

The Usefulness of Quantitative Methods
in Political Science:
The Case of Scaling and the
United States Supreme Court

Section IV: Scaling

Jonathan Pool

Utilika Foundation

WRF Venture Center
2815 Eastlake Avenue East, Suite 300
Seattle, Washington 98102-3086
Telephone (206) 336-5566
<http://utilika.org>

The Usefulness of Quantitative Methods in Political Science:
the Case of Scaling and the United States Supreme Court

A thesis presented

by

Jonathan Robert Pool

to

The Department of Government
in partial fulfillment of the requirements
for the degree with honors
of Bachelor of Arts
Harvard College
March, 1964

members study not "constitutional law", but "judicial behavior"; and quantitative methods are among their principal tools.

IV

Scaling

Quantitative methods have been used to study many aspects of the Supreme Court, since C. Herman Pritchett pioneered in their use in 1948.⁸¹ The summary decision-making of the Court has been quantitatively analyzed; the relation of Justices' behavior to their previous backgrounds and affiliations has been studied; the content of opinions has been mathematically analyzed; the statics and dynamics of intra-Court politics have been the object of much attention; and this is but a partial list.⁸²

The attitudes of the Justices have been treated by quantitative methods perhaps more than has any other topic. Three principal methods have been used for this purpose: content analysis, Guttman scaling, and multivariate analysis. Content analysis has been comparatively rarely employed, mostly by one scholar, and it represents not a very radical departure in theory from

non-quantitative methods, so we shall exclude it from consideration. For reasons to be explained below, we shall exclude also multivariate analysis. The remaining method, Guttman scaling, also known as cumulative scaling, is by far the most widely used and well established quantitative method for the investigation of Justices' attitudes, but, because of its theoretical and practical difficulties, its use for this purpose is often attacked. Scaling thus provides an ideal subject for our attention. As we indicated above, we shall study first the theory of scaling and then its operation in practice.

The theory of scaling is rooted, not in an esoteric mode of thought, but in common sense. The fact that the judicial process is based on the adversary system itself suggests that judges might think about their cases as problems of choice—each alternative being to favor one of two parties over the other. When precedent and the belief in rule by law rather than by men are firm, the choices involve deciding what precedents, laws, and constructions of law should apply to the facts of the case, as well as what facts should be taken into consideration. It is not unreasonable, then, to believe that judges translate the inter-party conflicts that they

face into conflicts between values, in an effort to assuage the impingements of their own arbitrariness on the course of justice. Thus it is quite conceivable that the Justices of the Supreme Court regard many of their cases as contests between two values. Obvious examples of such conflicting pairs include individual freedom versus national security, protection for workers versus economic freedom for management, and decision on Constitutional dilemmas by the politically semi-representative and semi-accountable Court versus by the popularly elected Congress.

Suppose, then, that some of the cases before the Court are viewed by all of the Justices as involving only, or predominantly, conflicts between pairs of values. If we can isolate a group of these cases involving conflicts between the same two values, then new possibilities for analysis should be opened. On the one hand, it should be possible to rank the Justices according to the intensity of their favoritism toward one value over the other. If in a certain case one Justice votes for value A and another for value B, we immediately know who is more willing to favor A over B. By making several such comparisons we can order all nine Justices,

from the most fervent partisan of A over B to the most avid supporter of B over A.

In precisely the same fashion we can rank the cases as well. It is common to conceptualize that one of the values in a case makes a claim against the other. Thus, if a certain Justice supports value A over value B in one case, but gives his vote to B in another case, we can conclude that for this Justice, with his particular attitude toward A and B, A's claim on B in the first case was mild enough to earn the Justice's support, but A's claim became more severe in the second case—so severe, in fact, that the Justice was no longer willing to support it. A series of comparisons can be made between cases, if we consider that a claim supported by a Justice is less severe than one denied by the same Justice. In this way we can rank the cases according to the severity of the claim of A on B.

It follows from what we have said that a Justice with a strong partiality toward value A will vote for A in more cases than will a Justice with cooler feelings toward A in comparison with B. And in a case in which A's claim on B is mild, A will get more votes than in a case where the claim is severe. It also follows that,

given a group of cases conforming to our specifications*, the clearest way to express our conclusions is to construct a diagram, in which the Justices, ranked, lie on one axis, and the cases, also ranked, lie on the other, and in which, at each point of intersection between a Justice and a case, the Justice's vote in the case is recorded. Such a diagram is named a "scalogram". A model of a perfect scalogram is shown in Table 1.

Case:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Justice:										
J1	x	x	x	x	x	x	x	x	x	-
J2	x	x	x	x	x	x	x	x	-	-
J3	x	x	x	x	x	x	x	-	-	-
J4	x	x	x	x	x	x	-	-	-	-
J5	x	x	x	x	x	-	-	-	-	-
J6	x	x	x	x	-	-	-	-	-	-
J7	x	x	x	-	-	-	-	-	-	-
J8	x	x	-	-	-	-	-	-	-	-
J9	x	-	-	-	-	-	-	-	-	-

Table 1

Model of a Perfect Scalogram

Symbols: x = vote for claim - = vote against claim

*That is, all cases viewed by all of the Justices as two-value conflicts, involving the same two values for every Justice and every case.

In this scalogram, the cases have been ranked from the one with the least severe claim on the left, C1, to the one with the most severe claim on the right, C10. The Justices have been ranked from the one most sympathetic to the claim at the top, J1, to the one least sympathetic to it on the bottom, J9. As is seen, for each Justice and for each case there is a "breaking point". Each Justice has voted for the claim in all cases left of his breaking point, and against the claim in all cases to the right. In each case, all the Justices above its breaking point have voted for the claim, and all those below the breaking point have voted against the claim. The breaking point of a Justice represents the most severe claim that he will support. We have shown that, if the Justices vote on a group of cases in accordance with their attitudes toward the conflict between two certain values, we can construct with the voting data a perfect scalogram. In practice, however, we do not begin with the Justices' motives and conclude from them a scalogram. It is the voting data that we are given, and with which we can construct a scalogram, and from this we seek to discover the attitudes of the Justices. However, it does not necessarily

follow that, if we can construct a perfect scalogram, we can also make firm conclusions about the Justices' attitudes. The reason is that there are other possible causes of the existence of a perfect scalogram than the fact that our assumption about the Justices' voting is correct. Even by pure chance the votes might form a pattern that could be perfectly scaled, as we shall see later. Therefore, in investigating attitudes by means of scaling, we must rename as our hypothesis what has heretofore been our basic assumption about the way the Justices decide how to vote.

Thus, the existence of a perfect scalogram may tend to confirm our hypothesis for the particular group of cases that is scaled. It is rare, however, that a perfect scalogram, such as in Table 1, can be produced from the real voting data of the Court, because of several characteristics of these data. It may be that two or more Justices have the same voting record, or that in two or more cases the same Justices voted for the claim and the same Justices against. In such a situation it is impossible to rank those Justices with the same voting record or those cases which elicited the same response from the Court. We are not justified in saying

that they are of the same rank: their breaking points may indeed differ. We know about each breaking point only that it lies somewhere in the region between the last positive and the first negative vote. Depending on the particular cases and Justices to whom these votes belong, the region of indeterminacy may be large or small, i.e. include a large or small variation in severity of claims and sympathy of Justices. If we cannot rank two cases, we shall never be able to rank them, for no new data will ever appear about the votes in those cases. If we cannot rank two Justices, however, there is still hope that in the future a case will arise whose claim splits the Justices, revealing who is more sympathetic to the claim.

The most serious reason for the frequent impossibility of constructing a perfect scalogram is that the votes cannot be arranged in such a way that every Justice has a breaking point which divides all his positive votes from all his negative votes. Some Justices have positive votes surrounded by negative ones, or vice versa: these are called by the scalers "inconsistencies". In order to eliminate the inconsistencies of one Justice, it would be necessary to re-arrange the cases in such a way that

inconsistencies would arise for other Justices. The presence of even one inconsistency forces us to re-examine our hypothesis that all of the Justices voted in all of the cases in accordance with their attitudes toward the conflict between one pair of values. Inconsistencies require the modification of this proposition, in one or more ways. One modification is to hypothesize that a Justice with inconsistencies in a scalogram changed his attitude toward the values at some time. If such a change is suspected, one can divide the group of cases into those decided before and those decided after the date of the alleged change, and then scale the two groups separately. If this operation does not work, however, and it becomes necessary to assume several changes of attitude in order to eliminate the inconsistencies, it then becomes reasonable to believe that the values about which the Justices' attitudes are being considered were not the only values affecting the votes. We are then led to conclude that there was more than one influential question in the minds of some of the Justices, making perfect scaling impossible. Another possible modification of the proposition tested by scaling is the addition of subjectivity. It is easily

conceivable that all the Justices could consider that a certain group of cases raised only one significant question and could all have a fixed attitude on that question, yet not vote in a scalable manner. All that is necessary is a difference in the severity of the claim perceived by one Justice from that perceived by another.

If, for a particular group of cases, the scalogram turns out to contain a large number of inconsistencies, we must conclude that the hypothesis is simply disproved for this group of cases. If, however, the number of inconsistencies is small, i.e. the scalogram is almost perfect, we can make it compatible with our hypothesis if we modify the hypothesis in one of the ways described above. We can also use a comprehensive, less specific modification, by making our hypothesis: the Justices in this group of cases voted generally, but not invariably, according to their attitudes toward the conflict between one pair of values. As long as there are only few inconsistencies, it is reasonable to consider this hypothesis, because the probability of a near-perfect scalogram occurring by chance is exceedingly small. In order to have some rule for deciding how imperfect a scalogram may be and still be considered helpful as an

almost perfect scalogram, scalers have established arbitrary criteria to decide whether a group of cases is scalable. The most sophisticated of these criteria is the coefficient of scalability, abbreviated "S", and it is equal to the fraction of the potentially inconsistent votes that is consistent. The commonly accepted minimum value of S for scalability is 0.60 or 0.65. In other words, if at least 60% or 65% of all the votes which might have been inconsistent (when nine Justices vote there is a maximum of four possible inconsistencies in each case) are consistent, the pattern is considered scalable. Another criterion is the coefficient of reproducibility (CR). This is the fraction of all votes cast that is consistent. If at least 90% of the votes recorded in the scalogram are consistent, the CR is 0.90 and the votes are considered scalable. The coefficient of scalability, the coefficient of reproducibility, and all other formulae for the measurement of scalability contain biases which become important when the given group of cases is characterized by one or another kind of voting pattern, such as many 5-4 votes. But our main concern is with the validity of the scaling process, and for our purposes the issue of the precision of the

indicators of scalability can be neglected.

Until this point we have discussed only one type of scalogram: that in which the same Justices voted in all of the cases in the group being scaled. Barring non-participations, such a scalogram contains a record of a vote at every available position in the diagram. Such completely filled scalograms are exceptional, however. Most groups of cases selected for scaling stretch over a period of time encompassing changes in the membership of the court or include cases in which not every Justice cast a vote. The result is a scalogram in which there are no indications of a vote at several positions, namely the intersections of each Justice with the cases decided when he was not on the Court or not participating. We may group these two situations together and call the result an unfilled scalogram. Unfilled scalograms have some different characteristics from filled ones and are worthy of separate consideration.

In discussing filled scalograms, we hypothesized that, according to scaling theory, a case presenting a more severe claim should bring fewer votes in support of the claim than a case in which the claim is milder. This principle is not valid for unfilled scalograms, however.

The claim in one case may receive more votes than the claim in another for the simple reason that one or more Justices sympathetic to such claims occupied positions on the Court during the one case, and less sympathetic Justices filled those positions during the other case. In fact, it may very well happen that a claim that receives many votes at one time is more severe than one receiving few votes at another time. Another principle used in constructing filled scalograms was that a Justice who gave more votes to a claim was more sympathetic to the claim than one who gave it fewer votes. This principle does not hold for unfilled scalograms, because Justices belonging to the Court at different times can be expected to vote on questions involving claims of different severities. Therefore the breaking point of a Justice is not related to the proportion of his votes that he gives to the claimant, but only, in accordance with the original definition, to the most severe claim that he supports and the mildest claim that he rejects. It is not even possible to locate exactly every Justice's breaking point in an unfilled scalogram, because the most severe claim voted for and the mildest one rejected by a Justice may be separated by several cases with inter-

mediate claims, on which the Justice did not vote at all. Since there is no way to determine how he would have voted in those cases, his breaking point cannot be defined, except by saying that it lies somewhere in the region between the most severe claim supported and the mildest one denied.

Unfilled scalograms not only have these special characteristics, but they also have a special value not possessed by filled ones. The value lies in the fact that unfilled scalograms bring together for comparison cases decided by different Justices, and Justices who decided different cases. Thus an unfilled scalogram can be used to support the hypothesis that one case raised a more severe claim than another case, even though an entire turnover in Court membership occurred between the two. Likewise, one Justice may be said to be more sympathetic to a certain kind of claim than another Justice, even if the two never shared a day on the Court. How such comparisons may be made is illustrated by the hypothetical scalogram in Table 2. For the sake of brevity, only a five-man Court has been assumed in Table 2, but the case of a nine-man Court is perfectly analogous. The cases are lettered in chronological order of decision,

Case:	C	B	F	G	E	K	I	H	A	D	J
Justice:											
J1									x		
J8			x	x	x	x	x	x		x	-
J4	x	x							x	-	
J10						x	x	-			-
J11						x	-				-
J7			x	x	x	-	-	-		-	-
J5	x	x	x	x	-				-	-	
J9			x	-	-	-	-	-			-
J6	x	x	-	-	-			-		-	
J2	x	-							-		
J3	-	-							-		

Table 2
Hypothetical Unfilled Scalogram

and the Justices are numbered in order of date of retirement from the Court. Among the many possible comparisons, it appears that case J raised a more severe claim than case A. We can say this in spite of the fact that they were decided by two entirely different Courts. The scalability of the voting pattern gives support to our general hypothesis. According to the hypothesis, case J presented a more severe claim than case D, because Justice J8 was willing to grant the claim in D but not in

J. And case D involved a more severe claim than did case A, since Justice J4 voted for the claim in A but denied it in D. If the claim in J was more severe than in D, and the claim in D more severe than in A, the claim in J was obviously more severe than the claim in A. Similarly, we can conclude that Justice J10 was apparently more sympathetic to the claim in question than was Justice J5, even though there was not one case on which they both voted, and even though J5 gave the claim 57% of his relevant votes and J10 gave it only 50% of his. The possibility of drawing such conclusions shows the special analytical power of unfilled as opposed to filled saalograms.

With this sketch of the theory of scaling as relevant to the Supreme Court, we turn to the ways in which this theory has been applied. We must pay the most attention to the work of Glendon Schubert. He has been the leading scaler of the Court, and many of the achievements of scaling, as well as most criticisms of it, are due to him.

One of Schubert's favorite techniques is to scale several groups of cases dealing with different variations

of the same subject. For example, he scaled all the cases involving aliens between 1950 and 1957, then scaled the ones between 1950 and 1953 (the Vinson Court) separately from the cases between 1954 and 1957 (the Warren Court), and finally separately scaled over the entire period those cases involving Communist and those involving non-Communist aliens. Each time, the resulting scalogram had an acceptably low number of inconsistencies by the conventional criteria. Naturally, the information contained in one scalogram could be translated into many paragraphs of verbal description, so the scaler must selectively interpret his scalograms for the reader. One thing noticed and called to the reader's attention by Schubert was that the Vinson Court was highly polarized, with four of the Justices voting for a large majority of the aliens and five Justices voting against the aliens almost every time. The scalogram for the Warren Court shows no such sharp division: the Justices were apparently rather evenly distributed over the continuum of attitudes toward the claims of aliens against the government. Regarding the possible differential treatment of aliens who were Communists, the scalograms showed no substantial difference, but Schubert observed that

it is entirely possible that part of the differences noted in comparing the Vinson and Warren Court periods may be attributable to the fact that a majority of the alien cases decided by the Vinson Court, but only four of the twenty decisions of the Warren Court, involved Communists. It would be necessary to make an independent scalogram analysis of all cases involving alleged Communists, decided by the Supreme Court during the period of the 1949-1956 Terms, in order to be more confident as to which of the two alternative interpretations should be preferred.⁸³

The rank of Frankfurter in attitude toward aliens is specifically noted by Schubert. Frankfurter, in both Courts and toward both Communists and non-Communist aliens, was the third most sympathetic to their claims, below only Douglas and Black.⁸⁴

Another topic subjected to scale analysis by Schubert was the right to counsel.⁸⁵ He found that right-to-counsel cases decided between 1940 and 1957 were scalable. He also managed to scale separately the sub-group of cases involving capital punishment and the cases in which lesser punishments were given, in order to test whether there was a difference in the Justices' attitudes toward the right to counsel in these two kinds of cases.

Examining cases involving alleged unconstitutional search and seizure,⁸⁶ Schubert scaled all of the cases between 1937 and 1957 and found enough inconsistencies to

make the scalogram only barely acceptable according to the coefficient of reproducibility. Then he separated the cases involving acts by Federal authorities from cases of search and seizure under states. When these groups were scaled separately, scalability was improved. Frankfurter was found to be the strongest supporter in the Court of protection against improper Federal search and seizure,⁸⁷ but to be relatively less enthusiastic about voting against alleged improper state search and seizure.⁸⁸

Schubert tried to scale "federal tax cases involving government-taxpayer conflicts decided during the 1953-1958 terms", but the resulting scalogram was unacceptable. He succeeded, however, with the subset including only the cases involving criminal charges.⁸⁹

When a selected group of cases proves to be scalable, the scaler usually concludes that the Justices voted on the cases according to their attitudes toward the claim which has been used as a criterion for choosing the group by the scaler. Thus S. Sidney Ulmer's studies of civil liberties cases in the 1956 and 1959 terms of the Court have concluded that the members of the Court voted on these cases according to "one dominant operating vari-

able....: deprivation of a claimed civil liberty."⁹⁰

Schubert has indicated that a high degree of scalability for the alien cases would be "persuasive evidence" for the consideration by the Justices of alien status itself as an important claim.⁹¹

Schubert has summed up the conclusions drawn from scaling as follows:

The research done thus far in cumulative scaling indicates that there is a high degree of consistency in the attitudes of Supreme Court justices toward the recurrent issues of public policy that characterize their work load. This consistency of response in individual judicial voting in such an area of public policy as civil liberties claims appears to provide a much better general explanation of how and why the Court makes its policy choices than does the alternative traditional theory of stare decisis, that consistency in the manipulation of precedential legal rules and principles is a function of legal craftsmanship.⁹²

Such conclusions are typical of scalars, and they are in contrast with what non-scalars usually conclude.

Those who use non-quantitative methods, whether unsophisticated or sophisticated, generally hold that a Justice's attitudes toward a few legal principles will go a long way in explaining his votes. Some, however, like Rodell, believe that the forces motivating the Justices are not few, but infinite, and not legal, but of diverse kinds. It might be anticipated that the users of quantitative

methods, supposedly on the frontier of political research, would represent an extension of the "realism" typified by Rodell. Such is not the case. The scalers view judicial behavior as explainable by few, not many, principles, and only some of these are non-legal, such as attitude toward aliens, while others are legal, e.g. attitude toward state versus Federal search and seizure. Thus, in their conclusions, the scalers tend to embody some of both Rodell and Mason.

The process of scaling and drawing conclusions from scalograms used by Schubert, Ulmer, and others is fraught with fallacies. We shall display this process in more detail as we expose the errors that it contains. One of the major faults lies in the manner in which scalograms are constructed from raw voting data. Scalers refuse to recognize the important differences between filled and unfilled scalograms, and use one set of rules for constructing both kinds. The only published comprehensive set of rules for scaling Supreme Court decisions⁹³ is too long and complicated to be described here. What is important is that these rules, in an apparent effort to conserve objectivity by eliminating any possible differ-

ences in the treatment of the same material from one investigator to another, elaborately specify how to arrange the cases and Justices in order, given any distribution of votes. In order to accomplish this arrangement, many arbitrary rules are employed. These have the effect of ordering cases according to the number of votes received by their claims, a principle which we have shown to violate scaling theory itself when applied to the construction of unfilled scalograms. The rules also redefine the breaking point of a Justice to be the most severe claim supported, i.e. the left end of the region of indeterminacy mentioned above, and the Justices are ordered accordingly. This procedure, too, violates scaling theory, by extracting from the voting data information more precise than the theory permits.

When these rules are applied to the construction of filled scalograms, the result usually (but by no means always) is the arrangement of cases and Justices which produces the least possible number of inconsistencies. Scaling theory offers no reason why a scaler should punish himself by making any arrangement that increases the number of inconsistencies, and yet such an increase is often the result of the construction of unfilled

scalograms with Schubert's rules.

Thus there are two major objections to the use of these rules in the construction of unfilled scalograms: more precision is claimed than the data warrant, and unnecessary inconsistencies are forced into the resulting scalograms. In illustration of the first objection, consider the following scalogram, constructed by Schubert for a group of Federal search and seizure cases from 1937 to 1948.⁹⁴

Case:			C	D	A	J	M	K	L	E	F	G	H	I	B
Rank:			1	2	3	4	5	6	7	8	9	10	11	12	13
Justice:															
1	12	Frankfurter	x	x		x	x	x	x	x	x	x	x	x	
2	12	Murphy				x	x	x	x	x	x	x	x	x	
3	12	Rutledge				x	x	x	x			x	x	x	
4	12	Jackson				x	x	x	x					x	
5	9	Stone	x	x	x					x	x				-
6	7	Douglas	x	x		x	x	x	x	-	-	-	-	-	-
7	5	Black	x	x	x	-	x	-	-	-	-	-	-	-	-
8	4	Burton				x	-	-	-			-	-	-	-
9	4	Reed	x			x	-	-	-	-	-	-	-	-	-
10	3	Hughes	x	x	x										-
11	3	Butler	x	x	x										-
12	3	Roberts	x	x	x					-	-				-
13	3	Cardozo				x									-
14	3	Brandeis				x									-
15	1	McReynolds	x	-	-										-
16	0	Byrnes								-	-				
17	0	Vinson				-	-	-	-						-
18	0	Sutherland				-									

Table 3
Unfilled Scalogram Constructed by Schubert

This scalogram shows only one inconsistency, and by inspection we can see that it would be impossible to reconstruct the scalogram without at least one inconsistency. The data on which the scalogram is based, however, are not complete enough to justify this arrangement of cases and Justices as the only possible one. There are many ways in which we could re-order them without increasing the number of inconsistencies, and the resulting new scalogram would be just as plausible, in terms of scaling theory, as is the one in Table 3. An example of how the same data could be rescaled is shown in Table 4. A comparison of Tables 3 and 4 will show that the number of inconsistencies has remained constant, but the order of cases and Justices has changed substantially. Case E, for example, is the eighth case in Table 3 and the eleventh in Table 4; case I moved from twelfth to eighth position; G went from tenth to thirteenth. Only two cases retained their positions. The change is more pronounced in the order of the Justices. Among the many changes of position, Brandeis moved from a low fourteenth place to a middle ninth, and, most spectacularly, Black dropped from an upper-middle seventh position to nearly the bottom of the list, at sixteenth place. As can be

Case:			C	A	D	J	M	L	K	I	H	F	E	B	G
Rank:			1	2	3	4	5	6	7	8	9	10	11	12	13
Justice:															
1	13	Jackson				X	X	X	X	X					
2	13	Rutledge				X	X	X	X	X	X				X
3	13	Murphy				X	X	X	X	X	X	X	X		X
4	13	Frankfurter	X		X	X	X	X	X	X	X	X	X		X
5	11	Stone	X	X	X							X	X	-	
6	7	Douglas	X		X	X	X	X	X	-	-	-	-		-
7	4	Reed	X			X	-	-	-	-	-	-	-	-	-
8	4	Burton				X	-	-	-	-	-				-
9	4	Brandeis		X											-
10	4	Cardozo		X											
11	4	Roberts	X	X	X							-	-	-	
12	4	Butler	X	X	X										-
13	4	Hughes	X	X	X										-
14	3	Byrnes										-	-		
15	3	Vinson				-	-	-	-	-					
16	3	Black	X	X	X	-	X	-	-	-	-	-	-	-	-
17	1	Sutherland		-											
18	1	McReynolds	X	-	-									-	

Table 4
Schubert's Scalogram (Table 3)
Reconstructed without His Rules

seen in the tables, either Black's vote in case J or his vote in case M must be considered an inconsistency in scaling terms. In Schubert's scalogram, the J vote is arbitrarily chosen as the inconsistency, and, in our revision, the M vote is instead chosen. These two alternatives make it possible to place Black almost anywhere between seventh and sixteenth positions. Thus indeterminacy may be an important characteristic of unfilled scalograms, but scalers tend to avoid acknow-

ledging it by injecting an artificial determinacy through their scaling rules.

Our second objection, that scalers include unnecessary inconsistencies in unfilled scalograms, is illustrated by the following pair of scalograms. Table 5 is

Case:			V	Q	U	N	T	P	R	O	W	S
Rank:			1	2	3	4	5	6	7	8	9	10
Justice:												
1	9	Frankfurter	x	x	x	x	x	x	x	x	x	-
2	8	Murphy				x				x		
3	8	Jackson		x		-		x	-	x		-
4	7	Douglas	x	x	x	x	x		x	-	x	x
5	7	Black	x	x	x	x	x	x	x	-	-	x
6	5	Warren	x		x		x				-	-
7	5	Clark	x	x	-		x	-	-		-	-
8	4	Rutledge				x				-		
9	3	Brennan	x		x						-	
10	3	Harlan	x		x		-				-	
11	2	Burton	x	x	-	-	-	-	x	-	-	-
12	1	Whittaker	x								-	
13	0	Vinson		-		-		-	-	-		
14	0	Minton					-	-	-			-
15	0	Reed				-	-	-	-	-		-

Table 5
Unfilled Scalogram Constructed by Schubert

Schubert's interpretation of a group of Federal search and seizure cases from 1949 to 1957.⁹⁵ This scalogram contains seven inconsistencies, but one of them is due to the fact that Schubert disobeyed his own rules for an unexplained reason, in calling Douglas's breaking point

seven instead of ten. Even if this error is corrected, however, the number of inconsistencies, now six, can be decreased by rearranging the cases and Justices, as in Table 6, for example. Here the number of inconsistencies

		Case:	V	Q	T	P	R	N	U	S	W	O
		Rank:	1	2	3	4	5	6	7	8	9	10
		Justice:										
1	10	Murphy						x				x
2	10	Frankfurter	x	x	x	x	x	x	x	-	x	x
3	9	Douglas	x	x	x		x	x	x	x	x	-
4	8	Harlan	x		-				x		-	
5	8	Black	x	x	x	x	x	x	x	x	-	-
6	7	Warren	x		x				x	-	-	
7	7	Rutledge						x			-	
8	7	Brennan	x						x		-	
9	4	Jackson		x		x	-	-		-		x
10	3	Clark	x	x	x	-	-		-	-	-	
11	2	Burton	x	x	-	-	x	-	-	-	-	-
12	2	Minton			-	-	-			-		
13	1	Reed		-	-	-	-	-		-		-
14	1	Vinson		-		-	-	-				-
15	1	Whittaker	x								-	

Table 6
Schubert's Scalogram (Table 5)
Reconstructed without His Rules

has been reduced from six to four. A few substantial changes of position have also taken place, such as of case U and Justices Harlan and Jackson.

The two illustrations above attempted to show that the potential value of unfilled scalograms has been largely negated by the substitution of arbitrary rules

for reasonableness in the scaling process.

Until this point, we have discussed some reasons why the model of a perfect scalogram in Table 1 cannot usually be reproduced in practice, and some failings in the ways in which scalers cope with the resulting imperfections and uncertainties. Even if all of these practical difficulties in constructing scalograms were eliminated, however, and every scalogram in practice turned out to be a perfect one, the conclusions drawn from them would still be open to doubt. In describing the theory of scaling, we noted one basic implication: if the Justices all vote on each of a group of cases in accordance with their fixed attitudes toward one question (the proper extent of the claim of one value on another), then a scalogram without inconsistencies can be constructed for that group of cases. In describing the conclusions drawn from scaling, however, we have seen that the scalers assume that the reverse implication, too, is valid. In fact, as we noted earlier, it is not. The scalability of a group of cases in no way implies that the cases were voted on with one and the same question in the minds of the Justices. If votes for each Justice are assigned at random to the cases in a group, there is a

certain chance that the group will be scalable without inconsistencies, and there is a greater probability that it will be scalable with few enough inconsistencies to meet the criteria for scalability used by most scalers. If the number of cases is large enough and/or the scalogram is sufficiently filled, the probability of a scalable voting pattern arising by pure chance is indeed small, but it is still there.

Moreover, if the voting of the Court on a certain group of cases is scalable, there is a greater probability that the votes of any single Justice would fit into the scale pattern, even if those votes were assigned to him at random, because of the small number of votes involved. The probability is especially great if the Justice was on the Court for only a part of the group of cases. In Table 6, for example, three of the Justices cast votes in only two cases each, and only one out of four possible combinations of votes, i.e. a negative vote left of a positive vote, could have produced an inconsistency for any of these three Justices.

We are not suggesting, of course, that any Justice votes by tossing a coin. But whatever the probability that random voting would have had of making a group of

cases scalable, voting on the basis of more than a single question would have at least the same probability of leading to an acceptable scalogram. Therefore, from the scalability of a group of cases we can conclude that they were voted on with one question in mind, but we must always attach the reservation that there is a certain statistically calculable probability that other considerations entered into the voting, perhaps making the scalogram more perfect or less perfect, and perhaps not.

Scalers seem to regard the perfectness or imperfectness of a scalogram as an indication of success or failure. This attitude obscures the fact that we can draw more certain conclusions from non-scalability than from scalability. Although we must make an allowance for chance, i.e. additional influences on the voting, when we conclude the dominance of a single question from the existence of an acceptable scalogram, we can say with absolute certainty, when a group of cases fails to scale, that the Justices' attitudes toward one question did not determine all of their votes. There may be groups of cases on which we would expect the Justices to vote on the basis of one question, and then such a negative conclusion would be quite important.

But if our conclusion is that one question did dominate in a group of cases, our difficulties have not ended when we have made the necessary reservation about chance. We must still ask ourselves what the dominant question was. The answer is by no means obvious, but scalars seem to think it is. Usually the group of cases selected for investigation by scaling is not picked by lot from all of the cases within a certain period of time, but is composed of those cases which, in the view of the investigator, have something important in common. Examples are civil rights cases, freedom of speech cases, and cases testing state regulation of business. If an investigator picks a group of cases according to what seems to him to be a common important characteristic, and if he finds that the votes on the cases in this group are scalable, he naturally tends to conclude that the single question which probably dominated the voting turned upon the criterion by which he originally chose the cases. The conclusion is not valid, because of two possibilities. First, it may be true that the investigator has chosen all of the cases, and only the cases, which share a certain characteristic, but the salient characteristic may be one other than that by which he made the choice.

The second, and more likely, possibility is that the cases selected are part of a larger group, distinguished by a different characteristic.

By looking at any perfect scalogram, it is easy to see that the elimination of any number of cases from consideration would not have any effect on the scalability of the remaining group. Hence the important principle that, if a group of cases is scalable, so is any subgroup thereof. If the group is only imperfectly scalable, but within the conventionally acceptable range, the removal from consideration of cases not responsible for any inconsistencies may leave a scalogram with a larger fraction of its votes inconsistent, and therefore with a poorer coefficient of scalability. In general, however, a subgroup is not likely to be much more or less scalable than the entire group from which it is taken, if that group scales well.

If a group of capital punishment cases proves to be scalable, it may be that the Justices voted according to their attitudes toward the death penalty. But perhaps the larger group of cases involving all criminal convictions would also produce an acceptable scalogram. If this group were originally chosen to be scaled, the

scaler would most likely conclude that the Justices' votes on all the cases, including the capital punishment ones, were based on their attitudes toward criminal defendants versus the state. As another example, Wallace Mendelson notes Ulmer's conclusion that the Justices voted on a group of cases according to their attitudes toward civil rights. But scaling, says Mendelson, can prove several contradictory things: if particular subgroups of the original group are scaled, it can be shown, in the same manner in which Ulmer "proved" his assertion, that the Justices voted on the 25 cases in one subgroup in accordance with their attitudes on Communism, and on the 15 cases in another subgroup according to their attitudes toward homicide.⁹⁶

Our objection to the validity of the conclusions derived from scaling is becoming serious indeed. We have seen that, if a group of cases is scalable, the scalability of a subgroup tells us almost nothing new, since it could be predicted from the fact that the full group is scalable. What must we then conclude if we find that all of the cases decided in one term produce an acceptable scalogram? The scaler would conclude that the Justices voted on all the cases according to some one dominant

question; indeed, this very conclusion is implied when we characterize a Justice as being to a certain degree liberal or conservative. Liberality-conservatism is conceived as a one-dimensional variable encompassing all, or almost all, of the particular policy issues which arise. According to this conception, the Justices can be ordered according to a scale of liberality, and this ordering implies an ordering of the cases as well, so that one case presenting a more severe liberal claim than another will receive fewer votes than the other, those voting for the claim being more liberal than those voting against it. The conception just described is what a scaler would conclude if he found all cases scalable together. And, for at least two terms (the 1936 and 1961), the group comprising all the cases decided within a term has been found to be scalable.

Pritchett described the entire period of 1931-6 as marked by "almost watertight" blocs in the Court, and by a fixed pattern of agreement and disagreement. "Locating the justices along a single attitude scale in terms of relative liberalism or conservatism would adequately account for the judicial disagreements manifested during that period."⁹⁷ Schubert confirmed this statement by

scaling the entire 1936 term.⁹⁸ To discover whether the adequacy of liberal-conservative ratings would apply to a recent term, we scaled the entire 1961 term, which is presented in Table 7. According to both of the standard criteria of scalability, this scalogram is acceptable, although not to as high a degree as the scalogram of the 1934 term.

This finding gives us reason to be very wary of bias on the part of the scaler. Whatever group of cases he picks out to scale, his results add little to our knowledge if the group composed of all cases is scalable. Clearly a reasonable procedure would be to begin by trying to scale entire chronological series of cases, and if these did not scale, to investigate how they might be divided into groups that would produce acceptable scalograms. Scalers, however, begin at once with small groups. They do not select, for scaling, groups comprised of those cases involving litigants one of whose middle names begins with "R", nor groups consisting of those cases arrived at by one or another manipulation of a random number table, for there is no reason even to suspect that such groups might be scalable. Scalers select, instead, groups of cases which have in common an

Case: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z a b c d e f g h i j k l m n o p q r s t u v w x y z
Justice:

Douglas	x x	x - - - -
Warren	x x	x - - - - - - - - - -
Black	x x x x x x x - x x - x x x x x x x x x x x x x x x x x x	x x x x x - x x a x x x x - - - - x - - - - x x x x
Brennan	x x	x x x x x - - - - - - - - - - - - - - - x - - - - - - - - - -
Clark	x x - x - x - x x x x - x x x - - x - x x x x x x x x	- - - - - - - - - - x -
Stewart	x x x x x - x x - x x x - x x x x x x x - - - - - - - -	- - - x - - - - - - - - - - * a x - - - - - - - - - - x -
Whittaker	- x x x x a - - - -	- - - - - a - - - - x
Harlan	- - x x x x x x x - - - - - - - - - - - - - - - - - - -	- - - - - x - - - - - - - - - - - - - - - x ; - - - - -
Frankfurter	a a a x a a a - - - - - a - a - - a x a a - a a a	- - a a - a - a a a - a - - a a a a - a a - - a - a
White	x x a - a a a a a a x a a	a a a a a - a a a - a - a a a

Table 7
Scalogram of All Non-unanimous Cases in 1961 Term
(Per Curiam Decisions Not Included)

Scalability:

$$CR: 1 - \frac{31}{413} = 0.92$$

CR (Schubert's method): $1 - \frac{22}{304} = 0.93$

$$S: 176 + 207 = 0.65$$

Legend:

x/- if the more numerous: voted with Court
if the less numerous: voted against Court

a did not participate

* partial dissent

: position ambiguous

(blank) nota member of the Court at time of decision

issue which the scalers think might well be the dominant issue in the minds of the voting Justices. When the groups scale, it is this issue which is assumed to be the dominant one, and the scalers do not consider the possibility that where the one issue occurred there coincided also another, to which the Justices paid more attention. Thus the scalers' conclusions depend closely on their presuppositions. One allegation as to what these presuppositions might be is made by Mendelson. Cases involving civil liberties, he says, involve also "the distinction between constitutional and statutory construction; between stare decisis in relation to constitutional, as against statutory, decisions; between judicial review of procedure and judicial review of substance—to mention only the obvious." "[O]nly an activist is inclined to ignore" such differences and to lump all civil liberties cases together without further distinction. Thereby "neo-behavioralism" (a term which subsumes scaling) "tends to reflect the dedicated libertarianism of most of its practitioners. Thus it generally applies to the judicial process an essentially political test; namely, the libertarian party line."⁹⁹ For Mendelson, the principal "weakness" of judicial

behavioralism is that "it assumes that every vote in a case that has some connection with civil liberty, for example, is necessarily a vote for or against that liberty."¹⁰⁰ Here Mendelson points to the existence of cases in which several Justices voted against the person claiming a civil liberty for procedural reasons, but simultaneously voted for a principle destined to have a wide pro-civil-liberties effect in the future.

There is evidence to support Mendelson's charge. On the trivial side, there is the fact that scalars use a language which seems to reflect, or which could easily lead to, the mode of thought that is alleged. Votes for the civil liberties claim or the civil rights claim or the individual's claim against the state or a business are invariably denoted as positive, and opposite votes as negative. And the term "inconsistency" is a very unfelicitous one, which, if not understood from a purely technical standpoint, implies the correctness of the proposition which scaling is designed to test, even before the results of the testing are known. It also implies that, if a Justice votes on a different basis from his fellow Justices, he does not vote according to any principle at all.

The crux of the question lies in the way in which scalers draw their conclusions. Mendelson is wrong in saying that scalers assume the irrelevance of jurisprudential, as opposed to political, principles. Such an assumption is not logically inherent in the scalers' work, nor, as we have seen, do scalers even conclude that legal principles are completely extraneous. Rather, the scalers combine an inattention to the implications of their own theory with a lack of rigor in their procedure, with the result that they are able to achieve almost any conclusions they might want, and they usually conclude, to a certain extent, that which Mendelson says they have assumed all along.

In order not to exceed the bounds beyond which scaling theory does not extend, scalers must choose one of two alternatives. They may select for investigation only those groups of cases which they wish, according to criteria of their own choice; then, when these groups prove scalable, the scalers may conclude, with a margin for the effect of probability, that the votes were based on a single issue, but the scalers have absolutely no basis for any speculation as to what that issue was. The second alternative is to scale every conceivable group of

cases, including groups randomly compiled. Then, if the scaler concludes from the scalability of, for example, a group of alien cases that the votes were based on relative sympathy for aliens, he must draw with equal certainty the conclusion that the Justices' attitudes toward some single question also decided their votes in any other group of cases that is scalable, whether the question is visible to the scaler or not. As stated, of course, this second alternative is impossible to follow, requiring as it does the construction of an infinite number of scalograms. In practice, what is called for is the open-minded, impartial scaling not only of groups of cases that are expected to be scalable, but also of groups expected not to be scalable, such as randomly selected groups and entire terms. Scalograms of the latter are necessary because what is important is not how scalable the cases involving, let us say, Communists are, but how much more scalable they are than any group of cases selected by lot. The scalers have followed neither of the alternatives outlined above. They have chosen groups of cases with the discretion of the first alternative, and drawn conclusions with the breadth allowed by the second.

Basically, the error of the scalers seems to have arisen from a misconception about the function of the scaling process. Scalers assume that a successful scalogram is a proof of something. It is really, however, nothing more than the expression in orderly form of an observed regularity. The name of this regularity is scalability. When we find that the votes in a certain collection of cases are characterized by this regularity, we may be inspired to formulate a hypothesis that would explain the regularity. Then it is our task to conduct further investigations, in such a way that our hypothesis is either proved or disproved. Scalers have fallen into the error of neglecting this last step. Instead of observing the regularity, making a guess about its cause, calling this guess a hypothesis, and proceeding to test their hypothesis, the scalers observe the regularity, guess as to its cause, but call their guess a definite explanation, and see no need to subject this explanation to any tests.

Not all scalers, of course, adhere to the procedure described above: some use more care at times, and some use less. Schubert notes, for example, that in one article¹⁰¹ Ulmer constructed a scalogram in which changes

in Court membership were handled by putting new Justices in the attitudinal positions of their predecessors.¹⁰² In the terminology of this paper, what should have been an unfilled scalogram was converted to a filled one by pretending that outgoing Justices and their successors were one and the same person. Clearly Ulmer was injecting a complicating question into the object of inquiry: Do Justices tend to vote according to the same attitudes toward the same questions as their predecessors? At a time when scalars are still struggling to prove that Justices vote on the basis of attitudes toward single questions in certain types of cases, the addition of another proposition to the burden of proof of scaling is premature.

On the side of greater prudence, some scholars of the behavioralist school do not always commit the fault noted above of taking a scalogram for a final explanation of the votes involved. Instead they use scaling as an indicator, to discover cases and particular votes worthy of more detailed examination. In one article,¹⁰³ Ulmer first scaled a group of civil rights cases and then looked non-quantitatively at those cases which evoked inconsistent votes and those which provided the breaking

points for the Justices. Similarly, Harold Spaeth scaled a group of cases and examined the contents and the written opinions of those cases which elicited inconsistent votes.¹⁰⁴

To at least two studies Mendelson's accusation is absolutely inapplicable, because scaling itself is used to test whether the principle of judicial self-restraint, believed by Mendelson to be supremely relevant, is really at work. In one article¹⁰⁵, Spaeth isolated, by a content-analytical process, 52 cases appearing to have been decided principally on the basis of Supreme Court activism versus self-restraint. He scaled the entire group, divided it into five subgroups, and scaled them. Although all of the scalograms were acceptable, scaling by subgroups reduced the total number of inconsistencies, suggesting that attitudes toward self-restraint differed with the kind of case, even if self-restraint was apparently the principal issue in all the cases. Spaeth tried to explain remaining inconsistencies by hypothesizing an attitudinal change by one Justice at a certain time and effects of other, intruding issues. Spaeth then, however, considered cases of the same time period which included, but were not dominated by, the issue of

self-restraint. In cases focusing on state regulation of business, for example, Douglas, Black, Warren, and Brennan, the activists in pure restraint cases, exercised deference to state power. Spaeth's conclusion was that, depending on the kind of case, the issue of self-restraint was treated as a first-, second-, or third-order consideration relative to the substantive issues involved.

Joel Grossman has taken fellow scalers to task for their superficial treatment of Frankfurter's voting on civil liberties, and has offered, through a more thorough investigation by scaling, his own interpretation.¹⁰⁶ The scalograms, said Grossman, that led other scalers to decide that Frankfurter was an anti-libertarian, were inconclusive. It was necessary to consider Frankfurter's own explanation of his voting behavior, but other scalers had cast this explanation aside at the outset. Grossman presented and analyzed Frankfurter's stated reasons and, on this basis, defined a new value, "denial of judicial responsibility", or DJR. To discover whether this value was relevant to Frankfurter's voting, Grossman scaled all the cases in the 1958 and 1959 terms in which DJR, as defined, was explicitly raised as an issue. Frankfurter

was the only Justice with a perfect pro-DJR voting record. Then Grossman scaled all the civil liberties cases in those two terms, in two groups: those in which DJR appeared as an issue, and those in which it did not. In the group involving DJR, Frankfurter voted against the civil liberties claim every time, that claim being in every case incompatible with the claim for DJR. In the civil liberties cases which did not involve DJR, however, Frankfurter voted sometimes for and sometimes against the civil liberties claim. In the scalograms for these cases, Frankfurter appears more sympathetic to civil liberties than Harlan and Clark in the 1958 term, and than Harlan, Clark, and Whittaker in the 1959 term, with 22% and 19%, respectively of his votes cast for the claim.

Grossman warned, in his conclusion, that Frankfurter could have raised the issue of DJR in some cases where he did not do so, but it seemed that he voted according to DJR where that issue was raised. If so, decided Grossman, Frankfurter indeed was more, but only slightly more, libertarian than his votes would suggest if the issue of judicial self-restraint were ignored by the analyst.

Schubert, too, has at times used an advanced proce-

dure.¹⁰⁷ His analysis then treats scaling more or less as a generator of hypotheses. A group of cases which is thought to be scalable is scaled, and certain inconsistencies arise. The Justices and cases responsible for the inconsistencies are studied non-quantitatively, in an effort to find a second issue which might have been operating, along with the issue originally presumed dominant, to determine the votes. Thus scaling ceases to be used as a test of the proposition that only one issue is significant. It becomes assumed that several issues enter into the votes, and a scalogram becomes merely a representation of the first approximation to reality, based on the most important of the issues. A second issue is hypothesized, and a second, more exact approximation is made. There is no necessary end to the series thus begun. Schubert has successively introduced up to five issues. At each stage, the cases and Justices are ranked on the new issue, in such a way as to help explain the hitherto inconsistent votes.

There are two ways in which this kind of analysis, multivariate as opposed to simple Guttman scaling, can be performed. One way is to cease reliance on the scaling process after the first issue has been studied. When

this method is used, the investigator reverts to traditional, non-quantitative examination. Heavy stress is laid on written opinions as a source of information on the subsidiary issues which have given rise to votes that do not fit in the scalogram. This is the technique used by Schubert in a study of cases involving the question of civilian versus military control.¹⁰⁸ When he scales the cases on the basis of this question and finds inconsistencies, he examines the opinions and concluded that a second issue was influential: stare decisis. Such an analysis, of course, is subject both to the criticisms of non-quantitative methods and to those of scaling, but to a modified degree in each case. Opinions are not accepted at face value, but are weighed along with votes, and the scalogram is not treated as a final explanation, but is supplemented by investigation of the opinions.

The second kind of multivariate analysis uses a different approach from anything so far described. Instead of beginning with a Guttman scalogram and adding issues one by one, the investigator assumes a certain number of issues (most commonly three) and, by a different process, scales the cases with respect to all of these issues simultaneously. When the number of issues

assumed is three, the results are represented as a cube of space, and the cases and Justices, instead of being ordered along one line, are arrayed at various points in the space. This type of analysis, mainly employed by Schubert¹⁰⁹, involves many complexities and problems, and to treat it would require as much space as is here devoted to Guttman scaling. The method is very new in its application to the Supreme Court, the entire literature consisting of two recent articles by Schubert, and it seems reasonable to let such a new method mature somewhat before subjecting it to critical analysis.

We have now reached the point at which an over-all evaluation of cumulative scaling as a method of analyzing Supreme Court Justices' attitudes is possible. For this purpose we must turn to those questions which we originally formulated as guides in this evaluation. We established four criteria for the usefulness of quantitative methods. The first was that the methods must improve knowledge or understanding. On the face, scaling has done so to a great degree. By its use, long-accepted notions about the composition of the Justices' attitudes, and the kinds of attitudes which influence their votes, such as support of judicial self-restraint, stare deci-

sis, and other legal principles, have been challenged, and the Supreme Court's voting patterns have been given quite new interpretations. The scalogram itself is an enlightening form for the presentation of information, even if we disregard the conclusions drawn from it. Some of the results of scaling, if they are accepted, constitute revolutionary additions to, and alterations in, our knowledge of the Supreme Court, while other scaling studies confirm the results of non-quantitative research. We must question, however, the validity of these results.

What is the relation between facts and values in the work done with scaling? Certainly there is nothing so blatant as statements about how Justices should have voted or what attitudes they should have had. Such normative expressions have been effectively excluded. Only more subtle biases, such as the libertarian activism alleged by Mändelson, might be present. We can not be sure, however, that such a bias is active, because the failure to distinguish among legally technical differences and the consequent collection of all civil liberties cases under one heading, while perhaps due to a bias of libertarianism, might also be a result of other things,

such as the belief, not that the Court ought to utilize legal technicalities merely as a tactic to advance ends dictated by values, but that it does so in fact. In other words, scalers may have drawn their conclusions partly because they investigated the facts in ways conducive to the achievement of the results that they expected. If this is so, we have a negative answer to the question whether scaling has eliminated the influence of subjectivity. Furthermore, scalers rarely acknowledge the subjectivity of their findings: it must be uncovered by the critic of the scalers' methodology.

As a further step in deciding whether scaling has added to knowledge or understanding, we must ask whether the method has led to a restriction of the problems or facts considered, or to the neglect of other methods. The most striking part of the answer to this question is the fact, already mentioned, that scaling makes use only of data on votes cast, and finds no room for opinions, nor for the insights offered by conditions surrounding the Court and the cases and by the ensuing effects of particular decisions. Indeed, scaling feeds not even on all votes, as has been shown, but only on those in non-unanimous cases. This severe limitation is prima facie

evidence that scaling is not a self-sufficient tool for the investigation of the Justices' attitudes. We should expect scaling to be one among several methods. Scalors, however, often fail to supplement their scalograms with a substantial amount of any other kind of research before announcing their conclusions. If these conclusions were expressed as hypothetical conjectures to be tested by reference to further scaling and to data inaccessible by scaling, we could only acclaim the scaling process as one fruitful first step in the acquisition of knowledge and understanding of the Court. Only when scalors make use jointly of scaling and of non-quantitative methods—and they often do not—does scaling approach what seems its proper role. There is no apparent quality in scaling which could be blamed for the fact that scalors often ignore the necessity of supplementing scaling with other methods. Not the method, but its users, over-enthusiastic about its value, must be blamed.

We must also ask ourselves whether the presentation of the results of scaling has been obscured by jargon, as critics charge. The potentially confusing word "inconsistency" has already been mentioned. This and all other technical words used in scaling have, however, been

clearly defined by their users, and the careful lay reader should have no difficulty with the terminology. Also, the results of scaling are generally presented in a clear, non-technical manner unlikely to give trouble to any reader. Whatever one's doubt as to the validity of the scalers' conclusions, there need be no uncertainty about what those conclusions are. Thus it is possible to say that scaling has indeed improved our knowledge and understanding of the attitudes of the Justices, but the method has been often mishandled, and as a result only a small part of the potential value of scaling has been realized.

Our second criterion for the usefulness of a method was that it must represent some improvement over other methods. How does scaling compare with non-quantitative methods with respect to the questions just discussed? Non-quantitative studies are in general not free from the influence of the scholar's values. Traditionally, scholars of the Supreme Court have not been satisfied with pure description of and explanation of the Court's behavior; "the dogma of some behavioral scientists, that value judgments are outside the pure stream of research, would be accepted by few legal scholars who had ever

stopped to think about what they do and why they do it.¹¹⁰ McCloskey calls Douglas and Black dogmatists because he is interested in seeing the development of legal doctrine. There is more warmth in Rodell's description of Black as a man motivated by human situations than in his portrayal of Frankfurter as one guided by fear of an imprudent, offensive extension of the Court's concern, perhaps because Rodell believes that Black behaves as a Supreme Court Justice ought to behave. For every influence by values that we can detect, there are no doubt many that are too deeply hidden for our notice. Scholars almost never explicitly warn readers about the values which underlie their work. The reader must do his best to find where facts and values have been confused, and, more often, where values have influenced the way in which the scholar perceived and presented his facts.

The role of subjectivity in fact perception is necessarily great in non-quantitative analysis. In the typical pattern, the belief that votes are cast according to legal principles, the belief that written opinions represent true opinions, and the belief that opinions are consistent, are influential in the conclusions drawn

about Justices' attitudes.

The vexing characteristic of non-quantitative studies is that it is impossible to determine the extent or the manner of the influence of the investigator's subjectivity, just as the influence of his values can never be clearly known. The writer states certain of his values and biases in perception and leaves others unmentioned. In his presentation he is forced to extract a few of the many relevant facts and to emphasize certain ones of those that he mentions. In the end, the reader can not trace the subjective sources of the author's conclusions, but must merely judge on their plausibility. Here lies a very important difference between non-quantitative methods and scaling. Scaling is a thoroughly defined process based on a limited fund of information. Every step in the process can be checked by the reader, and any fallacies can be detected. Furthermore, the reader can put the method to his own use: after deciding how much can be expected from scaling, he can apply its rules to any problem of his own choice. By contrast, no such explicit rules are available for the use of non-quantitative methods. Therefore the common accusation that quantitative findings are made incomprehensible to the layman

might well, in the case of Justices' attitudes, be turned back upon the users of traditional methods themselves. It is precisely because the process by which scalars reach their results is clear, that we have been able to detect the sources of the invalidity of those results and to recommend reforms in the process. No such recommendations could be made for the non-quantitative studies, because their processes are actually the occult thought processes of their authors, rather than an explicit set of rules. It is possible merely to conjecture about the sources of error in these studies. As has been suggested, non-quantitative methods seem characterized by a susceptibility to the intrusion of values and subjectivity, and the process of information gathering and reasoning, being implicit and obscure, is not likely to be immune from carelessness.

Neither non-quantitative methods nor scaling is perfect, either in theory or in practice. Non-quantitative methods are plagued by normative and observational bias; this can be reduced by more careful formulation of procedures and more explicit description of chains of reasoning, but the essence of non-quantitative study is the freedom which the scholar has to let his expert mind

roam over a huge range of information, fix on certain facts as important, and ignore others as trivial. Subjectivity cannot be entirely abandoned without destroying this essence. Scaling is beset by subjectivity also, but it is a product not of the intrinsic nature of the scaling process, but of the abuse of that process. If scaling theory is respected and rigor is added to the scalers' work, scaling can become a very valuable tool, because it will embody the objectivity that is necessarily lacking in non-quantitative work. On the other side of the coin, scaling relies on a small fraction of the information relevant to the Justices' attitudes, and non-quantitative methods are needed to make use of the vast majority of the information.

Two criteria remain to be considered. The third criterion is that the method give us information about interesting or important matters. In other comparisons of quantitative and non-quantitative methods, this criterion might be applicable. Here, however, we have chosen to study a quantitative method that has been used on a subject already of great interest to non-quantitative scholars. Scaling has therefore drawn attention neither toward nor away from the subject of the attitudes

of the Justices.

The fourth criterion is that the method should not be harmful. It is of course impossible to measure the effects of scaling on people's attitudes or on the Supreme Court. No doubt scaling has convinced a certain number of persons that the members of the Court adhere to legal principles far less than they profess. (This conclusion has not been proven by the scalers, as has been demonstrated, but it may nevertheless be true.) We can decide as we wish whether the growth of the opinion that the Court is a political body rather than an applier of the law is a beneficial or a harmful trend. The yardstick theory may give the Court a shield of sanctity against the attacks of those dissatisfied with its decisions, and scaling may be rendering that shield less effective.* As yet, however, scaling is employed by so

*"Let us suppose...that in studying the political process's relationship to the Supreme Court, political science eventually demonstrates..." empirically that "the judge is in fact nothing more than a glorious rationalizer of his own personal values. ...But what then? ...Is it not possible that the 'myth' of the 'objectivity' of the Court and courts is quite functional for stability and/or equilibrium? ...If Charles Black is correct in his opinion that the real genius of American politics has been [sic] in the establishment of the legitimizing and checking functions of the Supreme Court, then such a scientific discovery made, verified, and popularized by political scientists could be positively disastrous."lll

few scholars and is under such heavy attack that it seems unlikely that its effects on public opinion are at all significant. Likewise, there is no evidence that scaling has influenced the behavior of the Court itself. The Court continues to justify its decisions in legal terms, and seems to have made no effort to conform to the behaviorists' notion of consistency as opposed to the lawyers' conception.¹¹² Thus the effects of scaling outside the academic world have probably been miniscule up to now. As for the future, there is no guarantee that scalars will continue to arrive at the same conclusions about Justices' attitudes, particularly if scalars take heed of current criticisms of their methods and if scalars and non-scalars alike come to recognize the value of utilizing both methods coöperatively.

Two conclusions, in the way of advice to scalars and non-scalars, emerge. First, we have seen how extremely limited is the information put into the scaling process; what comes out is also necessarily limited. A non-quantitative scholar, given enough persistence, can discover all the patterns and trends that scaling might reveal (although perhaps not as efficiently). But scaling

has one redeeming quality that sets it off sharply from non-quantitative methods: objectivity. This quality is in part inalienable from the process: biases injected into scaling are discoverable, as they are not when operative in non-quantitative scholarship. But, for the rest, only the user of scaling can retain the objectivity of which the method is capable. If he does, he will make scaling; if he does not, he has destroyed its raison d'être and made the value of its use dubious indeed.

Finally, we have described non-quantitative studies as falling into various strata of methodological sophistication. The reader has perhaps inferred from our description of scaling that studies using this method, too, are sophisticated to different degrees in their methods. In each case, the more sophisticated work is characterized by a wider range of source materials and methods of analysis. Above the three levels of sophistication heretofore presented, there ought to be a fourth, but we have been unable to find any examples of it. A study at the highly sophisticated level would make use of both sophisticated scaling and sophisticated non-quantitative methods. The powerful interpretive skill of the expert mind, working intuitively, and the vast area

of information that mark non-quantitative scholarship at its best require that non-quantitative methods remain the core of study of the Justices' attitudes. The objective presentation of all votes by scaling, the possibility of formulating with complete precision the scaling procedure before it is applied, the convenience and clarity of scalograms, and the almost automatic tendency of scalograms to generate hypotheses—all these qualities should make scaling an almost indispensable aid in any thorough study in the area of Justices' attitudes. Non-quantitative methods and scaling possess nearly complementary features; their use together should produce more advanced knowledge of the Justices' attitudes than has so far been attained. In spite of its limitations and the faults in its application, scaling has been profitably used on the Supreme Court. Some studies, such as those of Grossman and Spaeth, have contribute to our understanding of Justices' attitudes in ways that would have been far more difficult without scaling. Other work with scaling has drawn conclusions that cannot be accepted as having been rigorously reached, but are at least made plausible enough to warrant further investigation. And one of the major contributions of scaling to the field has been to

stimulate reaction on the part of non-scalers, in the form of more detailed re-examination of their own beliefs and the careful defense of those beliefs against the attacks of the scalers.

Rashness, unsystematic procedure, and methodological solipsism have destroyed a large part of the potential fruit of the application of scaling to the Supreme Court, but we cannot therefore call scaling useless in this field of study. Scaling has been used with some profit, even if much of the profit is the by-product of the resulting intra-disciplinary debate. But the greatest contribution to Supreme Court scholarship of the scaling done so far will certainly be that it has laid a foundation for the more mature use of scaling in the future, closely allied to non-quantitative methods and freed from the abuses that have marked its use in the past.

Notes

1. For discussions of what constitutes scientific methods, see:

Heinz Eulau, "Segments of Political Science Most Susceptible to Behavioristic Treatment". In James C. Charlesworth (ed.), The Limits of Behavioralism in Political Science (Philadelphia, 1962), p. 32.

Arthur S. Goldberg, "Political Science as Science". In Nelson W. Polsby et. al., Politics and Social Life (Boston, 1963), pp. 30-3.

Samuel J. Eldersveld et. al., "Research in Political Behavior". In S. Sidney Ulmer (ed.), Introductory Readings in Political Behavior (Chicago, 1961), pp. 8 & 11.

Richard C. Snyder, "Experimental Techniques and Political Analysis". In Charlesworth, (op. cit., pp. 115-22.

Polsby, op. cit., pp. 4-8.

Robert A. Dahl, "The Behavioral Approach in Political Science". In Polsby, op. cit., pp. 18-21.

David Easton, "The Current Meaning of 'Behavioralism' in Political Science". In Charlesworth, op. cit., pp. 7-8.

2. Polsby, op. cit., pp. 8-14.

3. Ibid. See also Paul F. Lazarsfeld, "The American Soldier". In Polsby, op. cit., pp. 37-8.

4. Ibid.

5. Karl W. Deutsch, "The Limits of Common Sense". In Polsby, op. cit., p. 52.

6. Bertrand de Jouvenel, "On the Nature of Political Science". In American Political Science Review, vol. LV, no. 4 (Dec. 1961), pp. 773-4.

7. Mulford Q. Sibley, "The Limitations of Behavioralism". In Charlesworth, op. cit., pp. 70-1.

8. Leo Strauss, "An Epilogue". In Herbert J. Storing (ed.), Essays on the Scientific Study of Politics (New York, 1962), pp. 311-27. See also:

Walter Berns, "The Behavioral Sciences and the Study of Political Things". In American Political Science Review, vol. LV, no. 3 (Sept. 1961), pp. 558-9.

David Easton, The Political System (New York, 1963), ch. 9.

9. Goldberg, op. cit., pp. 27-8.

10. F.A. Hayek, The Counter-Revolution of Science,

(Glencoe, 1952), pp. 25-35.

11. Goldberg, op. cit., pp. 27-8.

12. Hayek, op. cit., pp. 44-51.

13. Sibley, op. cit., pp. 71-8.

14. Ibid., p. 70.

15. Russell Kirk, "Is Social Science Scientific?"

In Polsby, op. cit., pp. 60-4.

16. Strauss, op. cit., p. 312.

17. Ibid., p. 316.

18. Ibid., pp. 311-37.

19. Berns, op. cit., p. 556.

20. Kirk, loc. cit.

21. Strauss, op. cit., pp. 320 & 322-7.

22. Hayek, op. cit., pp. 50-1.

23. Kirk, loc. cit.

24. Wallace Mendelson, "The Neo-Behavioral Approach to the Judicial Process: a Critique". In American Political Science Review, vol. LVII, no. 3 (Sept. 1963), pp. 593-603.

25. Strauss, op. cit., pp. 312-3.

26. Hayek, op. cit., pp. 53-6.

27. Eulau, op. cit., p. 31.

Roger Hilsman, "The Foreign Policy Consensus". In Polsby, op. cit., p. 393.

H. Douglas Price, "Are Southern Democrats Different?" In Polsby, op. cit., pp. 740-4.

28. Thomas Morgan, "The People Machine". In Harpers Magazine,

29. Sibley, op. cit., pp. 89-92.

30. Hayek, op. cit., pp. 94-102.

31. Strauss, op. cit., pp. 318-9.

32. de Jouvenel, op. cit., pp. 773-4.

33. Ibid., p. 776.

34. Alpheus Thomas Mason, The Supreme Court: Palladium of Freedom (Ann Arbor, 1962), p. 175.

35. Ibid., pp. 175-6.

36. C. Herman Pritchett, The Roosevelt Court (New York, 1948), pp. 15-6.

37. Ibid., p. 19.

38. Ibid., p. 16.

39. For example:

David Fellman, "Constitutional Law in 1957-1958". In American Political Science Review, vol LIII, no. 1 (March 1959), pp. 138-80.

Fellman, "Constitutional Law in 1959-1960". In American

Political Science Review, vol LV, no. 1 (March 1961), pp. 112-35.

"The Supreme Court: 1958 Term". In Harvard Law Review, vol. 73, no. 1, pp. 840-940.

40. Francis P. McQuade et. al., "Mr. Justice Brannan and His Legal Philosophy". In Notre Dame Lawyer, vo. XXXIII, no. 3, pp. 321-49.

41. Ibid., p. 326.

42. Mason, op. cit.

43. Ibid., p. 169.

44. Ibid., p. 173.

45. Ibid., p. 169.

46. Helen Shirley Thomas, Felix Frankfurter: Scholar on the Bench (Baltimore, 1960).

47. Ibid., pp. ix-x.

48. Ibid., pp. 105-9.

49. Ibid., p. 218.

50. Ibid., pp. 256-8.

51. Ibid.

52. Ibid., p. 115.

53. Ibid., pp. 19-21.

54. Daniel M. Berman, "Mr. Justice Brennan: a Preliminary Appraisal". In The Catholic University of America Law Review, vol. VII, no.1 (Jan. 1958), pp. 1-15.

55. Ibid., p. 13.

56. Ibid., p. 14.

57. Ibid., p. 15.

58. Marlin M. Volz, "Mr. Justice Whittaker". In Notre Dame Lawyer, vol. XXXIII, no. 2 (March 1958), pp. 159-77.

59. Ibid., p. 176.

60. John M. Harlan, "The Frankfurter Imprint as Seen by a Colleague". In Harvard Law Review, vol. 76, no. 1 (Nov. 1962), pp. 1-2.

61. Ibid., p. 2.

62. Robert G. McCloskey, "Deeds without Doctrines: Civil Rights in the 1960 Term of the Supreme Court". In American Political Science Review, vol. LVI, no. 1 (March 1962), pp. 71-89.

63. 367 U.S. 203 (1961).

64. Clyde E. Jacobs, Justice Frankfurter and Civil Liberties (Berkeley, 1961).

65. Ibid., p. 14.

66. Ibid., pp. 35-6.

67. 341 U.S. 494 (1951)
68. Jacobs, op. cit., p. 127.
69. Ibid., pp. 120-1.
70. 354 U.S. 234 (1957)
71. Jacobs, op. cit., p. 127.
72. Ibid., p. 154.
73. Fred Rodell, "For Every Justice, Judicial Deference Is a Sometime Thing". In Georgetown Law Journal, vol. 50, no. 4 (summer 1962), pp. 700-8.
- 74.
75. Rodell, op. cit., p. 706-7.
76. Ibid., pp. 702-3.
77. Ibid., pp. 703-6.
78. Ibid., pp. 707-8.
79. Ibid., p. 708.
80. Wallace Mendelson, Justices Black and Frankfurter: Conflict on the Court
81. Pritchett, op. cit.
82. For a survey of the work that has been done, see Glendon Schubert, "Behavioral Research in Public Law". In American Political Science Review, vol. LVII, no. 2 (June 1963), pp. 433-45.
83. Glendon A. Schubert, Quantitative Analysis of Judicial Behavior (Glencoe, 1959; hereafter cited as Q), p. 305.
84. Ibid., p. 301.
85. Ibid., pp. 322-38.
86. Ibid., pp. 338-63.
87. Ibid., pp. 352-3.
88. Ibid., p. 354.
89. Harold J. Spaeth, "An Approach to the Study of Attitudinal Differences as an Aspect of Judicial Behavior". In Midwest Journal of Political Science, vol. V, no. 2 (May 1961), pp. 165-80.
90. S. Sidney Ulmer, "Supreme Court Behavior and Civil Rights". In Western Political Quarterly, vol. 13, no. 2 (June 1960), pp. 288-311.
91. Schubert, Q, p. 298.
92. Schubert, "From Public Law to Judicial Behavior". In Schubert (ed.), Judicial Decision-Making (New York, 1963), pp. 1-10.
93. Schubert, Q, pp. 280-90.
94. Ibid., p. 352.
95. Ibid., p. 353.
96. Mendelson, "The Neo-Behavioral...", p. 597.

97. Pritchett, op. cit., p. 33.
98. Schubert, Q, p. 314.
99. Mendelson, "The Neo-Behavioral...", p. 598.
100. Ibid., p. 596.
101. Ulmer, "A Note on Attitudinal Consistency in the United States Supreme Court". In Indian Journal of Political Science, vol. 22 (1961), pp. 195-204.
102. Schubert, "Behavioral...".
103. Ulmer, "Supreme Court...".
104. Spaeth, "An Approach...".
105. Spaeth, "Judicial Power as a Variable Motivating Supreme Court Behavior". In Midwest Journal of Political Science, vol. VI, no. 1 (Feb. 1962), pp. 54-82.
106. Joel B. Grossman, "Role-Playing and the Analysis of Judicial Behavior: the Case of Mr. Justice Frankfurter". In Journal of Public Law, vol. 11, no. 2 (1962), pp. 285-309.
107. See, for example, Schubert, "Civilian Control and Stare Decisis in the Warren Court". In Schubert, Judicial Decision-Making.
108. Ibid.
109. See Schubert, "Judicial Attitudes and Voting Behavior: the 1961 Term of the United States Supreme Court". In Law and Contemporary Problems, vol. XXVIII, no. 1 (winter 1963), pp. 100-42.
110. Ralph S. Brown, Jr., "Legal Research: The Resource Base and Traditional Approaches". In The American Behavioral Scientist, vol. VII, no. 4 (Dec. 1963), pp. 3-7.
111. Theodore L. Becker, "On Science, Political Science, and Law". In American Behavioral Scientist, vol. VII, no. 4 (Dec. 1963), pp. 11-5.
112. Hans W. Baade, "Forward to Jurimetrics, i.e. Law and Contemporary Problems", vol. XXVIII, no. 1 (winter 1963), pp. 3-4.

Selected Bibliography

On Scientific Methods in the Study of Politics:
 Charlesworth, James C. (ed.) The Limits of Behavior-
alism in Political Science, Philadelphia, 1962.

Hayek, F.A. The Counter-Revolution of Science,
 Glencoe, 1952.

de Jouvenel, Bertand. "On the Nature of Political
 Science". In American Political Science Review (Dec.
 1961).

Polsby, Nelson W. et. al. Politics and Social Life,
 Boston, 1963.

Storing, Herbert J. (ed.) Essays on the Scientific
Study of Politics, New York, 1962.

Ulmer, S. Sidney (ed.) Introductory Readings in
Political Behavior, Chicago, 1961.

Non-quantitative Studies of the Supreme Court:
 Jacobs, Clyde E. Justice Frankfurter and Civil
Liberties, Berkeley, 1961.

McCloskey, Robert G. "Deeds without Doctrines: Civil
 Rights in the 1960 Term of the Supreme Court". In
American Political Science Review, vol. LVI, no. 1
 (March 1962), pp. 71-89.

Rodell, Fred. "For Every Justice, Judicial Deference Is
 a Sometime Thing". In Georgetown Law Journal, vol. 50,
 no. 4 (summer 1962), pp. 700-8.

Thomas, Helen Shirley. Felix Frankfurter: Scholar on
the Bench, Baltimore, 1960.

Quantitative Studies of the Supreme Court:
 Grossman, Joel B. "Role-Playing and the Analysis of
 Judicial Behavior: The Case of Mr. Justice Frankfurter".
 In Journal of Public Law, vol. 11, no. 2 (1962), pp.
 285-309.

Schubert, Glendon A. (ed.) Judicial Decision-Making,
 New York, 1963.

Schubert, Glendon A. Quantitative Analysis of Judicial
Behavior, Glencoe, 1959.

Spaeth, Harold J. "Judicial Power as a Variable
 Motivating Supreme Court Behavior". In Midwest Journal
of Political Science, vol. VI, no. 1 (Feb. 1962), pp.
 54-82.

Miscellaneous:

Jurimetrics, i.e. Law and Contemporary Problems, vol. XXVIII, no. 1 (winter 1963).

Mendelson, Wallace. "The Neo-Behavioral Approach to the Judicial Process: A Critique". In American Political Science Review, vol. LVII, no. 3 (Sept. 1963), pp. 593-603.

Schubert, Glendon. "Behavioral Research in Public Law" (bibliographical essay). In American Political Science Review (June 1963), pp. 433-45.