Contemporary Supreme Court Agreement: What is a Better Representation — Categorical (Classificatory) or Continuous (Scalable)?

Lawrence Hubert and Douglas Steinley

February 23, 2014

Introduction. The *New York Times* major headline for Saturday, July 2, 2005, read as follows: "O'Connor to Retire, Touching Off Battle Over Court". The lead story that was attached to this headline, by Richard W. Stevenson, began — "Justice Sandra Day O'Connor, the first woman to serve on the United States Supreme Court and a critical swing vote on abortion and a host of other divisive social issues, announced Friday that she is retiring, setting up a tumultuous fight over her successor." Our purpose here is not to expand on this particular statement, because most would agree, irrespective of political sentiments, that "tumultuous" may end up being quite a gross adjectival understatement. Our interests are in the particular data set that the *Times* also provided on that same day, quantifying the (dis)agreement among the Supreme Court Justices over the decade they were together. We will provide two analysis of these data, and invite others to contribute their particular insights with whatever methodology is uppermost in their repertoires for dealing with square and symmetric proximity matrices. And I'd expect we will also invite all of our future Multivariate Statistical Analysis classes to do the same for some of their applied homework projects.

The information in the data set given by the *Times* is provided as Table 1 in the form of the percentages with which the justices *dis*agreed in non-unanimous cases from the 1994/95 term through 2003/4. The dissimilarity matrix (where large entries reflect less similar justices) is given in the same row and column order as did the *Times*, obviously ordered from "liberal" to "conservative":

1: St: John Paul Stevens
2: Br: Stephen G. Breyer
3: Gi: Ruth Bader Ginsberg
4: So: David Souter
5: Oc: Sandra Day O'Connor
6: Ke: Anthony M. Kennedy
7: Re: William H. Rehnquist
8: Sc: Antonin Scalia
9: Th: Clarence Thomas

Table 1: Dissimilarities Among the Nine Current Supreme Court Justices

	St	Br	Gi	So	Oc	Ke	Re	Sc	Th
1 St	.00	.38	.34	.37	.67	.64	.75	.86	.85
2 Br	.38	.00	.28	.29	.45	.53	.57	.75	.76
3 Gi	.34	.28	.00	.22	.53	.51	.57	.72	.74
4 So	.37	.29	.22	.00	.45	.50	.56	.69	.71
5 Oc	.67	.45	.53	.45	.00	.33	.29	.46	.46
6 Ke	.64	.53	.51	.50	.33	.00	.23	.42	.41
7 Re	.75	.57	.57	.56	.29	.23	.00	.34	.32
8 Sc	.86	.75	.72	.69	.46	.42	.34	.00	.21
9 Th	.85	.76	.74	.71	.46	.41	.32	.21	.00

We present two analyses of the proximity data of Table 1: (a) a unidimensional scaling of the Justices including the estimation of an additive constant that we can apply to the proximities as an augmentation; (b) a hierarchical (or categorical) classification through what is called an ultrametric, also obtained through a least-squares search strategy. Both of these representations as a best-fitting unidimensional scale and a best-fitting ultrametric, are generated from methods presented in a forthcoming monograph by Hubert, Arabie, and Meulman (2006), and using the available open-source M-files (within a MATLAB environment) that come with this text. This monograph is scheduled to appear within the joint ASA-SIAM Series on Statistics and Applied Probability (*The Structural Representation of Proximity Matrices With MATLAB*).

Unidimensional Scaling. The unidimensional scaling task can be formally phrased as follows: given the $n \times n$ (in this case, a 9×9) proximity matrix $\mathbf{P} = \{p_{ij}\}$ from Table 1, we wish to find an additive constant, c, and a set of coordinates $x_1, \ldots, x_n$ to minimize the least-squares criterion

$$\sum_{i \neq j} (p_{ij} + c - |x_j - x_i|)^2.$$

The best-fitting result was obtained for the set of coordinates we give below, with a "caret" ($\hat{}$) generally used in the sequel to indicate the best estimate. The additive constant $\hat{c}$ = $-.218$, Note that the coordinates are ordered in exactly the same way that the *Times* ordered the Justices in Table 1 (and without loss of generality, the sum of the estimated coordinate values is set to zero): $\widehat{x_{St}} = -.346$; $\widehat{x_{Br}} = -.216$; $\widehat{x_{Gi}} = -.200$; $\widehat{x_{So}} = -.177$; $\widehat{x_{Oc}} = .062$; $\widehat{x_{Ke}} = .113$; $\widehat{x_{Re}} = .160$; $\widehat{x_{Sc}} = .302$; $\widehat{x_{Th}} = .302$;

If we normalize the least-squares criterion to provide what is usually called a "variance-accounted-for" (VAF) measure as follows:

$$\text{VAF} = 1 - \frac{\sum_{i \neq j} (p_{ij} - [|\hat{x}_j - \hat{x}_i| - \hat{c}])^2}{\sum_{i \neq j} (p_{ij} - \bar{p})^2},$$

where $\hat{p}$ is the mean off-diagonal proximity measure in $\mathbf{P}$, the value we observe is 98.0%, and pretty close to a perfect 100%. In other words, the unidimensional scaling provides a very good representation for the data of Table 1, with the O'Connor coordinate being .062 — this is the coordinate nearest to zero and obviously the median of the nine coordinate values over the Justices. If we wished, a table could be provided for the reconstruction of the dissimilarities among the Justices using the values, $\{|\hat{x}_j - \hat{x}_i| - \hat{c}\}$; a comparison directly to Table 1 would reflect the high quality of reconstruction.

Although the O'Connor coordinate is the median value among the nine locations, a graphical representation in Figure 1 clearly shows that she groups very closely with Kennedy and Rehnquist; the gap between O'Connor and Souter, her closest colleague to her immediate left is rather large indeed. So, in choosing the next Justice, an equivalence would be more toward a Kennedy/Rehnquist conservative rather than to a Scalia/Thomas conservative.

Hierarchical Classification (Clustering). Instead of relying on a set of coordinates (and their absolute differences) as in unidimensional scaling to represent the elements in a proximity matrix, a best-fitting ultrametric tries to give a representation by constructing a second matrix to approximate $\mathbf{P}$ (say, $\mathbf{U} = \{\hat{u}_{ij}\}$) minimizing the least-squares criterion

$$\sum_{i \neq j} (p_{ij} - \hat{u}_{ij})^2,$$

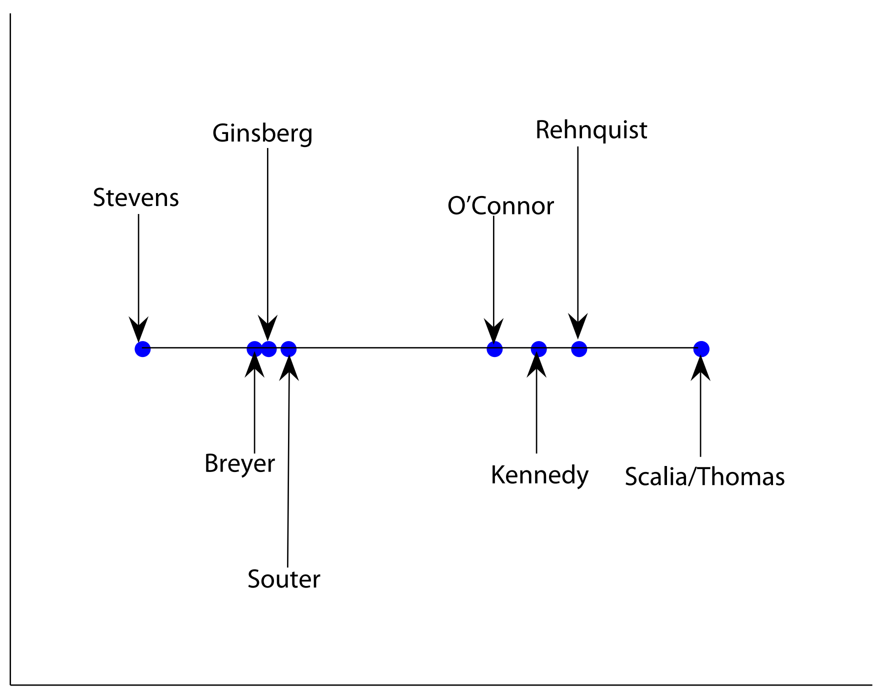
where the entries in $\mathbf{U}$ satisfy the ultrametric inequality: $\hat{u}_{ij} \leq \max\{\hat{u}_{ik}, \hat{u}_{kj}\}$ for all i, j, and k. In the case of the data of Table 1, the best-fitting ultrametric is defined by the eight nonzero distinct values that indicate how the hierarchical sequence of partitions is constructed:

Partition	*Level Formed*
{Sc, Th, Oc, Ke, Re, St, Br, Gi, So}	.641
{Sc, Th, Oc, Ke, Re}, {St, Br, Gi, So}	.402
{Sc, Th}, {Oc, Ke, Re}, {St, Br, Gi, So}	.363
{Sc, Th}, {Oc, Ke, Re}, {St}, {Br, Gi, So}	.310
{Sc, Th}, {Oc}, {Ke, Re}, {St}, {Br, Gi, So}	.285
{Sc, Th}, {Oc}, {Ke, Re}, {St}, {Br}, {Gi, So}	.230
{Sc, Th}, {Oc}, {Ke}, {Re}, {St}, {Br}, {Gi, So}	.220
{Sc, Th}, {Oc}, {Ke}, {Re}, {St}, {Br}, {Gi}, {So}	.210

The ultrametric representation has a less-adequate VAF (of 73.7%) than does the unidimensional scale. Figure 2 gives what is called a dendrogram for this best-fitting ultrametric. Note that the Justices are not ordered according to the unidimensional scaling in this particular graphical presentation. Although in this instance, the ordering of the justices in Figure 2 could have been given according to the unidimensional scaling, we did not do so to emphasize the basic "unorderedness" of the clusters implied by the construction of the ultrametric.

Conclusion. As a general summary, it appears that agreement among the Supreme Court Justices is better represented as a unidimensional scaling than as a categorical structure defined by a hierarchy of partitions through an associated ultrametric. Also, in terms of the identified Justice O'Connor, although she is placed in the middle of the scaling, there is a major tilt toward the conservative end in having a coordinate value very close to those

Figure 1: Unidimensional Scaling of the Justices Based on the Coordinates Given in the Text



for Kennedy and Rehnquist, and very discrepant from that for Souter, the colleague to her immediate left.

No other dissimilarity matrices were analyzed that might have resulted from different disaggregations, such as the type of case under consideration or the plurality of vote (e.g., in 5 to 4 resolutions). Also, we make no particular psychological interpretation of the strong unidimensionality observed in our aggregate analyses, or speculate about the underlying decision mechanisms. We obviously invite others to pursue these different analyses and interpretations.

Lawrence Hubert is Professor of Psychology and of Statistics, University of Illinois, Champaign, Illinois; Douglas Steinley is Assistant Professor of Psychological Sciences, University of Missouri, Columbia, Missouri.

Figure 2: Dendrogram Representation for the Best-Fitting Ultrametric Discussed in the Text