B.J.Pol.S. (2005) **35,** 731–744 *Copyright © 2005 Cambridge University Press*
doi:10.1017/S0007123405000372 *Printed in the United Kingdom*

Correcting for Small Group Inflation of Roll-Call Cohesion Scores

SCOTT W. DESPOSATO*

Roll-call cohesion scores are the most widely used measures of voting blocs in legislative studies, appearing in literally hundreds of studies since their introduction in 1924. Despite a staple of legislative studies, we know virtually nothing about the statistical properties of these scores. In this article, it is shown how such scores suffer a serious bias problem: scores are artificially inflated for small parties, especially those that are less unified. The problem is demonstrated and an intuitive solution proposed. It is illustrated with data from the United States and from Brazil.

Roll-call cohesion scores are one of political scientists' most important tools. Since their introduction by Rice,[1] cohesion scores have been used in literally hundreds of studies to measure voting unity in many diverse legislative and judicial bodies, including the US Congress, New York State Assembly, Russian Duma, Argentinean Chamber of Deputies, European Parliament, United Nations, Brazilian Chamber of Deputies, the US Supreme Court, Norwegian Storting and Confederate Congress.[2] Many theoretical debates have been fought with cohesion scores; many central findings rest on their backs. Scores have

* Department of Political Science, University of California, San Diego. Many helpful comments were provided by Ben Bishin, Brian Crisp, Kris Kanthak, Gary King, Bill Mishler, Rabinda Bhattaharya, Tom Volgy and three anonymous reviewers. Fernando Limongi and Argelina Figuereido generously shared their roll-call data from the Brazilian Congress.

[1] Stewart A. Rice, 'Farmers and Workers in American Politics' (doctoral dissertation, Columbia University, 1924).

[2] A few examples include Stewart A. Rice, *Quantitative Methods in Politics* (New York: Alfred A. Knopf, 1928); V. O. Key Jr, *Southern Politics in State and Nation* (New York: Alfred A. Knopf, 1949); David Truman, 'The State Delegations and the Structure of Party Voting in the United States House of Representatives', *American Political Science Review*, 50 (1956), 1023–45; Charles Press, 'Presidential Coattails and Party Cohesion', *Midwest Journal of Political Science*, 7 (1963), 320–35; Allan Kornberg, 'Caucus and Cohesion in Canadian Parliamentary Parties', *American Political Science Review,* 60 (1966), 83–92; Hugh L. LeBlanc, 'Voting in State Senates: Party and Constituency Influences', *Midwest Journal of Political Science,* 13 (1969), 33–57; Helmut Norpoth, 'Explaining Party Cohesion in Congress: The Case of Shared Policy Attitudes', *American Political Science Review*, 70 (1976), 1156–71; Hugh W. Stephens and David W. Brady, 'The Parliamentary Parties and the Electoral Reforms of 1884–85 in Britain', *Legislative Studies Quarterly*, 1 (1976), 491–510; David B. Meltz, 'Legislative Party Cohesion: A Model of the Bargaining Process in State Legislatures', *Journal of Politics*, 35 (1973), 647–81; Keith E. Hamm, Robert Harmel and Robert J. Thompson, 'Impacts of Districting Change on Voting Cohesion and Representation', *Journal of Politics*, 43 (1981), 544–55; E. Lee Bernick, 'The Impact of U. S. Governors on Party Voting in One-Party Dominated Legislatures', *Legislative Studies Quarterly*, 3 (1978), 431–44; Gary W. Cox and Matthew D. McCubbins, 'On the Decline of Party Voting in Congress', *Legislative Studies Quarterly*, 16 (1991), 547–70; David M. Wood and Jack T. Pitzer, 'Parties, Coalitions, and Cleavages: A Comparison of Two Legislatures in Two French Republics', *Legislative Studies Quarterly*, 4 (1979), 197–226; Hans-Peter Hertig, 'Party Cohesion in the Swiss Parliament', *Legislative Studies Quarterly*, 3 (1978), 63–81; Barbara Deckard Sinclair, 'Determinants of Aggregate Party Cohesion in the U. S. House of Representatives, 1901–1956', *Legislative Studies Quarterly*, 2 (1977), 155–75; Moshe Haspel, Thomas F. Remington and Steven S. Smith, 'Electoral Institutions and Party Cohesion in the Russian Duma', *Journal of Politics*, 60 (1998), 417–39; and Scott Morgenstern, *Patterns of Legislative Politics: Roll Call Voting in Latin America and the United States* (New York: Cambridge University Press, 2004). This is a very small sample; there are many others.

been used to address questions about party government, international alignments, racial solidarity, electoral systems and judicial fairness. But, in spite of their importance, there has been little effort to check cohesion scores' unbiasedness or even their sufficiency for some theoretically-motivated parameter.[3]

In this article I show that the frequently-used cohesion score can be a severely biased measure of voting bloc unity. Starting from a simple model of legislative behaviour, I prove a surprising result: small parties' cohesion scores are artificially inflated, making them appear more cohesive than they are. Even when all legislators' behaviour is driven by the same underlying utility functions, smaller parties' expected cohesion scores are greater than those of larger parties. At the extreme, bias can be as much as 0.50 – while cohesion scores only vary from 0 to 1. The bias fades away as voting bloc size becomes large or as groups approach perfect voting unity. The result holds under diverse models of legislative behaviour and for all standard cohesion and agreement measures.

These results will not affect studies of large parties, but do create a quandary for research where there are small voting blocs. This is especially problematic when party size varies dramatically across groups being compared. Results based on such comparisons will be biased and findings may reflect properties of cohesion scores, not the political phenomenon being studied. In fact, the small party inflation of cohesion scores may explain away a common finding in the comparative literature: that small parties are more homogeneous than large parties.[4] This finding also affects studies of other small voting blocs, including committees, courts and small caucuses.

[3] Occasionally a scholar will remind us that cohesion has never been grounded theoretically, but none have attempted such a grounding. For example, Aage R. Clausen, 'The Measurement of Legislative Group Behavior', *Midwest Journal of Political Science*, 11 (1967), 212–24, notes that scores lack an underlying model of individual behaviour and Lee F. Anderson, Meredith W. Watts Jr and Allen R. Wilcox, *Legislative Roll-Call Analysis* (Evanston, Ill.: Northwestern University Press, 1966) caution against blind use of traditional statistical methods to analyse cohesion scores. But no scholar I am aware of has investigated Rice cohesion scores' statistical properties.

Interestingly enough, a number of scholars nearly stumbled upon the problem and solution. Arend Lijpardt, 'The Analysis of Bloc Voting in the General Assembly: A Critique and a Proposal', *American Political Science Review*, 57 (1963), 902–17, p. 908, reports that a number of unanimity-based scores inflate small bloc values, but then erroneously concludes that the Rice cohesion score is free of this defect. Steven J. Brams and Michael K. O'Leary, 'An Axiomatic Model of Voting Bodies', *American Political Science Review*, 64 (1970), 449–70, advocate a measure similar to mine, but with an adjustment that actually magnifies small party inflation – increasing bias. Other scholars propose and examine cohesion measures' properties, but consistently in descriptive terms, without an underlying model of legislative behaviour.

[4] For example, studying the Swiss Parliament, Prisca Lanfranchi and Ruth Luthi, 'Cohesion of Party Groups and Interparty Conflict in the Swiss Parliament: Roll Call Voting in the National Council', in Shawn Bowler, David M. Farrell and Richard S. Katz, eds, *Party Discipline and Parliamentary Government* (Columbus: Ohio State University Press, 1999), pp. 99–120, at p. 108, find that 'the Rice Index is in most cases higher for non-governmental party groups than for the three "bourgeois" governmental parties (Radicals, Christian Democrats, People's Party).' Coincidentally, the non-governmental parties are also smaller than most governmental parties. Rapio Raunio, 'The Challenge of Diversity: Party Cohesion in the European Parliament', in S. Bowler, D. M. Farrell and R. S. Katz, eds, *Party Discipline and Parliamentary Government* (Columbus: Ohio State University Press, 1999), pp. 189–207, p. 199, examines cohesion in the parliament of the European Union. He observed that 'The high cohesion of the EDG, the ER, EUL, and the Left Unity (LU) result from these groups' rather homogeneous character.' Again, these four parties also happen to be the smallest in the European Union's National Council. Scott Mainwaring and Anibal Perez Liñán, 'Party Discipline in the Brazilian Constitutional Congress', *Legislative Studies Quarterly*, 12 (1997), 453–83, study party cohesion in Brazil's Constituent Assembly, finding that the biggest parties are less disciplined than small parties. In the United States, Roxanne L. Gile and Charles E. Jones, 'Congressional Racial Solidarity: Exploring Congressional Black Caucus Voting

I suggest a simple and flexible solution: making large parties small. There are many alternative approaches, but they require making strong parametric assumptions. The non-parametric approach I suggest will eliminate bias for all Rice-like cohesion scores across diverse theoretical models of legislative voting. The solution can be applied to any statistical use of cohesion scores. Finally, it shares one of cohesion scores' most important characteristics: simplicity and intuitiveness. The article is arranged in three sections. Below, I present and explain the bias problem. In the next section, I discuss the solution and apply it to examples from recent scholarship. A conclusion follows.

THE PROBLEM

I show in this section that the expected value of cohesion scores varies systematically with party size. Specifically, even when all legislators have identical underlying utility functions, expected cohesion is greater for small parties than for large parties.

What model of individual behaviour lies behind cohesion scores? None – there is no set of underlying parameters or a behavioural model from which cohesion scores are derived. Rice, the first political scientist to use them, considers the index of cohesion purely in descriptive terms.[5] And while scholars occasionally remind us that cohesion scores are effectively theoryless, no model has been suggested to fill the gap. This is problematic for two reasons. First, without a model of legislative behaviour, we effectively know nothing about the properties of the cohesion scores and cannot use them for hypothesis testing. Secondly, the lack of a model means that any solution should be very flexible and work for a broad class of models, not just one narrow understanding of legislative politics.

Because no specific model has previously been offered as a data-generating function for cohesion scores, I show that cohesion scores are inflated for small parties under a wide variety of theories of legislative voting. I first show the inflation in the simplest case – binomial voting – then generalize to other models. Under binomial voting, all legislators have independent and identical probabilities, p, of voting 'yes' on a bill, and $1 - p$ of voting 'no'. Since all legislators in this model have identical behavioural functions, we do not expect to see any differences in cohesion across parties. But as we will see, under this simple model, expected cohesion scores are greater for small parties than for large parties. The reason is that majority votes against the party's expected modal position push up expected cohesion scores, and such votes are much more likely in small parties than in large parties.

Throughout this article, I demonstrate the problem and solution using the Rice cohesion score. There are many ways of calculating cohesion, for example, the Index of Absolute Cohesion, the Index of Relative Cohesion and the Index of Agreement.[6] The bias problem described herein applies to *all* these cohesion scores, agreement scores and others not

(F'note continued)

Cohesion, 1971–1990', *Journal of Black Studies*, 25 (1995), 622–41, compare cohesion of the Congressional Black Caucus (CBC) with that of Northern and Southern Democrats. They find that the CBC is the most cohesive of the three. It is also easily the smallest. In general, scholars attribute these findings to the 'homogeneity' of small parties, but have not offered any theoretical explanation for small parties' consistently high cohesion.

[5] He describes the interpretation of cohesion scores as follows: 'Are the Republican members in a state senate more alike in their votes than are the members of the senate generally? If so, it may be inferred that they are more like-minded and the Republican group may be called more cohesive than the senate as a whole' (Rice, *Quantitative Methods in Political Science*, p. 208).

[6] Anderson *et al.*, *Legislative Roll-Call Analysis.*

listed. I base my examples on the Rice measure, however, because it is the most frequently used of group cohesion scores. As proposed by Rice, cohesion on a single roll-call vote is calculated as:

$$Rice = \frac{|Yes - No|}{Yes + No}, \tag{1}$$

where *Yes* is the number of favourable votes and *No* is the number of opposing votes cast.[7] Incorporating our simple random-utility model of legislative behaviour, we can write the expected Rice cohesion score on a bill as:

$$E(Rice) = \sum_{k=0}^{n} Rice(k)*P(k), \tag{2}$$

where k is the number of 'yes' votes and n is the number of members in the group for which we are calculating cohesion. Substituting k for *Yes*, and $n - k$ for *No*, and taking advantage of the fact that k is distributed according to a binomial distribution, we have:

$$E(Rice) = \sum_{k=0}^{n} \frac{|k - (n-k)|}{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}. \tag{3}$$

Assuming n is even,[8] we can rewrite this, eliminating the absolute value signs by adding back in the values that would be negative but for the absolute value sign:

$$E(Rice) = \sum_{k=0}^{n} \frac{2k-n}{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} + 2 \sum_{k=0}^{\frac{n}{2}} \frac{n-2k}{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}. \tag{4}$$

The first part of this sum reduces to

$$2p - 1^9. \tag{5}$$

The second part of the sum, however, is the problem. For $p > 0.5$, this can be intuitively thought of as the impact on expected cohesion of a majority voting the 'wrong' way. Voting the 'wrong' way means a majority of a group voting against the expected modal vote. For example, if the probability of voting 'yes' is 0.7, the modal expected vote is 'yes', and the second term adjusts for the impact of a majority voting 'no'. For large or disciplined parties, the probability of this happening is very small and the term has no significant impact on the cohesion score. But for small parties, voting the 'wrong' way is common and pushes up cohesion scores.

Table 1 illustrates the bias, comparing expected cohesion for groups of size 10 and size 3, where all legislators have independent and equal probabilities 0.6 of voting 'yes'. The table lists all possible vote outcomes, their probabilities and corresponding Rice cohesion scores. The fifth column lists the product of each vote outcome's probability and cohesion score. The sum of column five is thus the expected Rice cohesion score for the group.

With a group of size 10, the expected Rice cohesion score is 0.294, close to its asymptotic value ($2p - 1 = 2*0.6 - 1 = 0.3$). But with a smaller group of just three members, the expected Rice cohesion score is 0.456, more than 50 per cent greater than its asymptotic

[7] Reported scores usually represent the average cohesion score for a party across all roll-call votes cast during a year or legislative term. See Anderson *et al.*, *Legislative Roll-Call Analysis*, for a discussion of methods for averaging across multiple roll-call votes.

[8] Generalizing to n odd just requires changing the summation range's maximum value for the second term in (4) from $k = n/2$ to $k = (n-1)/2$.

TABLE 1 *Expected Cohesion Scores*

Yes	No	*P*(vote)	Cohesion	$P \times C$
Group Size = 10, *P*(yes) = 0.6				
10	0	0.006	1.0	0.006
9	1	0.040	0.8	0.032
8	2	0.121	0.6	0.073
7	3	0.215	0.4	0.086
6	4	0.251	0.2	0.05
5	5	0.201	0	0
4	6	0.111	0.2	0.022
3	7	0.042	0.4	0.017
2	8	0.011	0.6	0.006
1	9	0.002	0.8	0.001
0	10	0.000	1.0	0
P(no > yes) = 0.166 *E*(Rice) = 0.294				
Group Size = 3, *P*(yes) = 0.6				
3	0	0.216	1.000	0.216
2	1	0.432	0.333	0.144
1	2	0.288	0.333	0.096
0	3	0.064	1.000	0.064
P(no > yes) = 0.352 *E*(Rice) = 0.456				

value. The mechanism, again, is that (a) when a majority votes the 'wrong' way, cohesion is pushed upward and (b) such 'wrong' votes are more likely in small parties where there is more variance in vote outcomes. In this illustration, *P*(yes) = 0.6, so wrong votes are those where a majority voted 'no'. The probability of a majority voting the wrong way is reported at the bottom of the table. For a group of size 10, the probability of a majority voting 'no' is 0.166, or about one in six votes. For a group of size 3, the probability is much larger, 0.352, or more than one in three votes.

Figure 1 shows how expected cohesion varies with *p(yes)* and party size. The *Y* axis in the graph is the expected Rice cohesion score and the *X* axis is party size. Each of the lines represents the expected Rice cohesion score for a different value of *p* as labelled: 0.5, 0.6, 0.7, 0.8 and 0.9. For any value of *p*, increasing party size lowers the expected cohesion score. As *n* becomes large, the score converges to $2p - 1$.

The inflation is small and convergence fast when *p* is close to 1 or 0. Convergence is slowest and bias largest as *p* approaches 0.5. Intuitively, as *p* goes toward 1 or 0, the variance of the proportion voting 'yes' falls, so 'wrong' votes are less and less likely. At the other extreme, for $p = 0.5$, anything but a perfect split in the party's vote will raise the cohesion score above its asymptotic value (0), hence its convergence is especially slow.

This basic result holds under many different models of legislative behaviour, as explained below, and has broad implications for the study of legislative politics. In a context where group size varies substantially and includes some small groups, results could be biased by the artificial inflation of small party cohesion. Further, it applies not just to the Rice cohesion score, but can be generalized to all measures of cohesion that have a similar form, taking the absolute value or squaring the difference between the number of 'yes' votes, 'no' votes, abstentions or absences. In any such case, the extreme vote

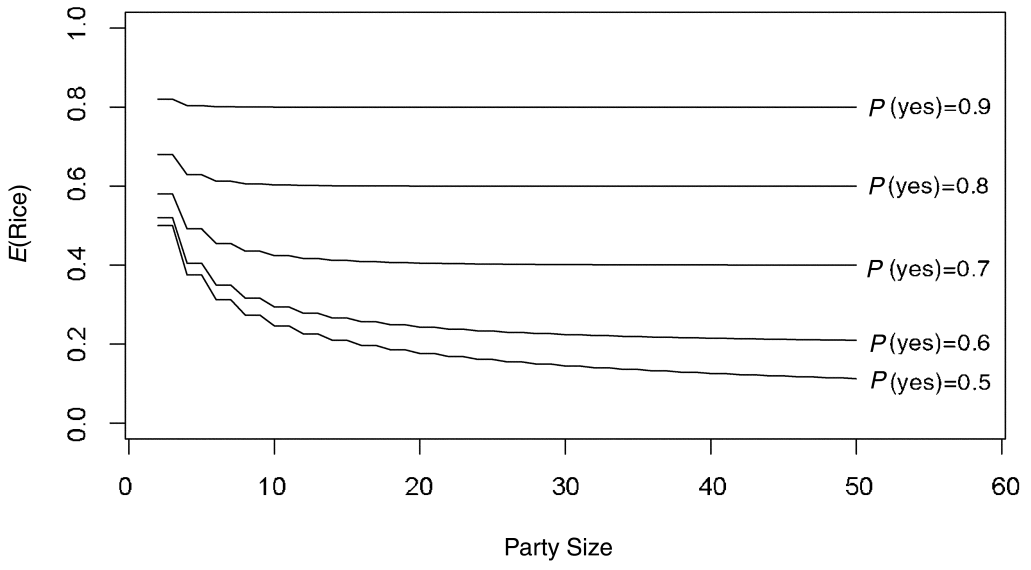

Fig. 1. Bias in cohesion scores by party size and P*(Yes)*

outcomes in the distribution of votes push the cohesion score up, and are more likely to have an impact in small parties. Consequently, all such measures suffer the same bias factor, though its precise form and magnitude will vary from one measure to another.

Practically, we expect the bias problem to be greatest when we study less disciplined and small voting blocks. Analyses of cohesion on committees, of inchoate party systems, or of weak multiparty systems are three examples where the impact will be greatest. Unaffected will be findings relying on analysis of large or very unified parties.

A SOLUTION

As with most problems, there are multiple solutions. One is simply to abandon cohesion scores and go to a spatial model, perhaps focusing on the variance of party members' ideal points. Recent scholarship includes many examples that build on spatial models of legislative behaviour to study aspects of party cohesion and discipline.[10]

[9] The algebra is in Appendix A.

[10] Keith T. Poole and Howard Rosenthal, *Congress: A Political-Economic History of Roll Call Voting* (New York: Oxford University Press, 1997); Gary W. Cox and Keith T. Poole, 'Measuring Partisanship in Roll-Call Voting: The U.S. House of Representatives, 1877–1999', *American Journal of Political Science*, 46 (2002), 477–89; James Snyder and Timothy Groseclose, 'Estimating Party Influence on Roll-Call Voting', *American Journal of Political Science,* 44 (2000), 193–211; Nolan McCarty, Keith Poole and Howard Rosenthal, 'The Hunt for Party Discipline in Congress', *American Political Science Review*, 95 (2001), 673–88; John Londregan, 'Appointment, Reelection and Autonomy in the Senate of Chile', in Scott Morgenstern and Benito Nacif, eds, *Legislative Politics in Latin America* (New York: Cambridge University Press, 2002), pp. 341–76; Simon Jackman, 'Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, and Model Checking', *Political Analysis*, 9 (2000), 227–41; Jeff Lewis and Keith Poole, 'Measuring Bias and Uncertainty in Ideal Point Estimates via the Parametric Bootstrap', *Political Analysis*, 12 (2004), 105–27.

However, several reasons might prompt a researcher to stick with cohesion scores. First, cohesion scores are simple and intuitive, and do not require advanced methods or programming to implement – important considerations for applied scholars. Secondly, ideal point estimates rest on strong assumptions about legislative behaviour. Cohesion scores provide an alternative approach when reality or data do not support such assumptions. Thirdly, even cutting-edge spatial methods continue to draw on cohesion scores. For example, Cox and Poole use a sophisticated estimation of ideal points to test for party effects using Rice's index of party agreement. They avoid the problems explored in this article because their group sizes (Democrats and Republicans) are large. This makes the probability of a majority 'wrong' vote practically zero, so they discard that possibility.[11] However, anyone trying to implement their method in a context with small parties will encounter the same problems described in this article. Finally, some recent work on legislative voting is explicitly non-spatial.[12] Testing the implications of such models will require non-spatial methods, like cohesion scores.

The solution I propose preserves the positive attributes of the cohesion score: it is simple, easy and avoids making strong assumptions. The inflation problem is eliminated by making large parties small. Specifically, one replaces each party's cohesion score (C_{ij}) with $E(C_{ijr})$. $E(C_{ijr})$ is the expected cohesion score of a sample of r votes taken without replacement from party i on bill j. Given Y 'yes' votes and $R - Y$ 'no' votes in party i on bill j, the probability of drawing k 'yes' and $r - k$ 'no' votes in a sample without replacement of size r is:

$$P(k \mid Y, R, r) = \frac{\binom{Y}{k}\binom{R-Y}{r-k}}{\binom{R}{r}}. \tag{6}$$

Thus the expected value of a cohesion score, when r votes are sampled from party i is:

$$E(C_{ijr} \mid Y, R, r) = \sum_{k=0}^{r} \frac{\binom{Y}{k}\binom{R-Y}{r-k}}{\binom{R}{r}} \frac{|2k - r|}{r}. \tag{7}$$

A proof that this expected value equals the expected cohesion score for a party of size r is in Appendix B.

By replacing large parties' cohesion scores with the expected value of the cohesion score of a sample of legislators of size r taken without replacement, parties of all sizes can be compared on the same metric. After making this adjustment, scholars can calculate any statistic of interest without bias. Intuitively, the adjusted cohesion scores reflect groups that are all the same size, so differences between them cannot be the result of the party size bias problem.

What is the appropriate sample size? r can be set to any value between 2 and the size of the smallest party voting on bill j, but the most general approach for interpretation and comparison across published results would be to set $r = 2$; $r = 2$ will work for every case, and has a intuitive interpretation: $E(C_{ij2})$ is the probability that two randomly-selected

[11] Cox and Poole, 'Measuring Partisanship in Roll-Call Voting:', p. 479.

[12] Brian Crisp, Kris Kanthak and Jennie Leijonhufvud, 'A Non-Spatial Understanding of Position-Taking and Vote Seeking: Cosponsorship as "Currency" for Purchasing Electoral Support', *American Political Science Review*, 98 (2004), 703–16.

members of party i voted together on bill j. It is also extremely simple to calculate. For $r = 2$, (2) reduces to:

$$E(C_{ij2} \mid Y, R) = \frac{Y(Y-1) + (R-Y)(R-Y-1)}{R(R-1)}, \tag{8}$$

which can easily be calculated with a single command in any statistical package. Further, given party size, a researcher could easily 'back out' $E(C_{ij2})$ from published cohesion scores for larger parties. Some simple algebra reveals that, given a party of size R and a reported cohesion score of C,

$$E(C_2 \mid Y, R) = \frac{RC^2 + R - 2}{2(R-1)}. \tag{9}$$

The primary advantage of this solution is that it preserves the flexibility of the cohesion score, working under multiple models of legislative behaviour. As demonstrated in Appendix B, it corrects the inflation problem under binomial voting. But more importantly, this solution is not limited to simple binomial models of legislative voting – it will correct bias when legislators' behaviour is driven by any of a wide variety of theoretical models.

Figure 2 shows how small party inflation persists across multiple theoretical models of legislative voting, and how my solution makes cohesion scores comparable in each case. The figure displays the results of 1,000 simulations of voting for binomial, beta-binomial, and spatial models of legislative behaviour. I conducted a similar analysis using actual roll-call votes from the 96th House of Representatives. Each graph shows results for a different model, as labelled. The X-axis shows simulated group size and the Y-axis the cohesion score. The solid line shows the mean cohesion score; the dotted line shows the mean corrected cohesion score for the simulations.[13]

Each figure tells the same basic story – confirming the robustness of the bias and of my proposed solution. The bias in cohesion scores persists across binomial, beta-binomial, spatial and actual votes from the House of Representatives. In each case, mean simulated cohesion is highest for small groups and falls as group size increases, although the underlying distribution for individual behaviour remains unchanged. At the same time, the solution proposed herein levels the playing field. The dotted lines are flat in all figures, showing how the solution eliminates any relationship between cohesion and group size.

An additional advantage of this solution is that it will correct not just Rice cohesion scores, but any cohesion or agreement score based on the absolute value or square of the difference in types of votes cast. One example is the unity score which counts abstentions as 'no' votes.[14] Another is the agreement score, reporting legislators' agreement with their party majority. In both cases, scores will be inflated for small parties and corrected by my solution.

The primary disadvantage of this solution is that the adjusted data are all inflated upward. Applying my solution to cohesion scores places all parties on an equal metric and allows for unbiased comparisons, but that field is inflated above cohesion scores' asymptotic (as party size goes to infinity) limit. This disadvantage does not limit hypothesis testing or analysis, it just requires that scholars use caution when comparing adjusted and unadjusted scores.[15]

[13] Additional details on the simulations are provided in the Appendices.

[14] See, for example, John M. Carey, 'Political Institutions, Competing Principals, and Party Unity in Legislative Voting' (unpublished paper, Dartmouth College, 2001).

[15] How does this solution affect the variance of a cohesion score? In simulations, it had no effect: $Var(C_{ij})$ was equal to $Var(C_{ijr})$.

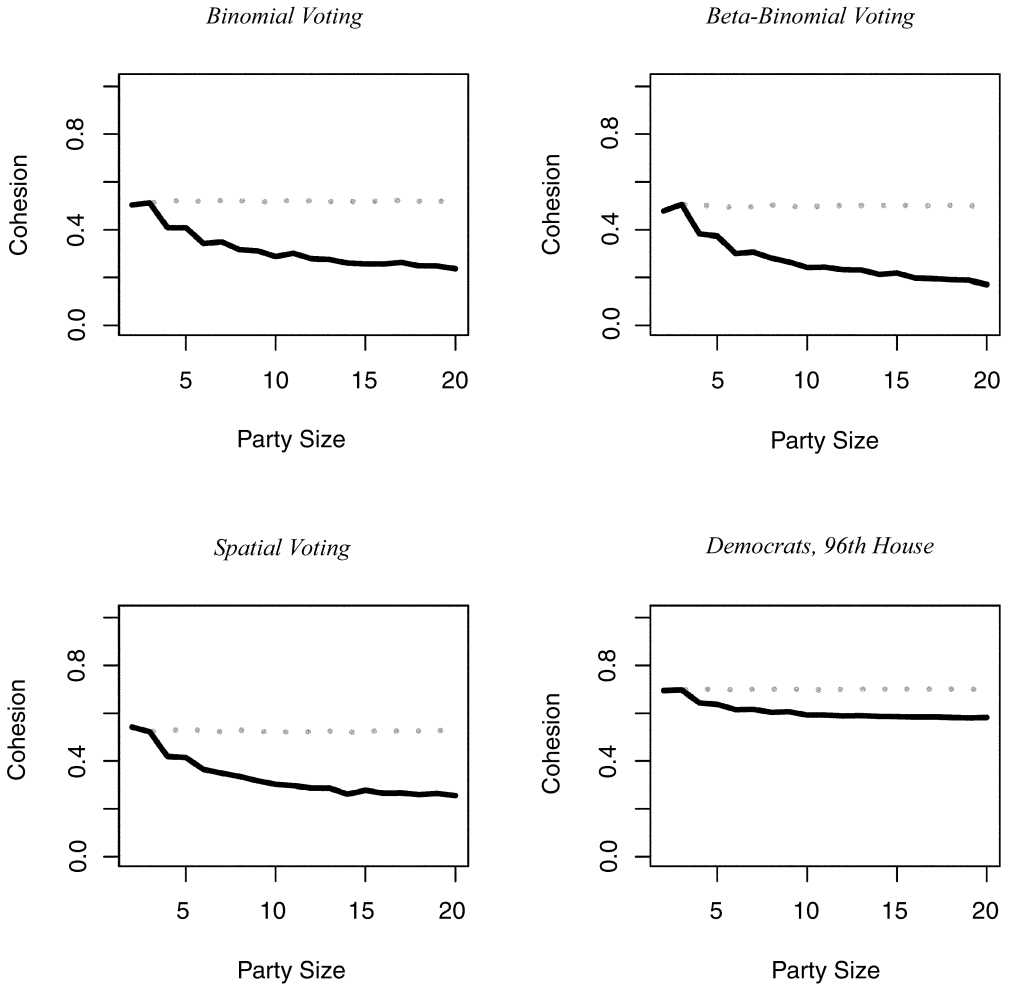

Fig. 2. Simulated cohesion and adjusted cohesion by group size under four different models.
Note: Solid lines show mean simulated cohesion; dotted lines show mean adjusted cohesion.

An alternative approach would involve estimating and subtracting the inflation factor – the second term in (4). Such an approach, however, would not be easily generalizable. A different adjustment factor would be required for each kind of cohesion measure and explicit assumptions about legislative behaviour would be required.[16] The solution proposed herein, in contrast, is completely portable across cohesion measures and works under diverse models of legislative behaviour.

[16] I explored several methods for subtracting an inflation factor. All but one approach failed to eliminate the relationship between party size and cohesion score. The approach that worked required assuming that p is constant across all roll-call votes – i.e., on all votes, legislators face exactly the same net pressures and positions. This assumption would be completely unrealistic for all legislatures I am familiar with.

Example 1: Are Black MCs More Cohesive than Democratic MCs?

I illustrate the solution with two examples. The first re-examines a finding from the American politics literature: black Democratic Members of Congress (MCs) vote more cohesively than other Democratic MCs. Gile and Jones compare the cohesion of black Democrats with that of Northern and Southern Democrats in the House of Representatives from the 92nd until the 102nd legislature. They find that black Democrats are consistently the most cohesive of the three groups. They conclude that, 'the African American members of Congress have sustained a high degree of racial solidarity'.[17] Black Democrats, however, are also the smallest of the three groups. From the 92nd to the 102nd House, blacks only held between fourteen and twenty-seven seats while Democrats as a whole averaged over 200 seats. Could the observed difference simply be a function of party size?

I replicated Gile and Jones's work by applying my solution to Black Democrats, Northern Democrats and Southern Democrats' voting during the same period they studied. The central hypotheses can be stated as follows:

H_0: $\theta_{BD-ND} = 0$ and $\theta_{BD-SD} = 0$
H_{A1}: $\theta_{BD-ND} > 0$
H_{A2}: $\theta_{BD-SD} > 0$

where θ_{BD-ND} is the mean difference between black Democrats' cohesion and Northern Democrats' cohesion, and θ_{BD-SD} is the mean difference between black Democrats' cohesion and Southern Democrats' cohesion. To test these hypotheses, I used adjusted cohesion scores.

TABLE 2 *Test: Black vs. Non-Black Democratic Cohesion*

	Mean	SE	95% CI
Rice Score			
$\theta_{\text{Black–N. Dems}}$	0.13	0.01	0.11–0.15
$\theta_{\text{Black–S. Dems}}$	0.20	0.01	0.18–0.22
Adjusted Score			
$\theta_{\text{Black–N. Dems}}$	0.11	0.01	0.09–0.13
$\theta_{\text{Black–S. Dems}}$	0.15	0.01	0.13–0.17

Table 2 compares hypotheses tests using the original cohesion scores and adjusted data. Using the original Rice cohesion scores, the mean difference between black Democrats and Northern Democrats is 0.13, and is easily significant at the 0.01 level. The difference is even larger when comparing black Democrats with Southern Democrats: 0.20, again easily significant at the 0.01 level. These differences shrink once corrected for small party bias. The difference between blacks and Northerners falls by over 20 per cent to 0.11 and that between blacks and Southerners falls similarly to 0.15 when the small group bias is

[17] Gile and Jones, 'Congressional Racial Solidarity', p. 638.

[18] Scott Mainwaring, 'Multipartism, Robust Federalism, and Presidentialism in Brazil', in Scott Mainwaring and Matthew S. Shugart, eds, *Presidentialism and Democracy in Latin America* (New York: Cambridge University Press, 1997), pp. 55–109; Barry Ames, *The Deadlock of Democracy in Brazil* (Ann Arbor: The University of Michigan Press, 2001); Octavio Amorim Neto and Fabiano Santos, 'The Executive Connection: Presidentially Defined Factions and Party Discipline in Brazil', *Party Politics*, 7 (2001), 213–34.

eliminated. Note in particular that the corrected cohesion is *outside* the confidence interval for the original cohesion score. But in both cases, the adjustment does *not* change the significance of the differences or the substantive conclusion: black Democrats really are significantly more cohesive than either regional group of Democrats.

Example: Are Small Brazilian Parties more Cohesive than Large Parties?

A second example comes from the literature on Brazilian politics. While that country's party system as a whole has been characterized as weak and inchoate by many scholars, observers have noted substantial variance across parties within the system.[18] Within this variance, scholars have observed that small political parties are significantly more cohesive than large 'catch all' parties. Such findings are often attributed to 'ideological homogeneity'.

I evaluated this finding by examining party cohesion in the Brazilian Chamber of Deputies during the 48th, 49th and 50th legislatures (1987–91, 1991–95, and 1995–99, respectively), using a simple model, regressing party cohesion on party size and including a dummy variable for each legislature:

$$\text{Rice} = \alpha + \beta\text{Size} + \text{I}_{49\text{th}} + \text{I}_{50\text{th}}.$$

The central hypotheses can be stated as follows:

H_0: $\beta = 0$
H_A: $\beta < 0$

Table 3 reports the results of regressing cohesion on party size and a set of indicator variables for legislative period, for both the original Rice cohesion scores and the adjusted scores.[19] Using the original cohesion scores, party size has a negative and significant impact on party cohesion. Every ten members lower cohesion, on average, by 0.014. At the extremes, the smallest parties (with just two members) should have Rice scores about 0.13 greater than the largest parties (with a hundred deputies). The corrected scores, however, reverse this finding. The magnitude of the coefficient on party size halves and is no longer significant. In other words, there is no evidence that small parties in Brazil are any more

TABLE 3 *Regression of Party Cohesion on Party Size*

	Rice Score		Adjusted Score	
	Coef.	SE	Coef.	SE
Size	− 0.0014	0.0007*	− 0.0006	0.0006
49th	− 0.02	0.04	− 0.02	0.03
50th	− 0.03	0.04	− 0.04	0.03
Constant	0.85	0.03*	0.87	0.03*
n	54		54	
R^2	0.09		0.05	

*Significant at $p < 0.10$.

[19] I excluded legislators that were not in a party, and a small party that only voted on two bills.

cohesive than large parties. The observed significant impact of party size on cohesion is apparently a statistical artefact produced by using the Rice score – not a significant indicator of small party homogeneity.

CONCLUSION

Cohesion scores are perhaps the most widely-used tool of legislative scholars. First applied to roll-call votes in the 1920s, scholars today still use cohesion scores to study the strength of legislative parties. Their advantages include being a widely-understood standard, intuitive and simple, and easy to calculate or program.

In this article I have shown that roll-call cohesion scores are inversely related to party size – all else equal, small parties will tend to appear more cohesive than large parties, even when they are not. I proposed a solution: equalizing party size, by making large parties small. This approach is simple, extremely flexible and eliminates the party-size bias problem. It can be applied to any use of cohesion scores, from simple difference tests to regression models, and will work under many different models and assumptions about legislators' behaviour. The problem and a solution were illustrated using data from the US House of Representatives and the Brazilian Chamber of Deputies. In the first case, correcting cohesion scores significantly reduced the difference between black Democrats and Southern Democrats as a whole, but did not eliminate the dramatic differences between them. In the second case, my correction did lead to different conclusions: small Brazilian parties are not significantly more cohesive than large Brazilian parties.

These findings will not affect work on large or highly disciplined parties, but will be a problem for the many studies that include small voting blocs. For example, many multiparty systems include several very small parties. Cross-party comparisons in such contexts will be biased by small voting bloc inflation. In fact, the phenomenon illustrated here might explain away a common finding in comparative legislative studies: that small parties are more cohesive than large parties. This has previously been attributed to 'ideological homogeneity', without any theoretical explanation, but might simply reflect bias in cohesion scores. Other possible applications include committee and judicial studies, where voting blocs are relatively small and results at risk of suffering the bias described here. In these and similar contexts, scholars should use caution to assure that their results are not simply artefacts of their methods.

APPENDIX A. ADDITIONAL DETAILS OF PROOF OF BIAS

This appendix shows that the first term in (4) reduces to $2p - 1$. The first term in (4) is:

$$\sum_{k=0}^{N} \frac{2k-n}{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \tag{8}$$

and can be reduced as follows:

$$\sum_{k=0}^{n} \frac{2k}{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} - \sum_{k=0}^{n} \frac{n}{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \tag{9}$$

$$\frac{2}{n} \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} - \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}. \tag{10}$$

The first term is just $2/n$ times the expected value of a binomial distribution (np), so it is $2np/n$, or $2p$. The second term is the sum of binomial probabilities times corresponding values of k, for all values of k. By definition, this sum is 1. Hence we arrive at $2p - 1$.

APPENDIX B. PROOF OF SOLUTION

This appendix demonstrates that the expected cohesion score of a sample of size r, taken with replacement from a party of size R, is equal to the expected cohesion score of a party of size r. The logic is simple. A sample of size r is equivalent to taking the first r legislators of a larger party. Their votes are independent Bernoulli random variables, so together they are equivalent to a party of size r under binomial voting.

A formal proof is easy. As previously, I illustrate using a simple binomial model of voting where each legislator has probability p of voting 'yes' and $1-p$ of voting 'no'. The voting in a party of size R follows a binomial distribution with parameters R and p and random variable Y representing the number of 'yes' votes. A sample of size r without replacement from that party has the following probability of drawing k 'yes' votes:

$$\frac{\binom{Y}{k}\binom{R-Y}{r-k}}{\binom{R}{r}}. \tag{11}$$

So the expected cohesion score of a sample of size r from a party of size R with probability p of voting 'yes', is:

$$E(C|p,R,r)=\sum_{Y=0}^{R}\binom{R}{Y}p^{Y}(1-p)^{R-Y}\sum_{k=0}^{r}\frac{\binom{Y}{k}\binom{R-Y}{r-k}}{\binom{R}{r}}\frac{|2k-r|}{r}. \tag{12}$$

Writing out the binomial coefficients, cancelling terms, and re-arranging slightly gives

$$\sum_{Y=0}^{R}\sum_{k=0}^{r}\frac{r!(R-r)!}{k!(Y-k)!(r-k)!(R-Y-(r-k))!}p^{Y}(1-p)^{R-Y}\frac{|2k-r|}{r}, \tag{13}$$

which can be rewritten as:

$$\sum_{k=0}^{r}\binom{r}{k}p^{k}(1-p)^{r-k}\frac{|2k-r|}{r}\sum_{Y=0}^{R}\binom{R-r}{Y-k}p^{Y-k}(1-p)^{R-Y-r+k}. \tag{14}$$

The second half of (14) is simply the sum of binomial probabilities over all possible values, or 1. Thus we can reduce this to:

$$\sum_{k=0}^{r}\binom{r}{k}p^{k}(1-p)^{r-k}\frac{|2k-r|}{r}, \tag{15}$$

which is identical to (4), proving the result. The simulations presented in the text show how the correction also works under other models of voting behaviour.

APPENDIX C. DATA AND CALCULATIONS

For the US House of Representatives, I used two data sources. Roll-call vote data for the US House of Representatives came from ICPSR study number 0004. Legislators' racial categories were coded as 'black' or 'non-black' following Amer.[20] Average cohesion scores per legislature were calculated for black Democrats, Northern non-black Democrats, and Southern non-black Democrats. For each group, average scores were calculated for each legislative session. These reported cohesion scores are the average of cohesion scores on each bill, weighted by the divisiveness of the bill. For a divisiveness weight, I used $1-C_{TOT}$, where C_{TOT} is the overall cohesion on the bill. For Brazil, roll-call vote data for the Brazilian Chamber of Deputies was generously provided by Figueiredo and Limongi.[21] For each party, I averaged cohesion scores across all bills in a session. These averages were weighted by how close the final vote was to passing.

[20] Mildred L. Amer, *Black Members of the United States Congress: 1789–2001* (Washington, D.C.: Congressional Research Service, The Library of Congress, 2001).

[21] Argelina Figueiredo and Fernando Limongi, 'Votações nominais na Câmara dos Deputados – 1989–1999', in dataset from Banco de Dados Legislativos Cebrap, 2000.

Simulations

In all cases, 1,000 simulations were run for each group size using R software (code is on my website). Legislators cast random votes following the distributions described below. From each vote, group cohesion and corrected cohesion were calculated. The average group cohesion and average corrected group cohesion were calculated across all 1,000 simulations, and are reported in Figure 2.

For binomial voting, all legislators cast random votes with a constant and independent probability $p = 0.6$ of voting 'yes'. Once simulated votes were cast, cohesion and corrected cohesion were calculated.

For the beta-binomial model, legislators cast random votes based on draws from a beta-binomial distribution. I first drew a different probability p_i for each legislator in the group or party – so each legislator has a different probability of voting 'yes'. These probabilities were drawn from a beta distribution with parameters $\alpha = 1$ and $\beta = 1$.[22] I Each legislator then cast a random vote with probability p_i of voting 'yes' and probability $1 - p_i$ of voting 'no'. A cohesion score and a corrected cohesion score were both calculated from these votes.

For the spatial voting simulation, legislators cast votes based on their own ideal points (θ_i), a cut point for each bill being considered (c_j), and a random error (ε_{ij}). For each simulation, ideal points, bill cut points and random errors were randomly drawn as follows. Legislators' ideal points θ were drawn from a uniform (0, 1) distribution. Cut points for each bill were also drawn from a uniform (0, 1) distribution. Random errors were drawn from a normal (0, 1) distribution.

Each legislator cast a 'yes' vote if her ideal point plus random error was greater than the simulated cutpoint, and a 'nay' if her ideal point plus random error was smaller than the simulated cutpoint. So legislator i voted 'yes' if $\theta_i + \varepsilon_{ij} > c_j$, and 'no' if $\theta_i + \varepsilon_{ij} < c_j$. From these simulated votes, cohesion scores and corrected scores were calculated and averaged over the 1,000 simulations run for each group size.

Finally, for the empirical simulation, random samples were drawn of Democrats from the 96th House. For example, to simulate group size of two, I randomly selected two Democrats and calculated their average cohesion and corrected cohesion over all votes (where both voted) in the 96th House. I repeated this 1,000 times and reported mean cohesion and mean corrected cohesion for each group size.

[22] A beta with parameters (1,1) is effectively a uniform distribution. I experimented with other parameters. Results were unchanged for any symmetric distribution ($\alpha = \beta$). Cohesion scores did increase for any skewed distribution, but the small party inflation persisted.