Artificial Extremism
In Interest Group Ratings

This paper shows that interest group ratings based on roll-call voting records tend to exaggerate the degree of extremism and bipolarity in Congress. That is, the scores assigned to relatively moderate members of Congress are typically more extreme than their true positions. The paper then examines some of the implications of this bias for empirical studies that use ratings.

Interest group ratings based on the roll-call voting records of members of Congress are used widely in political science and public choice studies, for a variety of purposes. Indeed, these ratings are used so extensively that virtually all students of U.S. politics must be familiar with them. Political scientists use ratings, both as dependent and as independent variables, to study the linkages between representatives and constituents (e.g., Markus 1974; Schwartz and Fenmore 1977; Erikson and Wright 1980; Johannes and McAdams 1981). Studies in the public choice tradition use ratings as proxies for ideology in estimating the relative importance of various determinants of congressional voting behavior (Kau and Rubin 1979, 1982; Chappell 1981, 1982; Kalt and Zupan 1984; Peltzman 1984; Coughlin 1985; MacArthur and Marks 1988; Dougan and Munger 1989; Davis and Porter 1989; Richardson and Munger 1990). Several recent papers use ratings to assess the extent to which congressional committees are representative of their parent bodies (Ray 1980; Weingast and Marshall 1988; Krehbiel 1990). Finally, some researchers apply factor analysis or scaling techniques to sets of ratings, in attempts to better estimate the ideological positions of members of Congress (Kritzer 1978; Poole 1981; Poole and Rosenthal 1984; Poole and Daniels 1985).1

One feature common to most interest group ratings is that they produce bimodal distributions of scores. That is, the ratings tend to assign extreme scores to a large fraction of members of Congress and moderate scores to relatively few members. This was true for the ratings compiled by the ADA, ACA, CCUS and COPE during the
1970s (see Fowler 1982) and continued to be true into the 1980s (see the histograms on the right-hand side of Figures 1–4). The scaling techniques that use ratings as inputs also produce bimodal distributions of estimated legislator positions (Poole 1981; Poole and Rosenthal 1984; Poole and Daniels 1985).

There are, however, at least two reasons to doubt that the true distribution of congressional preferences is bimodal. First, the median voter theorem implies that, in a one-dimensional world, relatively moderate candidates are more likely to be elected than extremists. Thus, the candidates who win elections should in most cases be those whose ideal points are close to the ideal points of the median voters in the districts they seek to represent (or they should be candidates who do not have strong policy preferences themselves, but who behave as if their ideal points were close to the median voter ideal points in their districts). It is unlikely that the distribution of voter medians across congressional districts is bimodal, especially if preferences are like many other individual characteristics, such as income and education. Second, on an empirical level, the one-dimensional D-NOMINATE scaling procedure of Poole and Rosenthal (1985, 1987) typically produces distributions of estimated ideal points that are unimodal. While not immune from criticism (see, e.g., Brady 1988), these estimated ideal points are probably superior to interest group ratings as estimates of the ideological positions of members of Congress, because of both the method and the set of votes used. Poole and Rosenthal use virtually all roll calls taken during each congressional session, rather than a small sample.

This paper shows that interest group ratings will often be bimodal even when the true distribution of congressional preferences is not bimodal. The reason is simple: if a large fraction of the roll calls used in constructing a group’s rating have cleavage points close to the center of the distribution of legislators’ ideal points, then many relatively moderate legislators will be assigned scores that are more extreme than their true ideal points. Under plausible assumptions, this effect will be so strong that the distribution of scores will be bimodal even when the distribution of legislators’ ideal points is not.

The tendency for interest group ratings to overstate the degree of extremism in Congress is likely to be an important problem in practice. Most interest groups base their ratings on samples of roll calls that include a disproportionate number of close votes—that is, roll calls in which the fraction of members voting on the majority side is close to 50%. A close vote generally means that the cleavage point on the roll
FIGURE 1
Distribution of ADA Ratings and Vote Margins, 1979 and 1980

Interest Group Ratings 321
FIGURE 2
Distribution of ACA Ratings and Vote Margins, 1979 and 1980

Distribution of Vote Margins of Roll Calls Used to Construct Ratings

1979

Distribution of Ratings

1979

1980

0 10 20 30 40 50 60 70 80 90 100
Percentage of Representatives Voting Aye

Fraction of Roll Calls

0 .1 .2 .3 .4 .5

0 .1 .2 .3 .4 .5

0 .1 .2 .3 .4 .5

0 .1 .2 .3 .4 .5

0 10 20 30 40 50 60 70 80 90 100
Rating

Fraction of Representatives

0 .1 .2 .3 .4 .5

0 .1 .2 .3 .4 .5

0 .1 .2 .3 .4 .5

0 .1 .2 .3 .4 .5

0 10 20 30 40 50 60 70 80 90 100
Percentage of Representatives Voting Aye

Fraction of Roll Calls
FIGURE 3
Distribution of COPE Ratings and Vote Margins, 1979 and 1980

Distribution of Ratings

1979

1980

Distribution of Vote Margins of
Roll Calls Used to Construct Ratings

1979

1980

Interest Group Ratings
FIGURE 4
Distribution of CCUS Ratings and Vote Margins, 1979 and 1980

Distribution of Ratings

1979

Distribution of Vote Margins of Roll Calls Used to Construct Ratings

1979

Percentage of Representatives Voting Aye

1980

Percentage of Representatives Voting Aye

1980

Fraction of Roll Calls

Fraction of Roll Calls
call is near the center of the distribution of members' ideal points. Thus, a deliberate effort to avoid lopsided roll calls produces artificially extreme scores for relatively moderate members. The histograms on the left-hand side of Figures 1–4 show the numbers of close and lopsided votes included in the ADA, ACA, COPE, and CCUS ratings in 1979 and 1980. For most of these, there is a strong bias towards close votes (I will have more to say about these histograms below). Another example is the rating constructed by Peltzman (1985), which excludes all roll calls for which the margin of victory was greater than 2 to 1.

Ratings are summary statistics that depend both on legislators' true preferences and on the location of the cleavage points (or, more generally, the locations of the aye and nay alternatives) associated with the roll calls used in constructing the ratings. Other factors may also affect the relationship between ratings and preferences, such as logrolling, vote trading, and leadership pressures. Lacking sufficient information about the distribution of cleavage points and about the importance of other potential complications, one can not be confident that ratings provide accurate estimates of legislators' preferences.

The fact that ratings are based on samples of roll calls makes them a special case of a general problem in the estimation of legislator preferences. To estimate legislators' preferences from roll-call data, one must make assumptions not only about how legislators make their voting decisions but also about how legislation reaches the roll-call stage. Just as it is unlikely that legislators' ideal points are randomly distributed throughout some feasible space, it is unlikely that roll calls on the House or Senate floor are randomly drawn from some set of all possible bills. For example, if most roll calls on important issues are the results of unanticipated shocks to policies or preferences (as in Snyder 1991a) and if most of these shocks are small relative to the range of ideal points among legislators, then the cleavage points on most important roll calls will be close to the median of the legislators' ideal points and few roll calls will have cleavage points near the tails of this distribution. Figure 5 shows that, across all roll calls, there are more close votes than lopsided votes, although the difference is less pronounced than in the small samples chosen by interest groups. The conceptual and methodological issues related to the endogenous roll-call agenda have recently received more attention (e.g., Schneider 1979; Koford 1990; VanDoren 1990; Snyder 1992a, 1992b) and will undoubtedly receive still more in the future.
A Simple Example

The intuition underlying the main argument in this paper is best illustrated by a simple example. Suppose there are 11 legislators, each with symmetric, single-peaked preferences over a one-dimensional policy space. Their ideal points are distributed uniformly over the interval [0,1], so one legislator's ideal point is at 0, one is at .1, one is at .2, and so on (see Figure 6, top).

Suppose an interest group uses 12 roll calls to calculate its ratings. If we assume that legislators always vote sincerely (i.e., that when faced with two options they always vote for the one they prefer), each roll call can be represented by a cleavage point on the line, with legislators whose ideal points lie to one side of the cleavage point voting aye on the roll call and those on the opposite side voting nay. Suppose the interest group's ideal point is far to the right, so that the group always labels those legislators whose ideal points lie to the right of the cleavage point as casting correct votes and those with ideal points to the left of the cleavage point as casting incorrect votes. Finally, suppose that
Interest Group Ratings

FIGURE 6
Uniform Distribution of Ideal Points, with Bimodal Distribution of Ratings

Correct votes are scored 1 and incorrect votes 0 and that a legislator's rating is simply his or her average score (the total number of correct votes divided by 12).

Suppose the 12 roll calls have cleavage points as follows: one each at .25 and at .75, two each at .35 and at .65, and three each at .45 and at .55 (see Figure 6, center). Thus the distribution of cleavage points is symmetric and unimodal, with its peak at around .5.

The result is a bimodal distribution of ratings, even though the distribution of ideal points is uniform (see Figure 6, bottom). All of the moderately left and moderately right legislators receive ratings that are more extreme than their ideal points. That is, legislators with ideal points in the interval (0,.5) are assigned ratings closer to 0 than their ideal points, and those with ideal points in the interval (.5,1) are assigned ratings closer to 1 than their ideal points. For example, a
A legislator whose true ideal point is at .4 votes correctly on 3 of the 12 roll calls and therefore receives a rating of .25. Moreover, there are mass points of legislators with ratings of 0 and 1. Thus, the ratings make the legislature appear to be more polarized than it actually is.

If the distribution of cleavage points is more uniform, the distribution of ratings will be less bimodal but will still exaggerate the extremism in the legislature, unless the distribution of cleavage points is exactly uniform. On the other hand, if the distribution of cleavage points is bimodal, then the ratings will tend to understate the degree of extremism in the legislature. In short, to tell whether the distribution of scores produced by a rating accurately reflects the underlying distribution of ideal points, one must know more about the distribution of cleavage points in the set of bills used in constructing the rating.5

A More General Result

I now present a somewhat more general result, for normal distributions of ideal points and cleavage points. The result is simply stated: if the distributions of ideal points and cleavage points are both normal, the distribution of legislators’ interest group scores will be bimodal if and only if the variance of the distribution of roll-call cleavage points is smaller than the variance of the distribution of legislator ideal points.

As above, assume the policy space is one dimensional. Let $F: \mathbb{R} \to [0,1]$ be a cumulative distribution function that describes the distribution of legislator ideal points, and let $f = F'$ be the probability density function corresponding to $F$. Similarly, let $G: \mathbb{R} \to [0,1]$ be a cumulative distribution function describing the distribution of cleavage points, and let $g = G'$ be the probability density function corresponding to $G$. Also, as above, assume that legislators always vote sincerely, that on each roll call the interest group assigns a 1 to all legislators who vote for the alternative to the right of the cleavage point, and that each legislator’s rating is equal to his or her average score. Then, since a legislator with ideal point at $x$ receives a score of 1 whenever the cleavage point is less than $x$ and a score of 0 otherwise, the rating of such a legislator is simply equal to the fraction of roll calls with cleavage points less than $x$, or $G(x)$. The following proposition is then easily proved.

**Proposition 1:** Suppose the distribution of legislator ideal points is $N(0,1)$ and the distribution of cleavage points is $N(0,\sigma^2)$. 
Interest Group Ratings

Then the resulting distribution of legislator ratings is bimodal if and only if \( \sigma^2 < 1 \). (For the proof, see the Appendix.)

When \( \sigma^2 \geq 1 \), the distribution of ratings is not bimodal, but it still exaggerates the degree of extremism in the legislature. For example, if \( \sigma^2 = 1 \), then the distribution of ratings is uniform over the interval \([0,1] \). However, the distribution of ideal points is normal and therefore strongly unimodal (and symmetric), with its mode in the center. Thus, the distribution of the ratings does not describe the distribution of ideal points very well—instead, it greatly understates the number of moderate legislators.

Clearly, if the distribution of ideal points is normal, then an interest group’s ratings accurately describe the distribution of ideal points only if the distribution of ratings is itself approximately normal. The example in the previous section suggests that, for the distribution of ratings to be normal, the cleavage points in the sample of roll calls used in constructing the rating should be approximately uniform over the range of legislators’ ideal points. Figure 7 shows that this is in fact the case. In the top half of the figure, the distribution of cleavage points is \( N(0,2) \), and therefore somewhat more diffuse than the distribution of ideal points. The left-hand side shows the distribution of cleavage points relative to the distribution of ideal points, and the right-hand side shows the resulting distribution of ratings compared to a normal distribution with the same mean and standard deviation (the mean is .5 and the standard deviation is approximately .179). The distribution of ratings is roughly normal, although the fit is not great. In the bottom half of Figure 7, the distribution of cleavage points is \( N(0,5) \). There, the distribution of cleavage points fits the distribution of ideal points quite well, and the distribution of ratings looks quite normal (with mean .5 and standard deviation approximately .079).

Figure 7 suggests that, given a distribution of legislator ideal points, the variance of the distribution of legislator ratings is inversely related to the variance of the distribution of cleavage point. In fact, it is easy to show that this is true. This result forms the basis for an empirical test in the next section, and may prove to be generally useful in empirical work.\(^6\)

**Proposition 2:** Suppose the distribution of legislator ideal points is \( N(0,1) \). Let \( g_1 \) and \( g_2 \) be probability density functions describing two distributions of cleavage points, and assume that these distributions are \( N(0,\sigma_1^2) \) and \( N(0,\sigma_2^2) \). Then the variance of the legislator ratings produced by \( g_1 \) is greater than the variance of the legislator ratings produced by \( g_2 \) if and only if \( \sigma_1^2 < \sigma_2^2 \). (For the proof, see the Appendix.)
FIGURE 7
Effect of More Uniform Distribution of Cleavage Points

Distribution of Ideal Points and Cleavage Points

Cleavage Points Distributed N(0,2)

Ideal Points, N(0,1) ———
Cleavage Points ———

Distribution of Ratings Relative to Distribution of Ideal Points

Cleavage Points Distributed N(0,2)

Distribution of Ratings

Normal Distribution ———
(in top figure variance is .179, in bottom figure variance is .079)
A Glance at Some Data

I have so far presented theoretical possibilities. This section presents empirical evidence indicating that these possibilities are likely to be important in practice.

Proposition 2 states that the variance in legislators' scores for a given group's rating is negatively related to the variance of the distribution of cleavage points in the sample of roll calls used to construct the rating. Of course, cleavage points are not observable in practice, but must be estimated. Vote margins are observable, however, and in the unidimensional model above there is a close relationship between the cleavage point on a roll call and the vote margin on that roll call. This relationship is straightforward: if close to 50% of the legislators vote aye on a roll call, then the cleavage point on that roll call must be near the center of the distribution of ideal points, while if 10% or 90% of the legislators vote aye, then the cleavage point must be near one of the ends of the distribution. By an argument similar to the proof of proposition 2, it is easy to show that the variance of the distribution of vote margins is an increasing function of the variance of the distribution of cleavage points. The intuition is simple: if the cleavage points are clustered, then all vote margins will be about the same, while if the cleavage points are spread out, then the vote margins will be spread out as well. Thus proposition 2 can be translated into a proposition about the relationship between the variance in scores and the distribution of vote margins in the sample of roll calls: namely, that the variance in legislators' scores for a given group's rating is inversely related to the variance of the distribution of vote margins in the sample of roll calls used to construct the rating. This is the proposition I will investigate.

The eight group ratings in Figures 1–4 show a pattern quite consistent with the prediction. For six of the eight ratings, the distribution of vote margins is densest near the center, with many more close roll calls than lopsided ones, while the distribution of legislators' scores is bimodal, with many more extremists than moderates. The ACA rating in 1979 is a good example (Figure 2). An interesting exception, also consistent with the proposition, is the CCUS rating in 1980 (Figure 4). For that rating, the distribution of vote margins is more widely dispersed, and the distribution of scores is strongly unimodal.

Figure 8 shows that the pattern in fact holds quite generally. The figure plots the variance in the distribution of scores across representatives against the variance in the distribution of victory margins (more precisely, the percentage of representative voting with the majority), for 47 ratings compiled in 1979 and 1980. As was
predicted, this relationship is negative, and it is surprisingly strong. Figure 8 also shows the regression line for a univariate regression of $y$ on $x$ (the coefficient on $x$ is $-1.28$, with a t-value of $-5.89$; the simple correlation between $y$ and $x$ is $-0.66$). Overall, Figures 1–4 and 8 strongly suggest that the interest group ratings most commonly used by congressional scholars tend to overstate the extent of extremism among legislators’ preferences and that the bimodality in these ratings may be quite artificial.

Applications

In this section, I discuss several applications of the results above. The first application concerns public choice studies that attempt to untangle the effects on congressional roll-call voting of constituency interests, ideology, shirking, interest group lobbying and campaign contributions, and logrolling. Stated simply, the implication of the discussion so far is this: since most interest group ratings make
fairly moderate members of Congress appear more extreme than they are, studies which use such ratings as measures of ideology tend to underestimate the effect of ideology on voting. This conclusion is somewhat surprising, since the studies in question virtually always find the ratings to be highly significant explanatory variables. It suggests that ideology may be an even more important determinant of roll-call voting than previous estimates have shown.9

To understand the result, suppose that a legislator’s ideology is his or her true ideal point on the left-right continuum. Also, suppose the true relationship between ideology and the propensity to vote for a particular bill is of the form \( p_i = \alpha_0 + \beta_0 x_i + \epsilon_i \), where \( p_i \) is legislator \( i \)'s propensity to vote for the bill and \( x_i \) is legislator \( i \)'s ideal point. Finally, suppose an observer sees \( \bar{x}_i \) rather than \( x_i \), where \( \bar{x}_i \) is legislator \( i \)'s ADA score (or some other rating). If \( \bar{x}_i \) is plagued with the artificial extremism studied above, then \( \bar{x}_i < x_i \) for relatively small values of \( x_i \) and \( \bar{x}_i > x_i \) for relatively large values of \( x_i \) (i.e., moderate liberals are misclassified as rather extreme liberals, and moderate conservatives are misclassified as rather extreme conservatives). It is then easy to see that estimating the equation \( p_i = \alpha + \beta \bar{x}_i + \bar{\epsilon}_i \) produces an estimate \( \hat{\beta} \) that is biased towards zero relative to \( \beta_0 \). This is shown graphically in Figure 9, where the true points are marked with circles and the observed points (i.e., the data) are marked with squares.10 The intuition is straightforward: if moderately liberal legislators receive extremely liberal ratings, then an observer who knows the ratings (but not the true positions) will expect those legislators to have a strong propensity to vote for the liberal alternative on any roll call. Since the legislators' actual votes will reflect their true, moderate positions, they will not vote for the liberal alternative as overwhelmingly as expected. Thus, the observer will wrongly conclude that ideology does not have a very strong effect on the legislator’s propensity to vote for the liberal alternative. The same argument holds for moderately conservative legislators. The extent of the bias clearly depends on the extent to which the scores are artificially extreme (and also on the distribution of true ideal points).

A closely related application concerns the studies of constituency/representative linkages that use interest group ratings to measure legislators’ roll-call records (e.g., Markus 1974; Schwartz and Fenmore 1977; Johannes and McAdams 1981). These studies typically regress representatives’ ratings on measures of constituency preferences, such as the district vote in presidential elections, demographic characteristics, survey responses, and voting on ballot propositions. (Some studies examine simple correlation coefficients rather than regression
coefficients). The underlying hypothesis is that a high degree of correlation indicates that representatives are paying close attention to their districts, while a low degree of correlation is evidence to the contrary. If interest group ratings are biased measures of representatives’ actual voting positions, then the regression r-squareds (and the simple correlation coefficients) found in these studies will typically understate the true degree of correspondence between constituency preferences and representatives’ voting positions. Of course, it is likely the measures of constituency preferences contain a large amount of error themselves, and some measures may also suffer from bias similar to that contained in ratings.

Another application concerns studies that use interest group ratings to determine which congressional committees are representative of their parent bodies and which are preference outliers (Ray 1980; Weingast and Marshall 1988; Krehbiel 1990; Hall and Grofman
These studies compare the mean score of committee members on some group's rating with the mean score of noncommittee members, using t-statistics to test the null hypothesis that the committee and noncommittee means are the same. Also, since it is arguable on theoretical grounds that a more relevant difference is that between the median ideal point of the committee and the median ideal point in the whole house, Krehbiel (1990) and Hall and Grofman (1990) report the difference between the median score in the committee and the median score in the house, as well as the percentage of legislators whose scores lie between the two medians.

Artificial extremism in the distribution of scores has three consequences for these studies that deserve attention. The first consequence is that, if the distribution of scores has artificially large tails, then the standard deviation of the distribution will be artificially large. The t-statistics of the difference between committee and noncommittee mean scores will then tend to be artificially small, and consequently there will be a bias towards accepting the null hypothesis that the two means are equal. That is, there will be a bias towards incorrectly concluding that committees are not significant preference outliers.

It is difficult to know by how much the standard deviations of various scores overstate the standard deviations in ideal points, because the latter are unmeasured. It is possible to gain some insight, however, by considering a uniform distribution as a base case. Suppose 435 representatives have ideal points distributed uniformly over some interval, and the mean ideal point is at $\bar{x}$. Assuming that all ideal points lie between 0 and 100 (as they do in most ratings), if $\bar{x} < 50$ then the largest possible interval over which the ideal points can be spread is $[0, 2\bar{x}]$. Similarly, if $\bar{x} > 50$, then the largest possible interval is $[2\bar{x} - 100, 100]$. In either case, the standard deviation of the representatives' ideal points is approximately $0.58\bar{x}$ (and the mean is $\bar{x}$). By comparison, of the 47 ratings in the sample used in Figure 8, 37 (almost 80%) have standard deviations greater than $0.58$ times their mean. Thus, for the vast majority of these ratings, the distributions of scores has a larger standard deviation than the widest possible uniform distribution with the same mean.

The second consequence of artificially bimodal distributions is that the percentage of representatives whose scores lie in the gap between the median score among committee members and the median score in the whole House is not a very meaningful statistic. Since bimodal distributions have relatively few scores near the center of the distribution, the percentage of scores between the medians will tend to
be small even if the medians are far apart. In fact, Krehbiel (1990) finds that for most committees this percentage is quite small, often even when the difference between the medians is rather large. For example, using the American Security Council (ASC) rating, Krehbiel reports that the difference between the House median and the median on the House Foreign Affairs committee is 30 points (30% of the possible range of scores), but that only 8% of all representatives scored between the medians. Again, a comparison with uniform distributions is illuminating. Suppose committee and noncommittee members' ideal points are both uniformly distributed, with medians (and means) at 65 and 35, respectively. That is, suppose the committee members' ideal points are distributed uniformly over the interval [30, 100] (the widest possible interval given a mean of 65), and the nonmembers' ideal points are distributed uniformly over the interval [0, 70]. Then the difference between the two medians (and the two means) is 30. However, the percentage of representatives whose ideal points lie in the gap between the two medians is about 43% (100x3/7), which is obviously much larger than 8%. Thus, if scores are bimodal one should probably pay more attention to the actual difference between the medians, not the percentage of members whose ideal points lie in the gap.

While the first two problems imply that artificial extremism in interest group ratings may lead researchers to underestimate the differences between committee preferences and floor preferences, the third problem is exactly the opposite is true. That is, for some distributions of ideal points, artificial extremism in the distribution of scores actually exaggerates the difference between a committee and the floor, leading one to conclude that a committee is more of an outlier than it really is. For example, a committee may contain a disproportionate number of moderate conservatives (or moderate high-demanders) while the House as a whole contains a disproportionate number of moderate liberals (or moderate low-demanders), or the reverse may be true. Then, since a bias in ratings gives moderate legislators scores that are more extreme than their true positions, a significant number of the committee members will receive ratings to the right of their true positions, and a significant number of noncommittee members will receive ratings to the left of their true positions. The net effect, clearly, is to exaggerate the difference in preferences between committee members and non-members.

A simple example shows how severe the problem can be. Suppose a committee has 40 members, with ideal points as follows: two each at 0, .1, .2, .3, .8 and .9; four at .4; six at .5; ten at .6; and eight at .7 (see Figure 10, upper left-hand side). Also, suppose there are 400 non-
members of the committee, with ideal points as follows: 20 each at .1, .2, .7, .8, .9 and 1.0; 80 at .3; 100 at .4; 60 at .5; and 40 at .6. Thus, the distribution of nonmembers is a mirror image of the committee, multiplied by 10 (producing a house with 440 members, which is close enough to 435). Then the mean and variance of the committee ideal points are .52 and .23 respectively, and the mean and variance of the noncommittee ideal points are .48 and .225. The t-statistic for a test of the difference in means is 1.07, which is not significant even at the 25% level. Now, suppose the ratings of the 440 representatives are generated by a set of 12 roll calls with cleavages as in Figure 6 (middle figure). The resulting distributions of scores for committee members and nonmembers are shown in the lower half of Figure 10 (left-hand side for committee members, right-hand side for nonmembers). The mean and variance of the committee members' ratings are .575 and .354.
respectively, and the mean and variance of the non-committee ratings are .425 and .350, and the t-statistic for a test of the difference in means is 2.58, which is significant at the 1% level.

Again, because the distribution of actual ideal points is unknown, it is difficult to guess how severe a problem this bias is likely to be in practice. On the one hand, the distributions of ideal points in the example above are not particularly odd or contrived. On the other hand, according to the high-demand-outlier hypothesis (see Weingast and Marshall 1988), outlier committees should be laden with extremely high-demand members, not with the moderately high-demand members in the example.

Overall, I suspect that the first two problems raised above are more important than the third, and therefore on balance the bias produced by artificial extremism in ratings is in the direction of finding no difference between committee and floor preferences. In any event, the problems raise doubts about the conclusions one can draw from empirical studies that focus exclusively on the means, medians, and variances of ratings.

A final application involves studies of roll-call voting behavior over time (see Sinclair 1981, 1982; Poole and Rosenthal 1984; Collie and Brady 1985; Rohde 1991). The basic problem is that one cannot necessarily infer a change in legislators' preferences from a change in their roll-call behavior, except in certain cases.

As an example, consider Rohde (1991), who uses indices of mean party cohesion to show that there was a dramatic increase in intraparty homogeneity in the House of Representatives, especially among Democrats, following the reforms of the early 1970s. For each party, these indices are computed as the absolute value of the difference between the percentages of party members voting aye and voting nay on each of a set of roll calls, with these absolute differences then averaged across the set of roll calls. Rohde tracks the changes in these indices over time, and notes that for Democrats the index increased sharply over the period 1971–88, especially from 1981 to 1988. He interprets this change as due at least in part to changes in the preferences of Democratic representatives (partly due to replacement, especially among the southerners). Another possibility, however, is that the change in the mean cohesion index is due primarily to a change in the cleavage points of the roll calls on the agenda and that preferences were relatively stable. For example, suppose Democratic preferences are given by the upper right-hand figure in Figure 10, and suppose that in one year most roll calls have cleavage points between .45 and .55 while in the next year most roll calls have cleavage points between .55 and
.65. The Democratic party cohesion index will then be higher in the second year than in the first, even if preferences are unchanged: Democratic cohesion is \( \frac{11}{20} - \frac{9}{20} = \frac{1}{10} \) on roll calls with cleavage points at .45, \( \frac{4}{10} \) on those with cleavage points at .55, and \( \frac{6}{10} \) on those with cleavage points at .65.

Looking at both parties' cohesion indices together sometimes helps in determining whether or not preferences have changed. In particular, there is some tendency for changes in the distribution of cleavage points to cause the two parties' cohesion indices to move in opposite directions. (Consider the case in which both parties' ideal point distributions are uniform, with Democrats generally to the left of the Republicans.) Thus, when the two indices move in the same direction, it is likely that the changes in these indices reflect some change in preferences.17

It is possible, however, for both parties' cohesion indices to move in the same direction or for one party's cohesion index to change while the other party's index remains constant, even when all legislators' preferences are unchanged. To see this, consider the example of the previous paragraph and suppose that Republican preferences are given by the upper left-hand figure in Figure 10. Republican cohesion is then \( \frac{14}{20} - \frac{6}{20} = \frac{4}{10} \) on roll calls with cleavage points at .45, \( \frac{1}{10} \) on those with cleavage points at .55, and \( \frac{4}{10} \) on those with cleavage points at .65. If half of the roll calls in one year have cleavage points at .45 and half have cleavage points at .55, then the Democratic and Republican indices of mean cohesion are both .25. If, in the next year, half of the roll calls have cleavage points at .55 and half have cleavage points at .65, then the Democratic and Republican indices of mean cohesion are .5 and .25, respectively. Thus, Democratic mean cohesion increases, while Republican mean cohesion remains constant. On the other hand, if 40% of the second year cleavage points are at .55 and 60% are at .65, then the Democratic and Republican indices of mean cohesion are .52 and .28, respectively—in this case, both indices increase.18

These examples raise doubts about the correct interpretation of Rohde's evidence on changes in House roll-call voting (although they do not question the overall importance of Rohde's analysis, which is insightful and rich in detail). As was noted above, there was a significant increase in Democratic mean cohesion over the period 1971–88. For Republicans outside the Northeast, mean cohesion increased over the period, suggesting that some change in preferences probably occurred. Among northeastern Republicans, however, mean cohesion exhibited little trend over the period; if anything, it decreased slightly. Careful consideration of the changes, if any, in the distributions of
vote margins on the roll calls taken during 1971–89 (as proxies for the distributions of cleavage points) would surely add to our understanding of the changes that took place.

**Concluding Remarks**

Proposition 1 above gives plausible conditions under which interest group ratings based on roll-call data will exaggerate the degree of extremism in the distribution of legislators’ ideal points. As was noted in the introduction, it is likely that most if not all ratings suffer from this problem, since the samples of roll calls used to construct these ratings tend not to include many roll calls with lopsided votes but are instead weighted heavily in favor of roll calls with close votes. The actual distribution of legislator ideal points may or may not be bimodal but, whatever the case, one cannot discover the truth simply by looking at interest group ratings.

What does this mean for the analysis of roll-call data? A few suggestions come to mind. First, the examples illustrated in Figures 6 and 7 suggest a way to reduce the bias in interest group ratings, without throwing away all of the information provided by the rating (in particular, without throwing away the groups’ judgments about what are important roll calls and what are correct votes on those roll calls). Recall that in order for a rating to accurately reflect the underlying distribution of ideal points, the sample of roll calls used in constructing the rating must contain roughly equal numbers of very close, fairly close, and lopsided votes. Thus, one way to correct for the bias in a group’s rating is to recalculate legislators’ scores after weighting the roll calls included in the rating to approximate a uniform distribution. Generally, lopsided roll calls will have to be weighted more heavily than close roll calls. In principle, the weighting should produce ideal point estimates that are less biased than the ratings themselves, although it remains to be seen how well this technique works in practice.

Another possibility, which is even simpler, is to convert interest group ratings into discrete variables, as in Johannes and McAdams (1981). If the information contained in most ratings is best viewed as ordinal-level data rather than interval-level data, then assigning low, middle, and high values to legislators may result in variables that are less biased. The choice of dividing points will always be somewhat arbitrary, of course, but it is possible to check for robustness. At a minimum, researchers using interest group ratings must analyze the distribution of scores and the distribution of vote margins in more detail.
On the basis of the results above, I suspect that the mean and variance of a group’s scores are rarely the only relevant summary statistics.

Better yet, researchers could rely less on interest group ratings than they have in the past and turn instead to estimates produced by scaling techniques, such as the techniques of Poole and Rosenthal (1985, 1991) and Brady (1989). These techniques can (indeed, should) be applied to large sets of roll calls, and they allow the researcher to choose which roll calls to include in his or her analysis, rather than relying on the choices of some interest group. Software to implement these scaling techniques is becoming more widely available (a one-dimensional version of Poole and Rosenthal’s NOMINATE scaling algorithm is now publicly available), and as they come into widespread use the strengths and weaknesses of various scalings will become more apparent.

Finally, Hall and Grofman (1990) are surely correct in advising researchers to use constituency characteristics as well as roll-call voting behavior to measure representatives’ preferences. I would add that carefully constructed data derived from interviews or drawn from other decisions made by legislators, such as bill sponsorship and committee votes, are also essential. Only when a variety of data lead to similar conclusions can these conclusions be stated with confidence.

James M. Snyder, Jr., is Assistant Professor of Economics, University of Chicago, Chicago, Illinois 60637.

APPENDIX

Proof of Proposition 1

Let \( \tilde{r}: \mathbb{R} \rightarrow [0,1] \) be the mapping from legislator ideal points into ratings, so \( \tilde{r}(x) \) is the rating of a legislator with ideal point \( x \). Then, as was noted above, \( \tilde{r}(x) = G(x) \) for all \( x \in \mathbb{R} \). Also, let \( \tilde{x} \) be the inverse of \( \tilde{r} \), so that for each \( r \in [0,1] \), \( \tilde{x}(r) \) satisfies \( \tilde{r}(\tilde{x}(r)) = r \), or \( G(\tilde{x}(r)) = r \). Evidently, \( \tilde{x} \) is well-defined on \([0,1]\), since \( G \) is strictly monotonic. Also, implicit differentiation yields \( \tilde{x}'(r) = 1/g(\tilde{x}(r)) \). Let \( h: [0,1] \rightarrow [0,1] \) be the probability distribution function describing the distribution of interest group ratings. Then for each \( r \in [0,1] \), \( h(r) = f(\tilde{x}(r))x'(r) = f(\tilde{x}(r))/g(\tilde{x}(r)) \). By assumption, \( f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \) and \( g(x) = \frac{1}{\sqrt{2\sigma^2}}e^{-x^2/2\sigma^2} \), so \( h(r) = \frac{1}{\sigma}\sqrt{\frac{2}{\pi}}e^{\frac{\sigma^2}{2}\tilde{x}(r)^2(1-r^2)/2\sigma^2} \). Differentiating yields \( h'(r) = \frac{\sigma(1-r^2)}{\sigma^2}\sqrt{\frac{2}{\pi}}e^{\frac{\sigma^2}{2}\tilde{x}(r)^2(1-r^2)/2\sigma^2} \). Clearly, \( h'(r) \) is positive if and only if \( \tilde{x}(r)(1-r^2) \) is.

Now, since \( \tilde{x} = G^{-1} \) and \( G \) is a normal cumulative distribution function with mean 0, \( \tilde{x}(r) > 0 \) if and only if \( r > 1/2 \), \( \tilde{x}(r) = 0 \) if and only if \( r = 1/2 \), and \( \tilde{x}(r) < 0 \) if and only if \( r < 1/2 \). Thus, if \( \sigma^2 < 1 \), then \( h'(r) > 0 \) if and only if \( r < 1/2 \), \( h'(r) = 0 \) if and only if \( r = 1/2 \), and \( h'(r) < 0 \) if and only if \( r > 1/2 \). Therefore, \( h \) is first decreasing and then increasing in \( r \); that is, \( h \) is bimodal with local maxima at both 0 and 1. On the other hand, if \( \sigma^2 > 1 \), then \( h'(r) > 0 \) if and only if \( r < 1/2 \), \( h'(r) = 0 \) if and only if \( r = 1/2 \), and \( h'(r) < 0 \) if and only if \( r > 1/2 \). Therefore, \( h \) is first decreasing and then increasing in \( r \); that is, \( h \) is bimodal with local maxima at both 0 and 1. On the other hand, if \( \sigma^2 = 1 \), then \( h'(r) = 0 \) for all \( r \), so \( h \) is the uniform distribution on \([0,1]\). Q.E.D.
Proof of Proposition 2

Let \( r_1: \mathbb{R} \rightarrow [0,1] \) be the mapping from legislator ideal points into ratings produced by \( g_1 \), and let \( r_2: \mathbb{R} \rightarrow [0,1] \) be the mapping from legislator ideal points into ratings produced by \( g_2 \). Then \( r_1(x) = G_1(x) \) and \( r_2(x) = G_2(x) \), for all \( x \in \mathbb{R} \). The mean of \( r_1 \) is then \( E(r_1) = \int_{-\infty}^{\infty} G_1(x)f(x)dx \), where \( f \) is the distribution over ideal points. By assumption, \( G_1 \) is the cumulative distribution function of a \( N(0, \sigma_1^2) \) random variable, so \( G_1(-x) = 1 - G_1(x) \) for all \( x \in \mathbb{R} \). Also, \( f \) is the probability distribution function of a \( N(0,1) \) random variable, so \( f(-x) = f(x) \) for \( x \in \mathbb{R} \). Thus, \( E(r_1) = \int_{-\infty}^{\infty} G_1(x)f(x)dx + \int_{0}^{\infty} G_1(x)f(x)dx = \int_{0}^{\infty} [1 - G_1(x)]f(x)dx + \int_{0}^{\infty} G_1(x)f(x)dx = \int_{0}^{\infty} f(x)dx = \frac{1}{2} \). Similarly, \( E(r_2) = \frac{1}{2} \).

Since \( E(r_1) = \frac{1}{2} \), the variance of \( r_1 \) is \( \text{var}(r_1) = E(r_1^2) - E(r_1)^2 = \int_{-\infty}^{\infty} [G_1(x)]^2f(x)dx - \int_{0}^{\infty} [G_1(x)]^2f(x)dx - \int_{-\infty}^{0} [G_1(x)]^2f(x)dx = \int_{0}^{\infty} [1 + 2G_1(x)(G_1(x) - 1)]f(x)dx = \frac{1}{4} \). Similarly, \( \text{var}(r_2) = \frac{1}{4} \) if and only if \( \int_{0}^{\infty} [2G_2(x)(G_2(x) - 1)]f(x)dx > \int_{0}^{\infty} [2G_1(x)(G_1(x) - 1)]f(x)dx \). Since \( G_1 \) and \( G_2 \) are the cumulative distribution functions of normal random variables with variances \( \sigma_1^2 \) and \( \sigma_2^2 \), respectively, if \( \sigma_1^2 < \sigma_2^2 \) then \( \frac{1}{2} < G_1(x) < G_2(x) < 0 \) for all \( x > 0 \), and conversely. Thus, \( 2G_2(x)(G_2(x) - 1) < 2G_1(x)(G_1(x) - 1) \) for all \( x > 0 \), and therefore \( \text{var}(r_1) > \text{var}(r_2) \), if and only if \( \sigma_1^2 < \sigma_2^2 \). Q.E.D.

NOTES

This paper was completed while the author was a postdoctoral fellow in political economy at the Graduate School of Industrial Administration, Carnegie-Mellon University.

1. This cursory survey of the literature is hardly complete, but it makes the point.

2. Snyder (1991) presents evidence that voters' preferences in California state Assembly districts are not bimodal. Note, however, that gerrymandering might produce more extremist legislators than one would normally expect: if districts are drawn to be safe for incumbents and are therefore packed whenever possible with loyal Democrats or loyal Republicans, then many districts will tend to favor relatively extremist candidates.

3. Fowler (1982) found that interest groups tend to select roll calls that enable them to distinguish easily their friends from their enemies. Fowler raises other troublesome questions about the way groups select roll calls to include in their ratings; in particular, the selection criteria of some groups may not be consistent over time, and groups may sometimes tailor their ratings to appease certain members of Congress.

4. Some formal voting models, including the behavioral model underlying the scaling procedure used by Poole and Rosenthal (1985, 1991), assume there are random, idiosyncratic factors that affect voting, so that legislators do not always vote their preference. The intuition developed here clearly carries over to models with probabilistic voting.

5. Of course, the interest groups producing the ratings may not be particularly interested in accurately measuring ideal points but may have other goals in mind, as is discussed in Fowler 1982.

6. A referee pointed out that proposition 2 can also be viewed as a comparative static for an interest group contemplating the choice of roll calls to use in constructing its ratings. If the group is primarily interested in dividing legislators into supporters and opponents, then it should select roll calls with cleavage points concentrated around the center of the distribution of legislators' ideal points. On the other hand, if the group...
wishes to spread out the distribution of ratings, so that it can distinguish more easily among close supporters, moderate supporters, and so on, then it should use a set of roll calls with a less concentrated distribution of cleavage points.

7. In fact, cleavage points do not really exist, at least not as defined earlier in this paper—as the point that divides the legislators who vote aye from those who vote nay. This point exists in theory by the assumption of perfect, sincere voting. In actual roll calls, however, voting is almost never so perfect, so there are almost never perfect cleavage points.

8. The data were generously provided by Keith Poole. Putting the percentage of representatives voting aye along the x-axis, rather than the percentage of representatives voting with the majority, produces almost identical results.

9. Of course, representatives may still do a good job representing their districts. There may still be a strong correspondence between the representative's ideology and that of a majority of the constituents.

10. Note that this bias is not the same as the bias due to simple measurement error, in which \( \bar{x} \) is assumed to be equal to \( x \) plus some random error, although it is similar in effect.

11. Hall and Grofman (1990) argue that interest group ratings are inappropriate for the task of finding outlier committees because the roll calls included in most ratings (even those of rather narrowly focused interest groups, such as farm organizations) cover a wide range of issues. They suggest that, to better test whether a particular committee is a preference outlier, one should eliminate from the ratings those roll calls that are not relevant to the committee's jurisdiction. The results here are independent of this critique and would clearly apply to modified ratings as well as unmodified ratings.

12. That is, the interval \( [0, 2\bar{x}] \) is the largest possible support for distribution of ideal points.

13. The standard deviation is approximately \( \bar{x}/435 \left[ \sum_{i=1}^{435} (2i-435)^2/435 \right]^{1/2} \).

14. I thank the editor and an anonymous referee for pointing out the importance of this application.

15. The discussion here draws from Rohde 1991, chaps. 3 and 5, and deals with only one of the book's many arguments.

16. Rohde is well aware of the problem of the changing agenda and studies it by examining different issues and different types of roll calls (final passage, amendments, procedural votes) in some detail. He does not address the point made here, however. The problem I raise applies to restricted sets of roll calls as well as to the set of all roll calls.

17. Interestingly, exactly the opposite is true for interest group ratings: changes in the distribution of cleavage points tend to cause the average rating within each party to move in the same direction. Thus, when the average party ratings move in opposite directions, it is likely that preferences have changed.

18. Under more modest changes in the distribution of ideal points, the Democratic and Republican indices will generally move in opposite directions. For example, if 60% of the second year cleavage points are at .55 and 40% are at .65, then the Democratic and Republican indices of mean cohesion are .48 and .22, respectively.

REFERENCES


