such measures. But even simpler scaling techniques can deliver measures with graded scores that might satisfy researchers ostensibly interested in partial membership.

Thus, in the case of executive-legislative relations, a researcher may be tempted to eschew classification—even a classification system as familiar as the dichotomous presidentialism versus parliamentarism—and instead opt for a continuous scale that taps a *primary* dimension of the distinction. Some very accomplished scholars have built exactly these sorts of scales, which are central to the executive-legislature relationship (e.g., Shugart and Carey 1992, Fish and Kroenig 2009, Tsebelis 2002, Lijphart 2012). To unpack some of the issues involved in this sort of scaling, consider *another* scale: scope of executive authority (power) as encoded in written constitutions (Elkins, Ginsburg, and Melton 2013).

The data for this scale include measures of a comprehensive range of executive and legislative powers. However, constructing a valid scale from these characteristics requires special attention to aggregation and weighting. Because powers are substitutable, it is not especially meaningful to simply add powers, such as the executive veto or the executive's initiation of legislation. While executives may win in one arena (say, the budget), this may come only after threatening the legislature in another arena (say, military action); in general, power in one domain will likely provide clout in another domain.

The implication of substitutability for measurement is that these various powers will not necessarily correlate, which limits their utility in standard measurement models.[3] More precisely, these powers are not typical "reflective" indicators (manifestations, or reflections, of the latent construct); rather, they are "formative" indicators, in that they are causes of (or routes to) to a latent variable.[4] The set-up for formative indicators is, importantly, quite different from standard reflective models. The weighting for a given item in the scale is derived not from its intercorrelation with other items, but from its prediction of variables that are manifestations of the latent construct. So, in order to identify the measurement model, one needs to specify at least one reflective indicator (i.e., the outcome) of the latent variable along with the host of formative indicators (i.e., the potential causes).

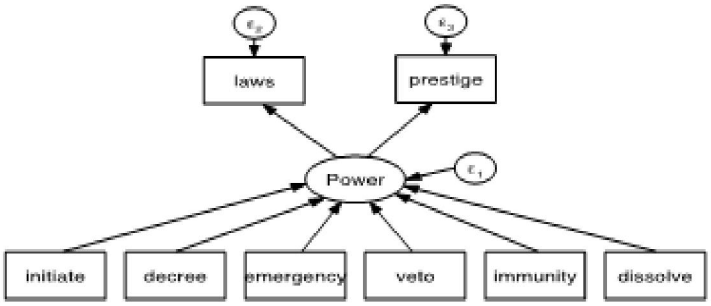
Figure 1 depicts the relationships we theorize in a measurement model of constitutional executive power. In measurement modeling terms, this structure is known as a MIMIC (Multiple indicator, multiple cause) model. The outcome variables are (1) the executive's success rate in passing bills that originate from the executive and (2) the relative prestige of the office (labeled as *laws* and *prestige* in the figure). The substitutable power variables (which, again, are treated as causal) all relate to a constitution's degree of executive power, shown in the six attributes in the figure.

[3] Standard measurement models develop weights based on an item's intercorrelation with other items.

[4] The idea of a latent variable is most useful if not over-interpreted—in the reified sense of an underlying phenomenon in the real world; but rather is treated as a metaphor for the conceptual understanding of how and why these items (i.e., characteristics) may be interrelated.

Figure 1: MIMIC (Multiple Indicators Multiple Causes

Measurement Model of Executive Power



Legend
laws = success rate of executive-initiated bills
prestige = expert coding of executive importance
initiate = executive empowered to initiate legislation
decree = executive empowered to enact legislative decrees
emergency = executive empowered to declare state of emergency
veto = executive empowered to veto legislation
immunity = executive granted immunity from prosecution
dissolve = executive empowered to dissolve the legislature

Clearly, this sort of measurement scheme sidesteps classification and is therefore not helpful in answering the three evaluative questions I raise above. However, it may also be possible to approximate, or at least to explore, partial-set membership using a continuous measure. One approach would be to test, iteratively, the effect of any discontinuities by employing a combination of continuous and dichotomous variables in a regression and identifying, diagnostically, any discontinuity in the association between the outcome and the continuum under consideration.

It may also be possible to combine the continuous measure with a related classification system in order to assess internal diversity and, even, approximate degrees of membership. So, in our example, one could pair a continuous measure of executive power together with an existing classification of parliamentary and presidential systems. Assuming that executive power constitutes a dominant dimension of presidentialism and parliamentarism, one may be able to describe degrees of category membership.

Of course, this approach is probably not too dissimilar from traditional fuzzy set practices, as described above, and carries with it some of the aforementioned limitations. For example, one must make the strong assumption that the continuous measure is a dominant dimension of presidentialism and parliamentarism. As we suggest above, it seems more likely that categories would be defined by a mix of traits and that a single continuum would describe only one of multiple dimensions of the concept. Still, it may be instructive to calculate the variation with respect to the one dimension (executive power) within categories (presidentialism and parliamentarism) and even to use the interaction of the two in analytic models.

One would want to be clear that these are not, strictly speaking, partial membership scores. Still the joint effect of membership and variation on a primary dimension of membership would amount to something close to partial membership. Nevertheless, one might more satisfactorily measure partial membership with a more multidimensional approach.

Strategy Two: Similarity-Based Measures of Family Resemblance

An approach more consistent with Wittgenstein's idea of family resemblance and Rosch's prototype analysis is to group (or at least measure similarities among) cases based on a set of relevant characteristics. As in various algorithms used in cluster analysis, the idea here is to calculate quantities of similarity or distance among cases in light of their scores on a set of presumably multidimensional characteristics. Further specifying a prototypical case allows researchers to calculate an explicit measure of degree of membership: the distance between each case and the prototype.

These kinds of methods are often employed in a more exploratory fashion. They are useful for identifying units that flock together, exploring alternative classification strategies, and discovering different types or "species." However, as I suggest above, it is is possible to calculate partial membership scores in a straightforward manner with these methods.

A first step might be to classify cases based on one or more definitional attributes. In the case of presidentialism and parliamentarism, such clasifications abound: here we use an authoritative coding by Cheibub, who classifies cases based on the selection and survival properties of the executive, as is conventional. The next step is to identify a set of secondary characteristics that are associated with membership in one or more of the classes. It is then possible to build partial membership scores for each case's "resemblance" to identifiable "families," based on their values on these secondary characteristics.

For example, Table 1 identifies seven attributes typically associated with presidentialism, semi-presidentialism, and parliamentrism. From these I calculate a simple measure of similarity (the Pearson correlation *between* cases and *across* the seven attributes) for each dyad in the data. The similarity to a prototypical case constitutes a measure of partial membership.

For the 108 constitutions included in this analysis,[5] Fig-

[5] The sample includes all independent states for which the constitution specifies executive-legislative relationships to a sufficient de-

Table 1: Characteristics of Executive-Legislative Systems

	System		
	Presidential	Semi-Presidential	Parliamentary
Assembly Confidence	No	For head of govt	Yes
Executive decree	No	Depends	Yes
Emergency powers	Strong	Strong	Weak
Initiation of legislation	Legislature	Depends	Executive
Legislative oversight	Yes	Depends	No
Executive veto	Yes	Depends	No
Cabinet appointment	Executive	Depends	Legislature

ure 2 depicts the distribution of the measure of similarity to the United States, grouping cases according to whether they are categorized as presidential, semi-presidential, parliamentary based on the defining attributes. A score of 1.0 would represent perfect similarity (with the same values on all component attributes); 0.0 would reflect the absence of any shared attributes.

One way to interpret these scores is as a measure of the degree to which cases belong in the "presidential" category, defined by the U.S. prototype. The results are startling, though they confirm something that institutional researchers (e.g., Shugart and Carey 1992) have suspected for years: there is enormous heterogeneity within the classic categories.

Specifically, and counter-intuitively, the mean similarity scores vis-à-vis the United States differ relatively little across the three system types. Indeed, Parliamentary systems are, on average, more similar to the U.S. prototype than are presidential systems—a rather shocking result. One could also analyze the variance, by category, of each of the secondary characteristics to determine which ones are more or less responsible for the lack of family resemblance. It turns out that all of these characteristics vary substantially (though to varying degrees) within the three categories (Cheibub, Elkins, Ginsburg 2013).

All of this to say that this measure appears to provide satisfactory answers to the three evaluative questions identified above. The scores allow us to assess the degree of heterogeneity and the sources of diversity within categories, and to measure degree of membership. One could even imagine further analyses in which one employed these family resemblance scores in statistical models that use the classic typology, perhaps by substituting the "degree of presidentialism" for a binary variable of presidentialism in a regression analysis.

So what do you do when you identify a family that does not appear to exhibit much resemblance among its members? Do you stop speaking of those families altogether? Do you reject the findings and seek other, overlooked, characteristics that are more related to the family line? Or do you use these new measures of familial distance to speak more accurately of the family's members—say, of siblings versus second, third, and fourth cousins? In the case of executive-legislative relations, the second and third responses seem worth pursuing. The family still matters: presidentialism and parliamentarism still connote remarkably important differences between systems.

Of course, it is possible that I tested family resemblances with characteristics that are not connected to the family's "DNA." To evaluate this potential problem, it is useful to revisit our expectations regarding the secondary characteristics and test the diversity question with *other* known concomitants of presidentialism and parliamentarism. In fact, it really does make sense to speak of close and distant relatives, given that in further testing we do not find new evidence of homogenous families. Also, we are convinced that the secondary characteristics originally tested really *are* attributes closely associated with the family. It is therefore plausible, as in Figure 2, to suggest that Liberia (LBR) is a close Presidential member, while Nicaragua (NIC) is something of a distant cousin.

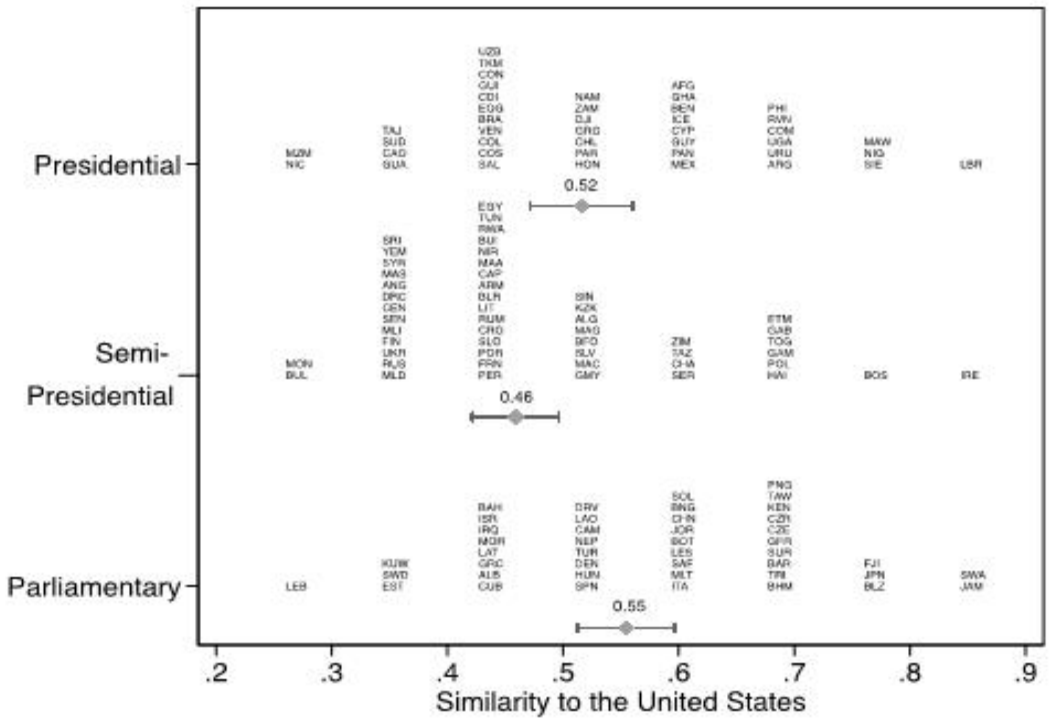
One will note that these families are only plausible if we are able to identify defining attributes with which to make such designations. That makes sense in the case of the presidential and parliamentary families. But what can we do if we are in a purely Wittgenstein/Rosch-like world (as opposed to a Linnean taxonomical world), in which we do not have the luxury of necessary and sufficient conditions to define categories?

Strategy Three: Latent Class Analysis

The cluster-analytic techniques of Strategy Two are illuminating, but they are for the most part exploratory. They lack the statistical properties that would justify more precise statements about the degree to which particular units belong to categories. Latent Class Analysis (LCA), by contrast, provides somewhat more precise answers in this regard. It is a productive third strategy, given that we are concerned with categorical distinctions. LCA is a version of cluster analysis in which one analyzes the attributes that characterize each category, and uses estimates of the clustering of categories to sort cases

gree as to allow classification with respect to the three systems.

Figure 2: Surprising Similarity to the United States of Presidential, Semi-Presidential, and Parliamentary Systems



Note: N=137 national constitutions as of June 2013. Table includes independent states with constitutions that specify procedures for executive selection and survival. A score of 1.0 signifies perfect similarity to the United States. Average similarity scores: Presidential 0.52; Semi-Presidential 0.46; and Parliamentary 0.55.

into appropriate groupings. This analysis can estimate the probability that cases belong to a particular category, as well as the association of the various items with each category.

In order to understand its uses, think of an everyday concept, *autism*, the diagnosis of which many modern parents puzzle over at some point. Like most of the learned concepts under consideration here, autism is highly multidimensional, and characterized by a variety of social, cognitive, and emotional symptoms. It is also regularly treated as a category with *partial* memberships within this multidimensional space (cases are said to be somewhere on the "spectrum"), but memberships seem to matter nonetheless. "Having it" triggers certain treatments, certain accommodations, and certain sympathies. But how to assign membership given these multiple continuous dimensions? LCA offers one approach, which not only allows for the estimation of membership in a single category (e.g., Autism) but also the estimation of membership in subcategories (e.g., Aspergers).

The parallel to executive-legislative relations is striking. Presidentialism, parliamentarism, and semi-presidentialism are as multidimensional and graded as any other category: however, the categories themselves matter. What can LCA tell us about these classes and their members? An initial answer is suggested by building on the characteristics listed in Table 1 and performing an LCA analysis. Table 2 reports a critical set of quantities from this analysis: the probability of membership in the three broad categories, for a selected set of countries. These memberships, then, are *graded* and *multiple*. At the same time, the probabilities suggest that cases may be assigned to one of the categories, based on the highest probability in each row (shown in bold in Table 2). The number of categories analyzed was fixed at three but that number can be permitted to vary and, like the number of dimensions in confirmatory factor analysis, should be subjected to close scrutiny.

The labeling of each category requires interpretation. In Table 2, I have assigned labels to the categories, based on (a) the clustering of cases; and (b) case scores on the defining attributes of the categories.[6] The results in Table 2 suggest some intriguing answers with respect to partial membership. In general, membership scores seem to corroborate those calculated in the cluster analysis above. The difference, now, is a much more precise sense of how and why they do and do not fit well. So, is Liberia presidential? Yes, unequivocally so. It belongs to that category with a probability of 0.91 and to the others at less than 0.30. Brazil, however, might just as easily be categorized as semi-presidential ($p = 0.56$) as presidential ($p = 0.60$). And so on.

With respect to the criteria identified, we might think of LCA as something like a more precise version of the family

[6] The conditional probability of the items for each category is not reported here. This probability essentially maps the relationship of the items with the categories.

Table 2: Latent Class Analysis: Probability of Membership in Three Derived Categories for Selected Cases

	Categories, with Interpretation Shown in Brackets		
	1 [Presidential]	**2** [Semi-Presidential]	**3** [Parliamentary]
Guatemala	**0.46**	0.32	0.22
Brazil	**0.60**	0.56	0.33
Peru	**0.87**	0.34	0.17
Liberia	**0.91**	0.28	0.12
Belarus	**0.68**	0.54	0.34
Ukraine	**0.56**	0.45	0.23
Russia	0.38	**0.65**	0.13
Denmark	0.21	0.39	**0.69**
Spain	0.15	0.34	**0.95**

Note: Numbers in bold indicate the highest category in each row. Substantive interpretation of categories is based on visual inspection of: (a) clustering of cases; and (b) case scores on the defining attributes of the categories.

resemblance measures in Strategy 2. That is, LCA allows us to describe the diversity within categories and to assign partial membership scores to individual cases. It also facilitates an investigation of the architecture of the various categories through an analysis of the correspondence between the various attributes and category membership (though that analysis is not shown here). The advantage over more informal clustering methods is a nuanced one. The LCA results, like those generated in the cluster analysis, have a tight connection to the idea of partial membership. The difference is that the LCA results have a stronger, or at least more widely understood, grounding in statistical and measurement theory.

Further Observations on Partial Membership in Categories

A final consideration that emerges from our evaluation of partial membership strategies concerns the interpretation of scores. Partial membership is not directly observable and the scores generated from any of the methods under discussion will be scientific constructs. Some of these constructs, however, are simply more meaningful than others. Typical fuzzy set methods rely upon a calibration approach that depends upon some rather aggressive assumptions about the location of set-membership boundaries, which then serve as reference points. Ultimately, it is not entirely clear what the scores surrounding these boundaries mean, exactly. 0.4 may mean that a case is slightly more out than in (if the set cutoff is 0.5), but that relative judgment is not itself particularly easy to grasp, or to convey to others.

In other methods, however, 0.4 may well have a more comprehensible, or at least more *established*, meaning. In a measure of family resemblance, that score—depending upon how similarity is measured—may mean that a case shares 40 percent of some group of characteristics with the category's prototype (if similarity is measured as percent matching) or that it correlates at 0.4 with the set's prototype (if similarity is measured as the correlation between two cases across their characteristics). 0.4 in a LCA model suggests that a case belongs to a particular category with a probability of 0.4. Any of these interpretations are just as unobserved as are those in the fuzzy set context. The difference is that these units are constructed as mathematical concepts (probabilities, correlations, percentage) that need no introduction and have well understood properties. Ultimately, that sort of resonance will be important, at least in a descriptive endeavor.

Conclusion

The idea of building taxonomies with partial membership is compelling. The idea makes even more sense once we understand insights from cognitive psychology about how our minds process stimuli. But how to operationalize the idea of partial membership? The concept of fuzzy sets is helpful, but the measurement tools associated with that approach in the social sciences, at least, are quite underdeveloped. Still, identifying the shortcomings of extant fuzzy-set measurement practices focuses our attention on some desirable properties of partial-membership measures—and establishes a basis for evaluation. Helpful measurement properties are found in some alternatives to fuzzy sets. In particular, clustering and latent-class analytic methods generate family resemblance scores that seem to deliver the punch that we expect from partial membership. The illustrations in the domain of executive-legislative relations help us describe and diagnose the bounded nature of some well-established categories, presidentialism and parliamentarism.

The illustration raises a set of questions about what, exactly, to do with partial membership scores. Description is a worthy end in itself. However, once acquainted with a set of valid partial membership scores, one will also be inclined to put partial membership scores to work in explanatory models. This essay does not tackle that problem. However, it is easy to imagine statistical tests in which a binary measure of, say, presidentialism is replaced with a family resemblance measure of that class. Indeed, the evaluation of partial membership in explanatory models would be a productive next step.

References

Cheibub, Jose Antonio. 2007. *Presidentialism, Parliamentarism, and Democracy*. Cambridge: Cambridge University Press.

Cheibub, Jose Antonio, Zachary Elkins, and Tom Ginsburg. 2013. "Beyond Presidentialism and Parliamentarism." *British Journal of Political Science*. Published first online, November 14, 2013. http://dx.doi.org/10.1017/S000712341300032X.

Collier, David and James E. Mahon, Jr. 1993. "Conceptual 'Stretching' Revisited: Adapting Categories in Comparative Analysis." *American Political Science Review* 87 (4): 845–855.

Collier, David and Steven Levitsky. "Democracy with Adjectives." *World Politics* 49 (3): 430-451.

Diamond, Jared. 1997. *Guns, Germs, and Steel: The Fates of Human Societies*. New York: Norton.

Elkins, Zachary. 2000. "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations." *American Journal of Political Science* 44 (2): 293–300.

Elkins, Zachary, Tom Ginsburg, and James Melton. 2013. The Comparative Constitutions Project. Datasource. http://comparativeconstitutionsproject.org/ (Accessed April, 22, 2014).

Fish, M. Steven and Mathew Kroenig. 2011. *Handbook of National Legislatures*. Cambridge: Cambridge University Press.

Harnad, Stevan R., ed. 1990. *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press.

Lijphart, Arendt. 2012. *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. New Haven: Yale University Press.

Murphy, Gregory L. 2004. *The Big Book of Concepts*. Cambridge: MIT Press.

Ragin, Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.

Rosch, Eleanor. 1975. "Cognitive Representations of Semantic Categories." *Journal of Experimental Psychology: General* 104 (3): 192–233.

Seawright, Jason. 2013. "Warrantability." Presented at the Annual Meeting of the American Political Science Association, Chicago.

Shugart, Matthew and John Carey. 1992. *Presidents and Assemblies: Constitutional Design and Electoral Dynamics*. Cambridge: Cambridge University Press.

Tsebelis, George. 2002. *Veto Players: How Institutions Work*. Princeton: Princeton University Press.

Wittgenstein, Ludwig, 1953 [2001]. *Philosophical Investigations*. Oxford: Blackwell Publishing.

Zadeh, Lotfi. 1965. "Fuzzy Sets." *Information and Control* 8 (3): 338–353.

Analyzing Interactions: Four Alternative Models

Bear F. Braumoeller
Ohio State University
braumoeller.1@osu.edu

A Car-Buyer's Guide

To a consumer of methods in political science, the act of choosing an appropriate model can resemble the process of buying a new car. The journal article that introduces the method generally plays up its strengths while giving short shrift to its potential weaknesses, and other users often have little incentive to dwell on its potential shortcomings. "Check out this year's new model!" the author seems to say. "It lets you make asymptotically unbiased estimates with fewer observations than your existing model—which," and here the voice drops to a whisper—"can provide *really* terrible answers in circumstances like these. And nothing could be simpler to use! Just download this Stata package and add a single line of code to your batch file."

Practitioners are generally looking for a tool to solve a particular problem, not an in-depth discussion of the pros and cons of a particular method or set of methods. They often don't stop to take a close look under the hood or to ask the hard questions. "Asymptotically unbiased, you say... but what about the precision?" "Ah, you say you want *precision*? Perhaps you'd like to take a look at this model over here...."

As a result, many scholars doing substantive research tend to hop from one flashy new model to another without fully exploring the capabilities and limitations of each. When next year's model comes along, they jump on that, every bit as disdainful of last year's methods as they were of those that came the year before. Should they come across a crosstab or a chi-squared test in a published paper, they shake their heads sadly at the author's methodological naïveté. It rarely occurs to them that the chi-squared test has been chugging along reliably for more than a century, while newer, flashier models have ended up in the ditch.

Rectifying this situation mainly involves more, and better, methods training for practitioners. In the short run, however, we can offer some straightforward advice to applied researchers to help get to the heart of the issue. Perhaps the most important of these is this: You rarely get something for nothing. More inferential oomph generally comes at a cost, and it is important to know what that cost is before adopting the model.

To illustrate this point, I will discuss four different ways to model interactions: fs/QCA, multiplicative interaction terms, a stochastic frontier model, and Boolean logit. They are located, roughly, on a spectrum between assumption-intensive (fs/QCA) and information-intensive (Boolean logit). Each has some advantages vis-à-vis the others, but in every case those advan-

Acknowledgement: I am grateful to David Collier, Kimberly Twist, Katherine Michel and Alisan Varney for comments on earlier drafts.