A Pattern Analysis of the Second Rehnquist U.S. Supreme Court

Lawrence Sirovich

Laboratory of Applied Mathematics
Department of Biomathematical Sciences
Mount Sinai School of Medicine
One Gustave L. Levy Place
New York, NY 10029

Pages: 20 pages
Figures: 1
Tables: 4

Word and Character Count:
Abstract: 134 words
Character Count: 30,414

Classification: Social Science: Political Science

Correspondence: Lawrence Sirovich, Laboratory of Applied Mathematics, Biomathematical Sciences Dept., Box 1012, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY 10029. Tel: 212-241-3994, Fax: 212-426-5037, email: chico@camelot.mssm.edu .

Abstract

The second Rehnquist Court has remained unchanged in composition for eight years, resulting in a large, temporally stable, database. This paper reports on a mathematically objective analysis of this ensemble of rulings aimed at extracting key patterns and latent information. Although the rulings of a nine-justice court requires representation in nine dimensions, smaller spaces describe the court's actions. E.g., a two dimensional subspace describes the margins of all decisions, and use of Shannon information shows that the court acts as if composed of 4.68 ideal justices. Comparison is also made with the 1959-61 and 1967-69 Warren courts. Both of these Warren courts have remarkable parallels with the Rehnquist Court. In each instance we present an optimal mapping of the justices between the courts which underscores the similarity in the workings of seemingly dissimilar courts.

1. Introduction

The 'second Rehnquist court' begins with the Supreme Court appointment of Stephen Breyer, by William Jefferson Clinton, on August 3, 1994. Since then, the nine Justices, which comprise the court, have remained the same. To the extent that the decision-making process of an individual justice does not change with time, the present Court has been temporally stable for more than eight years. The last time the Court could boast of comparable stability was in 1823. (Linda Greenhouse, New York Times, Oct. 6th, 2002, sec. 4, p.3, column 1)

The present court hands down roughly eighty cases per year. We approach this relatively abundant database of decisions in the spirit of a physicist or an applied mathematician, and seek to find structural patterns and latent information. Singlular Value Decomposition (SVD)[1], a key tool in our investigation, has furnished objective and mathematically optimal pattern and information in diverse scientific areas[2].

Our view is that the ensemble of rulings may be regarded phenomenologically, without reference to the merits of the corresponding case issues. Students of the Court may believe, with some justification, that this is 'throwing out the baby with the bath water'. However, avoidance of underlying legal issues is dictated by the author's (lack of) background in such matters. It is hoped that the treatment of the data by an objective observer from another discipline offers value. Neither mathematical voting strategy [3] nor judicial analysis[4] will play a role here.

Data

Supreme Court cases and decisions can be located on a number of websites.*
Although about eighty cases are handed down annually , other considerations reduced the case selection to roughly 70% of these, which we term *admissible*. Nearly 30% of the cases were discarded because the vote was incomplete or ambiguous (*per curiam*, 'by the court', decisions furnished no details of the vote and were deemed *inadmissible*, as were cases in which a justice was absent or voted differently on the parts of a case). The two guides by Kermit L. Hall[5,6] were valuable sources for understanding the data, as were popular accounts of the Court by Rehnquist[7], Starr[8] and others.

2. Geometry of Decision Space

To quantify the decision making process the justices are arranged in alphabetical order:

R=[Breyer, Ginsburg, Kennedy, O'Connor, Rehnquist, Scalia, Souter, Stevens, Thomas]

In obvious notation, a vector of nine entries specifies a decision

$$\boldsymbol{n} = [n_B, n_G, n_K, n_O, n_R, n_{Sc}, n_S, n_{St}, n_T]\,. \tag{1}$$

Each entry n_i can take on the value of ± 1 depending on agreement. For example

$$\boldsymbol{U} = [1,1,1,1,1,1,1,1,1] \tag{2}$$

specifies the unanimous decision. Another example is

* Two sites used by me are http://supct.law.cornell.edu:8080/supct/ and http://www.findlaw.com/casecode/supreme.html.

$$\boldsymbol{P} = [-1,-1,1,1,1,1,-1,-1,1], \qquad (3)$$

the 5-4 majority characterizing, perhaps, the most famous decision of this Court, *viz*., that handed down in connection with the 2000 US Presidential election.

In total there are $2^9 = 512$ possible decisions that a full court of nine justices might render. In keeping with the decision not to consider issues, we associate +1 (-1) with a vote that agrees (disagrees) with the majority. This reduces the possible decisions by half to $2^8 = 256$. For later reference note that the *margin* by which a majority is carried is restricted to the first 5 odd integers M = 1 (5-4), 3 (6-3), 5 (7-2), 7 (8-1), 9 (9-0). In geometric terms the ensemble of decisions are embedded in 9-dimensional Euclidean space, and which is restricted to the half-space $M = \sum_j n_j > 0$. Each decision belongs to the locus $\sqrt{\sum_j n_j^2} = 3$, the *decision sphere*. Each decision, (1), is a lattice point that lies on the decision sphere.

Court Models

Progress in the physical sciences has proceeded in large part by the invention of simpler, but adequate models. In the less exact sciences where matters tend to be more complex and messier, mathematical models are viewed with suspicion[†]. Nevertheless for comparison purposes it will be useful to introduce two idealizations.

Omniscient Court. Under this idealization each justice is all-knowing and omniscient and therefore each always makes the *right* decision. Further, since all the judges are equally *god-like* each opinion will be unanimous, and given by ***U*****,** (2).

[†] In keeping with the spirit of this paper we do not consider models based on psychology, behaviour, economics and other such considerations[9]. (Added in proof: A quantitative approach to such issues can be found in : Martin, A.D. & Quinn, K.M. (2002) *Political Analysis* **10**,134-153)

Although *court space* is nine dimensional, in this idealization a one dimensional subspace suffices. (Justices are clones, and only one is needed).

Platonic Court. Under this idealization each justice is free of ideology and sees equally compelling arguments on both sides on each issue. From the point of view of an outside observer the vote of a platonic justice is as predictable as the toss of a fair coin. Under this construct all nine dimensions are necessary to specify decisions that are handed down and all 256 possible decisions are equally likely‡.

Decisions handed down under the two idealizations can be characterized by means of Shannon's definition of information[11]. In the present context this states that if $\{p_n\}$ represents the probability set of possible outcomes, then the information (entropy) conveyed by a decision is

$$I = -\sum_n p_n \log_2 p_n \;, \tag{4}$$

where the logarithm is base two. Information is said to be measured in bits. I is also said to measure the *surprise* or novelty of an outcome. For the omniscient court there is just one outcome, which therefore has probability unity and $I_u = 0\,\text{bits}$ for this court. There is zero surprise or novelty, since the outcome of the judicial issue does not figure in our deliberations. On the other hand the platonic court has 2^8 possible outcomes, all equally probable, and therefore, in agreement with (4), $I_P = 8\,\text{bits}$ of information are revealed when an opinion is handed down. More generally we will take $I+1$ as determining the *effective* number of justices in the operation of the court.

‡ Although "Platonic" is being used in the sense of lofty or idealistic, mention might be made of "Platonic solids", which refer to the five perfect solids in 3-space, of which one is the cube. In 9-dimensions regular polytopes, the generalization of Platonic solids, are three in number, of which one is the (hyper) cube with 2^9 vertices given by (1)[10].

3. The Second Rehnquist Court

Statistics on court decisions can be found in a number of locations[12,13] For example the Harvard Law Review furnishes tables on voting alignments and average actions of individual justices on a term basis. The 'Law Review' includes among its concerns the opinion making process, e.g. their tables do "not treat two justices as having agreed if they did not join the same opinion even if they agreed in the result". For present purposes such distinctions will be overlooked. Our sole criterion will be whether a justice does or does not join in the Court's opinion.

The ensemble of decisions for the eight year period 1995-2002 derives from the 468 *admissible* opinions. Nine dimensional decision space contains just 256 points (1), and some decisions therefore must be visited more than once. Table 1 accounts for the 12 most frequent decisions, 377 in number. All other decisions occurred less than 1% of the time.

Unanimous decisions denoted by ***U***, occur 47% of the time, while the particular five-four majority, denoted by ***P***, account for almost 10% of the rulings. Next most frequent accounting for almost 4.5% of the decisions is noteworthy. Of the nine possible eight-one decisions (margin 7), in which one justice dissents, justice Stevens was the sole dissenter 21 times. Rehnquist, Scalia and Thomas were each sole dissenters three times, and Breyer, Ginsberg, Kennedy and O'Connor, each, sole dissenters just once. Justice Souter was never the sole dissenter: the decision $[1,1,1,1,1,1,-1,1,1]$ was never visited in the eight-year period. In fact, a total 181 possible decisions were never visited in the course of the eight-year period, 1995-2002. Furthermore, 45 decisions were visited only once, 0.2% of the time, and if we regard these as (ignorable) outliers, just 30 decisions can be regarded as significant in the sense that they were visited more than once.

On the basis of the probabilities of occurrence of the decisions of the present court we can calculate the information, (4). From Table 1 and the additional vote probabilities, it is determined that each time a ruling is handed down by the present Court, $I_R = 3.68$ bits of information is conveyed. This value lies between $I_O = 0$ bits and $I_P = 8$ bits, and implies that in effect the court acts as if composed of 4.68 Platonic justices.

Venturing into interpretation, one might suppose that in the 220 unanimous decisions, some (abstract) threshold was not reached so ideology did not play a role and, the Court then behaved according to the Omniscient model. Another supposition (not in conflict with the first) might be that the justices did not really rise to omnipotence in the ***U*** cases, but that these cases were "no brainers", which in a more efficient system would not have reached the Supreme Court. To pursue this further requires the reasons, possibly manifold, why the Court decides to rule on what will become a ***U*** decision.

Some insight into the likelihood of decisions can be gleaned from the joint probabilities that two justices will agree in a decision. Since Table 1 informs us that any two justices agree at least 47% of the time, joint probabilities are displayed in complementary form, *viz*., the probability that two justices disagree. This is shown in Table 2. Thus , the least probable event is that justices Scalia and Thomas disagree, 6.6%, and the next most unlikely event is that justices Ginsburg and Souter disagree, 9.6% of the time.

Alternatively, justices Scalia and Thomas agree more than 93% of the time, and justices Ginsburg and Souter more than 90% of the time. Column sums, shown, are an index of dissent. Thus, Judge Stevens is the most likely to disagree with the other justices, while justices Kennedy and O'Connor are the most likely to be in agreement with their colleagues. The total number of dissents of each justice, is given by $\boldsymbol{D} = [88,100,53,52,78,107,87,136,102]$, under the convention given in (1). Thus justice Stevens cast, by far, the most, while justices Kennedy and O'Connor cast the fewest dissents. Another interpretation of the latter remark might be that Kennedy and O'Connor are the likeliest to determine the majority opinion, a view supported by Table 1. Of the 72 margin 1 (5-4) decisions shown there , one or both of these justices might be regarded as casting the deciding vote.

Singular Value Decomposition

Each decision has been depicted as a point in 9-dimensional *court* space, (1), but this may not be the *best* representation. By well-defined mathematical criteria SVD furnishes the optimal coordinate system with which to view data. What this means will become clearer in the following. More detailed mathematical considerations appear in the Appendix.

The ensemble of all decisions can be put into the form of a matrix

$$(\boldsymbol{S})_{ij} = S_{ij} = n_j(i) \tag{5}$$

where the rows i = 1,2,..,468 index the decisions and the columns, j = 1,2,..,9, follow the convention adopted in (1) for voting. An SVD analysis subsumes the calculation of all correlations of voting alignments among all justices. This information generates new coordinate directions, dictated by the data, which are ordered by degree of importance. The first direction reflects the most frequent voting alignment, the next direction is the second most likely alignment, under the condition, that it be orthogonal to the first, and so forth, thus leading to a full set of *characteristic* directions or vectors.. It is also conventional to specify directions by vectors of unit length. Thus if we denote the *characteristic vectors* by $\{\boldsymbol{V}_j\}, j = 1, 2, .., 9$, and if the elements of $\boldsymbol{V}_j$ are denoted by $V_j(k)$ then $\boldsymbol{V}_i \cdot \boldsymbol{V}_j = \sum_{k=1}^{9} V_i(k) V_j(k) = \delta_{ij}$, where the inner product, denoted by a dot, is defined by the summation and δ_{ij} is zero for $i \neq j$ and 1 for $i = j$. These characteristic vectors, ordered by decreasing importance, are the columns of Table 3. Any decision, **n**, can be *exactly* expressed in these terms by $\boldsymbol{n} = \sum_{j=1}^{9} (\mathbf{n} \cdot \boldsymbol{V}_j) \boldsymbol{V}_j$.

Above each vector (column) of Table 3 is the weighting, w_j, j = 1,..9, this gives the probability with which a decision lies in the corresponding direction $\boldsymbol{V}_j$, and hence measures its importance. The third highest probability, w_3, is more than an order of magnitude smaller than w_1, which implies that we might approximate *decision space*, as embodied by ***S***, by just two directions,

$$\boldsymbol{n}(k) \approx \sum_{j=1}^{2} (\boldsymbol{n}(k) \cdot \boldsymbol{V}_j) \boldsymbol{V}_j = \boldsymbol{n}_2(k), k = 1, ..., 468. \tag{6}$$

The implication is that decision space of the Rehnquist court requires only two dimensions for its description. If true, the *will of the Court* is embodied in the space spanned by the first two columns of Table 3. ($\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are close, but not the same, in direction as ***U*** and ***P***; correlation between $\boldsymbol{V}_1$ and ***U*** is .996, and between $\boldsymbol{V}_2$ and ***P*** .949, the latter implies an angular separation of roughly 18°.) This implies that each justice's vote can be regarded, up to a sign, depending on agreement or disagreement, as a fixed admixture of two voting patterns, $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$. Each vote in this approximation represents a balance of these two basic voting patterns.

As a criterion for the evaluation of the two dimensional approximation, (6), we calculate the margin by which a majority is carried. The tenth column of Table 1 gives the true margin, which for the k^{th} decision is $M(k)=\sum_{i=1}^{9} n_i(k)$.

The two-term approximation to the margin is

$$M_2^{(k)}=\sum_{i=1}^{9}\sum_{j=1}^{2}\left(n(k)\cdot \boldsymbol{V}_j\right)V_j(i), \tag{7}$$

where $V_j(i)$ is the i^{th} component of $\boldsymbol{V}_j$. The elements of $\boldsymbol{V}_j$ have decimal form, which is awkward, and we round (7) to the nearest integer,

$$\tilde{M}_2 = \text{round}\ (M_2)\,. \tag{8}$$

If we carry out this calculation and form the difference, $M(k)-\tilde{M}_2$, we find that this is *zero in all but four of the 468 cases*. By this criterion, (6) is an excellent approximation. (In the appendix we demonstrate, that this goal of "goodness of fit" in fact can be used as a criterion for generating the characteristic directions or votes.) The exceptional cases are shown in Table 4.

The middle case in Table 4 occurred twice. In two cases rounding gives a margin of 2, and the other a margin of 4 violating the rule that the margin must be an odd number. In each the error is small enough to preserve the correct outcome. In two instances of Table 4, Rehnquist breaks with Scalia and Thomas, and in the other, Breyer breaks with Ginsburg and Souter. As implied in Table 2, these are low-probability occurrences. Alternatively, two of the votes were visited once and the other twice. The two-term approximation, (7), is not expected to approximate the class of unvisited decisions, as discussed in the Appendix.

4. Comparison with Two Warren Courts

The analysis just presented implies that the US Supreme Court functions in a subspace smaller than 9-dimensional space. Over the eight-year period, followed here, only a small fraction of the 256 possible decisions was visited. Information theory implies that the court operates in effect with 4.86 ideal justices. Decision margins suggest that an essentially two-dimensional description expresses the will of the court.

It is therefore of interest to make comparisons and we consider the Warren Courts of 1959-1961 and 1967-69.

A principal reason for choosing the Warren Courts for comparison is the widely held view of a strong contrast in the inclinations of the Rehnquist and Warren courts[8]. Amongst other distinctions, the present court is said to be conservative (versus liberal) in Constitutional interpretation, and inclined toward weak (versus strong) federal controls. In point of fact the first of the Warren courts was not really that dissimilar ideologically, and, as will be seen, was astonishingly similar to the Rehnquist court. The greater contrast is really with the 2nd Warren court, where interesting parallels again exist, but are of a more subtle sort.

The 1959-1961 Court was composed of:

W1= [Black, Warren, Stewart, Clark, Whittaker, Harlan, Brennan, Douglas, Frankfurter]

(For reasons that will become clear, these justices have not been arranged in alphabetical order.)

For the indicated period, in which these justices served, we obtained 233 (out of 650) admissible decisions.§

This ***W1*** court voted unanimously about 33% of the time in contrast to the 47% of the Rehnquist court. The 5-4 majority, earlier denoted by ***P*****,** was visited almost 13% of the time, which is comparable to the 10% for the Rehnquist Court. Clearly, a reason for the above arrangement of the Warren Court justices, ***W1*****,** was to facilitate this comparison. However, there are 2880 permutations of the entries of ***W1*** that give ***P***, the second most frequent Warren Court ruling. The actual choice of ordering of ***W1*** was dictated by the subsequent SVD. This analysis again reveals that there are just two dominant voting directions, denoted by $\boldsymbol{V}_1^{W1}$ and $\boldsymbol{V}_2^{W1}$. The ordering in ***W1*** was chosen to maximize the correlation of these with their counterparts from the Rehnquist court. In fact with our choice we find that $\boldsymbol{V_1} \cdot \boldsymbol{V_1^{W1}} = .994$ and $\boldsymbol{V_2} \cdot \boldsymbol{V_2^{W1}} = .986$, exceptionally high correlations. The chosen ***W1*** provides a mapping of the justices of the two courts, ***R↔W1*** that reveals a similarity in their complexions, and in the workings of the two courts**.** Inspection of ***R↔W1*** makes for some curious identifications which we don't comment on.

A difference in the courts is that 73% of the variance is captured by the first two components for the Warren Court, in contrast to the 79% for the Rehnquist court. As a result a two-dimensional approximation for Warren court decisions does not do as well as for the Rehnquist court. Eight minor errors, of the previous sort, now occur in calculating the 2-term approximation to the margin, a 3.5% error compared to the 1% error of Table 4. Along similar lines the Warren court visited 56 different

---

§ This period contained a substantially larger fraction of *per curiam* and ambiguous cases than the Rehnquist collection.

decisions in handing down 233 decisions, proportionally more, by 33% than the 75 visited by the Rehnquist court in their 468 decisions. Finally, the *novelty* of a ruling by this court was $I_{W1} \approx 4.1632$ bits, larger than $I_R = 3.68$ bits, implying that this Warren Court operated in effect with 5.16 Platonic justices.

Certain other similarities also might be of interest. The third most frequent vote, 4.7%, was the eight-one majority in which Douglas was the sole dissenter, closely paralleling the eight-one majority, 4.5%, occurring when Stevens was the sole dissenter. In total, Douglas dissented about 35% of the time, while Stevens did so about 30% of the time. Another interesting parallel, already foreshadowed by the ***R***↔***W1*** mapping, is that Clark and Stewart each voted with the majority about 86% of the time, thus playing a role similar to that of Kennedy and O'Connor who voted with the majority about 90% of the time.

The 67-69 Warren Court was composed of

W2=[White, *Stewart*, *Warren*, Marshall, *Brennan*, Fortas, *Black*, *Harlan*, *Douglas*]

The six carryover justices have been italicised. For the period in which the court was composed of these justices, we obtained only 85 admissible cases. (In addition to what has been said before, absences of one or more justices was a very significant factor for this period.) Again ***U*** was the most frequent vote, 40%, however,the second most frequent vote, 7%, was a 7-2 vote in which Harlan and Stewart dissented, and the third most frequent vote, 6%, was a 6-3 vote in which Black, Harlan and White dissented. 29 decisions were visited by this court, which is proportionately more than the ***R*** or ***W1*** courts. *A 5-4 majority occurred infrequently, only 4 times, and each occurrence had a different composition*. In spite of this divergence with the ***R*** and ***W1*** courts, SVD analysis reveals two dominant components, $\boldsymbol{V}_1^{W2}$ and $\boldsymbol{V}_2^{W2}$, which capture 73% of the variance. $\boldsymbol{V}_1^{W2}$ is similar to a ***U*** vote and even though single 5-4 decisions were seldom or not visited, $\boldsymbol{V}^{W2}$ is of this the general form (with Black, Harlan, Stewart and White in the minority). In fact $\boldsymbol{V}_1 \cdot \boldsymbol{V}_1^{W2} = .988$ and $\boldsymbol{V}_2 \cdot \boldsymbol{V}_2^{W2} = .919$, so that $\boldsymbol{W}_2 \leftrightarrow \boldsymbol{R}$ is well correlated. Thus, from the perspective of voting margins each vote that was cast again appears as an admixture of a ***U***-like and a ***P***-like pattern. For this Warren court $I_{W2} \approx 3.717$ bits, and 4.717 platonic justices.

It is both diverting and interesting to directly compare the two Warren Courts, ***W1*** and ***W2***. In fact the mapping $\boldsymbol{W1} \leftrightarrow \boldsymbol{W2}$ which is the optimal transformation of the two courts, shows that each carryover justices play new roles in the altered court. Metaphorically (and literally) under this map *black* goes to *white*. Douglas, Warren and Brennan, all belonging to the minority in 5-4 decisions of ***W1*** become part of the '5-4' majority of ***W2***, and Harlan, Stewart and Black go from the majority to the minority. Harlan went from dissenting 20% to 34%, while Douglas went from 35% to 14%. A very remarkable aspect of the ***W2*** court is the fact that Marshall dissented only once (which might be interpreted as indicating a leadership role), and Warren and Brennan did so only three times each.

5. Comments

The three courts we have focused on all share the feature that their decisions, in terms of margins, are well described by a two dimensional space that bears a strong correlation to ***U*** and ***P***. At the risk of extrapolating from *small statistics*, one can speculate that the strong correlations of these dominant patterns might, in part, be dictated by a sameness in the overall quality of cases percolating up to the Court through the judicial substructure; and also perhaps, a dynamic that is generated by the court size itself.

In another vein both SVD and information theory suggest that court *coalitions* reduce the dimension of the court from its potential of 9. The information dimension, which is the better measure of judicial independence, appears to lie between 4.5 and 5. While this is much smaller than 9, it is significantly higher than a dimension of one, which would be the case if all decisions only depended on a liberal versus conservative *axis*. By contrast in considering the U.S. Congress, Poole and Rosenthal[14] demonstrate, by different methods, that each of our (435) congressmen's vote is located in a two-dimensional space. (Also see Paul Krugman, New York Times, October 20, 2002, sec. 6, p.62, col. 1.) From the perspective of information theory there is much less 'novelty' in the outcome of a congressional vote than there is in a Supreme Court decision. Information theory, (4), states that the former is potentially enormously larger than the novelty of the latter. The notion of novelty should be balanced by the observation that 9 monkeys, trained to flip coins, would render decisions on this basis having the highest novelty.

Appendix. We briefly review some elements of SVD in the context of this paper.

The matrix of decisions, $S_{ij} = n_i(j)$, is defined by (5). The search for a unit vector ***V***, such that $\|\boldsymbol{SV}\|^2$ is a maximum leads to the eigen (characteristic) equation, $\boldsymbol{S}^{\dagger}\boldsymbol{SV} = \lambda\boldsymbol{V}$, where $\dagger$ denotes adjoint. E.g., this generates the 9-orthonormal *characteristic* voting directions (eigenvectors) $\{\boldsymbol{V}_n\}$, shown in Table 3. The weightings given in that table are given by $w_j = \lambda_j / \sum_k \lambda_k$. All eigenvalues are non-negative, and the solution to the stated maximization problem is $\boldsymbol{V}_1$, if eigenvalues are arranged in descending order. $\boldsymbol{V}_2$ solves the same maximization problem, with the added constraint that $\boldsymbol{V}_1 \cdot \boldsymbol{V}_2 = 0$, and so on.

We can connect SVD to the demand that an approximate form, (6), of $\{n_i(k)\}$ should closely fit the true voting margin, $M(k)$. For this purpose it will suffice for us, to consider a one term approximation to $\boldsymbol{n}(k)$. If $\boldsymbol{V}$ is an as yet unknown unit vector, we approximate each $\boldsymbol{n}(k)$ by $\boldsymbol{n}(k) \approx (\boldsymbol{n}(k) \cdot \boldsymbol{V})\boldsymbol{V}$, where we have made use of the fact that projection gives the best coefficient. Therefore we demand that the average, over all *k*, of $\sum_{j=1}^{9}\left\{n_j - (\boldsymbol{n}(k) \cdot V)V_j\right\}$, for $\|\boldsymbol{V}\| = 1$ be close to zero. A straightforward minimization is complicated by the fact that each sum can be negative which can lead to an erroneous (negative) minimum. Therefore we replace this criterion with minimization of $\sum_j \left(n_j(k) - (\boldsymbol{n}(k) \cdot \boldsymbol{V})V_j\right)^2$, subject to $\|\boldsymbol{V}\| = 1$. But this summation is equal to $\|\boldsymbol{n}(k)\|^2 - (\boldsymbol{n}(k), V)^2$ and therefore minimization of the summation is equivalent to maximizing $(\boldsymbol{n}(k) \cdot \boldsymbol{V})^2$. But this is equivalent to maximizing $\|\boldsymbol{SV}\|^2$ which is just the condition that yields SVD.

Another approach to treating the data looks at the departure of each decision from the averaged vote. When the same SVD analysis is applied to the mean subtracted data, the procedure is called principal components analysis (PCA), Stewart[15]. For this reason the two terms are sometimes used interchangeably. Besides the awkwardness of speaking of departures from an average vote, the resulting PCA analysis is less efficient, it requires an additional characteristic vector to achieve the same margin criterion, (6). (This is due to the fact that the lead SVD direction and the mean are not sufficiently close.)

Next we consider the 2-term approximation $\boldsymbol{n}_2(k)$, (6), for the Rehnquist court. Figure 1 contains the projection of all points of the decision sphere onto the $(\boldsymbol{V}_1, \boldsymbol{V}_2)$ plane, where $\boldsymbol{V}_1$ is the vertical and $\boldsymbol{V}_2$ the horizontal direction. The coordinates of each decision are given by $\boldsymbol{n}_2(k) = (\boldsymbol{n}(k) \cdot \boldsymbol{V}_2, \boldsymbol{n}(k)\boldsymbol{V}_1), k = 1,..,256$. The semi-circles represent $R_2 = 2$ and $R_2 = 3$, where $R_2 = \sqrt{(\boldsymbol{n} \cdot \boldsymbol{V})^2 + (\boldsymbol{n}_1\boldsymbol{V}_2)^2}$. This is a nominal choice for an annulus in which $\boldsymbol{n}(k)$ is well approximated by $\boldsymbol{n}_2(k)$. The asterisks, *, mark the 18 most frequent voting alignments, accounting for 397, or

85%, of the rulings. The uppermost * corresponds to ***U*** and the rightmost to ***P***. The circles, $\circ$, denote the next 12 most frequent rulings, and pluses, +, denote once visited votes. The 181 dots, $\bullet$, represent unvisited voting alignments. There are some striking features of this figure and we comment on a few.

1. Many votes, $\circ$ and +, lie inside $R_2 < 2$, where $\boldsymbol{n}_2(k)$ is a poor approximation to $\boldsymbol{n}(k)$, but for which the approximate margin, (8), is correctly given. To account for this we recall that the margin is obtained by summing the individual votes of the justices, and so is like an average. The process by which characteristic directions are obtained, as discussed above, overlooks individual errors in favour of obtaining a good approximation to the average. The dashed arrow to the upper $*$ is significant, since the inner product of this vector, with a vector to any symbol gives $M_2(k)$, (7), of the corresponding ruling.

2. There are a number of once visited rulings, +, that lie well in the annulus $2 < R_2 < 3$. Four of these are 8-1 rulings and another is $[1,1,-1,-1,1,-1,1,1,-1]$. Analysis implies that such rulings are well within the framework of the court. The unanswered question is "why are there not more cases which produce these alignments?"

3. Similarly in considering the unvisited votes, dots, we find that $[-1,1,1,1,1,1,-1,-1,1],[-1,-1,1,1,1,1,1,1,1]$ and $[1,1,1,1,1,1,-1,1,1]$ are all within the annulus. Why were there no cases to produce such alignments when analysis suggests that they are within the court's framework?

References

1. Strang, G. (1988) *Linear Algebra and its Applications, Third Edition.* (Harcourt Brace Jovanovich, Publishers, San Diego).

2. Sirovich, L., Everson, R. (1992) Management and analysis of large scientific datasets *The International Journal of Supercomputer Applications,* **6**, 50-68.

3. Taylor, A.D. (1995) *Mathematics and Politics: Strategy, Voting, Power and Proof* (Springer-Verlag, New York).

4. Epstein, L. & Knight, K. (1997) *The Choices Justices Make* (Congressional Quarterly, Inc.),

5. Hall, K.L., Ed. (1992) *The Oxford Companion to the Supreme Court of the United States* (Oxford University Press, Inc.).

6. Hall, K.L. Ed.(1999) *The Oxford Guide to United States Supreme Court Decisions* (Oxford University Press, Inc.).

7. Rehnquist, W.H. (2002) *Supreme Court* (Knopf Publishing Group).

8. Starr, K.W. (2002) *First Among Equals: The Supreme Court in American Life* (Warner Books, Inc., New York).

9. Segal, J.A. & Spaeth, H.J. (2002) *The Supreme Court and the Attitudinal Model Revisited* (Cambridge University Press).

10. Hilbert , D. & Cohn-Vossen, S. (1999) *Geometry and the Imagination* (American Mathematical Society).

11. Pierce, R. (1980) *An Introduction to Information Theory* (Dover Press).

12. Quain, A. Ed., *The Political Reference Almanac, 2001-2002 edition* (Keynote Publishing Co.).

13. Harvard Law Review, The Harvard Law Review Association, Vol. 108, (November, 1994).

14. Poole, K.T. & Rosenthal, H. (2000)*Congress: A Political-Economic History of Roll Call Voting* (Oxford University Press).

15. Stewart, G.W. (1993) On the early history of the singular value decomposition, *SIAM Review* **35**, 551-566.

Acknowledgement: The author thankfully acknowledges his debt to Ellen Paley, who with humor, patience and intelligence took on the job of inputting the data, and suffered through all the revisions of this manuscript. Mention should also be made of colleagues at the NYU Law School, who allowed me use of the law library and made useful comments, as did many friends. I thank them all.

TABLE 1. Most Frequent Decisions of the 2nd Rehnquist Court

Br	Gi	Ke	O'C	Re	Sc	So	St	Th	Margin	Frequency
1	1	1	1	1	1	1	1	1	9	47% (220)
-1	-1	1	1	1	1	-1	-1	1	1	9.6 %(45)
1	1	1	1	1	1	1	-1	1	7	4.5% (21)
1	1	-1	1	-1	-1	1	1	-1	1	3.8% (18)
1	1	1	1	1	-1	1	1	-1	5	3% (14)
1	1	1	1	-1	-1	1	1	-1	3	2.6% (12)
1	-1	1	1	1	1	1	-1	1	5	2.4% (11)
-1	1	1	1	1	1	1	-1	1	5	1.9% (9)
1	1	1	-1	-1	-1	1	1	-1	1	1.9% (9)
1	1	1	-1	1	-1	1	1	-1	3	1.5% (7)
1	-1	1	1	1	1	-1	-1	1	3	1.3% (6)
1	1	-1	1	1	-1	1	1	-1	3	1.1% (5)

Table 2. Joint Probability for disagreement for the 2nd Rehnquist Court

	Br	Gi	Ke	O'C	Re	Sc	So	St	Th
Br	0	0.11966	0.25	0.2094	0.29915	0.35256	0.11752	0.16239	0.35897
Gi	0.11966	0	0.2671	0.25214	0.30769	0.36966	0.09615	0.1453	0.36752
Ke	0.25	0.26709	0	0.15598	0.12179	0.18803	0.24786	0.32692	0.17735
O'C	0.2094	0.25214	0.156	0	0.16239	0.20726	0.22009	0.32906	0.20513
Re	0.29915	0.30769	0.1218	0.16239	0	0.14316	0.29274	0.40171	0.13675
Sc	0.35256	0.36966	0.188	0.20726	0.14316	0	0.33761	0.43803	0.06624
So	0.11752	0.09615	0.2479	0.22009	0.29274	0.33761	0	0.1688	0.3312
St	0.16239	0.1453	0.3269	0.32906	0.40171	0.43803	0.1688	0	0.4359
Th	0.35897	0.36752	0.1774	0.20513	0.13675	0.06624	0.3312	0.4359	0
Dissent	1.86966	1.92521	1.735	1.74145	1.86538	2.10256	1.81197	2.40812	2.07906

Table 3. Characteristic Voting Vectors and Weightings for the 2nd Rehnquist Court

w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9
0.570925	0.218552	0.051724	0.040329	0.033949	0.025719	0.024656	0.019724	0.014422
V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
0.341083	-0.327401	-0.122481	0.242601	0.142766	0.818428	0.037204	-0.020455	-0.102976
0.332609	-0.367567	0.110215	0.073447	-0.427827	-0.193473	-0.212646	0.685597	-0.031553
0.363366	0.174192	-0.346174	-0.579431	-0.101478	0.039381	0.583564	0.165764	0.04661
0.362987	0.104212	-0.527083	0.382865	0.51364	-0.384338	-0.103697	0.103361	-0.000622
0.346547	0.304502	-0.221575	-0.255128	-0.335368	0.10267	-0.665052	-0.327666	0.018063
0.312679	0.403145	0.458959	0.168381	0.114938	0.115622	0.047714	0.122879	0.675836
0.348709	-0.3127	0.075215	0.295861	-0.366213	-0.300389	0.330139	-0.587015	0.0975
0.264479	-0.445911	0.34732	-0.511823	0.509594	-0.166624	-0.191138	-0.155812	0.018946
0.315957	0.405752	0.434937	0.102405	0.054563	-0.039616	0.109552	0.007046	-0.720612

Table 4. Aberrant Approximations. The middle now occurs twice.

Br	Gi	Ke	O'C	Re	Sc	So	St	Th	Margin	$M(k)-\tilde{M}_2$	$M(k)-M_2$
-1	1	1	1	1	-1	1	-1	-1	1	2	1.562244
1	1	1	1	1	-1	1	-1	-1	3	4	3.641294
1	1	1	1	-1	-1	1	-1	-1	1	2	1.607662

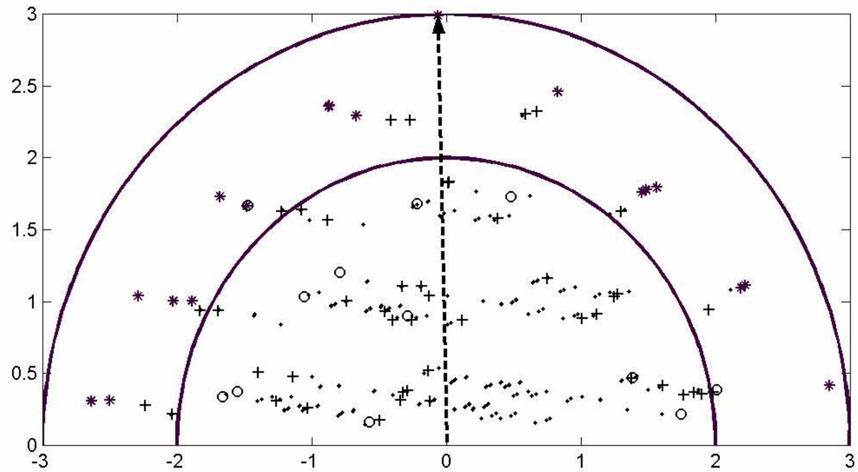
3
2.5
2
1.5
1
0.5
0
-3
-2
-1
0
1
2
3

Figure Legend

Figure 1: Two Dimensional Projection of the 2nd Rehnquist Court. $\boldsymbol{V}_1$ is the vertical and $\boldsymbol{V}_2$ the horizontal direction. $*, \circ, +$ represent visited rulings, and $\bullet$, unvisited rulings (see text). Inner product of heavy dashed vector with vector to any point generates $M_2(k)$, (7).