B.J.Pol.S. **34,** 137–192 *Copyright © 2004 Cambridge University Press*
DOI: 10.1017/S0007123403000395 *Printed in the United Kingdom*

Notes and Comments

Democracy and Gender Inequality in Education: A Cross-National Examination

DAVID S. BROWN*

This Research Note examines the role democracy plays in explaining the disparity in educational attainment between men and women in a cross-national context. Policies designed to improve education figure prominently in recent attempts by governments, international institutions and non-governmental organizations (NGOs) to stimulate economic development. The recent emphasis on education is grounded in a well-established literature in economics on the rate of return on investment in education[1] and in a fairly new stream of research in economics – endogenous growth theory – that suggests knowledge is the generative force of economic growth.[2] Although a majority of the work in both literatures focuses on aggregate levels of educational attainment (usually enrolment ratios or literacy rates), a growing body of evidence identifies women's educational attainment relative to men's as a crucial variable in explaining the wide variation in economic development throughout the world.[3] In addition to its direct impact on economic growth, women's education may indirectly affect economic performance in a number of important ways: through its impact on health, fertility and infant mortality. Previous empirical work shows that women's education has a strong negative effect on fertility and infant mortality.[4] Moreover, family health practices improve in direct proportion to female education. According to the World Bank, countries that achieved universal primary education for boys in 1965 but lagged far behind in educating girls had about twice the infant mortality and fertility rates in 1985 of countries with smaller gender gaps.[5]

* Department of Political Science, University of Colorado, Boulder. The author thanks Margaret Levi, Ilgu Ozler and Randy Stevenson, as well as the Editor and anonymous referees of the *Journal*, for their advice and helpful comments.

[1] Moses Abramowitz, 'Catching Up, Forging Ahead, and Falling Behind', *Journal of Economic History*, 46 (1986), 385–406; Robert J. Barro, 'Economic Growth in a Cross Section of Countries', *Quarterly Journal of Economics*, 106 (1991), 407–44; Gary S. Becker, 'Investment in Human Capital: A Theoretical Analysis', *Journal of Political Economy*, 70 (1962), 9–49; Edward F. Dennison, *The Sources of Economic Growth in the United States* (New York: Committee for Economic Development, 1962); Theodore W. Schultz, 'Investment in Human Capital', *American Economic Review*, 61 (1961), 1–17.

[2] Robert J. Barro, 'Economic Growth in a Cross Section of Countries', *Quarterly Journal of Economics*, 106 (1961), 407–44; Gary S. Becker, K. M. Murphy *et al*., 'Human Capital, Fertility, and Economic Growth', *Journal of Political Economy*, 98 (1990), 12–37; Paul M. Romer, 'Dynamic Competitive Equilibria with Externalities, Increasing Returns and Unbounded Growth' (doctoral dissertation, University of Chicago, 1983); Paul M. Romer, *Human Capital and Growth: Theory and Evidence* (Washington, D.C.: National Bureau of Economic Research Working Paper No. 3173, 1989).

[3] Aaron Benavot, 'Education, Gender, and Economic Development: A Cross-National Study', *Sociology of Education*, 62 (1989), 14–32; Susan H. Cochrane, *Fertility and Education: What Do We Really Know?* (Washington, D.C.: World Bank Occasional Working Paper, 1979); Nelly P. Stromquist, Elizabeth M. King and M. Anne Hill, 'Women's Education in Developing Countries: Barriers, Benefits, and Policies', *Comparative Education Review*, 40 (1996), 450–3; Audrey C. Smock, *Women's Education in Developing Countries: Opportunities and Outcomes* (New York: Praeger, 1981); Nelly P. Stromquist, *Determinants of Educational Participation and Achievement of Women in the Third World: A Review of Evidence and a Theoretical Critique* (Palo Alto, Calif.: CERAS, 1988); Kalanidhi Subbarao and Laura Raney, *Social Gains from Female Education: A Cross-National Study* (Washington, D.C.: World Bank, 1993).

[4] World Bank. *World Development Report 1991: The Challenge of Development* (Oxford: Oxford University Press, 1991).

[5] World Bank, *World Development Report 1991*.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

There is relatively little cross-national work on the politics of gender inequality in education. This seems odd given the inherently political nature of gender inequality. Analysts with the United Nations Development Program (UNDP) argue that moving towards gender equality is a distinctly political process.[6] Preliminary evidence indicates there is a strong correlation between democracy and gender equality in education.[7] The evidence linking democracy to women's education, however, is limited to a simple bivariate regression of women's educational attainment on Gastil's index of political rights and civil liberties.[8] The results from the bivariate regression are less than convincing since there may be a number of confounding factors other than democracy driving the result; the obvious culprits are religion, important regional factors, income and economic growth. A broad empirical analysis of the relationship between democracy and gender inequality in education does not exist. The following is an attempt to provide one.

I use Robert Barro and Jong-Wha Lee's data on educational attainment to examine the relationship between democracy and the gap in education between women and men.[9] Three broadly gauged measures of regime type are used to establish whether democracy influences the gender gap in educational attainment.[10] Broad-based measures of democracy, I find, are not correlated with women's educational attainment. The aggregate measures of regime type combine several dimensions of democracy, making it difficult to determine whether specific democratic institutions influence the stock of women's human capital. I use Gurr's Polity III data to isolate and identify specific components of regime type that may be associated with women's education. Of the three dimensions that constitute Gurr's definition of democracy, only one (executive recruitment) shows a strong, consistent correlation with gender inequality in education. The results suggest that the institutional constraints under which a nation's political leader is placed do not make a significant difference; nor do those aspects of democracy associated directly with political participation. My findings imply that the processes nations use to recruit and select their chief executives strongly influence opportunities for women. A more regulated, open and competitive process of executive recruitment compels candidates to address concerns voiced by previously ignored groups in order to build electoral support. The results lend support to Schumpeter's argument that the manner in which a country selects its chief executive is the most salient feature of the democratic process.

This Note is organized as follows. The first section provides the theoretical motivation for examining the impact of political institutions on women's education. The second section ties the theoretical literature to specific measures of democracy and an array of control variables; the third section lays out the specification of the model used to estimate democracy's impact on women's educational attainment. The penultimate section presents the results and is followed by the conclusion.

THEORY

Political Institutions and Gender Inequality in Education

Although a literature on democracy and its impact on women's educational attainment has yet to be developed fully, a well-established literature on democracy and its consequences holds important

[6] UNDP, *Human Development Report* (New York: Oxford University Press, 1995), p. 1.

[7] World Bank, *World Development Report 1991*, p. 50.

[8] The measure of education used in the World Bank study is the percentage of the female population that has had more than one year of school divided by the percentage of the male population that has had more than one year of school.

[9] Robert J. Barro and Jong-Wha Lee, 'International Comparisons of Educational Attainment', *Journal of Monetary Economics*, 32 (1993), 363–94.

[10] Michael Alvarez, José Antonio Cheibub, Fernando Limongi and Adam Przeworski, 'Classifying Political Regimes', *Studies in Comparative International Development,* 31 (1996), 3–36; Raymond D. Gastil, 'The Comparative Survey of Freedom: Experiences and Suggestions', *Studies in Comparative International Development*, 25 (1990), 25–51; Keith Jaggers, Ted R. Gurr *et al*., 'Transitions to Democracy: Tracking Democracy's Third Wave with the Polity III Data', *Journal of Peace Research*, 32 (1995), 469–82.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

implications for the theory of democracy and women's education.[11] The theoretical discussion below is not exhaustive, it simply identifies some plausible causal connections that link political institutions to gender inequality in education.

Democracy can affect the supply and demand of educational opportunity for women through a number of different channels. Democratic institutions can influence the ability of women to organize and express their views in addition to influencing their ability to obtain and disseminate information. Arguments focusing on the ability of women to organize and express their views emphasize the importance of autonomous organizations (for example, non-governmental organizations and trade unions). Autonomous organizations are important because they provide individuals with the organizational wherewithal to challenge public officials. Without the rights and liberties associated with democratic governance, the kinds of demands associated with more open forms of government are circumscribed. States that limit the freedom of expression and association also limit the flow of information between the state and society.[12] Maintaining open lines of communication between state and society is important because it encourages government efficiency.[13] Voting represents an important form of communication. Participating in free and fair elections allows women to express their preferences. When participation by women is limited or non-existent, societal preferences are not accurately expressed, resulting ultimately in the under-provision of public goods: educational opportunities for women.[14]

Arguments grounded in property rights have similar implications for women's education. Predatory rulers – politicians left unchecked by constitutional constraints on their authority – are less likely to provide the optimal level of public goods.[15] Here the emphasis is on politicians and whether limits are placed on their rule in the form of legislatures, courts or other constitutional mechanisms that encourage power-sharing. Under these constitutional constraints, predatory rulers will be more subject to the demands made by a wider segment of the population, a segment that includes women.

There are reasons to doubt, however, that democracy affects women's education in such a straightforward manner. Olson provides an important reason to be sceptical: as an interest group, women constitute a very large and heterogeneous array of interests, making collective action very difficult.[16] Because large groups with interests that cut across a number of different issues complicate collective action, participation in the electoral game may not represent a viable institutional tool that can be used by women to increase educational opportunity. Unless women can overcome significant collective action problems, it is doubtful whether democracy can influence women's education through voting in elections. As countries democratize, it is not clear that demands for increased educational opportunity for women will emerge automatically. Instead, some institutional features

[11] For one of the most recent contributions that discusses the literature on democracy and its consequences, see Adam Przeworski, Michael Alvarez, José Antonio Cheibub and Fernando Limongi, *Democracy and Development* (Cambridge: Cambridge University Press, 2000). The authors find that find there is a strong correlation between democracy and women's fertility, a relationship that could be influenced by educational opportunity for women.

[12] William McCord, *The Springtime of Freedom* (New York: Oxford University Press, 1965).

[13] James C. Strouse and Richard P. Claude, 'Empirical Comparative Rights Research: Some Preliminary Tests of Development Hypotheses', in Richard P. Claude, ed., *Comparative Human Rights* (Baltimore, Md.: Johns Hopkins University Press, 1976), pp. 51–68.

[14] This is the line of argument a number of rational choice models use in their attempts to show that democracies stand the best chance of providing the optimal level of public goods. See Adam Przeworski, *The State and the Economy under Capitalism: Fundamentals of Pure and Applied Economics* (Chur: Harwood Academic Publishers, 1990); and Adam Przeworski and Fernando Limongi, 'Political Regimes and Economic Growth', *Journal of Economic Perspectives*, 7 (1993), 51–71, for a description of rational choice theories designed to explain regime type and its impact on economic performance.

[15] Margaret Levi, *Of Rule and Revenue* (Berkeley and Los Angeles: University of California Press, 1988); Mancur Olson, 'Dictatorship, Democracy, and Development', *American Political Science Review*, 87 (1993), 567–76; Przeworski, *The State and the Economy under Capitalism*.

[16] Mancur Olson, *The Logic of Collective Action* (Cambridge, Mass.; Harvard University Press, 1965).

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

associated with democracy may be more important than others for explaining the variance in women's education.

Elections influence the degree of insulation politicians enjoy *vis-à-vis* the electorate, but so do a number of other rules and regulations that affect the process of selecting a leader. Placing less emphasis on participation, individual rights and civil liberties, Joseph Schumpeter argues the defining feature of democracy is how individuals acquire the power to make political decisions by means of a competitive struggle for the people's vote.[17] Schumpeter's argument underscores the supply side in the emphasis he places on competition at the elite level for a free vote. Particularly relevant is the importance he attributes to choosing a nation's leader. The institutional constraints leaders face while in office are less important than what Schumpeter believes to be the ultimate constraint, the ability of the electorate to determine who occupies the office of the chief executive. How that process unfolds and whether it allows free competition for a free vote represents the most critical aspect of democracy. According to Schumpeter, classical definitions of democracy 'attributed to the electorate an altogether unrealistic degree of initiative which practically amounted to ignoring leadership'.[18] Collectives or interest groups, 'act almost exclusively by accepting leadership – this is the dominant mechanism of practically any collective action which is more than a reflex'.[19] Even if these collectives or interest groups are strong and readily identified,

they remain latent, often for decades, until they are called to life by some political leader who turns them into political factors. This he does, or else his agents do it for him, by organizing these volitions, by working them up and by including eventually appropriate items in his competitive offering.[20]

Put simply, political entrepreneurs will help activate latent political interests by offering policies that address their concerns. Note that under Schumpeter's conception of democracy, the collective action problem may be resolved by a political entrepreneur who bears the cost of organization in order to gain political support.

A regularized and competitive process of elite recruitment may not only encourage politicians to seek electoral support from previously ignored groups, a more open and competitive process may produce a wider range of candidates. In particular, under more open and institutionalized forms of elite recruitment, women candidates may be more likely to run for office. Consequently, more competitive forms of elite recruitment not only encourage politicians to address the concerns of women, but they increase the level of participation by women in politics. The impact of a growing number of women in the political arena is two-fold. First, women candidates may be more likely to address the demands articulated by women voters. Secondly, women who achieve high political office demonstrate to other women that success in a traditionally male-dominated profession is possible. Ultimately, then, the increased presence of women in the political arena may have not only a direct impact on policy that affects the number of women receiving education, it may encourage other women to choose career paths that require obtaining an education.

Economic and Sociological Determinants

As societies industrialize and income levels rise, female enrolment rates will tend to increase. Higher incomes may affect attitudes towards educating women as well as the demand women express for education. Modernization theory predicts that industrialization will challenge antiquated notions about women's role in society, increasing the number of women who enrol in school. Income could also have a direct impact on a woman's decision to enrol: families who rely on the extra income

[17] Joseph A. Schumpeter, *Capitalism, Socialism, and Democracy* (New York: Harper Torchbooks, 1975), p. 269.
[18] Schumpeter, *Capitalism, Socialism, and Democracy,* p. 270.
[19] Schumpeter, *Capitalism, Socialism, and Democracy*, p. 270.
[20] Schumpeter, *Capitalism, Socialism, and Democracy*, p. 270.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

generated by domestic labour (usually performed by girls) cannot afford to send their children to school.[21]

A rapidly improving or deteriorating economy can have a substantial impact on the demand for education.[22] During rapid growth, employment opportunities for males may allow females to attend school. When economic growth proceeds at high levels, the incentives to stay at home change. Although the greatest incentive may be for females to enter the workforce directly in times of economic prosperity, some will find the return to schooling attractive.

A quick glance at the regional differences in women's education leave little doubt that important regional effects account for a large portion of the variance in women's education. Norms and customs that arise from living in a particular physical environment can have a tremendous impact on decisions to attend school. A number of other variables associated with region may explain a greater proportion of the variance in women's education. The availability of mineral resources, the kind of agricultural goods produced in the region, the region's proximity to industrialized countries and to important trade routes can all influence socio-economic factors that ultimately influence the decision to attend school.

Religious norms and practices could have the most important impact on women's education. Some faiths discourage women's education while others offer no resistance. Many of the largest gaps in educational attainment between men and women are found in countries where the Islamic faith is prevalent. Given its deep-rooted nature in society, religion may be much better at explaining women's educational attainment than regime type.

External forces may also overshadow individual regime characteristics. Colonization had a tremendous impact on economic development in countries throughout Africa, Latin America and Asia. Whether countries were colonized by the British or by the French is strongly correlated with subsequent enrolment ratios for primary and secondary school in sub-Saharan Africa.[23] Similar findings may hold for women's educational attainment.

Finally, perhaps more important than democracy in terms of women's political power is whether they have been given the right to vote; this is particularly important since most measures of democracy fail to account for the inception of women's suffrage. With some exceptions (France 1945, Belgium 1948, Switzerland 1971), many of the Organization for Economic Cooperation and Development (OECD) countries granted women's suffrage at the turn of the century (Norway 1913, Denmark 1908, Australia 1902). In Latin America, suffrage was granted between the 1930s and 1950s (Brazil 1932, Costa Rica 1949, Mexico 1953). In Africa, suffrage was granted on or around the date of independence (Malawi 1961, Senegal 1956, Sierra Leone 1961). The date women gained the right to vote in each country may be more important for putting pressure on politicians to expand educational opportunity for women.

MEASUREMENT

Dependent Variable (Gender Inequality Ratio)

To compare educational attainment between the genders, I constructed a measure that records the average number of years women attended school divided by the average number of years men attended school.[24] The resulting ratio provides an indication of the educational disparity that exists between men and women. The mean value of the Gender Inequality Ratio for 1990 is 0.78 with a minimum of 0.13 (Nepal 1980) and maximum of 1.35 (Lesotho 1970). To give an indication of the

[21] George Psacharopoulos, 'Returns to Education: An Updated International Comparison', *Journal of Human Resources*, 20 (1985), 321–41.

[22] Psacharopoulos, 'Returns to Education'.

[23] David S. Brown, 'Democracy, Colonization, and Human Capital in Sub-Saharan Africa', *Studies in Comparative International Development*, 35 (2000), 20–41.

[24] The variables I used are called TYRF15 and TYRM15 which record the average number of years females over the age of 15 and males over the age of 15 have attended school.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

variance throughout the world, the mean score for Sub-Saharan Africa is 0.60, the mean score for Asia is 0.65, the mean for Latin America is 0.93, the mean score for Europe is 0.98, and the mean for the Middle-East is 0.58. A normal probability plot indicates the distribution of the variable is not badly skewed.[25] A statistical summary of the variables included in the analysis is presented in Appendix A. A list of the countries for which data were available is presented in Appendix B. The sample includes 105 countries, which both constitutes a representative cross-section of the world's nations as well as incorporating all of the available data.

Political Variables (Aggregate Measures)

Although most theories of democracy and its policy consequences focus on fairly specific institutional features of regime type, scholars generally rely on broad measures of democracy in their empirical analyses. Virtually all regime type measures attempt to account for several basic features of democracy: political rights, civil liberties, competitive elections and constitutional constraints. Most measures combine two or more of these components to provide a single indicator.

Three measures of democracy commonly found in the literature are used to estimate democracy's effect. The measure developed by Michael Alvarez, José Antonio Cheibub, Fernando Limongi and Adam Przeworski makes a dichotomous distinction between authoritarian and democratic regimes.[26] Alvarez *et al*. operationalize democracy based on Robert Dahl's concept of polyarchy.[27] Polyarchy is founded on two essential features of democratic governance: inclusiveness and public contestation. To be classified as an authoritarian regime, at least one of the following conditions must hold:

Rule 1. 'Executive Selection'. The Chief Executive is not elected.
Rule 2. 'Legislative Selection'. The Legislature is not elected.
Rule 3. 'Party'. There is no more than one party. Specifically, this rule applies if 1) there were no parties, or 2) there was only one party, or 3) the current term in office ended in the establishment of a non-party or one-party rule, or 4) the incumbents unconstitutionally closed the legislature and rewrote the rules in their favor.
Rule 4. 'Type II Error'. A regime passes the previous three rules, the incumbents will have or already have had continuously held office by virtue of elections for more than two terms or without being elected for any duration, and until today, or the time when they were overthrown, they have not lost an election.[28]

The second measure of political institutions I use is drawn from Ted Gurr's Polity III dataset.[29] To construct the measure of democracy from Gurr's Polity III data, I followed the work of Londregan and Poole by subtracting Gurr's AUTOC score from the DEMOC score (usually referred to as the D-A score), producing a measure of democracy that ranges from -10 to $+10$ ($+10$ being the most 'democratic' score).[30] Finally, I used Gastil's measure which provides two indices: (1) political rights; (2) civil liberties.[31] I combined both indices to best approximate a continuous measure. The measure derived from the Gastil indices ranged from 2 to 14; I reversed Gastil's scale so that 14 represents the highest level of democracy.

[25] Note that there is no theoretical limit at 1: women can – at least in theory – accumulate more human capital than men (although this rarely happens).

[26] Alvarez, 'Classifying Political Regimes'.

[27] Robert Dahl, *Polyarchy; Participation and Opposition* (New Haven, Conn.: Yale University Press, 1971).

[28] Alvarez, 'Classifying Political Regimes'.

[29] Ted R. Gurr, 'Polity III' (May 1996), see http:///www.colorado.edu/IBS/GAD /spacetime/data/Polity.html.

[30] John B. Londregan and Keith T. Poole, 'Does High Income Promote Democracy?' *World Politics*, 49 (1996), 1–30. See Jaggers, 'Transitions to Democracy', for a description and analysis of the Polity III data. See Ted R. Gurr and Harry Eckstein, *Patterns of Authority: A Structural Basis for Political Inquiry* (New York: Wiley, 1975), for a theoretical presentation of the Polity I and Polity II data.

[31] Gastil, 'The Comparative Survey of Freedom', p. 25.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

Employing broadly-based measures of democracy is a well-established practice. Using broad measures of democracy can tell us if being democratic is correlated with women's educational attainment. Broad measures are, however, less useful at telling us why the two might be correlated. If the results are insignificant, it could be that some components of democracy affect women's education positively while other institutions affect women's education negatively. If the results are positive or negative, we do not know whether the estimates are generated by political rights, civil liberties, free and fair elections or constitutional constraints on the chief executive. Gleditsch and Ward note there are a number of possible institutional combinations that produce the same score on Gurr's D-A index. In their study of the Gurr Polity III data, they show that fifty-four combinations are observed for the democracy scale and fifty-seven combinations are observed for the autocracy scale. Illustrating the complexity of the Gurr democracy measure, Gleditsch and Ward find ten different combinations in which polities have received the value of 6 on the democracy score.[32] The tendency towards using broad aggregate measures is surprising given a theoretical literature that makes fairly specific claims about how democracy should work. Since each aggregate measure may be conflating several different dimensions of democracy, it is difficult to determine whether specific democratic institutions matter or not. Without going beyond the broad measures of regime type, we simply cannot know.

Political Variables (Refined Measures)

Gurr's Polity III data provide more refined measures to test the individual claims put forward in the theoretical literature. The conceptual framework Gurr uses to construct measures of democracy and autocracy maps onto the theoretical literature described above. As noted earlier, a significant stream of research focuses on political participation and how free and fair elections allow citizens to express their demands. Gurr's measures PARCOMP – the competitiveness of political participation – and PARREG – the regulation of political participation – are designed to capture two different dimensions of political participation. Gurr's measure XCONST – constraints on the chief executive – registers the degree to which politicians are constrained once in office, corresponding nicely to the theoretical literature on the theory of predatory rule. Finally, Gurr's measures XRREG, XROPEN, XRCOMP – the regulation, openness and competitiveness of executive recruitment – closely resemble the institutional features emphasized in Schumpeter's work. Let me briefly describe in more detail what these more refined indicators are designed to measure.

The variables PARCOMP and PARREG are designed to capture two separate dimensions of political participation: regulation and competitiveness. According to Gurr *et al*., 'Participation is regulated to the extent that there are binding rules on when, whether, and how political preferences are expressed.'[33] Highly regulated participation can be found both in one-party states and in Western democracies. Unregulated participation is characterized by polities with no 'enduring national political organizations and no effective regime controls on political activity. In such situations, political competition is fluid and usually characterized by recurring violent conflict among shifting coalitions of partisan groups.'[34] The competitiveness of participation variable records the degree to which alternative preferences for policy and leadership can be pursued in the political arena.[35] The measure indicates whether oppositional activity is permitted outside the ranks of the regime or ruling party. To achieve the highest score, countries must have stable and enduring political groups that regularly compete for political influence and position with little use of violence or disruption.[36]

[32] Kristian S. Gleditsch and Michael D. Ward, 'Double Take: A Reexamination of Democracy and Autocracy in Modern Polities', *Journal of Conflict Resolution*, 41 (1997), 361–83.

[33] Ted R. Gurr, Keith Jaggers *et al*., 'The Transformation of the Western State: The Growth of Democracy, Autocracy, and State Power Since 1800', *Studies in Comparative International Development*, 25 (1990), p. 78.

[34] Gurr *et al*., 'The Transformation of the Western State', p. 78.

[35] Gurr *et al*., 'The Transformation of the Western State', p. 79.

[36] Gurr *et al*., 'The Transformation of the Western State', p. 79.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

Gurr's measure XCONST – constraints on the chief executive – registers the extent to which the chief executive is constrained by constitutional rules once in office. This measure records the degree to which politicians are constrained by power-sharing arrangements. At the low end of the scale, politicians enjoy unlimited authority: only coups and assassinations limit rulers at this end of the scale.[37] On the high end of the scale – the more democratic end – there is executive parity or subordination of the executive to a specified body of equal or greater effective authority.[38]

The third and final dimension of Gurr's conceptual framework is the selection and recruitment of the chief executive. The three variables that constitute this dimension are the regulation of recruitment (XRREG), the competitiveness of recruitment (XRCOMP), and the openness of recruitment (XROPEN). The regulation of executive recruitment variable indicates how institutionalized or regulated the process of recruitment is in a given polity. Leaders who take office by force receive the lowest score. Leaders who achieve office through heredity or through open and competitive elections receive the highest scores. Although at first this seems contradictory, it is necessary to remember that the regulation of executive recruitment variable simply attempts to establish whether or to what extent a polity has institutionalized procedures for transferring power.[39]

The competitiveness of executive recruitment variable indicates the extent to which the selection of the chief executive occurs through popular elections matching two or more parties or candidates. The lowest score is assigned to polities in which chief executives are determined by hereditary succession, designation or by a combination of both. At the other extreme are chief executives who are chosen in popular elections matching two or more major parties or candidates

The openness of recruitment variable attempts to gauge whether all individuals have the opportunity, in principle, to attain the position of chief executive through a regularized process. The lowest scores are assigned to polities that select their leaders strictly through heredity. Higher scores are assigned to polities that choose leaders through elite designation, competitive election or transitional arrangements between designation and election.

Economic and Sociological Variables

A number of different explanations put forward in the theoretical section may account for the disparity between men and women in educational attainment. Below is a brief description of how they were operationalized.

A nation's level of income and its record of economic performance (the annual change in per capita gross domestic product (GDP)) are included in the analysis. The GDP per capita data are taken from Robert Summers and Alan Heston's dataset (Mark 5.6) which bases its estimates of output on purchasing power parities.[40] The variable used in the analysis (RGDPCH) is designed to allow for more accurate cross-temporal as well as cross-sectional comparisons. So that linear methods of estimation can be applied to the data, I will use a logarithmic transformation of the GDP per capita variable. I use the same income data from the Summers and Heston dataset to calculate the annual percentage change in GDP per capita.

To account for the important regional factors that might explain the wide variance in women's educational attainment throughout the world, I included dummy variables for each region of the world.[41] A number of different factors associated with region may not be reflected in the other control variables. For example, although many countries in Eastern Europe have income levels that are comparable to some countries in Latin America and Asia, the Gender Inequality Ratio looks

[37] Gurr *et al.*, 'The Transformation of the Western State', p. 80.

[38] Gurr *et al.*, 'The Transformation of the Western State', p. 80.

[39] The results reported below are not generated by cases that score well on the regulation of recruitment variable because executive office is inherited.

[40] Robert Summers and Alan Heston, 'The Penn World Table (MARK 5): An Expanded Set of International Comparisons, 1950–1988', *Quarterly Journal of Economics*, 106 (1991), 327–68.

[41] The coefficient for each regional dummy variable indicates how that region differs from the Middle East, which serves as the baseline for comparison.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

somewhat different in Eastern Europe compared to other geographic regions because of the different roles women play in a communist economy. So that we can gain a more accurate estimate of democracy's association with gender inequality in education, regional differences not captured by the independent variables are accounted for by using regional dummy variables. Although loosely correlated with region, religion crosses regional boundaries. Simply relegating the effects of religion to regional dummy variables fails to provide an accurate assessment of religion's importance. To guard against that problem I constructed a dummy variable that designates countries in which the majority of society practises the Islamic faith (1 if Islamic; otherwise 0).[42] In addition to the region and religion variables, I included a dummy variable that registered whether the country was a former British colony (1 if British; otherwise 0). The British administered education in their colonies very differently from their French, German and Dutch counterparts. Their emphasis on basic education may have an important impact on the number of girls attending school. Included in the model to control for women's suffrage is a variable that registers the year in which women gained the right to vote in each country.[43] Not only is the right to vote important, but women's suffrage is associated with the ability of women to stand for election. To the extent that a longer history of political enfranchisement has resulted in policies that increase educational opportunity for women, it is important to account for the varying dates upon which women formally received the right to vote.

MODEL

Barro and Lee's data record the level of educational attainment already accumulated in society, providing a measure of stock rather than a measure of flow. To measure democracy's effect on the stock of human capital, we need to allow for a significant time lag. Correlating democracy with the stock of human capital in the same year does not render an accurate estimate of the regime type's influence since the stock of any society's human capital has accumulated over many years. To address the problem, observations for the dependent variable are taken from the last year in the dataset – 1990. Consequently, the dataset is cross-sectional. The independent variables represent the mean values for each country between 1960 and 1990. Taking averages over a thirty-year time period introduces some interpretation problems. For example, by taking averages, countries scoring in the middle of the D-A scale can obtain the same score in one of two ways. Countries that swing wildly from authoritarianism to democracy register scores similar to those countries that maintain stable regimes not considered fully authoritarian or democratic. The problem of interpretation, however, does not exist in these data: there were no cases located in the middle of the D-A scale able to hold power for long periods of time. In fact, regimes scoring close to the mean D-A score all experienced significant change over the thirty-year period (see Figure 1 which plots the mean D-A score against the standard deviation of D-A in each country).[44] Interpretation problems do not exist as one moves to either end of the D-A scale since countries at either extreme were either solidly authoritarian or democratic.[45] Taking averages is also justified given the nature of the dependent variable. Since the Gender Inequality Ratio in 1990 was calculated for all men and women above the age of 15, the measure records the human capital accumulated by both genders over a significant time period (15–40 years). A more practical justification dictates the use of averaging: since data

[42] Information on religion in each country was obtained from John Paxton, *The Stateman's Year-Book* (New York: St Martin's Press, 1985).

[43] The dates when women obtained the right to vote in each country were obtained from Francisco G. Ramirez, Yasemin Soysal and Suzanne Shanahan, 'The Changing Logic of Political Citizenship: Cross-National Acquisition of Women's Suffrage Rights, 1890–1990', *American Sociological Review*, 62 (1997), 735–45.

[44] Figure 1 indicates that the countries recording a D-A score around the mean for the period in question all experienced considerable variance in regime type. Averaging the D-A scores does not, therefore, confuse different scenarios with similar D-A scores.

[45] I limited the analysis to those cases that were either authoritarian or democratic during the vast majority of the period (cases where the standard deviation in the D-A score was less than 2.5). Limiting the regression analysis to those cases did not change any of the estimates reported in Table 1.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

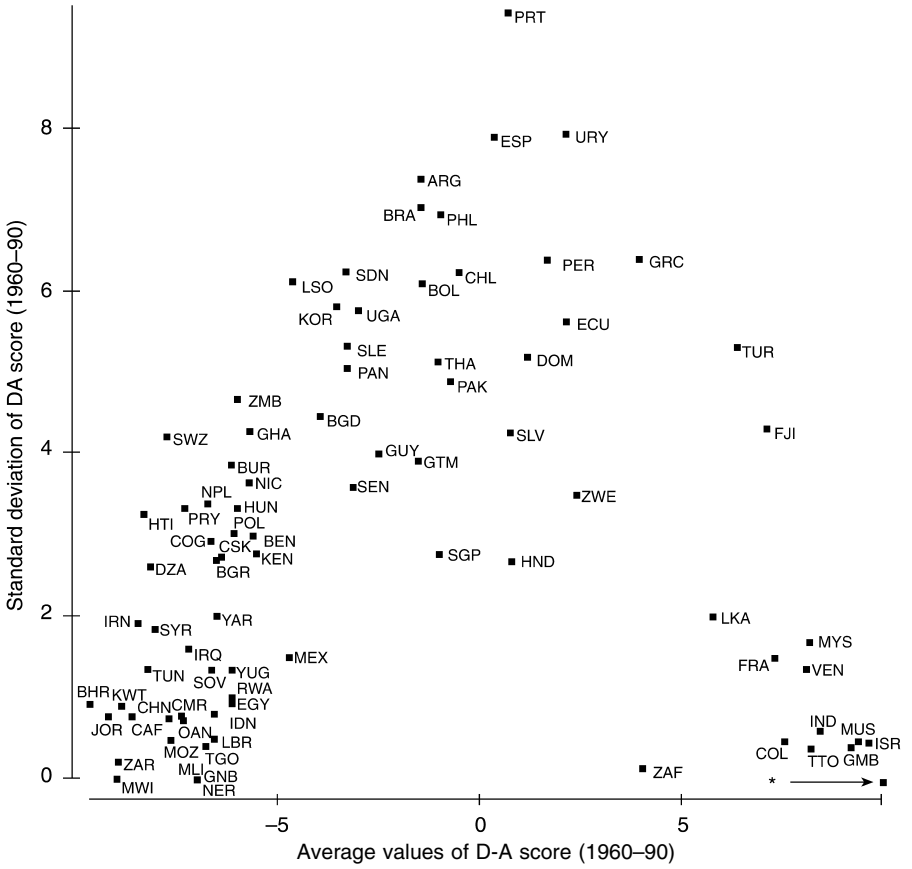


Fig. 1 Scatterplot of standard deviation in D-A score against the mean D-A score

**The following countries share the same value of 10 with a standard deviation of 0: Australia, Austria, Belgium, Botswana, Canada, Switzerland, West Germany, Denmark, Finland, Great Britain, Ireland, Iceland, Italy, Jamaica, Japan, Netherlands, Norway, New Zealand, Papua New Guinea, Sweden, and the United States. A guide to the country codes is given in Appendix B.*

for the dependent variable are only available in five-year increments, adding a time-series component that accurately models the dynamics of the relationship between political institutions and the accumulation of human capital is simply not feasible when there are a maximum of seven cases per country (1960–90 at five-year increments).

The model is designed to test whether the mean values of democracy for each country over the thirty-year period are correlated with the Gender Inequality Ratio in 1990. The regression model is as follows:

$$\begin{aligned}\text{Gender Inequality Ratio}_{1990} = \quad & a + b_1\,(\text{GDP/capita}_{(\mu\,\text{of country}\,i)}) \\ & + b_2\,(\Delta\text{GDP/capita}_{(\mu\,\text{of country}\,i)}) \\ & + b_3\,(\text{Religion}) \\ & + b_4\,(\text{British Colony}) \\ & + b_5\,(\text{Democracy Measure}_{(\mu\,\text{of country}\,i)}) \\ & + b_{6\text{–}11}\,(\text{Regional Dummy Variables}) \\ & + b_{12}\,(\text{Year Women's Suffrage is granted}) \\ & + e.\end{aligned}$$

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

I used this basic model to test the effect of the broadly based regime type variables and the more refined indicators found in the Gurr Polity III data. Given the normal distribution of the dependent variable and the cross-sectional structure of the data, I used ordinary least squares estimation.[46]

RESULTS

Measures of democracy that combine various aspects of electoral competition, civil rights and institutional design fail to explain much of the variance in the Gender Inequality Ratio. Disentangling the various dimensions, however, reveals a distinct pattern: the process through which a country's chief executive is selected has a substantively and statistically significant effect on the Gender Inequality Ratio. Table 1 reports the results for four regressions that estimate the correlation between broadly based measures of democracy and the Gender Inequality Ratio.

TABLE 1 *OLS Regression of Gender Ratio on Competing Measures of Democracy*

	(1)	(2)	(3)
Constant	0.566 (1.559)	1.654 (1.599)	− 0.007 (1.536)
Year suffrage is established	− 0.000 (0.001)	− 0.001 (0.001)	− 0.000 (0.001)
British colony (1/0)	0.071 (0.029)**	0.055 (0.030)*	0.061 (0.030)**
Muslim majority (1/0)	− 0.109 (0.033)***	− 0.090 (0.034)***	− 0.099 (0.032)***
Annual change in GDP per capita$_{(\mu \text{ of country } i)}$	0.642 (0.412)	0.483 (0.434)	0.778 (0.388)**
GDP per capita$_{(\mu \text{ of country } i)}$	0.286 (0.046)***	0.279 (0.051)***	0.261 (0.044)***
Przeworski's Measure$_{(\mu \text{ of country } i)}$	0.020 (0.044)		
Gastil's Measure$_{(\mu \text{ of country } i)}$		0.000 (0.006)	
Polity III (D-A)$_{(\mu \text{ of country } i)}$			0.002 (0.002)
Nepal dummy†	− 0.343 (0.116)***	− 0.358 (0.111)***	− 0.323 (0.114)***
Lesotho dummy†	0.632 (0.110)***	0.661 (0.106)***	0.649 (0.110)***
Observations	103	94	105
Adjusted R^2	0.72	0.74	0.72

Notes: Standard errors in parentheses: *significant at 10 per cent; **significant at 5 per cent; ***significant at 1 per cent. Region dummies were included in every regression but are not reported above for presentation purposes. Variables with the subscript (μ of country *i*) represent the mean values calculated over the years 1960–90 (except Gastil's measure, which is for 1974–90). †Using significant values of DFFITS as a criterion, two cases stand out: Lesotho and Nepal. The estimates reported in both Tables 1 and 2 were based on regressions that included those two countries as dummy variables, providing a more accurate estimate of the underlying empirical pattern.

[46] The results reported below withstood a number of different tests including altering the specification of the model. Concerned with possible simultaneity (gender inequality in education and democracy), I ran several 2SLS models along with investigating the hypothesis that the gender inequality measure is directly correlated with the stock of human capital (an important correlation which would indicate whether simultaneity is a concern). Neither line of inquiry indicated the results suffered from bias caused by simultaneity.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

Before proceeding to the variables of interest, it is important to note that the control variables are all strong and significant in the direction posited by previous theory. Growth in GDP per capita along with income levels (GDP per capita) are positively and significantly associated with women's education. The British colony dummy variable and the religion dummy variable (Muslim = 1) record signs in the expected directions: (1) the British colony dummy is positive, showing that former British colonies tend to have higher levels of equality relative to the former French, German and Dutch colonies; (2) the sign on the religion variable is negative, indicating there is less parity among the sexes in Islamic countries. The only surprise was the lack of association between the suffrage variable and the Gender Inequality Ratio. An interpretation consistent with the results below suggests that a woman's right to vote is not a sufficient condition leading to equal educational opportunity.

Whether using the mean of Gurr's D-A score (1960–90) for each country, the mean of Gastil's index (1974–90), or the mean of Alvarez *et al.*'s measure of regime type (1960–90), democracy seems to have little or no association with educational opportunity for women relative to men. One of two interpretations is possible: (1) the underlying process of democratization has no impact on the Gender Inequality Ratio; (2) the measures used to capture democracy's effects are too broadly gauged and fail to register the impact that specific democratic institutions have on women's education. If the former holds, we need to look elsewhere for answers. If the latter holds, using more refined measures of democracy may help.

To determine which dimension, if any, has an effect on the Gender Inequality Ratio, I ran separate regressions for each of Gurr's indicators. The estimates reveal a clear pattern (see Table 2). Coefficients for the political participation variables and the variable tracking constraints on the chief executive could not be distinguished from zero.[47] The results suggest that political participation is not a sufficient condition for achieving educational equality between the sexes. The coefficient for the women's suffrage variable provides further evidence that the act of voting is by itself no panacea. In no regression could the estimate for the suffrage variable be distinguished from zero.

However, each of the variables associated with the selection and recruitment of the chief executive is strongly correlated with the Gender Inequality Ratio. To illustrate the substantive effect of the regulation of recruitment variable, I generated predicted values and 95 per cent confidence intervals from the model by holding the independent variables constant at their means and by varying the regulation of recruitment variable between 1 and 3. At the lowest actual score (Honduras, 1.3), the model predicts a Gender Inequality Ratio of approximately 74 per cent (95 per cent confidence interval, 0.67 to 0.80). When countries achieve a score of 3 (most of the OECD countries), the model predicts 90 per cent (95 per cent confidence interval, 0.85 to 0.94). The variable measuring the competitiveness of executive recruitment produces similar results. For countries with the lowest score on the competitiveness of executive recruitment scale (Cuba, Bangladesh, 0), the model predicts a Gender Inequality Ratio of 78 per cent (95 per cent confidence interval, 0.72 to 0.85). For countries that score the highest (4), the model predicts 92 per cent (95 per cent confidence interval, 0.84 to 0.99).

The results suggest that a specific set of democratic institutions is associated with greater gender equality in education. Broad-based measures of regime type, the results imply, conflate a number of different dimensions of democracy, producing estimates that show no correlation between democracy and women's education. The wrong conclusion to draw is that democracy does not matter. Democracy does have an important impact in a very specific way under specific institutional arrangements. The estimates confirm theoretical arguments that emphasize the importance of the recruitment and selection of the chief executive.

To illustrate the distinction, consider Argentina and Indonesia. Although both have similar scores in terms of the competitiveness of participation – between 1960 and 1990 Indonesia and Argentina averaged 2 – Argentina scores much higher on the regulation of executive recruitment – Argentina

[47] Again, there is additional evidence consistent with this claim. The variables Przeworski uses to draw distinctions among authoritarian regimes are based on criteria similar to that used in the construction of the XCONST variable. The Gurr and Przeworski measures generate similar estimates.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

TABLE 2 *OLS Regressions of Gender Ratio on Components of Gurr's Polity Data*

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.105 (1.51)	0.381 (1.49)	−0.232 (1.54)	−0.051 (1.47)	−0.408 (1.51)	−0.111 (1.43)
Year sufferage was granted	−0.000 (0.00)	−0.000 (0.00)	0.000 (0.00)	−0.000 (0.00)	0.000 (0.00)	−0.000 (0.00)
British colony (1/0)	0.065 (0.03)**	0.072 (0.03)**	0.054 (0.03)*	0.060 (0.03)**	0.047 (0.03)	0.034 (0.03)
Muslim (1/0)	−0.100 (0.03)***	−0.104 (0.03)***	−0.097 (0.03)***	−0.099 (0.03)***	−0.093 (0.03)***	−0.095 (0.03)***
Annual GDP growth$_{(\mu \text{ of country } i)}$	0.770 (0.39)*	0.758 (0.39)*	0.806 (0.39)**	0.765 (0.38)**	0.828 (0.38)**	0.990 (0.38)**
GDP per capita$_{(\mu \text{ of country } i)}$	0.260 (0.04)***	0.268 (0.04)***	0.252 (0.05)***	0.247 (0.04)***	0.238 (0.05)***	0.218 (0.05)***
PARCOMP (Competitiveness of participation$_{(\mu \text{ of country } i)}$)	0.010 (0.01)					
PARREG (Regulation of participation$_{(\mu \text{ of country } i)}$)		0.007 (0.02)				
XCONST (Constraints on chief executive$_{(\mu \text{ of country } i)}$)			0.010 (0.01)			
XROPEN (Openness of executive-recruitment$_{(\mu \text{ of country } i)}$)				0.025 (0.01)**		
XRCOMP (Competitiveness of Executive Recruitment$_{(\mu \text{ of country } i)}$)					0.033 (0.02)*	
XRREG (Regulation of executive recruitment$_{(\mu \text{ of country } i)}$)						0.091 (0.03)***
Nepal dummy†	−0.325 (0.11)***	−0.336 (0.11)***	−0.324 (0.11)***	−0.314 (0.11)***	−0.330 (0.11)***	−0.409 (0.11)***
Lesotho dummy†	0.641 (0.11)***	0.634 (0.11)***	0.658 (0.11)***	0.640 (0.11)***	0.665 (0.11)***	0.689 (0.11)***
Adjusted R^2	0.72	0.72	0.72	0.73	0.73	0.74

Notes: Standard errors in parentheses: *significant at 10 per cent; **significant at 5 per cent; ***significant at 1 per cent. Dummy variables for each region were included in every regression but were not reported above for presentation purposes. Variables with the subscript (μ of country i) represent the mean values calculated over the years 1960–90. $N = 105$. †See Table 1.

averaged 2.5 while Indonesia averaged 2. In terms of participation, both countries witnessed dictatorships during significant parts of the 1960–90 period. However, the two countries differed significantly in the selection and recruitment of the executive. In Indonesia, power rested with one individual, Suharto, from 1965 through the end of the data's coverage (twenty-five of the possible thirty years). In Argentina, competition at the elite level – during the authoritarian and democratic years – was somewhat more competitive, regularized and open. The empirical results suggest that the difference in the process of executive recruitment between countries like Argentina and Indonesia is important for gender equality in education. Predicted values generated by Model 6 in Table 2 suggest the difference along the one dimension is important. The model generates a predicted Gender Inequality Ratio of 1.01 for Argentina (95 per cent confidence interval 0.96 − 1.06) while it predicts 0.64 for Indonesia (confidence interval 0.57 − 0.72).[48] The actual values for Argentina and Indonesia are 1.01 and 0.79 respectively. Other countries whose competitiveness of participation score approximates Argentina's, but score much lower in the regulation of executive recruitment, have much lower Gender Inequality Ratios: Thailand, Bolivia, Senegal, Liberia and Nigeria to name just a few.

CONCLUSION

Previous work on democracy and women's education attempts to show that a positive relationship exists between democratic institutions and educational opportunity for women. To test rigorously whether democracy influences women's education, I specified a regression model that controlled for income, economic growth, region, religion, colonial history and the date when women gained suffrage. Although broadly gauged measures of democracy fail to explain the variance in women's educational attainment relative to men, more refined and specific indicators that track whether the process of executive recruitment is competitive, regulated or open have an observable impact on gender inequality in education. The results were consistent across a number of variables that measure different dimensions of democracy. Only the measures of executive recruitment registered statistically significant results.

The pattern described in the statistical analysis adheres to Schumpeter's conception of democracy. Particularly relevant is the emphasis he places on choosing a nation's leader. How that process unfolds and whether it allows free competition for a free vote represents the most critical aspect of democracy for Schumpeter. Variables that register the degree to which society can freely organize, associate and participate in the act of voting seem to matter much less than the degree to which the competitive process among elites for the peoples' vote is regulated and open. That the same processes are evident across different measures of democracy and across different measures of the gender gap in education suggests there may be an important underlying process at work.

There are a number of possible causal mechanisms underlying the results. First, as competition at the executive level becomes regularized, more open and more competitive, politicians are compelled to build electoral support among previously marginalized groups: politicians may specifically target women as an electoral strategy. To build support among women, politicians can propose and implement a number of different programmes that directly or indirectly affect female enrolment. Policies that benefit women come in a number of different forms: family planning, subsidized daycare, campaigns to educate parents on the benefits of sending their daughters to school, self-employment programmes that provide credit to women and other specialized employment programmes all increase the incentives for women to attend school. Any policy that increases the return on investment in education for women can alter the incentive structure and help close the existing gap in the gender inequality ratio. The results suggest that more open forms of political participation are not as effective in bringing these sorts of changes about. A more effective mechanism involves the role chief executives play in organizing groups with an interest in improving

[48] To generate the predicted values and their confidence intervals I used Gary King's CLARIFY procedure: Gary King, Michael Tomz and Jason Wittenberg, 'Making the Most of Statistical Analyses: Improving Interpretation and Presentation', *American Journal of Political Science*, 44 (2000), 347–61.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

women's education. Put simply, as competition among political leaders increases, the incentives to implement the policies that benefit women's education increases.

Secondly, in addition to the more instrumental set of causal mechanisms, increased competition at the executive level results eventually in successful women candidates. As women become more visible in the political arena, other women will see the possibility of achieving success at the highest levels. The increasing presence of women at the highest levels of government provide very recognizable signals to society and more specifically to parents that there are significant payoffs to educating all of their children. Finally, female political leaders may be more sensitive to the problems that produce a wide gap between men and women in education, resulting in more rigorous attempts to address the problem.

At this point, it is important to recognize what, exactly, this study has accomplished and what it has not. The analysis identified a strong pattern that suggests there are important differences that exist among the various dimensions of democracy. At least in the case of women's educational attainment, only a very specific aspect of democracy seems to matter. The findings point to a new line of investigation; they do not settle the issue. Additional work is required in order to draw the connection between executive recruitment and gender equality more tightly. Until further research is completed, however, there is strong evidence implying that in the fight for equality, some avenues for change may be more effective than others. Perhaps political freedoms and human rights are not sufficient to diminish the gap between genders. Only when political competition exists at the highest level or when women occupy high political office – implying a high degree of openness and competition in executive recruitment – can real change occur.

APPENDIX A: SUMMARY STATISTICS OF VARIABLES

Variable	Mean	Std. Dev.	Min.	Max.
Democracy (Przeworski) $N = 101$	0.644	0.425	0.000	1.000
Democracy (Gastil) $N = 92$	7.299	3.919	2.056	14.000
Democracy (Gurr)	− 1.331	7.082	− 10.000	10.000
GDP/capita (logged)	3.460	0.459	2.601	4.257
Gender ratio (1960)	0.685	0.305	0.036	1.222
Gender ratio (1990)	0.792	0.224	0.180	1.328
Annual change in GDP/capita	0.051	0.176	− 0.075	1.090
Regulation of executive recruitment (XRREG)	2.337	0.513	1.333	3.000
Competitiveness of executive rec. (XRCOMP)	1.591	0.943	0.000	3.000
Openness of executive recruitment (XROPEN)	3.214	1.138	0.000	4.000
Constraints on the chief executive (XCONST)	3.591	2.171	1.000	7.000
Regulation of participation (PARREG)	3.887	0.706	2.000	5.000
Competitiveness of participation (PARCOMP)	2.401	1.494	0.710	5.000

Note: $N = 105$ for each variable unless otherwise indicated.

APPENDIX B: COUNTRIES INCLUDED IN REGRESSIONS*

Argentina-ARG	Australia-AUS	Austria-AUT
Belgium-BEL	Benin-BEN	Bangladesh-BGD
Bulgaria-BGR	Bahrain-BHR	Bolivia-BOL
Brazil-BRA	Burma-BUR	Botswana-BWA
Central African Rep.-CAF	Canada-CAN	Switzerland-CHE
Chile-CHL	China-CHN	Cameroon-CMR
Congo-COG	Colombia-COL	Costa Rica-CRI
Czechoslovakia-CSK	West Germany-DEU	Denmark-DNK
Dominican Rep.-DOM	Angola-DZA	Ecuador-ECU
Egypt-EGY	Spain-ESP	Finland-FIN
Fiji-FJI	France-FRA	Great Britain-GBR
Ghana-GHA	Gambia-GMB	Guinea Bissau-GNB
Greece-GRC	Guatemala-GTM	Guyana-GUY

Cont.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

Honduras-HND	Haiti-HTI	Hungary-HUN
Indonesia-IDN	India-IND	Ireland-IRL
Iran-IRN	Iraq-IRQ	Iceland-ISL
Israel-ISR	Italy-ITA	Jamaica-JAM
Jordan-JOR	Japan-JPN	Kenya-KEN
South Korea-KOR	Kuwait-KWT	Liberia-LBR
Sri Lanka-LKA	Lesotho-LSO	Mexico-MEX
Mali-MLI	Mozambique-MOZ	Mauritius-MUS
Malawi-MWI	Malaysia-MYS	Niger-NER
Nicaragua-NIC	Netherlands-NLD	Norway-NOR
Nepal-NPL	New Zealand-NZL	Taiwan-OAN
Pakistan-PAK	Panama-PAN	Peru-PER
Philippines-PHL	Papua New Guinea-PNG	Poland-POL
Portugal-PRT	Paraguay-PRY	Rwanda-RWA
Sudan-SDN	Senegal-SEN	Singapore-SGP
Sierra Leone-SLE	El Salvador-SLV	Soviet Union-SOV
Sweden-SWE	Swaziland-SWZ	Syria-SYR
Togo-TGO	Thailand-THA	Trinidad & Tobago-TTO
Tunisia-TUN	Turkey-TUR	Uganda-UGA
Uruguay-URY	United States-USA	Venezuela-VEN
Yemen Arab Rep.-YAR	Yugoslavia-YUG	South Africa-ZAF
Zaire-ZAR	Zambia-ZMB	Zimbabwe-ZWE

*See Tables 1 and 2.

Definition and Measurement of Tactical Voting: The Role of Rational Choice

STEPHEN D. FISHER*

There has been substantial debate on the measurement of tactical voting in this *Journal*, much of which has actually been concerned with the definition of a tactical vote. For this reason it is necessary to go 'back to basics' to examine the notion of a tactical vote within rational choice theory and follow its implications for the measurement of tactical voting. This Note has four aims: to elucidate the formal theoretical concepts behind the notion of tactical voting; to provide a practical definition of tactical voting based on those concepts; to examine and use the implications of the theory and definition to evaluate existing measures of tactical voting; and to further appraise measures of tactical voting from general methodological considerations. This Note focuses on tactical voting in single-member simple-plurality electoral systems. The following section argues that the range of situations where it may be optimal to vote tactically is both broader and less well defined than previous authors have generally realized. This implies that both the definition, and therefore the measurement, of tactical voting rest mainly on the criteria of voting for a party other than the first choice and doing so in order to best influence who wins. The empirical section examines various approaches to the measurement of tactical voting and argues that the Heath *et al.* measure is *a priori* the most attractive available.[1] Furthermore, detailed analysis using British Election Study (BES) data from English voters in 1987, 1992 and 1997,[2] shows that whilst the Heath *et al.* measure is

* Department of Sociology, University of Oxford. Thanks to David Firth, Anthony Heath, Iain McLean, David Myatt, David Sanders and two anonymous referees for their comments on earlier versions of this *Note*. I am grateful to them, to the ESRC (award no. R00429634167), and to Nuffield College for their support for this work.

[1] A. Heath, R. Jowell, J. Curtice, G. Evans, J. Field and S. Witherspoon, *Understanding Political Change: The British Voter 1964–1987* (Oxford: Pergamon Press, 1991), chap. 4.

[2] Details of the surveys can be found in Heath *et al.*, *Understanding Political Change*; A. Heath, R. Jowell and J. Curtice with B. Taylor, eds, *Labour's Last Chance? The 1992 Election and Beyond* (Aldershot, Surrey: Dartmouth, 1994); G. Evans and P. Norris, eds, *Critical Elections, British Parties and Voters in Long Term Perspective* (London: Sage, 1999).

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

relatively robust, it can be usefully revised to improve consistency with the strength-of-feeling scores. A corresponding method for identifying a voter's preferred party is also presented.

DEFINITION

Tactical voting is generally understood through a rational choice framework, if only because no other theory predicts tactical voting. In formal rational choice models voters are assumed to be short-term instrumentally rational and those for whom it is utility-maximizing to vote for a party other than the preferred party are said to be tactical voters.[3] What does this mean in practice? Since no one fulfils the abstract conception of a short-term instrumentally rational voter in real life, tactical voters should be identified by their predominant motivations and behaviour conforming to rational choice specifications. Three criteria define instances of tactical voting. First, a tactical voter has short-term instrumental motivations for their vote choice, i.e. the voter wants to influence who wins in their constituency at that election. Secondly, a tactical vote is a vote for a party other than the first preference. Thirdly, the vote choice should be consistent with the expectations of the constituency result, the utilities for the parties and the principal of utility maximization.

The first criterion, having instrumental motivations, implies that the voter's utility from the outcome of the election is affected only by who wins the election. Instrumental voters do not care about the margin of victory or the order in which the candidates finish, or any other aspect of the contest. The phrase 'short-term' is used to emphasize that they care only about who wins in the present election and that election only. It follows that a voter with short-term instrumental motivations wants to influence who wins in their constituency in the current election. This does not preclude other factors affecting voting behaviour, so long as their influence can be reflected in the utilities associated with who wins. For instance, preferences between potential coalitions affect how much you want any particular party to win at the constituency level and so can be factored into the utilities associated with each party winning. Likewise, the utility associated with a party winning this time may be influenced by the effect that such a win might have on policy outcomes or some future election either national, regional or local. The treatment of tactical voting here does not exclude the possibility that voters have strategic considerations as a result of the relationships between different branches of government and the temporal context of elections, such as those described in Alesina and Rosenthal.[4] However, when people derive satisfaction simply from the act of voting for a party, they are clearly not instrumental. Such issues can be incorporated into formal models and they influence the willingness to vote strategically. But this does not change the fact that short-term instrumental motivations must be decisive for tactical voters.

The second criterion is that a tactical vote is never a vote for the first preference party. This is probably the most consistent feature of tactical voting definitions.[5] But it does create problems. First, it suggests that people are not instrumentally rational if they vote for their favourite party, even when it is clearly optimal to do so. Secondly, describing someone as a tactical voter has an artificial air

[3] For example, G. Cox, *Making Votes Count* (Cambridge: Cambridge University Press, 1997). Note that 'tactical voting' is the same as the US term 'strategic voting'. Although the terms are synonymous and this author is indifferent between them, there is a case for using the word 'tactical' rather than 'strategic', based on etymology. Whilst strategy suggests a sequence of actions, a tactic is a single manœuvre. Since strategic voting is utility maximizing for a single election only, the term 'tactical' is probably more appropriate. The term 'sophisticated voting' has also been used, though as Shepsle argues, it is more appropriate for voting on bills and amendments when there are well-defined agenda rules; K. A. Shepsle, *Models of Multiparty Electoral Competition* (Chur: Harwood, 1991), pp. 62–3.

[4] A. Alesina and H. Rosenthal, *Partisan Politics, Divided Government and the Economy* (Cambridge: Cambridge University Press, 1995).

[5] One exception to this rule is Niemi *et al.* who define tactical voting solely by having tactical considerations; R. G. Niemi, G. Whitten and M. N. Franklin, 'Constituency Characteristics, Individual Characteristics and Tactical Voting in the 1987 British General Election', *British Journal of Political Science*, 22 (1992), 229–54, pp. 229–30. In later work they went on to distinguish between instrumental and expressive tactical voting.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

TABLE 1 *Utilities from Voting in Pivotal Situations*

Event	Utility from a B Vote	Utility from an A Vote	$U_{(B)} - U_{(A)}$
$x_A = x_B > x_C$	u_B	u_A	$u_B - u_A$
$x_A = x_C > x_B$	$u_A/2$	u_A	$-u_A/2$
$x_C = x_B > x_A$	u_B	$u_B/2$	$u_B/2$
$x_A = x_B + 1 > x_C$	$(u_A + u_B)/2$	u_A	$(u_B - u_A)/2$
$x_B = x_A + 1 > x_C$	u_B	$(u_A + u_B)/2$	$(u_B - u_A)/2$
$x_A = x_C + 1 > x_B$	u_A	u_A	0
$x_C = x_A + 1 > x_B$	0	$u_A/2$	$-u_A/2$
$x_B = x_C + 1 > x_A$	u_B	u_B	0
$x_C = x_B + 1 > x_A$	$u_B/2$	0	$u_B/2$

of permanence, with unwarranted implications for their behaviour in other elections. Thirdly, confusion can occur if people with instrumental motivations, 'have convinced themselves that the "best" party was one that had a chance of winning.'[6] So there may be tactical voters who, in a sense, feel that they voted sincerely. However, if they are genuinely tactical voters they should still be able to say there was another party they really preferred. The concept of party preference can be separated from that of the chance of winning. Some people want to back a winner and others are drawn to successful parties, but they are only tactical voters when they would rather see another party win. Although the term 'tactical voter' is potentially confusing and on face value suggests nothing more than having instrumental motivations, this is not how the term is used and there is little to be gained by changing terminology.

The third and most interesting criterion is that the tactical voter should maximize utility given the expectations of the constituency result. Using the McKelvey and Ordeshook approach, this condition can be shown to constrain when voting other than the first choice can be considered to be tactical.[7] Suppose a voter prefers party A to party B and party B to party C. Define u_A, u_B and u_C as the utilities the voter gains if A, B or C, respectively, is elected and without loss of generality let $u_C = 0$. Let x_i be the total number of votes for party i cast by all the other voters in the constituency. A voter only affects the result of the election when they are pivotal, i.e. when the votes of all the other voters result in a tie or a near tie for the lead between two or more parties. Table 1 gives the utility from a vote for party B, a vote for party A and the difference between the two in each pivotal situation. Ties and near ties between all three parties have been left out because the probability of such an event is virtually zero compared with that for a tie or near tie between two parties.

Since rational choice theory shows that it is irrational to vote because the costs of voting easily outweigh the potential influence of the vote on the outcome, for present purposes it is necessary to assume that people do vote. Given this assumption, the hypothetical voter making a short-term instrumentally rational calculation will vote tactically for party B if and only if the expected utility gain from voting for B rather than A is greater than zero.[8] This is equivalent to saying that the voter will vote tactically if and only if,

[6] R.G. Niemi, G. Whitten and M. N. Franklin, 'People Who Live in Glass Houses: A Response to Evans and Heath's Critique of our Note on Tactical Voting', *British Journal of Political Science*, 23 (1993), 549–63. Also on this point see S. Bowler, 'Reasoning Voters, Voter Behaviour and Institutions: The Decision Dependence of Voting Behaviour', in D. M. Farrell, D. Broughton, D. Denver and J. Fisher, eds, *British Elections and Parties Yearbook 1996* (London: Frank Cass, 1996).

[7] R. McKelvey and P. Ordeshook, 'A General Theory of the Calculus of Voting', in J. F. Herndon and J. L. Bernd, eds, *Mathematical Applications in Political Science, vol.6* (Charlottesville: University Press of Virginia, 1972).

[8] Whether or not it is sensible to assume that people vote and are short-term instrumentally rational at the same time is a separate question. The assumption is simply necessary to elucidate the nature of tactical voting under rational choice theory.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

$$\begin{aligned} &2(u_B - u_A).P(x_A = x_B > x_C) - u_A.P(x_A = x_C > x_B) + u_B.P(x_B = x_C > x_A) \\ &\quad + (u_B - u_A)[P(x_A = x_B + 1 > x_C) + P(x_B = x_A + 1 > x_C)] \\ &\quad - u_A.P(x_C = x_A + 1 > x_B) + u_B.P(x_C = x_B + 1 > x_A) \ > 0, \end{aligned} \tag{1}$$

where $P(\ )$ is the voter's perceived probability of each event. Now it is reasonable to assume that $P(x_i = x_j > x_k) = P(x_i = x_j + 1 > x_k) = P(x_j = x_i + 1 > x_k)$, for all i, j and k. Using this assumption, Condition (1) can be reduced to

$$\frac{u_B}{u_A} > \frac{[2P(x_A = x_B > x_C) + P(x_A = x_C > x_B)]}{[2P(x_A = x_B > x_C) + P(x_B = x_C > x_A)]}, \tag{2}$$

which is essentially the same voting rule as in McKelvey and Ordershook, Hoffman, Palfrey, Cox, Myatt and other formal models.[9]

Since $u_A > u_B$ the left-hand side of Condition (2) is less than 1. The right-hand side is less than 1 if and only if $P(x_B = x_C > x_A)$ is greater than $P(x_A = x_C > x_B)$. However, it will only be optimal to vote strategically for B if $P(x_B = x_C > x_A)$ is sufficiently greater than $P(x_A = x_C > x_B)$ relative to the ratio of u_A and u_B. But since the left-hand side may be arbitrarily close to 1, tactical voting is possible whenever the right-hand side is less than 1. Further information on the utilities and the expectations of the voter is necessary to ascertain precisely when it is optimal to vote tactically. Authors such as Palfrey, Cox and Myatt describe models for the distribution of utilities and the expectations of the voters, which generate predictions for the pattern of tactical voting. However, these models require strong assumptions and if one is concerned with the definition of tactical voting it is undesirable to make strong assumptions regarding either the utilities or beliefs of voters. If one is ever to test these models against each other we need some definition of tactical voting which is neutral to all models. However, some clarification of the voting rule is possible with very weak assumptions. To this extent, two claims can be made about tactical voting in three-party competition. First, only if the second choice party is expected to win more votes than the first preference can it be optimal to vote tactically. Secondly, it can be optimal to vote tactically even when the most preferred party is expected to come second.

Formally, the first claim is that Condition (2) is only satisfied if $P(x_B > x_A)$ is greater than $P(x_A > x_B)$. This is not logically true, but true under weak and reasonable assumptions about voter expectations. As observed above, Condition (2) can only be satisfied if $P(x_B = x_C > x_A)$ is greater than $P(x_A = x_C > x_B)$. If $P(x_B = x_C > x_A)$ is greater than $P(x_A = x_C > x_B)$, the expected chances of a tie for the lead with A coming third are greater than the expected chances of a tie for the lead with B coming third. This suggests that the expected order of parties is one in which A does worse than B, since A is expected to come third to a tie more often than B. Although it does not logically follow, $P(x_B = x_C > x_A) > P(x_A = x_C > x_B)$ does imply $P(x_B > x_A) > P(x_A > x_B)$ for a range of probability distributions for compositional data. More precisely the result holds for any distribution where the probability of an outcome is a monotone function of the distance (using any metric symmetric in $\{x_A, x_B, x_C\}$) from a single point in the outcome space. The single point represents the outcome the voter thinks is most probable. Other possible outcomes are less likely the further they are away from this point. The Dirichlet and Gaussian kernels have these properties[10] and are therefore natural and commonly used distributions. Even the weak assumptions here are not all required for the result to hold. For example, if symmetry is violated because, say, votes lost by one party are not equally likely to go to each of the other two parties, it still true that $P(x_B > x_A) > P(x_A > x_B)$ whenever $P(x_B = x_C > x_A) > P(x_A = x_C > x_B)$.

[9] McKelvey and Ordeshook, 'A General Theory of the Calculus of Voting'; T. R. Palfrey, 'A Mathematical Proof of Duverger's Law', in P. C. Ordeshook, ed., *Models of Strategic Choice in Politics* (Ann Arbor: University of Michigan Press, 1989); D. T. Hoffman, 'A Model for Strategic Voting', *SIAM Journal of Applied Mathematics*, 42 (1982), 751–61; Cox, *Making Votes Count*; D. P. Myatt, 'A New Theory of Strategic Voting', Nuffield College Economics Discussion Paper 1999-W21.

[10] D. P. Myatt, 'Multi-Party Tactical Voting Incentives', Nuffield College Economics Discussion Paper 2000-W19.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

Approaching the question from the opposite direction, it is quite difficult to find any example distribution with a substantive interpretation where $P(x_B = x_C > x_A)$ is greater than $P(x_A = x_C > x_B)$ but $P(x_B > x_A)$ is not greater than $P(x_A > x_B)$.[11] This difficulty is not surprising since it would be odd if the first claim were not true, and it could be optimal to vote tactically for the second choice B even when the first choice A was expected to do better than B.

The role of the first claim is to establish a necessary condition for tactical voting. But it is not particularly limiting to require that the second choice be expected to do better than the first choice. Whether or not Condition (2) is actually fulfilled remains contingent on the particular values of the terms in the expression. Knowledge of the party preference ordering and the expected order of finish alone is not sufficient to determine whether someone should vote tactically.

To establish the second claim it is sufficient to show that if $P(x_B > x_A > x_C)$ is greater than the perceived probability of any other strict ordering of parties, then it is possible that $P(x_B = x_C > x_A)$ is greater than $P(x_A = x_C > x_B)$. Suppose that $P(x_B = x_C > x_A)$ is less than $P(x_A = x_C > x_B)$. By the argument for the first claim, A is seen as likely to win more votes than B. So it would therefore be strange to expect $x_B > x_A > x_C$ to be the most likely order of parties. Rather, $P(x_B = x_C > x_A)$ being greater than $P(x_A = x_C > x_B)$ is much more consistent with the idea of $x_B > x_A > x_C$ as the expected outcome. So if the second choice B is expected first, and the favourite A is expected second, then it is not only possible that $P(x_B = x_C > x_A)$ is greater than $P(x_A = x_C > x_B)$ but highly likely that this is so. Another argument for the same proposition can be based on the assumption that people think that their expectations are more likely to be slightly wrong than very wrong. If you knew you had made a mistake you would think it more likely that you had mistaken first and second place and guessed third place correctly, or mistaken second and third place with first place correct, than you mistook first and third place with second place correctly identified.

The significance of the second claim is that it can be utility maximizing to vote for one's second preference party even when the first preference party is expected to come second. The extent of such tactical voting is likely to be very low. If party B is expected to come first and party A second, the probability of being pivotal between A and B will be far greater than the probability of being pivotal between B and C or between A and C. Hence Condition (2) will only be satisfied if the voter is nearly indifferent between parties A and B. Nonetheless, this result is important because it shows that second-placed parties could suffer from strategic desertion. The term 'Duvergerian' tactical voting can be used to describe tactical voting when the preferred party is expected to come third or lower. This is in the spirit of Duverger's law, which suggests that third or lower placed parties are deserted by voters in the single-ballot simple-plurality electoral system.[12] Correspondingly, 'non-Duvergerian' tactical voting refers to tactical voting when the preferred party is expected to come second in the constituency.

The possibility of non-Duvergerian tactical voting means tactical voting is not simply motivated by the desire to 'avoid a wasted vote', as suggested by Franklin *et al.* amongst others.[13] Moreover, deserting a third or lower placed party is not necessarily optimal. Strictly speaking, every vote is wasted (or at least expected to be wasted) because voters should never expect there to be a tie. But there remains a question as to whether the wasted vote thesis still holds as it is actually thought of. The common interpretation is that people should not vote for parties they think are likely to come third or lower in their constituency. Returning to the example in Table 1 with the voter that prefers A to B and B to C, suppose that the preferred party A is expected to come last with parties B and C ahead in any order. It is clear that Condition (2) is not binding in either direction. Depending on

[11] One possibility is that a voter may be extremely uncertain what share of the vote A will get (so that any share between 0 per cent and 100 per cent is equally likely) but very sure that B and C will get almost the same share of the vote. Whilst not impossible, this would be an odd set of beliefs and there is no election that this author knows of where this could have been expected.

[12] M. Duverger, *Political Parties* (London: Methuen, 1954).

[13] M. N. Franklin, R. G. Niemi and G. Whitten, 'The Two Faces of Tactical Voting', *British Journal of Political Science*, 24 (1994), 126–36, p. 549.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

the actual assessed probabilities it may or may not be optimal for the voter to vote tactically when they expect their preferred party to come third. Thus, avoiding voting for third parties is not always utility maximizing.

It is not just the wasted vote characterization of tactical voting that falls foul of careful scrutiny. In fact, most attempts to sum up the driving force behind tactical incentives in common parlance are inadequate. Tactical voting is not necessarily targeted at defeating the most likely winner as described by the *Collins English Dictionary*.[14] In fact it may be perfectly consistent to vote tactically for the likely winner. Tactical voting is not characterized solely by possessing tactical motivations as Niemi *et al.* suggest.[15] Evans and Heath are wrong to describe tactical voting as, 'any case where the voter is aiming to reduce the chances of a disliked party'.[16] Whilst tactical votes should have this effect it is not always utility maximizing to reduce the chances of a disliked party. Likewise, although tactical votes should always go to a party perceived as having a greater chance of winning, the fact that the second preference has a better chance than the favourite is not a sufficient reason to vote tactically as Blais and Nadeau suggest it is.[17] With the exception of a few misleading cases, the problem with the commonly quoted motivations for tactical voting is not that they are wrong, but they are insufficiently broad on their own to account for all situations where tactical incentives can exist.

Tactical voters should be identified by having short-term instrumental motivations, voting for a party other than their preferred party and by utility maximization given preferences for parties and expectations of the constituency result. But this is not a meaningful working definition. I therefore propose the following:

A tactical voter is someone who votes for a party they believe is more likely to win than their preferred party, to best influence who wins in the constituency.

This definition covers all three criteria for a tactical vote. The short-term instrumental motivations are described by the phrase, 'to best influence who wins in the constituency'. The vote is explicitly not for the preferred party and the utility maximization principle is expressed in the requirement to vote for a party that is believed to be more likely to win. Whilst this definition may still seem ambiguous it has the virtue of generality. There are other ways of expressing short-term instrumental motivations, but they are individually too specific. For example, avoiding a wasted vote, reducing the chances of a disliked party or wanting to vote for a party with a better chance, are all examples of people wanting to cast their votes effectively, but none of them are comprehensive. Similarly, it is difficult to express the implications of utility maximization more precisely. For instance, to say that a tactical voter wants, 'to vote effectively by breaking a tie', may reflect the focus on pivotal probabilities in the voting rule, but it wrongly suggests that tactical voting requires a tie, or that tactical voters believe a tie is relatively likely. Since Condition (2) depends on a ratio of pivotal probabilities, it can be optimal to vote tactically even though the absolute probability of a tie is extremely small. So the utility maximization principle does not imply that tactical voters think that a tie is at all likely.

The definition of tactical voting developed here is designed to express the insights from rational choice theory in a language that can be commonly understood. This means translating (at the cost of precision) the complex probability terms in Condition (2) into criteria that can be used to identify real-world instances of tactical voting. To the extent that it draws on formal theory, it cannot easily be adapted for different electoral systems. Strictly speaking, it is not even applicable to four or more

[14] *Collins English Dictionary* (Glasgow: Harper Collins, 1998).

[15] Niemi *et al.* 'Constituency Characteristics, Individual Characteristics and Tactical Voting in the 1987 British General Election', pp. 229–30.

[16] A. Heath and G. Evans, 'Tactical Voting: Concepts, Measurement and Findings', *British Journal of Political Science*, 24 (1994), 557–61, p. 558.

[17] A. Blais and R. Nadeau, 'Measuring Strategic Voting: A Two-Step Procedure', *Electoral Studies*, 15 (1996), 39–52.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

party competition within the simple plurality system. The idea that it is only ever optimal to vote for a lower preference when that party is believed to have a better chance of winning, has only been defended for the three-party case. It is probably true in general, but difficult to prove. Four-party competition raises the question as to which party to vote for tactically, the second or the third choice. There probably are circumstances where a vote for the third choice is optimal, for example if that party is much preferred to and best placed to defeat the most disliked party. However, it is unlikely that it is ever optimal to vote for the third choice when the second choice is believed to have a better chance of winning. So a more exact definition of tactical voting may read: 'a tactical voter is someone who votes for a party they believe is more likely to win than *each of one or more* preferred parties, in order to best influence who wins in the constituency'. However, this condition has not been formally demonstrated.

The translation of the definition of tactical voting to other electoral systems is more complicated. It should still be true that tactical voters are short-term instrumentally rational and do not vote sincerely. For systems with ordinal voting, such as single transferable voting (STV), alternative voting (AV) and supplementary voting (SV), the 'vote' is an ordering of parties, and people have a chance to misrepresent their preference ordering in many ways.[18] So tactical voting should be conceived of as any misrepresentation of party preference order to best influence the outcome. In mixed systems, it is often tempting to see split-ticket voting as tactical, but it can simply reflect the different preference orderings voters may have over parties and candidates in different parts of the system. When the electoral system gives people the opportunity to represent complex preferences, tactical voting occurs when people misrepresent those complex preferences to best influence who is elected.

The utility maximization calculation for alternative electoral systems is again in terms of pivotal probabilities, because it is still true that the voter only affects the outcome when they are pivotal. For the most part, the pivotal situation involves the allocation of the last seat in a multimember district, however large. Even when the system is relatively simple, such as a list or quota proportional representation (PR) system, it is unclear whether anything can be said about the competition for the last seat allocated from the expected order of the vote shares. The relative chances of different pairs of parties being involved in a tie for the final seat may bear little relationship to the overall party ordering. So it may not be possible to say anything at all about when tactical voting is optimal solely in terms of the expected order of finish. Whether it is helpful to describe rules with respect to the competition for the final seat depends on whether people have reasonable knowledge or beliefs about this aspect of the election. Most of the time the answer is probably no, but there may be some cases where the number of seats allocated to each party in a district is fairly stable over time and the characteristics of the competition for the last seat are well known. This is especially likely when there is a minor party which is in contention for the single final seat. Nonetheless, tactical voting outside simple plurality electoral systems does not seem to be subject to any non-trivial necessary conditions that are straightforward or easily understood. The only fully general definition of tactical voting is misrepresenting preferences to best influence who wins.

MEASUREMENT

If one accepts the definition of tactical voting in the simple plurality system, there are implications for the measurement of tactical voting. A good indicator of tactical voting should identify as tactical voters those who deserted their preferred party for another in order to influence who won in their constituency. Thus identification is ideally on the basis of information regarding party preferences, vote choice and the motivations for vote choice, not on the strategic situation in the constituency. The possibility of non-Duvergerian tactical voting does not seem to have been seriously considered

[18] J. G. Neal, 'A Sophisticated Voter Model of Preferential Electoral Systems', in S. Bowler and B. Grofman, eds, *Elections in Australia, Ireland, and Malta under the Single Transferable Vote* (Ann Arbor: University of Michigan Press, 2000), pp. 59–78.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

by previous authors on tactical voting in Britain, despite some definitions that make allowance for it. So this section will largely take for granted that it is only tactical voting away from parties expected to come third or lower in the constituency that is being measured. Nonetheless, the definition developed above provides a basis for judgement on the criterion validity of existing measures of tactical voting. Beyond this, different measures are evaluated on their statistical properties and according to more general methodological considerations. Furthermore, direct comparison between measures based on the BES data is used to understand their similarities and differences. This leads to the development of a more robust measure of tactical voting that is a refinement of the Heath *et al.* measure.

Inference of tactical voting from aggregate data suffers not only from the classic ecological inference problem[19] but also rests on assumptions of party preference structure and motivations for vote choice. Whether it is tactical voting at a single election or new tactical voting between elections, inference is solely on the basis of vote choice. Since no information on party preferences or motivations for the vote choice are available in aggregate data, it would seem that there is only justification for aggregate data studies in the absence of good quality individual-level data. Even then, there are still substantial technical difficulties involved in the analysis of tactical voting with aggregate data.[20]

Another approach is to infer tactical voting from models of party preference. For example, Blais and Nadeau build a series of logistic regression models of vote choice to identify the preference orderings of voters.[21] They define tactical voters to be those who voted for their predicted second preference when they believed it had a better chance of winning in the constituency than their preferred party. So the measure employs information on vote choice and beliefs about the expected order of finish, but there is no information about motivations for vote choice and the party preferences have to be inferred. The inference of tactical voting is essentially from the residuals of a regression analysis, which suggests that there is actually an omitted variable bias.[22] A more significant problem is that a particular interpretation has been given for one set of respondents for whom the predicted vote choice differed from the actual vote. Not only does this exclude the possibility of random error but it begs the question as to how we should interpret the behaviour of others who did not fit the model predictions. The lack of information on instrumental motivations is clearly a problem.

Alvarez and Nagler have a similar approach but seek to avoid the problem of imputing a substantive interpretation to all respondents who do not fit the model.[23] Rather than using actual votes, tactical voting is estimated by comparing model predictions both with and without variables designed to capture the effects of the strategic situation ($W1$ and $W2$). This method infers not only the preferences and motivations of voters, but also the vote itself. So the measure is an aggregate one and does not seek to identify tactical voters within the sample of survey respondents. Not only is the measure therefore of limited use, but the implementation is poor for various reasons. First, $W1$ is supposed to measure how far a given party is behind the leader in the constituency, but it takes the same numerical value whether the party is ahead of or behind the leader of the other two by some amount. Secondly, $W2$ does not measure the closeness of the race between the leading contenders, rather it measures the similarity in the share of the vote for the two other contenders, whatever their position. The use of the interaction term between $W1$ and $W2$ in the model does not solve these two problems. Thirdly, Scottish and Welsh constituencies seem to be modelled as if the nationalist parties did not exist. Only Conservative, Labour or Alliance voters are included. This goes some

[19] W. S. Robinson, 'Ecological Correlation and Behaviour of Individuals', *American Sociological Review*, 15 (1950), 351–7.

[20] Fisher, *Tactical Voting in England 1987–1997*, chap. 7.

[21] Blais and Nadeau, 'Measuring Strategic Voting'.

[22] G. King, 'How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science', *American Journal of Political Science*, 30 (1986), 666–87, pp. 667–9.

[23] R. M. Alvarez and J. Nagler, 'A New Approach for Modelling Strategic Voting in Multiparty Elections', *British Journal of Political Science*, 30 (2000), 57–75.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

way to explaining why there are only 2,080 respondents in the analysis, when there were actually 3,204 people answering the vote question in the 1987 BES.[24] Much of this difference is probably due to list-wise deletion of cases with item non-response, which on such a scale could lead to serious selection bias. Fourthly, and far more worrying, is the method of inference for tactical voting.[25] The tactical votes are predicted from the model and compared to predictions from the same model assuming $W1$ and $W2$ are equal to zero. The latter are labelled as 'predicted sincere votes', but they are totally meaningless because they are predictions under logically impossible conditions. Since $W2$ is defined as the reciprocal of the absolute value of the difference in two vote shares, it cannot be zero by definition. Moreover, suppose the three major parties each had an equal share of the vote (the situation that the authors claim they are simulating), then $W2$ would not be defined.

Although it may be possible to improve the measure proposed by Alvarez and Nagler by changing the model specification, the approach is desirable only when there is no information on motivations for vote choice. Models of vote choice are not accurate enough. Even a model with an extremely good classification rate (say 90 per cent) is inadequate for estimating which respondents, or just how many, deserted their preferred party for strategic reasons. Moreover, the method relies on a theory of strategic incentives to capture all the tactical voting effect on vote choice. This makes the resultant measure useless for actually testing theories of tactical voting. To test theories we need a measure that does not presuppose the strategic incentive structure.

The Heath *et al.* measure of tactical voting is a direct measure which uses the following question from the British Election Studies:[26]

A. Which one of the reasons on this card comes *closest* to the main reason you voted for the party you chose?

1. I always vote that way
2. I thought it was the best party
3. I really preferred another party but it had no chance of winning in this constituency
4. Other (write in)
5. None of these/Don't know

Tactical voters are those who gave option 3 or volunteered other tactical motivations for their vote under option 4. Tactical voters identified by response option 3 in question A were asked a follow up question.

B. Which was the party you really preferred?

People who said they had no party of first preference or the party they voted for was the same as the party they declared as their party of first preference are considered to be inconsistent and excluded. Those who offered tactical reasons for voting under option 4 were not asked this question – a small problem that is returned to later.

The Heath *et al.* measure accords well with the definition of tactical voting. It seeks to identify people who did not vote for their first preference party to influence who won in their constituency, thereby fulfilling two of the three criteria. Since the phrase, 'had no chance of winning' is normally interpreted as meaning that the party was likely to come third or lower, the measure is primarily one of Duvergerian tactical voting. But this is as expected and not a problem for reasons discussed above. Despite its attractions, the measure has come under some criticism. Alvarez and Nagler claim that there is some systematic bias in measures based on self-declared motivations. In particular, they claim that there is a 'postelection bias in favour of finding increased levels of strategic voting the further the interview is conducted from election day.' As Evans points out, their own analysis shows

[24] Heath *et al.*, *Understanding Political Change*, Table III.2A.

[25] Alvarez and Nagler, 'A New Approach for Modelling Strategic Voting in Multiparty Elections', p. 73. Actually, the method as described in the paragraph above Table 3 on p. 73 is ambiguous. I assume that there is a mistaken reference to 'row entries' instead of 'column entries' in the third sentence.

[26] Heath *et al.*, *Understanding Political Change*, chap. 4.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

that this does not apply at all to the Heath *et al.* measure.[27] Footnote 14 in Alvarez and Nagler shows that there was essentially no measurable association between the interview time and the level of tactical voting using the Heath *et al.* measure. So even by their own standards, Alvarez and Nagler have presented no evidence to suggest that the Heath *et al.* measure suffers from any bias. Even supposing that there were an association between tactical voting and the time of interview, it may be that tactical voters are less frequently available for interview and so more likely to be interviewed later in the fieldwork period.

Niemi, Whitten and Franklin also criticize Heath *et al.* and argue that instrumental tactical voters should be identified as people whose preferred party came third or lower in their constituency, and who voted for a party which came first or second.[28] The preferred party for Niemi, Whitten and Franklin is given either as a response to question B above, when this was asked, or it is inferred from three open-ended questions about the parties.[29] In many cases this procedure did not produce a preferred party different from the party voted for, and this is a serious failing. More importantly, as Heath and Evans argue, the measure of instrumental tactical voting is inadequate because it does not allow for the possibility that people may be mistaken in their perceptions of the constituency result.[30]

Despite a good case for arguing strongly for their measure over the Niemi, Whitten and Franklin measure, Heath and Evans note that, 'disagreements about measurement do not appear to have major consequences either for our estimates of the amount of instrumental tactical voting or for our understanding about the characteristics of the instrumental tactical voters and the conditions under which such voting occurs.'[31] So could it be that the two measures are essentially the same? No, 47 per cent of those identified as tactical in the Heath *et al.* measure in 1987 did not prefer a party that came third or lower in the constituency and voted for a party which came first or second.[32] The overlap between the two measures of tactical voting is only about 50 per cent of each, so they are clearly not measuring the same phenomenon. Rather, it seems that there were many tactical voters in the Heath *et al.* measure who strategically deserted a party that actually came first or second in the constituency. Whether this is due to mistaken expectations or whether there is a problem with the Heath *et al.* measure cannot be said. The similarity between the reported levels of tactical voting for the two measures does not tell us they are similar measures.

Evans makes another comparison at the aggregate level between the Heath *et al.* measure of tactical voting and the number of people not voting for their first preference on the strength-of-feeling scores.[33] The strength-of-feeling score, or approval rating, for a party is given by the response coding to the following question:

C. Please choose a phrase from this card to say how you feel about the (Conservative Party/Labour Party/Liberal Democrats/…)?

1. Strongly in favour
2. In favour
3. Neither in favour nor against
4. Against
5. Strongly against

Evans notes that the proportion of voters who gave their best approval rating to some party other

[27] G. Evans, 'On the Validation of Measures of Strategic Motivations: A Critical Comment on Alvarez and Nagler', *British Journal of Political Science*, 32 (2002), 185–8.

[28] Franklin *et al.*, 'The Two Faces of Tactical Voting'.

[29] For details, see Niemi *et al.*, 'People Who Live in Glass Houses', pp. 550–1.

[30] Heath and Evans, 'Tactical Voting', p. 559.

[31] Heath and Evans, 'Tactical Voting', p. 561.

[32] Franklin *et al.*, 'The Two Faces of Tactical Voting', Table 2, p. 551.

[33] G. Evans, 'Tactical Voting and Labour's Prospects', in Heath, Jowell and Curtice with Taylor, eds, *Labour's Last Chance?*, pp. 65–84, at fn. 5, p. 80.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

than the one they voted for, is similar to the proportion of tactical voters under the Heath *et al.* measure. This would seem to add support to the Heath *et al.* measure, even though, as Evans argues, not voting for the favourite on the strength-of-feeling scores is not really an alternative measure of tactical voting. A measure of tactical voting based on strength-of-feeling scores can misclassify respondents in two ways. First, tactical voters who are indifferent between two parties on the strength-of-feeling scores will be classified as non-tactical. The high incidence of ties for first preference (between 15 and 22 per cent) and the strong association between tactical voting and indifference between one's top two parties[34] suggests that a strength-of-feeling score measure should produce a substantially lower estimate. Secondly, voters may vote for a party other than their preferred party for non-tactical reasons. With no prior expectations of how frequent such non-tactical non-myopic voting can be, it is impossible to say whether the level of tactical voting in the strength-of-feeling score measure should be similar to the level under the Heath *et al.* measure of tactical voting or not.

As before, simply to compare the aggregate levels of tactical voting provides no check on the validity of the Heath *et al.* measure. A better test is provided by inspection of the strength-of-feeling scores for those respondents identified as tactical voters by the Heath *et al.* measure. Ideally, one should find that tactical voters abandoned their favourite on the strength-of-feeling scores to vote for another less favoured party. If the respondent has a tied first preference for two parties on the strength-of-feeling scores, then one should be the party voted for and the other should be the party 'really preferred' as given in response to question B. In practice there are many ways in which voters can and do give responses that are not consistent with their self-assessment as a tactical voter. Moreover, the situation is considerably more complex in Scotland and Wales than it is in England. So for simplicity's sake the following discussion is restricted to voters in England.

The first possible inconsistency is that the party voted for is the clear favourite on the strength-of-feeling scores. Likewise, when a voter has a clear first preference for a party on the strength-of-feeling scores, but this is not the 'party really preferred' in answer to question B, there is reason to doubt the tactical nature of the vote. If a voter has a tied first preference for two or more parties on the strength-of-feeling scores and voted for one of them, then they are only inconsistent if they also declared their 'really preferred party' to be other than one of the parties involved in the tie for first preference. However, when a preference for a minor party is involved, the voter cannot be shown to be inconsistent because the strength-of-feeling questions cannot provide the same check.

It is also impossible to apply some of the consistency tests to respondents who volunteered 'Other' tactical reasons for their vote (via response option 4 to question A), because they were not asked which party they really preferred (question B). However, the lack of a preferred party is problematic in itself for further analysis. How to ascertain a first preference party for these tactical voters is a serious problem that has been addressed previously by Niemi *et al.*[35] Their paper refers to the 1987 BES and makes use of an open-ended question which was not repeated in the 1992 BES. For 1992 and 1997 there are two possible methods for inferring a really preferred party.

The solution recommended by Evans is to use answers to the following question:[36]

D. If the voting paper had required you to give *two* votes, in order of preference, which party would you have put as your second choice?

This question was asked of all respondents and if someone voted tactically for party X when they really preferred party Y it is not unreasonable to think that they would vote for party Y if they had a second vote. However some tactical voters said the party they voted for would be the one that they would put second, because it is their sincere second choice. Furthermore, with no knowledge of the electoral system and a preference for a minor party, some tactical voters may think to use their second vote tactically too and give their third preference party in answer to question D. So the 'second vote'

[34] Heath *et al.*, *Understanding Political Change*, Table 4.6 n.
[35] Niemi *et al.*, 'People Who Live in Glass Houses', pp. 550–1.
[36] Evans, 'Tactical Voting and Labour's Prospects', fn. 6, p. 81.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

TABLE 2 *Scrutiny of Tactical Voters Using Strength-of-Feeling Scores*

	1987	1992	1997
1. 'Really preferred' equal to vote	7.5	3.6	4.8
2. Clear favourite on strength-of-feeling scores equal to vote or otherwise not equal to 'really preferred'	21.7	17.1	13.3
3. Joint favourites on strength-of-feeling scores do not include 'really preferred'	1.1	1.1	1.6
4. Impossible to infer a 'really preferred party'	0.0	2.7	4.4
5. Remainder	69.8	75.7	75.8
Total	100	100	100
N	(189)	(205)	(237)

Notes: Tactical voters in England only weighted to the share of the vote.

question for tactical voters is ambiguous and should not be relied on to infer a 'really preferred' party.

The strength-of-feeling scores can be used instead. If someone demonstrates a clear preference on the strength-of-feeling scores for a party they did not vote for, this could be taken as their 'really preferred' party. Those who were indifferent between two or more parties as a first preference and voted for one of them could reasonably be thought to have 'really preferred' the other if they said that they were tactical. For some, the attempt to infer a preferred party leads to inconsistency. For others it is impossible to allocate a preferred party using the strength-of-feeling scores. In testing models of tactical voting there can be little choice but to exclude these tactical voters from consideration, because it is unhelpful to maintain them without knowledge of the preferred party.

Given the above discussion it seems appropriate to revise the Heath *et al.* measure of tactical voting by making consistency adjustments on the basis of strength-of-feeling scores. Simultaneously, these scores can be used to impute a preferred party for those tactical voters who did not answer question B, for whatever reason. Table 2 is designed to show the effect of this process on the measure of tactical voting. For each election the table takes all those who, in response to question A above, gave option 3 or volunteered other tactical motivations for their vote under option 4. The first row shows what proportion of these gave the party they voted for as the 'really preferred' party. Heath *et al.* already recognize these as inconsistent. The second row gives the proportion of tactical voters who had a clear preference for a party, but this was the same as the party voted for, or it was not the 'really preferred' party.[37] Row three highlights those who had a tie for first preference on the strength-of-feeling scores that did not include the 'really preferred' party. For those who did not answer question B it is often possible to infer a 'really preferred' party. Those for whom this is impossible are included in row four. Finally, row five represents the remainder of the tactical voters. This group constitutes a revision of the Heath *et al.* measure.

Under the standard Heath *et al.* measure, tactical voting in England was at 6.6 per cent in 1987, 9.8 per cent in 1992 and 10.8 per cent in 1997. Under the revised measure the corresponding figures are 5.0 per cent in 1987, 7.7 per cent in 1992 and 8.5 per cent in 1997. This constitutes a substantial reduction in the level of tactical voting. However, it could also be said that Table 2 shows that the Heath *et al.* measure stands up well to the strict tests, with at least a 70 per cent success rate.

Nevertheless, there are several benefits to be gained from using the strength-of-feeling scores, both for imputation of preferred parties and as a consistency check. Firstly, it improves confidence in some of the '4. Other' responses to question A that have been coded as tactical or similar in the BES. For instance, those who said they voted primarily against a party or candidate, without specifying that they did not vote for their preferred party, can be coded as tactical with confidence if they did not vote for their preferred party on the strength-of-feeling scores. Also, those

[37] These two criteria are not the same but since they can be fulfilled simultaneously it is sensible to combine them in Table 2.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

tactical voters who gave response option 3 in question A, but did not give a 'really preferred party', can be checked for consistency. However, the main benefits are increased confidence that all those coded as tactical are indeed tactical, and knowledge of the preferred party, which is essential to test theories of tactical voting.

In line with previous tests of construct validity, it should be established that those respondents in the Heath *et al.* measure who have been rejected for the revised measure do not behave as tactical voters. To do this it is first necessary to define a first preference party for all voters and then to look at the relationship between tactical voting and distance from contention. Previous authors have stipulated that a respondent's preferred party is the party voted for unless they voted tactically, in which case, it is the party they 'really preferred'.[38] However, a few non-tactical voters (roughly 3 per cent) did not vote for their first preference party according to the strength-of-feeling scores. These voters may have particular reasons for not voting for their preferred parties, and our analysis should respect these and account for them in the definition of the preferred party. Protest voting is an example of what Franklin *et al.* term 'expressive tactical voting', which covers 'all instances in which individuals might vote not so as to alter the winner in their constituency but instead to send some message or signal, to their party or to some other.'[39] Since the 3 per cent represents an upper limit on the extent of non-sincere expressive tactical voting, the phenomenon seems to be much less frequent than Franklin *et al.* suggest. The existence of this group is also an indication of the importance of information on motivations for vote choice in a measure of tactical voting.

Identification of a unique preferred party solely from the strength-of-feeling scores alone is impossible because of the prevalence of ties. The pragmatic solution is to define the preferred party as the party voted for, except when the vote is tactical or when the voter has a clear (unique) favourite on the strength-of-feeling scores for a party other than that voted for. The exception to this rule occurs for minor party voters who cannot be judged by the strength-of-feeling scores, which are only recorded for the three main parties. When the vote is tactical the preferred party is provided by the response to question B or is inferred from strength-of-feeling scores as described above. The second preference party can then be defined as the party with the highest strength-of-feeling score that is not the first preference party. The third preference is defined likewise. Where there is a tie for second place on the strength-of-feeling scores for non-tactical voters, the second vote question (D) can sometimes show which of the two is really the second preference. This scheme for identifying first, second and third preference parties has two important properties: the first preference party is unique and the preference ordering is consistent with the strength-of-feeling scores. This can be achieved in all but a couple of odd cases.[40]

Once the preferred party has been identified it is possible to measure the distance from contention for the voter. This is the gap between the share of the vote for the preferred party and the party that came second in the constituency. If this is positive it implies that the preferred party is in third place or lower. If the distance from contention is zero, the preferred party is the second placed party and if it is negative the preferred party is the winner. Most theories of tactical voting suggest that it will increase with distance from contention. The relationship is very strong and is well documented in the literature and therefore useful to test the construct validity of tactical voting measures.[41]

[38] For example, Evans, 'Tactical Voting and Labour's Prospects', and E. A. Fieldhouse, C. J. Pattie and R. J. Johnston, 'Tactical Voting and Party Constituency Campaigning at the 1992 General Election in England', *British Journal of Political Science*, 26 (1996), 403–39.

[39] Franklin *et al.*, 'The Two Faces of Tactical Voting', p. 552.

[40] A few respondents were found to have a negative gap between the first and second preference parties. This is possible only when a non-tactical voter gives a lower strength-of-feeling score to the party they voted for than they give to the other main two parties and when these two parties receive the same score. This is a bizarre set of responses that suggests coding error.

[41] Niemi *et al.*, 'Constituency Characteristics, Individual Characteristics and Tactical Voting in the 1987 British General Election', p. 223; G. Evans and A. Heath, 'A Tactical Error in the Analysis of Tactical Voting: A Response to Niemi, Whitten and Franklin', *British Journal of Political Science*, 23 (1993), 131–7; Evans, 'Tactical Voting and Labour's Prospects', pp. 74–5.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

TABLE 3 *Percentage Rejected from and Accepted into the Revised Measure by Distance from Contention*

Distance from contention	1987			1992			1997		
	Reject	Accept	*N*	Reject	Accept	*N*	Reject	Accept	*N*
Less than −15	1.1	0.7	(1,000)	1.4	1.4	(761)	1.3	1.2	(562)
−15 to 0	2.3	2.5	(359)	1.3	1.0	(295)	2.5	1.3	(319)
Zero	2.4	5.6	(784)	2.6	6.5	(573)	1.6	6.0	(470)
0 to 15	1.1	11.5	(305)	4.1	18.6	(151)	3.3	22.5	(214)
15 or more	1.9	18.1	(205)	2.2	33.0	(230)	1.5	28.2	(238)
Total	1.7	5.0	(2,653)	2.0	7.7	(2,010)	1.8	8.5	(1,802)

Notes: Base for 'Reject' is English voters not in the revised measure of tactical voting. Base for 'Accept' is all English voters. Cases weighted to the share of the vote.

Table 3 shows the distribution of those rejected from and accepted into the revised Heath *et al.* measure across various categories of the distance from contention calculated on contemporary election results. Note that it is only those who were excluded because of inconsistency with the strength-of-feeling scores that are displayed below, not those for whom it was impossible to infer a really preferred party. The construct validity of the revised measure is clear from the columns labelled 'Accept'. The probability of tactical voting increases with the distance from contention and there is a marked increase when the preferred party is in third place. Now, if the strength-of-feeling scores are misleading, and tactical voters have been wrongly excluded from the revised measure, then there should be a similar relationship for the 'rejects' as there is for those accepted. But the rejects actually follow a very different pattern with respect to distance from contention. In fact, there is no consistent monotonic association, so the decision to exclude these respondents from the measure of tactical voting appears to be a good one.[42]

This test of construct validity adds weight to the argument that the checks for consistency are necessary. However, the value of this test requires agreement on the theory of tactical voting. Other variables were thought to be good candidate predictors of tactical voting by Niemi, Whitten and Franklin, and also Heath and Evans, but they were not always found to be so. It is also true that some of the predictor variables are not so well founded in theory as may be supposed. For example, Niemi *et al.* argue that tactical voting should increase when the margin between the top two candidates narrows, given that the preferred party is in third place.[43] But after controlling for the distance from contention effect, Myatt predicts the opposite.[44] What is and is not 'theoretically prescribed' is sometimes contested. With no consensus on the theory, the scope for construct validity is limited. The major argument for the revised measure is therefore based on the need for internal consistency.

CONCLUSION AND SUMMARY

By rooting the definition of tactical voting in formal rational choice theory, it is possible to understand the criterion validity of tactical voting measures more thoroughly. Rational choice theory

[42] The association between the probability of being rejected and the distance from contention is never statistically significant. Also a chi-squared goodness of fit test easily shows that the rejects do not fit the observed distribution of those in the revised measure. Note that it is not possible to test whether the rejects and the remaining tactical voters were drawn from the same distribution because the observations are not independent.

[43] Niemi *et al.*, 'Constituency Characteristics, Individual Characteristics and Tactical Voting in the 1987 British General Election', p. 232.

[44] Myatt, 'A New Theory of Strategic Voting'.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

suggests that tactical voters should be identified by short-term instrumental motivations (i.e. wanting to influence who wins in the constituency) and a vote other than for the preferred party. Further characterization of tactical voting is difficult. For example, tactical voting is not simply avoiding a 'wasted vote' since it can be utility maximizing to vote tactically when the first preference party is expected to come second in the constituency. Also, it is not necessarily optimal to vote tactically when the preferred party is expected to come third or lower in the constituency. The only clear rule is that it is never optimal to vote for a party thought less likely to win than the preferred party. A tactical voter is someone who votes for a party they believe is more likely to win than their preferred party, to best influence who wins in the constituency.

The emphasis on both voting other than for the first preference and doing so for instrumental reasons should be reflected in the measurement of tactical voting. The relative silence of the definition on when it is and is not optimal to vote tactically implies that measures of tactical voting should not rely on theories of strategic incentives. Strong assumptions about the information and beliefs of voters are required for such theories. Moreover, the definition and measurement of tactical voting should be neutral between theories of strategic incentives in order to test between their predictions. So the formal concepts behind the definition of tactical voting suggest that measures of tactical voting using information on motivations and evidence of voting for a party other than the first choice should be preferred to those which infer strategic voting from constituency characteristics. Thus, tactical voting should ideally be measured directly, rather than indirectly.

By these criteria the Heath *et al.* measure of tactical voting is *a priori* the best available. Failure to use direct evidence on motivations and voting other than for the first choice to identify tactical voters can be justified when such information is not available. However, other measures are also problematic for statistical or more general methodological reasons. Within the Heath *et al.* measure, close inspection of tactical voters in England from the 1987, 1992 and 1997 BES suggests that a test of consistency using strength-of-feeling scores is necessary. Whilst over 70 per cent of tactical voters in the Heath *et al.* measure pass the consistency test, those who fail can be shown to have been justly rejected using a test of construct validity. The strength-of-feeling scores are also useful in identifying the preferred party of tactical voters who were not asked, and other voters who did not vote for their preferred party for non-tactical reasons. However the main benefit of the analysis of tactical voters using the strength-of-feeling scores is the increased confidence in the Heath *et al.* measure of tactical voting.

Electoral Systems and Proportional Tenure of Government: Renewing the Debate

JACK VOWLES*

Debates between advocates of the two major alternative forms of electoral system often proceed at cross-purposes. In parliamentary systems, governments not only wield executive power, they strongly influence the agenda of the legislature and in some cases dominate it. The core literature on electoral systems focuses almost entirely on the relations between vote shares and legislative seat

* Department of Political Studies, University of Auckland. An earlier version of this Note was delivered at the Annual Meeting of the American Political Science Association, Boston, 1999. Clare Williams assisted with the data preparation for this article. The research of which it is a part was supported by the New Zealand Election Study and by a James Cook Fellowship, administered by the Royal Society of New Zealand. For helpful comments, criticisms, advice, data and/or encouragement, thanks are due to Geoff Cupit, Dimitri Margaritis, Jeffrey Karp, Arend Lijphart and the anonymous referees for this *Journal*. All remaining errors are the author's.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

shares, leaving government formation aside.[1] Yet the main debate about democratic accountability focuses on representative government, of which a representative legislature is at most only one part.

Advocates of first-past-the-post (FPP), plurality electoral systems and a majoritarian model of government argue that their main virtue is a tendency to translate the preferences of a plurality of voters into a majority of legislators and therefore into a clear choice of government. While elections under proportional representation (PR) frequently lead to alteration in the composition of government, the precise influence of voter choices, case by case, is less certain than under FPP, and there is normally a much greater role for elite bargaining. In contrast, PR encourages a more representative legislature, which tends to be more fragmented between parties and in which there is a lower likelihood that one party alone can form a majority government. Thus advocates of proportional representation and for a proportional or consensus model defend the virtue of a more accurate translation of voter preferences into legislative seats. Governments can be less powerful and more responsive to parliamentary pressure, particularly where this is institutionalized in a strong legislative committee system.[2] Some PR advocates argue that direct voter influence on government composition is not the key issue, because under PR power tends to be more widely shared.

While this position has merit, the composition of governments can hardly be dismissed as trivial. Furthermore, the claim that there are closer links between vote choice and government composition under FPP has not been thoroughly tested.[3] Government composition under PR tends to be examined in the context of elite bargaining, and vote shares are taken, not so much as insignificant, but certainly as given.[4] According to cynics, in coalition bargaining under PR parliamentary elites take little or no account of voter preferences, seeking office regardless of policy, ideology or vote shifts.[5] Yet many such claims cite anecdotal rather than systematic evidence, and can be rebutted on the same basis.[6]

How effective are voter choices in determining the composition of government in FPP and PR systems? Are there significant differences? These are questions still relatively neglected in the literature. This Note addresses one aspect of the debate, proportional shares or tenure of government over time (PT). Here the focus has moved from the analysis of single elections as cases to the expectation that 'the amount of time during a given period that a party is in government should be proportional to the votes it receives over the same time period'. In short, we consider not proportional representation in a legislature, but instead 'proportional tenure' of government. This is a concept developed in Britain to defend FPP against arguments for PR.[7] It was later applied to eighteen

[1] Rein Taagepera and Matthew Shugart, *Seats and Votes: The Effects and Determinants of Electoral Systems* (New Haven, Conn.: Yale University Press, 1989); Gary Cox, *Making Votes Count: Strategic Coordination in the World's Electoral Systems* (Cambridge: Cambridge University Press, 1997).

[2] G. Bingham Powell, 'Constitutional Design and Citizen Electoral Control', *Journal of Theoretical Politics*, 1 (1989), 107–30; G. Bingham Powell, *Elections as Instruments of Democracy* (New Haven, Conn.: Yale University Press, 2000).

[3] But see Arend Lijphart, *Patterns of Democracy: Government Forms and Performance in 36 Countries (*New Haven, Conn.: Yale University Press, 1999), pp. 290–1; and Richard Katz, *Democracy and Elections* (Oxford: Oxford University Press, 1997), pp. 166–7.

[4] See, for example, Ian Budge and Hans Keman, *Parties and Democracy: Coalition Formation and Functioning in Twenty States* (New York: Oxford University Press, 1990); Michael Laver and Kenneth Shepsle, *Making and Breaking Governments: Cabinets and Legislatures in Parliamentary Democracies* (Cambridge: Cambridge University Press, 1996).

[5] William Downs, *Coalition Government Subnational Style* (Columbus: Ohio State University Press, 1998), p. 147; Michael Pinto-Duschinsky, 'Send the Rascals Packing: Defects of Proportional Representation and Virtues of the Westminster Model', *Representation*, 36 (1999), 117–26.

[6] David Farrell, *Electoral Systems: A Comparative Introduction* (Basingstoke, Hants.: Palgrave, 2001), p. 196; Michael Gallagher, Michael Laver and Peter Mair, *Representative Government in Modern Europe: Institutions, Parties, and Governments*, 3rd edn (Boston, Mass.: McGraw Hill, 2000), p. 341.

[7] P. J. Taylor, 'The Case for Proportional Tenure: A Defence of the British Electoral System', in Arend Lijphart and Bernard Grofman, eds, *Choosing an Electoral System: Issues and Alternatives* (New York: Praeger, 1984).

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

countries between 1945 and 1979.[8] The findings of this exercise indicated that neither FPP nor PR countries performed very well in terms of PT. Biases were towards the more dominant party of two in FPP systems, and towards smaller centre parties and away from larger ideologically extreme parties under PR. More recently, the concept has been revived, with an explicit claim once more that PT is more effective under the British system of FPP.[9] From a critical perspective, this article explores the theoretical assumptions behind PT, widens the debate and refines and partly expands the range of evidence. Using data from parliamentary systems with either FPP or relatively pure PR systems consistent over a sufficient time period, I retest the claim that FPP systems enhance PT against the null hypothesis that electoral system differences have no effect on PT. To do this, one must consider the various reasons why proportional tenure might be minimized or enhanced. The theoretical foundations of arguments about proportional tenure also need examination.

WHAT DETERMINES PROPORTIONAL TENURE?

Electoral System

The small amount of analysis of PT so far conducted focuses on the role of electoral systems and elite strategies. Taylor and Pinto-Duschinsky appear to argue that PT is more likely in FPP systems. However, their theoretical reasons for expecting PT under FPP are undeveloped. One normally expects exponents of a theory to develop it. One does not expect this task to be left to critics. However, one could start from the well-supported Duvergerian theory that FPP tends to manufacture a two-party system.[10] One could further extend that position to assume that power will alternate on a reasonably equal basis between parties that are equally favoured by voters, on average, over time. This appears to be Taylor's thinking. However, neither Taylor nor Pinto-Duschinsky make a theoretical argument to this effect. Taylor found an 'empirical relationship' in Britain between government majorities and governments' time in office,[11] and argued that this implies that elections allocate time in office to a party. The smaller the majority, the shorter the time in office. However, this empirical relationship was based on seats won, not votes cast. Taylor then slipped into the claim that this implied a relationship between votes and PT over time, quite a different matter. Pinto-Duschinsky also adopts a British-centred thesis, simply arguing that higher PT could be found in Britain than in six other countries, most with PR.[12]

To construct a theory to justify the claim that FPP generally promotes PT, one would presumably argue that FPP systems have a built-in tendency to two-party equilibrium, all other things being equal. Some sort of formal model might be elaborated, but one would also need evidence that such a tendency can be strong enough to generate such equilibrium in the real world, consistently in time and space. Formal models confirm a tendency to Duvergerian equilibrium at the level of single district competition,[13] but not at the national level over multiple districts. As Cox puts it, 'there was no formal argument to this effect in Duverger's work, nor has there been any since'.[14] Taylor's implicit model fitted the British case empirically to some degree between 1945 and 1979. Eighteen subsequent years of Conservative government then skewed the distribution of government, although Labour under Blair has since somewhat redressed the balance.

[8] P. J.Taylor and Arend Lijphart, 'Proportional Tenure vs. Proportional Representation: Introducing a New Debate', *European Journal of Political Research*, 13 (1985), 387–99.

[9] Pinto-Duschinsky, 'Send the Rascals Packing'.

[10] Maurice Duverger, *Political Parties: Their Organization and Activity in the Modern State* (New York: Wiley, 1954); Maurice Duverger, 'Duverger's Law: Forty Years Later', in Bernard Grofman and Arend Lijphart, eds, *Electoral Laws and Their Political Consequences* (New York: Agathon, 1986).

[11] Taylor, 'The Case for Proportional Tenure', p. 55.

[12] Pinto-Duschinsky, 'Send the Rascals Packing', p. 124.

[13] Roger Myerson and Robert Weber, 'A Theory of Voting Equlibria', *American Political Science Review*, 87 (1993), 102–14.

[14] Cox, *Making Votes Count*, p. 237.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

Meanwhile experience of FPP systems outside of Britain suggests that, even if they exist, tendencies to two-party equilibrium can be outweighed by the effects of social cleavages[15] and, in particular, by those that produce regional concentrations of votes for parties other than the two that are nationally most dominant. Indeed, the role of regional concentration goes further than this: a degree of regional concentration of votes is required to produce a two-party system in the first place: otherwise, the party with a national plurality of the vote would win all districts.[16]

The ideal of a FPP system in two-party equilibrium both in voter support and government tenure still underpins much advocacy of that system. If so, this is a theory that is both under-specified and under-tested. It is the responsibility of those who wish to defend the theory to elaborate it more fully. But such assumptions provide the most logical theoretical reason to expect PT over time in FPP systems where PT comes through alternation in government between two parties. Over time, such alternation will reflect vote shares, averaged over the longer term. However, if there is no long-term equilibrium, but instead either a long-term and consistent skew or an unstable party system, PT is threatened. Where there is a long-term skew, the party with the advantage is likely to monopolize government, preventing the minority party from enjoying a fair share of government tenure, unless sufficiently sustained bursts of short-term volatility can compensate for the long-term bias. In an unstable party system, PT will be unlikely, if only because significant parties may not exist for long enough.

However, majoritarians have a more fully developed argument about PT under PR. They argue that under PR tenure of government will be skewed towards small pivotal parties that are needed to form coalitions. These parties will stay longer in power over a given period than their share of votes over that period would justify. This claim constitutes the strongest argument for those using PT to attack PR but, as will be argued later, a key measurement assumption is questionable. First, however, we move on to discuss the intervening party system variables that tend to be downplayed by the exponents of PT.

Party System Variables

There is reciprocal influence between electoral systems and party systems.[17] Electoral systems far from fully determine the shape of party systems, and party system variables are therefore likely to shape the extent of PT independent of electoral systems. The association between two-party systems under FPP and multi-party systems under PR is not as close as some theory predicts, at least at the 'elective' level, where voter choices are expressed. Under PR *party system fragmentation* should influence PT of government over time, but unpredictably, because it increases uncertainty about government formation and expands the range of choices of party elites. For this reason, it is difficult to hypothesize direct effects of fragmentation on PT under PR. As party fragmentation increases, elites may be more disposed to engage in co-operative strategies, but may choose otherwise. Under FPP, meanwhile, party system fragmentation is likely to be a source of disproportionality in government tenure, because significant parties other than the dominant two competitors will receive little or no representation and therefore few opportunities to enter government.

One-party domination in vote shares can be measured both by the average vote share for the largest party and by its average distance from its major competitor. Duvergerian theory would suggest that the psychological effect of FPP electoral systems should lead to higher levels of average vote share for the two biggest parties in FPP as compared to PR systems.

[15] Rein Taagepera and Bernard Grofman, 'Rethinking Duverger's Law: Predicting the Effective Number of Parties in Plurality and PR Systems: Parties Minus Issues Equals One', *European Journal of Political Research*, 13 (1985), 341–52; Peter Ordeshook and Olga Shvetsova, 'Ethnic Heterogeneity, District Magnitude, and the Number of Parties', *American Journal of Political Science*, 38 (1994), 100–23.

[16] Jae-On Kim and Mahn-Guem Ohn, 'A Theory of Minor-Party Persistence: Election Rules, Social Cleavages, and the Number of Political Parties', *Social Forces*, 70 (1992), 575–99.

[17] Arend Lijphart, *Electoral Systems and Party Systems* (Oxford: Oxford University Press, 1994).

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

This should mean a more competitive system, at least between two parties. But if one major party dominates over all others, disproportionality in government tenure is likely to be far more potent in FPP systems. Such tendencies may be offset in PR systems by elite strategies for government formation. These form the next set of variables to consider.

Elite Strategies and Types of Government

Party systems and electoral systems shape elite strategies in the formation of governments but do not determine them. The existence of *minority governments* in some countries has three implications. First, one could be willing to tolerate greater disproportionality in government tenure if those parties over-represented in government could not dominate the legislature. If the government is weak and non-government parties are stronger, tenure of government is less valuable to incumbents. Secondly, the formation of minority governments is presumably tolerated by other parties that might otherwise form a majority. A willingness of some parties to defer to others may denote a political culture that facilitates and accommodates disproportionality of government tenure because such governments cannot achieve all they want without negotiation and, sometimes, not even then. Some parties may be too ideologically extreme to consider themselves, or be considered by others, as able to take part in government. Thirdly, a propensity to form minority governments will obviously decrease proportionality of government tenure because minorities govern and majorities do not.

A propensity to form *surplus coalitions or grand coalitions* is harder to interpret. It may, on the one hand, denote a co-operative or consensus approach to government or, on the other, a desire by elites in a highly fragmented party system to reduce the ability of a single small party to exert veto power. Indeed, the wish to avoid such veto power may both contribute to surplus coalitions and to grand coalitions between major parties. It is important to note that Taylor and Lijphart explicitly identified consociationalism as a strategy for promoting proportional tenure. This is a model of government in which, among other things, grand and surplus coalitions are featured.[18] It is obvious that grand coalitions, in particular, will tend to enhance PT as a means of promoting proportional tenure. Whether or not surplus coalitions should reduce or enhance proportional tenure is an empirical question. Elites may choose more popular surplus parties to please voters and broaden the support base of a government, leading to higher PT, or they may choose less electorally popular parties which occupy a pivotal or other strategic position in the party system, leading to lower PT.

DESIGN AND MEASUREMENT ISSUES

Because rigorous comparison requires a 'most similar' strategy,[19] the countries from which data are drawn are eighteen parliamentary democracies with populations of over 400,000, which have either a pure FPP system, or PR that meets a reasonable standard of proportionality, and have had a long enough continuous history of democratic elections to provide meaningful data. The initial selection of countries is derived from Lijphart,[20] but countries that have changed their electoral systems significantly (Greece) or have a system of PR with low proportionality (Spain) are excluded. The population cutoff is higher than that of Lijphart, but excludes very small states where there are tendencies towards a smaller number of parties that might skew analysis, as somewhat more of them use FPP.[21] Presidential systems are excluded, as in that context one expects that voters will behave differently as their vote for a legislature does not necessarily confirm or remove the executive.

[18] Arend Lijphart, *Democracy in Plural Societies: A Comparative Explanation* (New Haven, Conn.: Yale University Press, 1977).

[19] M. Dogan and D. Pelassy, *How to Compare Nations*, 2nd edn (Boston, Mass.: Chatham House, 1990), pp. 132–40.

[20] Lijphart, *Patterns of Democracy*, p. 50.

[21] Lijphart, *Patterns of Democracy*, p. 291–2.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

Taylor and Lijphart tested data from elections in eighteen countries.[22] These included some countries that will be excluded from this study on the grounds outlined above. Specifically, the cases included here are Austria 1945–98; Belgium 1946–99; Botswana 1965–99, Canada 1945–2001; Denmark 1945–98; Finland 1945–95; Germany 1949–98; India 1952–99, Israel 1949–96; Italy 1946–93; Jamaica 1949–97; Luxembourg 1945–95; Netherlands 1946–98; New Zealand 1946–95; Norway 1945–97; Sweden 1946–98; Trinidad and Tobago 1961–2001; United Kingdom 1945–2001. In each case, the last year is the final year of a government rather than the first year of a new one (for which the end date and therefore duration is as yet unknown). Data from three countries are closed off in the early to mid-1990s after significant electoral system changes – in the case of New Zealand, from FPP to PR, in Italy from PR to a semi-proportional system, and in Israel to the direct election of a prime minister, which has significantly altered the context of the parliamentary vote. Seven countries have FPP systems, the rest have PR.

The small selection of countries raises questions about how robust any conclusions can be. For reasons to be explored further below, it is assumed that at least twenty years of continuous democracy are required for a minimum time dimension. Other majoritarian, mixed or semi-proportional countries that qualified in this way were considered for inclusion, particularly among the thirty-six covered by Lijphart in 1999. However, there is little or no available theory about what one might expect for PT outside the two sets of pure cases identified here. Australia's preferential voting and France's Second Ballot system are forms of majoritarianism that nonetheless foster coalition government and greater small party participation in government. They form a category of cases in themselves, although the two systems also differ substantially in how and when further preferences are expressed. Adding such cases would weaken the ability to generalize, not strengthen it. Japan's system of single non-transferable votes (SNTV) up to 1996 is another idiosyncratic case. Malta and Ireland's systems of single transferable votes (STV) and Spain's regional list PR have relatively low district magnitudes that take them out of the camp of pure PR. Switzerland is excluded because the composition of the government is determined by voting in the two Houses of the Federal Assembly, one of which is non-proportional. More cases of FPP would be useful. The extreme regional diversity and party system instability found in Papua New Guinea has given rise to a highly fragmented party system in constant flux, with many independent members and 'party hoppers'. Given this, linking vote shares to Cabinet shares is virtually impossible except for brief snapshots. Mauritius was considered but its system of three-member 'at-large' plurality districts forms an extreme case of FPP, which seemed better excluded.

While disproportionality of votes to government tenure will be examined on a party by party basis, for systematic analysis an overall index of proportionality is required for each country case. In what follows, Gallagher's least squares index of disproportionality, normally applied to seats and votes, is applied to vote and government shares. This index is generally agreed to be technically the most precise.[23] This analysis follows Lijphart in excluding very small 'other' parties (normally 1 per cent or less) or independents from the calculations. The index measures disproportionality of government tenure: in other words, applied to vote and Cabinet shares, it is the opposite of PT.

Continuing to explore methodological issues, how do we interpret the governmental tenure of parties under coalition conditions? This issue is crucial to the evaluation of the claim noted earlier that smaller parties under PR have an excessive share of government tenure. Is the variable unweighted or binary (in government or out of government) or numeric, weighted by share of Cabinet seats? Taylor and Lijphart chose the binary or unweighted approach, which means that all parties in government have equal weights, regardless of their share of Cabinet seats. They justified their position as follows:

> Should a party's tenure be weighted by the number of cabinet posts it holds in a government? The problem is that some posts are more important than others. We have decided to treat a party's tenure

[22] Taylor and Lijphart, 'Proportional Tenure Versus Proportional Representation'.
[23] Lijphart, *Electoral Systems and Party Systems*, pp. 60–1.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

as simply the number of months it held one or more cabinet posts in a government. This may be criticised for overestimating the importance of small parties in coalition government, but we would defend our measure on the grounds that it reflects the veto power common to nearly all parties that join a government. Small parties may not often provide a Prime Minister, but if they have the power to bring the government down by withdrawing support, their influence in government will be far more than the one or two cabinet posts that they receive.[24]

While some of these points have force, on balance the position they defend is problematic. The unweighted approach means treating the long periods of office held by the German Free Democratic Party as a minority coalition partner as amounting to the same as long periods of office held as a single party by, say, the British Conservative party. It is more reasonable to assume that influence in government is roughly proportional to a party's share in a Cabinet. The veto power of withdrawal is a double-edged sword for small parties, who can have more at risk and more to lose from precipitating an election. Junior coalition parties can usually claim one or two significant Cabinet posts, but their ministers remain in a minority in a Cabinet dominated by the larger party. The longer coalitions remain in place, the more stable they become, and the less likely that junior partners will engage in brinkmanship. And as Table 4 below indicates, a significant number of governments under PR are based on surplus coalitions, and in these circumstances the ability of one coalition partner to exert veto power is more limited, so much so that Taylor and Lijphart's assertion that such power is 'common to nearly all parties that join a goverrnment' is, at the least, questionable. However, to recognize the points made in advocacy of the unweighted approach, data based on its assumptions will also be generated in most of the analysis that follows.

However, Taylor and Lijphart's calculations on this basis are a little problematic. Vote shares are computed by simply adding votes cast for parties over successive elections, and taking proportions of the total. This means that voter preferences for each election are weighted equally, and differences in the time periods between elections are not taken into account. However, Taylor and Lijphart measure parties in government by time in office, with the total time spent in office by all government parties as the base for calculations of tenure. To make proper comparison between vote shares and tenure, time should be included in both estimates. The method of calculating government tenure presents another problem. Using total time in office by parties rather than by governments has the effect of weighting coalition governments more than single-party governments, to the tune of the number of parties in the coalition. This has the effect of underestimating the tenure of single-party governments as compared to that of parties in coalitions.

The equal weighting of parties in government is estimated differently here. Tenure of government and associated vote and cabinet shares are weighted by days in government: in other words, for any one government the vote data from the previous election and its distribution of cabinet seats are weighted by government duration. A government of two years, therefore, is weighted to have twice the impact on overall average shares than a government of only one year. Votes cast for a parliament that lasted for four years are similarly weighted as twice those cast for a two-year parliament. And where government parties are weighted as having equal shares of government, calculations are made on the assumption that all governments are equal. In other words, for a three-party coalition each party is measured as having one third of the government. A weighted estimate simply substitutes the actual shares of Cabinet seats in an alternative set of calculations.

One more concern remains – over what length of time should the extent of proportional tenure of government be judged? On what grounds should one assign importance to one length of time rather than others? Again, one needs a theoretical justification that is lacking in those who promote the idea of PT. Taylor and Lijphart's period was thirty-four years, because that was the length of time for which uninterrupted data for a sufficient number of countries was available when they did their analysis. Data is now available for somewhat over fifty years. In a two-party system, consider one party in government that serves for a continuous twenty years, followed by the alternative party

[24] Taylor and Lijphart, 'Proportional Tenure Versus Proportional Representation', p. 389.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

that serves in government for the next twenty. Assume that the average vote shares cast over each period amounted to a little less than half for each 'out' party. On a twenty-year basis, the picture is one of high disproportionality of government tenure; on a forty-year basis, one would identify high proportionality.

If the ideal behind the concept of proportional tenure is the democratic elitist concern for dismissability (very obvious in Pinto-Duschinksky's Schumpterian argument), twenty years in continuous power is perhaps too long. But perhaps ten or twelve years might not be. Other than such intuition, there is no obvious answer to this measurement problem, as to solve it one would have to examine expectations of democratic accountability over time, both in normative theory and in public perceptions, another area of inquiry that is under-theorized. To proceed empirically, the problem is addressed by adopting twenty years as an appropriate period, and thus constructing overlapping twenty-year cross-sections of elections for each country for the following periods, overlapping by ten years: for example, 1945–64, 1955–75, 1965–84 and 1975–94. In other words, a twenty-year period is assumed to provide an appropriate test of proportionality of government through alternation. The periods overlap, in case two different governments each rule for twenty-year periods which exactly match periods chosen to define the cross-sections.

DATA AND DISCUSSION

First, let us address the implications of weighted and unweighted government shares in terms of examination of their implications for biases that are associated with particular parties in particular countries. We begin with data averaged for each country over the entire period of somewhat more than fifty years. The distributions of biases have been calculated for ninety-seven parties, and the extreme cases of 10 per cent or more appear in Tables 1 and 2, Table 1 on a basis weighted for Cabinet shares, and Table 2 assuming all government parties possessed equal power.[25]

Parties are ranked in order of bias from positive to negative. Shaded rows indicate which parties compete in the FPP systems. In Table 1 weighting for Cabinet shares highlights a bias towards some larger dominant parties in PR systems: Norwegian and Israeli Labour (and its allies), the Italian, German and Belgian Christian Democrats, and the Danish and Swedish Social Democrats. Negative biases under PR accrue to German Social Democrats, Italian Communists, the Finnish left and Swedish Moderates. In Table 2, not weighting for Cabinet shares propels three smaller parties from PR systems into the positive bias category: a Finnish ethnic party and two Italian centre parties join the German FDP, which had a significant advantage in Table 1 and a more substantial one in Table 2. The German FDP was significantly over-represented in government on a weighted basis but unweighted for its lower share of cabinet seats, it moves from fourteenth to fifth in positive bias. Because all parties in the three plurality systems formed single-party governments, their extent of bias is the same in both tables. The same applies to the Norwegian Labour party, which has always formed single-party governments.

The data in Tables 1 and 2 give little indication of systematic differences in biases between FPP and PR systems. However, large parties from the FPP systems take significant places in the rankings in both tables. Comparison of the two sets of rankings suggests that much of the apparent bias in

[25] Thomas Mackie and Richard Rose, *The International Almanac of Electoral History*, 3rd edn (Basingstoke, Hants.: Macmillan, 1991); Thomas Mackie and Richard Rose, *A Decade of Election Results: Updating the International Almanac* (Glasgow: Centre for the Study of Public Policy, University of Strathclyde, 1997); Dieter Nohlen, Bernard Thibaut and Michael Krennerich, eds, *Elections in Africa* (Oxford: Oxford University Press, 1999). I am grateful to the Electoral Commissions of Jamaica, of Trinidad and Tobago and of Botswana for sending me the necessary data for their countries, and to Arend Lijphart for letting me share his unpublished data on India (see also Lijphart, *Electoral Systems and Party Systems*, pp. 169–72). The most comprehensive source for Cabinet data is Jaap Woldendorp, Hans Keman and Ian Budge, eds, *Party Government in 48 Democracies* (Dordrecht: Kluwer, 2000). Where necessary, the most recent Cabinet data was found in *Keesing's Record of World Events* (London: Longman, 1987–). Also see the *European Journal of Political Research's* annual *Political Data Handbook* for other recent data.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

TABLE 1 *PT Bias on a Weighted Basis, by Political Party, 10 Per Cent and Above*

		Vote share	PT share	Bias
India	Congress	41.9	84.0	42.1
Norway	Labour	41.9	76.6	34.7
Canada	Liberal	40.4	72.1	31.7
Botswana	BDP	69.6	100.0	30.4
Italy	Christian	38.9	68.5	29.6
Sweden	Social Dem.	45.0	74.6	29.6
Trinidad	PNM	52.2	76.5	24.3
New Zealand	National	44.7	68.4	23.7
UK	Conservative	42.4	62.2	19.8
Denmark	Social Dem.	35.9	53.2	17.3
Belgium	Christian	34.7	51.1	16.4
Austria	Socialist	44.5	57.5	13.0
Israel	Labour	39.9	51.7	11.8
Germany	CDU/CSU	44.8	55.0	10.2
Germany	FDP	9.1	19.3	10.2
Finland	Centre	21.0	31.2	10.2
Finland	SKDL (Left)	16.2	5.8	− 10.4
Trinidad	DLP	10.8	0.0	− 10.8
New Zealand	Labour	43.6	31.6	− 12.0
Sweden	Moderate	17.2	4.8	− 12.4
UK	Lib/SDP/All/LD	12.7	0.0	− 12.7
Germany	SPD	36.9	22.4	− 14.5
Canada	NDP	14.9	0.0	− 14.9
Botswana	BNF	17.7	0.0	− 17.7
Italy	Communist	25.6	0.4	− 25.2

proportional tenure towards smaller parties of the centre reported by Taylor and Lijphart is probably an artefact of constructing PT on a basis unweighted by shares of Cabinet seats. Even so, there remains a small benefit in bias towards some larger parties in PT even on this basis, and even in PR systems. It appears that some PR systems may manufacture government tenure for larger parties over time, even if they do not do so significantly election by election.

The systematic effects of bias in government tenure for broad categories of parties can also be briefly examined over the period since about 1945 to approximately 1995–2001.[26] Parties of the extremes and the margins able to flourish in parliaments under PR suffer most from bias against them in government tenure. Christian and Liberal parties do consistently better (see Table 3). In the FPP countries, Conservatives gain from a positive bias and Social Democrats lose. Liberals appear to gain slightly, but only because of the strong bias towards the Canadian Liberal party. Under PR Conservative parties are mainly found in Scandinavia, and have narrower support bases than in the FPP countries. They gain less tenure, and Social Democrats more. Christian Democratic parties, found only in the PR countries, are more broadly-based, and have a positive bias.

In order to indicate more clearly the ideological and potential policy implications of bias in government tenure, some of the parties represented are also aggregates of relatively close alliances of parties – for example, Belgian Catholic parties are put together despite their ethnic division in

[26] Definitions are taken from J. E. Lane, D. McKay and K. Newton, *Political Data Handbook, OECD Countries*, 1st edn (Oxford: Oxford University Press, 1991).

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

TABLE 2 *PT Bias, Unweighted Basis, by Party, 10 Per Cent and Above*

		Vote share	PT share	Bias
India	Congress	41.9	83.9	42.0
Norway	Labour	41.9	76.6	34.7
Canada	Liberal	40.4	72.1	31.7
Botswana	BDP	69.6	100.0	30.4
Germany	FDP	9.1	38.8	29.7
Trinidad	PNM	52.2	76.5	24.3
New Zealand	National	44.7	68.4	23.7
UK	Conservative	42.4	62.2	19.8
Sweden	Social Dem.	45.0	64.0	19.0
Italy	Social Dem.	4.1	17.3	13.2
Finland	SWPP	6.1	19.0	12.9
Italy	Republican	2.4	15.1	12.7
Belgium	Christian	34.7	45.9	11.2
Finland	KOK	17.5	7.2	− 10.3
Finland	SKDL (Left)	16.2	5.4	− 10.8
Trinidad	DLP	10.8	0.0	− 10.8
New Zealand	Labour	43.6	31.6	− 12.0
Norway	Conservative	20.8	8.4	− 12.4
UK	Lib/SDP/All/LD	12.7	0.0	− 12.7
Netherlands	Labour	28.8	15.9	− 12.9
Sweden	Moderate	17.2	3.0	− 14.2
Canada	NDP	14.9	0.0	− 14.9
Botswana	BNF	17.7	0.0	− 17.7
Israel	Likud/Right	28.3	10.5	− 17.8
Germany	SPD	36.9	15.3	− 21.6
Italy	Communist	25.6	0.4	− 25.2

the late 1960s. In some cases, like that of Israel's Likud and Labour, ideologically aligned parties that coalesced during the period are counted as one unit throughout.

Reporting data by party gives some idea of possible country-wide biasing tendencies, but is partial. Further, within countries some parties' positive biases are the other side of the coin of others'

TABLE 3 *Biases by Parties in Ideological Groupings and Electoral Systems*

Type of party	Votes	Weighted bias in government	Unweighted bias in government
Protest	2.5	− 1.8	− 0.7
Ethnic	3.7	− 1.0	2.1
Christian	17.5	4.9	0.2
Conservative (PR)	20.1	− 7.0	− 12.0
Conservative (FPP)	41.6	13.2	13.2
Liberal (PR)	9.3	0.7	3.0
Liberal (FPP)	17.1	0.9	0.9
Social Democratic (PR)	26.7	6.6	2.1
Social Democratic (FPP)	26.9	− 8.1	− 8.1
Communist/Green/Left	6.5	− 5.6	− 4.8

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

TABLE 4 *PT, Party System Characteristics and Types of Government*

	Disproportionality of government tenure		Party system			Type of government		
Country	Weighted	Not weighted	EFFEP	Dominant party votes %	Dominant – second party %	Minority govt	Surplus coalition	Grand coalition
India	30.4	30.4	4.6	42.0	37.1	16.7	4.9	0.0
Italy	28.1	22.7	4.1	38.9	13.3	34.8	61.1	0.0
Norway	26.2	26.7	4.0	41.9	21.1	54.2	0.0	0.0
Botswana	25.6	25.6	1.9	69.6	51.9	0.0	0.0	0.0
Canada	25.4	25.4	3.1	40.4	6.2	16.4	0.0	0.0
Sweden	24.1	19.0	3.4	45.0	27.8	76.6	0.0	0.0
New Zealand	19.6	19.6	2.5	44.7	1.1	0.0	0.0	0.0
Trinidad	18.5	18.8	2.3	53.6	39.0	0.0	0.0	0.0
Luxembourg	18.4	14.4	3.4	36.6	6.6	0.0	2.4	53.6
Finland	16.3	15.5	5.6	24.4	3.4	10.4	65.2	22.7
UK	15.4	15.4	2.8	42.4	4.7	1.1	0.0	0.0
Germany	14.6	27.3	3.6	44.8	7.9	2.8	20.8	5.9
Denmark	14.0	7.8	4.7	35.9	18.2	83.9	0.0	0.0
Austria	11.1	9.0	2.6	44.5	2.5	2.8	7.3	40.6
Netherlands	9.3	12.4	5.0	28.9	0.1	0.0	48.3	46.4
Israel	8.8	21.4	4.8	35.6	20.9	1.9	69.8	30.3
Belgium	8.8	8.0	5.9	34.7	4.9	0.8	9.3	36.4
Jamaica	1.5	1.5	2.1	50.7	4.2	0.0	0.0	0.0
Average FPP	19.5	19.5	2.7	49.0	20.6	4.9	0.7	0.0
Average PR	16.3	16.7	4.3	37.4	11.5	24.4	25.8	21.4

Note: FPP countries in italics.

negative biases. A more telling indication of overall bias is on a country by country basis. Table 4 reports the index of disproportionality discussed above, setting PT over the whole period covered for each country against the average vote shares for each political party for which votes are reported. Countries are ranked from the most disproportional to the most proportional, on a weighted basis. For these estimates, the 10 per cent exclusion rule is abandoned and all significant parties are included in the calculations. Regardless of the weighting or otherwise, most FPP countries fall in the upper half of the range, reflecting the relative dominance of one party in India, Botswana and Canada. The best ranking for PT is Jamaica, with an FPP system; but the worst is India, another FPP country.

Party system fragmentation is measured by the effective number of elective parties.[27] It is worth noting that only two, or, at the most, three of the plurality systems were, on average, pure two-party systems during the period under examination. Indeed, a PR system, Austria, was somewhat closer to a pure two-party system than Britain or Canada, and only very slightly higher on the effective parties scale than New Zealand. Belgium and Finland register as the most fragmented systems, but India, with a plurality system, has a very high level of fragmentation.

The extent of one-party domination in vote shares is measured both by the average vote share for the largest party and by its average distance from its major competitor. The psychological effect of FPP electoral systems should lead to higher levels of average vote share for the biggest party in FPP than in PR systems. There is evidence for this in the data but it is weak. While the average for dominant party share is 49 per cent in the FPP systems and 37.4 in the PR systems, this is not a statistically significant difference given the number of cases. As for the difference between first and second most popular parties, the FPP countries have an average gap of nearly 20 per cent between the two major parties' average vote shares, while in the PR systems the gap is smaller at nearly 12 per cent. This goes against the expectation that party competition should be tighter under FPP, and reflects the extent of one-party dominance in some FPP systems.

Tendencies shaped by party system characteristics may be partly offset in PR systems by elite strategies. Belgium, the Netherlands and Austria have among the lowest levels of disproportionality in government tenure. Of these, only the Netherlands has a high proportion of surplus coalitions. The main elite strategy influencing PT appears to involve grand coalitions. Italy, Finland, Israel and the Netherlands also have high proportions of surplus coalitions, and Germany has some experience of them. Minority governments are highly significant in Scandinavia, although much less so in Finland. Italy also has a high rate. Canada and India are the FPP countries that have had significant minority government from time to time.

The correlations in Table 5 indicate the bivariate relationships between disproportionality of government tenure and the other variables identified. The correlations are run weighted and unweighted on a country-basis, and on a weighted basis on twenty-year cross-sections (usually four per country). The relationship between the difference of average vote shares between the dominant and second most dominant parties in the system over the period shows up consistently as the variable most closely associated with proportional tenure. The more competitive the party system between the two largest parties, the greater likelihood of PT. Grand coalitions most facilitate PT on a weighted basis. The effect of minority government on PT is not significant, but is nearly so for the twenty-year period cross-sections, and runs in the expected direction of a positive association with disproportional tenure. Most important, none of the other party system or electoral system variables correlate significantly with PT. The non-significant negative correlation with PR indicates that, if anything, PR might facilitate proportional tenure. However, the relationship is weak and thus the null hypothesis is confirmed.

Further more sophisticated analysis of this data would be problematic. First, the number of cases is small, even using the overlapping cross-sections. But in this form of the data, not all observations

[27] This is calculated for each election as 1 divided by the sum of the squares of the proportional shares of votes cast for each party, and then averaged over the relevant period, weighted by days in government and between elections.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

TABLE 5 *Correlations with Key Variables*

Variables	Disproportionality of government tenure		
	50-year cases		20-year cases (weighted)
	Weighted	Unweighted	
Minority governments	0.36 *0.15*	0.06 *0.83*	0.22 *0.08*
Grand coalitions	− 0.47 *0.05*	− 0.40 *0.10*	− 0.41 *0.00*
Surplus coalitions	− 0.17 *0.51*	0.07 *0.80*	− 0.03 *0.81*
First minus second	0.53 *0.02*	0.50 *0.03*	0.42 *0.00*
Vote for biggest party	0.24 *0.35*	0.23 *0.36*	0.17 *0.16*
Effective elective parties	− 0.10 *0.68*	− 0.04 *0.87*	− 0.05 *0.70*
New democracies*	0.10 *0.69*	0.09 *0.73*	− 0.11 *0.39*
PR	− 0.20 *0.43*	− 0.18 *0.49*	− 0.08 *0.52*
N	18	18	68

Note: The values shown are Pearson's *r* correlation coefficients, with *p* values in italics below them, indicating statistical significance (at less than 0.5).
*India, Botswana, Jamaica,Trinidad and Tobago.

are independent. There are technical problems. Vote shares are used to construct all independent party system variables as well as the dependent variables of PT, putting the same data on both sides of most conceivable regression equations. Only the system-level characteristics and type of government data are immune from this problem.

CONCLUSIONS

There is no evidence that FPP systems produce higher proportional tenure of government when comparing data from eighteen countries over fifty years. Three key findings deserve emphasis. First, unweighted methods of measuring government in previous work should be rejected. They are excessively influenced by parties in surplus coalitions, which lack the veto power that is used to justify failing to take into account their shares of Cabinet seats. Second, evaluation of PT by country should take into account the prevalence of minority government in some countries with PR, which reduces the force of the argument for PT because those governments wield reduced power and opposition parties may have significant influence on the legislative process. Thirdly, biases in PT under PR exclude from government the more extreme parties that can gain seats more easily under that system. Given this, at least since 1945 democratic elitists' fears about extremism in government under PR seem more weakly grounded than many have assumed. Social Democrats benefit from PR and are disadvantaged by FPP, while for Conservatives it is the other way about. There is a bias towards larger parties gaining greater shares of government under both electoral systems. While some smaller centre parties gain greater shares of government than votes under PR, their advantage is most noteworthy only on a basis unweighted by their shares of Cabinet seats. The closest correlate

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

with level of PT is the average gap between the first and second most popular parties. Regardless of electoral system, which has little association with it, this is the party system variable that is most closely linked with level of PT. It is this form of party system competitiveness that best fosters PT.

In terms of elite strategy, grand coalitions very obviously provide the most effective way to minimize disproportionality of government tenure, not a move towards FPP. Under an ideal model of FPP, two parties compete in relative equilibrium in government tenure and long-term voting support. But this ideal is achieved through contingency, and is not systemic. Under FPP proportional tenure of government through alternation in power between two parties may occur, but may not. Jamaica provides the best example in the data to hand. Where this does not apply, under FPP an obvious problem emerges. One party might be consistently dominant, and able to exclude all other parties from office, as in Botswana.

Bringing normative theory into play, in such a scenario it is not at all clear why a democratic elitist such as Pinto-Duschinsky would wish to apply proportional tenure as a yardstick of democratic judgement. If an opposition party never gains enough votes to form a government at a single election, why should a democratic elitist ever contend that it should gain office? By contrast, a moderate advocate of PR favouring responsible party government is in a far better position to make such an argument. A PR electoral system makes it possible to recirculate power away from a persistent plurality winner toward a majority coalition of singularly less but collectively more popular parties. Germany provides an example. Proportional tenure is a useful addition to the variables that may be used to evaluate democratic performance, but its limits must be appreciated. In particular, its implications for debates about electoral systems have been greatly over-stated.

Malapportionment and Income Inequality: A Cross-National Analysis

YUSAKU HORIUCHI*

Samuels and Snyder in their recent article in this *Journal* presented an index of malapportionment (i.e., the discrepancy between seat shares and population shares by electoral districts) in national legislatures for seventy-eight countries.[1] This short Research Note merges their index with the Gini index, a commonly used measure of income inequality,[2] and explores cross-nationally the relationship between inequality in political representation and inequality in economic conditions.

This relationship is expected to be significant and positive. The logic is simple: if there is a considerable disparity in economic conditions within a country, people gradually move from lower-income to higher-income areas (i.e., from rural to urban areas) expecting to earn higher wages. As a result, rural electoral districts with lower average incomes tend to be over-represented in a national legislature, while urban districts tend to be under-represented. In other words, the degree of political representation tends to be biased in favour of less economically advantaged citizens. In theory, such malapportionment of seats can be alleviated by regular reapportionment. In reality,

* Department of Political Science, National University of Singapore. An earlier version of this Note was presented at the annual meeting of the American Political Science Association, San Francisco, 2001. I would like to thank Stephen Ansolabehere, Benjamin Earl Goldsmith, David Samuels, Jonathan Rodden and James M. Snyder Jr for their comments on earlier drafts.

[1] David Samuels and Richard Snyder, 'The Value of a Vote: Malapportionment in Comparative Perspective', *British Journal of Political Science*, 31 (2001), 651–71. For detailed explanations about their index and alternative measures of malapportionment, see their article.

[2] It measures 'the extent to which the distribution of income (or, in some cases, consumption expenditures) among individuals or households within an economy deviates from a perfectly equal distribution' (World Bank, *World Development Report 2000* (New York: Oxford University Press, 2000), p. 69).

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

however, reapportionment decisions in most democracies are often insufficient to allow all votes to count equally. For these reasons, the positive correlation between political and economic inequalities is expected to persist in most countries. In what follows, I present empirical results, which show that the level of income inequality has a significant and positive effect on the degree of malapportionment of seats. I then discuss some avenues for future research.

DATA ANALYSIS

The dependent variable that I employ is the natural log of Samuels and Snyder's malapportionment index for lower chambers of national legislatures.[3] It should be noted that, unlike Samuels and Snyder's regression models, which used the raw values of the index, I used the logged values. Since the distribution of the malapportionment index is highly skewed to the right, it is methodologically more appropriate to transform the variable so that it fits the normality assumption better. The drawback of taking the natural log is to drop several perfectly-apportioned countries. I would argue, however, that dropping these cases is substantively not problematic, because they are special cases in that certain electoral rules (for example, a single, national at-large district in Israel) do not allow citizens' voting weights to vary at all. This means that for these five cases, the value of the dependent variable is *defined to be constant* regardless of the values of independent variables. Therefore, it is meaningless to include them in regression analyses.

The main independent variable is the natural log of the Gini index.[4] The index has been calculated by many organizations and scholars for various countries for various years. Recently, the United Nations University, World Institute for Development Economics Research (UNU/WIDER) compiled all the existing data, evaluated their reliability, and publicized them on a website.[5] I employed their database and selected data rated as 'reliable' and referring to the entire population (i.e., not the entire urban or rural populations) of the country. I also selected data based on net income or earnings (i.e., income or earnings before tax transfers), if available. This was because gross income or earnings may be partly influenced by the degree of malapportionment within a country. Namely, citizens in over-represented rural districts with higher political power in a legislative body may be taxed less than those in under-represented urban districts. If this were indeed the case, regression estimates using the Gini index based on after-tax income data would suffer from an endogenous bias problem.

To control for the effects of other variables and to examine a *ceteris paribus* effect of income inequality, I included the two independent variables that Samuels and Snyder showed had a significant effect on the degree of malapportionment. They are a dummy variable for a single-member district (SMD) system (= 1 if some or all representatives are elected from single-member districts, = 0 otherwise), and a dummy variable for Latin American countries (= 1 if a Latin American country, = 0 otherwise).[6] The other variables that Samuels and Snyder included in their regression models but found had insignificant effects are dropped. They include the size of a country (in square kilometres), dummy variables for Asia and Africa, a dummy variable for federalism, the level of democracy (measured by Freedom House's democracy score) and a dummy variable for tiered electoral systems.

I also used a measure of population movements within a country: the *change* in urban population ratio between 1980 and 1998.[7] This variable may serve as an intervening variable. (It may be influenced by income inequality within a country but it may also have an influence on the degree

[3] Samuels and Snyder also presented the degree of malapportionment in upper chambers, but as the number of samples is relatively small, I only focus on malapportionment in lower chambers.

[4] I use the logged Gini index, because it is also highly skewed. Since no country registers zero on the Gini index, no sample is dropped by this transformation.

[5] United Nations University, World Institute for Development Economics Research (UNDP/WIDER), 'UNU/WIDER-UNDP World Income Inequality Database', http://www.undp.org/poverty/initiatives/wider/wiid.htm, Version 1.0, 12 September 2000.

[6] The SMD dummy is from Samuels and Snyder, 'The Value of a Vote', Table 3. The Latin America dummy is based on the classification in World Bank, *World Development Report 2000*.

[7] World Bank, *World Development Report 2000*.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

TABLE 1 *Determinants of Malapportionment of Seats in National Legislatures*

Independent Variables	Model 1	Model 2	Model 3	Model 4
Gini index (in log)		1.396**		1.124**
		(0.358)		(0.381)
Change in urban population ratio			0.041*	0.033
			(0.019)	(0.019)
SMD dummy	0.562**	0.565**	0.694**	0.646**
	(0.207)	(0.204)	(0.209)	(0.199)
Latin America dummy	0.672**	0.156	0.723**	0.263
	(0.191)	(0.218)	(0.203)	(0.202)
Constant	1.180**	− 3.874**	0.811**	− 3.141*
	(0.144)	(1.316)	(0.180)	(1.364)
Number of samples	73	62	64	60
F-statistics	9.37	9.87	9.49	8.34
R^2	0.180	0.299	0.316	0.372
Square root of MSE	0.837	0.762	0.760	0.737

Note: The dependent variable is Samuels and Snyder's malapportionment index for lower chambers (in log). Robust standard errors are in parentheses.
**Significant at 1 per cent level (two-sided). *Significant at 5 per cent level (two-sided).

of malapportionment. As I explained earlier, people may move from rural-poor to urban-rich areas, which may result in the malapportionment of seats.) If income inequality has only an *indirect* effect through population movements and has no *direct* effect on malapportionment, then the effect of the Gini index should disappear by adding this variable.

The regression estimates are presented in Table 1. Model 1 is a simplified version of Samuels and Snyder's model. As expected, the two independent variables significantly affect the degree of malapportionment. Countries using a single-member district system and/or Latin American countries tend to have a higher degree of malapportionment, as compared to others. Model 2 adds the Gini index to Samuels and Snyder's model, and Model 3 adds the change in urban population ratio. The regression estimates in Models 2 and 3 suggest that income inequality and population movements have significant effects on malapportionment.

It is worth noting that when the Gini index is included, the positive effect of the Latin American dummy becomes statistically insignificant. This is because Latin American countries, compared to other countries, tend to have more unequal income distribution. Thus, they have a high degree of malapportionment because of these income inequalities, not because of some other unobservable factors common to Latin American countries. This result suggests that Samuels and Snyder's model, which excludes the Gini index and shows the significant effect of the Latin American dummy, suffers from omitted variable bias.

Model 4, the full model, includes both the Gini index and the change in urban population ratio. It shows that the Gini index still has a highly significant and large effect on the malapportionment index, even after controlling for the effect of population movements and other factors. The estimated coefficient suggests that a 1 per cent change in the Gini index yields a 1.124 per cent increase in the malapportionment index; that is, when the degree of income inequality doubles, the degree of inequality in representation more than doubles. This quite 'elastic' effect is statistically very significant at the 1 per cent level.

There are two possible interpretations of this finding in Model 4. First, the change in urban population ratio may not be a good indicator of population movements within a country (more specifically, between electoral districts). Other than rural-to-urban migration, there are other patterns of migrations, including rural-to-rural migration, urban-to-urban migration, urban-to-rural

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

migration, and country-to-country migration. If these other patterns of migration are substantively important and have significant implications for malapportionment, the change in urban population ratio suffers from systematic measurement error, which produces biased estimates of all effect parameters.

The second interpretation is that income disparities do indeed have direct effects on the degree of malapportionment. Population movements may still be the main cause of the *change* in malapportionment, but once the malapportionment is institutionalized, citizens in rural and over-represented areas may justify the current *level* of malapportionment based on some economic rationale. For example, politicians from economically less privileged but politically over-represented districts may claim that poor, depopulated, and rural areas deserve to have more political weight. Under this scenario, the effect of the Gini index should be significant, even after controlling for the effect of population movements.

The results reported here show that the overall degree of income inequality, measured by the Gini index, has a significantly positive effect on the overall degree of malapportionment of seats in national legislatures, measured by Samuels and Snyder's recently presented index. This suggests that we need to add further items to the list of topics for future research on malapportionment that Samuels and Snyder presented. First, to examine in detail how income inequalities and population movements affect the level and change of malapportionment, we need longitudinal data. Without longitudinal data, we cannot fully assess the validity of the two possible interpretations of the findings that were presented above. Secondly, we should perhaps pay more attention to the constitutional requirements for reapportionment in the countries included in the analysis and to the length of time since the last reapportionment in their countries. Thirdly, more use needs to be made of disaggregated data. The highly aggregated cross-national data showed correlation between the overall levels of political and economic inequalities across countries, but they do not, by themselves, indicate whether economically less-advantaged areas are over-represented or under-represented in a national legislature. To answer this question, we must compare the level of income per capita and the number of seats per capita across certain subdivisions of a country – ideally, across electoral districts.[8]

Finally, we should investigate the causal mechanisms that link income inequality and malapportionment. In this Note, I have argued that countries with wide income disparities are more likely to have malapportioned legislatures because of population movements from rural-poor areas to urban-rich areas, and because of the residual political power of rural electorates. Yet, as I briefly discussed, the degree of malapportionment may conversely affect the level of income inequality. Citizens in over-represented, poor and rural districts may use their political clout to secure policy interventions that improve their economic welfare. In such cases, malapportionment may contribute to *reduce* income inequality over time. In the analysis here, I used the before-tax Gini index, if available, to control for this potentially important endogenous effect of malapportionment on income inequality. However, this may not remove the simultaneity bias completely, because there are public policies other than tax policy that can improve the economic conditions of over-represented rural districts, such as constructing public infrastructures and providing various types of subsidies to stimulate the economy.

With cross-national cross-sectional data, an alternative method by which to control for possible simultaneity bias is to find an appropriate set of instrumental variables and to use a statistical technique, such as a two-stage least square regression. The problem with this approach is that it is difficult to find valid instruments that *only* affect either the level of income inequality or the degree of malapportionment. Probably, a better way to investigate reciprocal causality between malapportionment and income inequality thoroughly is to focus on some selected democracies and

[8] The problem with examining these variables across electoral districts is that the geographical units for elections and for censuses (or other household surveys) are usually different. At the time of writing, I could obtain data showing the average income of citizens in each district for only three countries: the United States, Canada and Japan. Readers may request data and data sources from the author.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

to examine carefully both (a) how malapportionment affects economic conditions across geographical units over time, and (b) how economic conditions affect population movements and malapportionment over time. Cross-national data of the sort used in Samuels and Snyder's article and in this Note are useful as a first-cut analysis. However, these are not sufficient to explore fully the causes and consequences of the malapportionment of seats in legislatures.

Analysing Bureau-Shaping Models: Comments on Marsh, Smith and Richards

KEITH DOWDING AND OLIVER JAMES*

In a recent article in this *Journal* Marsh, Smith and Richards (MSR) note the massive recent changes in the organization of British government and the attention the bureau-shaping model has received both at a theoretical level and as an explanation of changes.[1] They suggest that the model has breathed new life into debates about the behaviour of officials and is important in the context of the 'Next Steps' agency reform. They state two aims of their article: 'First, it is a critical contribution to the literature on the bureau-shaping model', and secondly it examines 'the model's utility as an explanation of the changes that have occurred in British central government in the past decade'.[2] They also use their arguments as part of an assault upon rational choice and empirical political science more generally in favour of interpretative sociology. However, in this Note, we respond to their work on the bureau-shaping model and rational choice.

First, we specify the bureau-shaping model a little more clearly than do MSR. From Dunleavy's original version a number of important variants have been developed.[3] To the detriment of their 'test', MSR do not distinguish between the original and the later models. Indeed, most of their evidence is actually directed at the James variant.[4] Few of their criticisms are original and the later bureau-shaping models have taken many of the points on board. Secondly, we examine MSR's critique of the bureau-shaping model as an explanation of the Next Steps reform process. We point out that whilst they provide some interesting interview evidence their findings are more ambiguous than they suggest. Thirdly, we examine the methods underlying their test that rely exclusively on elite interviews and statements by those involved in the reform. We suggest that an examination of the bureau-shaping model requires a broader set of evidence. Finally, we suggest that whilst the bureau-shaping model does not provide the final word on the Next Steps changes in Britain, it does provide some valuable insights that, in lieu of any other theorized rival, should not be dismissed lightly.

* Department of Government, London School of Economics and Political Science; and Department of Politics, University of Exeter, respectively.

[1] David Marsh, M. J. Smith and D. Richards, 'Bureaucrats, Politicians and Reform in Whitehall: Analysing the Bureau-Shaping Model', *British Journal of Political Science*, 30 (2000), 461–82 (hereafter MSR).

[2] MSR, pp. 461–2.

[3] The original model appears in Patrick Dunleavy, 'Bureaucrats, Budgets and the Growth of the State: Reconstructing an Instrumental Model', *British Journal of Political Science*, 15 (1985), 299–328; 'Explaining the Privatization Boom: Public Choice Versus Radical Approaches', *Public Administration*, 64 (1986), 13–34; 'The Architecture of the British State, Part One: Framework For Analysis', *Public Administration*, 67 (1989), 249–75; 'The Architecture of the British State, Part Two: Empirical Findings', *Public Administration*, 68 (1989), 3–28; *Bureaucracy, Democracy and Public Choice* (Hemel Hempstead, Herts.: Harvester Wheatsheaf, 1991). The variant forming the main basis for this article is Oliver James, 'Explaining the Next Steps in the Department of Social Security: The Bureau-shaping Model of Central State Organization', *Political Studies*, 43 (1995), 614–29, and Oliver James, *The Executive Agency Revolution in Whitehall: Public Interest versus Bureau-shaping Interpretations* (Basingstoke, Hants.: Palgrave/Macmillan, 2003). Another variant is Stephanie Hoopes, *Oil Privatization, Public Choice and International Forces* (London: Macmillan, 1997).

[4] James, 'Explaining Next Steps'.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

BUREAU-SHAPING MODELS

The original bureau-shaping model developed by Dunleavy was designed to broaden the claims of budget maximizing by widening the utility function of senior civil servants to include factors other than the pecuniary rewards that may (or may not!) come with larger budgets. He argued that only elements of the total departmental budget were unambiguously related to bureaucrat welfare. Indeed, he argued some elements of a larger departmental budget may be inversely related to bureaucrat welfare. Dunleavy's bureau-shaping model is a transactions costs model, where bureaucrats reduce the costs to themselves by changing the institutional structure of their department in response to a changing environment. The relationship between the four types of budget he identifies specifies the incentives for change for senior civil servants within each department. This, together with his categorization of agencies, leads to his predictions about the new shape different types of agency will take. He does not take much account of the activities of politicians in civil service change. These constitute exogenous changes in the nature of the constraints upon the bureaucrat's maximand. He has been criticized from within the rational choice paradigm for ignoring the role of politicians.[5] Other scholars using the bureau-shaping model, notably James's work on 'Next Steps' and Hoopes's on oil privatization, have not ignored the role of politicians, and more precisely tried to examine the intentions and behaviour of top officials.[6]

THE NEXT STEPS PROCESS

The second aim of MSR's article is to 'examine the model's utility as an explanation of the changes that have occurred in British central government in the past decade', in particular the Next Steps reform. They criticize the discussion of the model in Dunleavy's 1991 account.[7] However, the main aim of the 1991 book was to develop models as alternatives to Niskanen's budget-maximizing account and suggest possible applications rather than to embark on a full empirical investigation. In this context, MSR's investigation is timely. Their analysis drew on seventy-seven interviews with retired and serving senior civil servants.[8] They make several empirical claims based on this evidence, which we consider in turn.

They claim politicians were important in the reform. We agree with this point but suggest that the evidence they use in support of this statement is sparse. They claim to have found evidence that politicians had a 'crucial' role in the 'evolution and development of Next Steps', stating 'More broadly [than just Next Steps], it is widely accepted that the reforms which the public service underwent during the last Conservative administration were driven by politicians and not bureaucrats. The most senior ex-Conservative ministers' memoirs do indeed pay clear testimony to this fact'.[9] They cite three former senior politicians' autobiographies. The first, by Michael Heseltine, was published before the reform began. Whilst Heseltine argued that ministers were key in initiating many policy initiatives and suggested government needed to be more business-like, he did not present any blueprint for 'Next Steps'.[10] The second book is Margaret Thatcher's autobiography but that mentions 'Next Steps' only as a short footnote in a book of 862 pages of text.[11] The third, by Nigel Lawson, is scarcely more forthcoming: the only substantial reference to Next Steps is on two pages out of 1,036.[12] If politicians saw their role as significant, they certainly do not seem to advertise the fact in the evidential manner MSR see as meaningful.

[5] Keith Dowding, *The Civil Service* (London: Routledge, 1995), pp. 79–81.
[6] Hoopes, *Oil Privatization*; James, 'Explaining Next Steps'.
[7] MSR, p. 472.
[8] MSR, p. 464
[9] MSR, p. 470.
[10] Michael Heseltine, *Where There's a Will* (London: Hutchinson, 1987).
[11] Margaret Thatcher, *Margaret Thatcher: The Downing Street Years* (London: Harper Collins, 1993), p. 49.
[12] Nigel Lawson, *The View From Number 11: Memoirs of a Tory Radical* (London: Corgi Books, 1993), pp. 391–2.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

The real question here is to what extent the reform proposals and their implementation configured departments in the way the bureau-shaping models suggest. Since the claim of the Dunleavy model is a *comparative* one – the organizational form of the department in terms of its functional and budgetary structure will determine the precise form in which each department (or more precisely bureau) will be shaped – only a comparison of different types of departments as defined by Dunleavy can test this claim. MSR instead largely discuss the James model as it is applied to a specific department – the Department of Social Security.

In James's model, officials prefer higher budget levels to lower ones and higher levels of policy work time in total work time (and by implication lower levels of routine management work) to lower levels of policy work time. Bureaucrats are constrained in their activities by the directives of politicians. (In other words the activities of the politicians are built into the model.) Politicians' concern with 'management' in the 1980s changed the constraint by increasing the amount of management attention they demanded for executive activities. This shift 'triggered' bureau-shaping strategies by officials attempting to achieve their most preferred organizational form by hiving off routine executive activities to agencies and concentrating on policy work instead.[13] The proportion of time officials ended up spending on management is less than *it would have been* had the agency process not been introduced and had integrated departments instead remained under their direct management supervision. MSR comment, 'The weight of evidence does not support any suggestion that the reform process was bureaucratically driven. Rather it suggests that even if they wanted to, senior officials were not in a position to forward their own preferences at the expense of those of their political masters'.[14] If this statement means that bureaucrats were constrained in what they could do, then that is part of the James model. If it means that Next Steps was directed at every stage by politicians and civil servants were simply neutral cogs in a machine carrying out their masters' wishes, then we do not feel the weight of evidence demonstrates this at all.

MSR argue that much of what senior officials were involved in was 'management' rather than 'policy' work. James's model makes a distinction between policy work and *routine* management work. Following Dunleavy, policy work is valued because it 'involves innovation and often entails working in small staff units in close proximity to political power sources', by contrast management work often entails 'working in a large extended hierarchy rather than in small staff units and tends to consist of work at the point of delivery, remote from political power sources'.[15] At one point, MSR seem to suggest that this aspect of the models cannot be evaluated in a meaningful sense. Their criticism centres on the difficulty of separating 'policy' work from 'management' work.[16] Of course, not all work tasks can be neatly categorized and both the Dunleavy and James models only suggest that a substantial portion can be. Yet MSR implicitly recognize that such a distinction can be made, stating: 'To the extent that a simple dichotomy between the policy advice and the management function is possible, some civil servants certainly appear to prefer policy work, others management work and others a balance between the two.'[17] For this empirical claim to have much meaning they must accept the possibility of some form of distinction for a substantial part of officials' work. MSR split the tasks of permanent secretaries into managing the policy process, administering the department, and devising and implementing specific policies.[18] In the conclusion they suggest that, 'In our view, borne out by our interviews, PSs [permanent secretaries] manage in two senses. They administer the department and they manage the policy-making process.'[19]

[13] James, 'Explaining Next Steps', pp. 626–7.
[14] MSR, p. 471; Dowding, *Civil Service*, chap. 5; James, 'Explaining Next Steps'.
[15] Dunleavy, *Bureaucracy, Democracy and Public Choice*, p. 202; James, 'Explaining Next Steps', p. 617.
[16] MSR, p. 467.
[17] MSR, p. 468. In the James model these different preferences require drawing distinct indifference curves for each type of civil servant.
[18] MSR, p. 466.
[19] MSR, p. 480.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

The three-fold split in permanent secretaries' work does not assist the evaluation of the model. This would have been more straightforwardly achieved by directly evaluating the categories defined in the model and considering the work of all senior officials. However, several of the tasks that MSR define as management work have the characteristics of policy work by the definition in both the 1991 and 1995 models. MSR state that the administration of the department involves 'managing senior staff and resources in the department, being the department's representative/spokesperson inside and outside the Westminster/Whitehall community and acting as department accounting officer'.[20] These activities involve proximity to political power sources in the 'Whitehall village' and beyond and seem to exclude routine management of non-senior staff and resources in the department. The second feature suggested by MSR is managing the policy process, ensuring 'long-term policy strategy is being effectively managed', choosing officials to deal with policy issues, ensuring progress on policy issues and 'attempting to prevent political problems resulting from "failures" in the policy process'.[21] These tasks appear to involve innovation, working in small staff units or dealing with political power sources. Finally, MSR identify direct involvement in policy work as a task. They acknowledge that this work is consistent with 'policy' work in the bureau-shaping model.

MSR anticipate these objections by commenting: 'Of course, Dunleavy might respond by saying that all these tasks are of high status involving innovative work and close contact with the centres of power.' An alternative approach to that followed by MSR would be to test the models as they were originally set out rather than setting up a different categorization.[22] Several 'management' activities they identify fit the definition of 'policy' work offered in the bureau-shaping models. Furthermore, MSR do not find evidence of officials claiming to like working in a large extended hierarchy rather than in small staff units, and working at the point of delivery, remote from political power sources. That would constitute more of a puzzle for the models. Far from contradicting the behavioural assumptions of the bureau-shaping models, much of MSR's evidence provides empirical support.

Even to the limited extent that statements by officials preferring management to policy work refer to management as defined by the bureau-shaping model, there appears to be much conflicting evidence. MSR quotes an Efficiency Unit official who stated that some civil servants were interested 'in all aspects of management'.[23] But there is plenty of conflicting interview-based evidence. For example, Hennessy quotes the former cabinet secretary who, in the latter half of the 1980s, commented on civil servants' reluctance to get involved in management.[24] In a study of Next Steps, Zifcak noted the comments of a deputy secretary, who said: 'Management is a tiresome business, nobody goes into it unless they have to.'[25] We suggest that, on balance, elite interview-based evidence would seem to lend more support to the bureau-shaping model than MSR suggest. However, we further suggest, in the next section, that interview material is by itself not sufficient to draw clear conclusions about the model and MSR's reliance on interview material presents the danger of academic debate by exchange of anecdote.

METHODS: STATED AND REVEALED PREFERENCES

The major thrust of the Dunleavy bureau-shaping model is the relationship between four types of budgets and the shaping processes he predicts. Talking to civil servants cannot produce evidence for or against this model one way or another. If the relationships Dunleavy predicts are to be found

[20] MSR, p. 466.
[21] MSR, p. 466.
[22] MSR, p. 467.
[23] MSR, p. 468.
[24] Peter Hennessy, *Whitehall* (London: Fontana, 1990), pp. 737–8.
[25] S. Zifcak, *Administrative Reform in Whitehall and Canberra* (Buckingham: Open University Press, 1994), p. 40.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

the model is corroborated, if they are not then it is falsified.[26] However, interviewing can produce evidence for the behavioural assumptions or micro foundations of the model;[27] and provide evidence for or against the James variant of bureau shaping. The sort of exercise MSR conducted is thus useful, but the relationship between different sorts of behavioural evidence is not straightforward. Seeking the arguments in individuals' utility functions is problematic and controversial.

There are many ways in which individuals' preferences may be revealed. We may broadly divide them into two categories: stated preferences and 'revealed' preferences. The former are the statements made by people and come in many forms: anonymous answers to questionnaires, face-to-face interviews, public statements, and so on. Some of these are more public than others. And individuals do not always state the same views in private as they do in public. Most interviewers know that some information unavailable in a formal interview may be gathered once the tape recorder is switched off, and even more 'off the record' information may be gathered if one is lucky enough to buy the interviewee a few drinks later.[28] The conventional objections to relying on interview material come not just from rational choice but also from sociological perspectives.[29] Interviews are not always valid indicators of underlying motivations. An interview response may be self-serving, or reflect what an individual thought he or she was supposed to say in an interview, rather than revealing important factors for explaining what was going on. Similarly, respondents do not always give the same answers in face-to-face surveys as they do when the questionnaire is completed privately and returned anonymously. We also know that individuals are not always completely consistent when stating their preferences. This does not occur simply because people change their minds, or forget incidents and reconstruct them, nor through lack of honesty. Individuals may be contemporaneously inconsistent as shown in the well-known framing problems demonstrated by social-psychologists.[30] Furthermore, those conducting surveys have found that the order in which questions are asked can have important effects upon the answers given.[31] The same questions may elicit different sets of answers when asked in a different order. MSR seem to think that those who point out these problems with the stated-preference approach think we should not run surveys or conduct elite-level interviews.[32] Most do not, rather they believe we need to be very careful in the ways in which we interpret the results. Interviews are a good source of information about the processes that bring about an outcome provided that the material is triangulated with evidence from other sources. In particular, rational choice approaches suggest that actions involving costs to actors in terms of forgone alternatives should be given particular attention. Stated-preference evidence is low cost for the actors involved and is vulnerable to strategic misrepresentation or inaccuracy because the actors do not have much of a stake in trying to recollect what happened in

[26] It is worth repeating that corroboration and falsifiability mean increasing and lowering the *odds* that a model is true in relationship to a rival model (in the example here the null model). According to Popper no model is ever corroborated or falsified with certainty.

[27] See Peter John, Keith Dowding and Stephen Biggs, 'Residential Mobility in London: A Micro-Level Test of the Behavioural Assumptions of the Tiebout Model', *British Journal of Political Science*, 25 (1995), 379–97, for more on this technique.

[28] Of course, only the recorded or transcribed formal interview material is 'hard' data in the sense that it can be examined later by other researchers. The softer 'gossip' may be used by the researcher to search further or to confirm evidence collected by other means.

[29] D. Silverman, *Interpreting Qualititative Data: Methods for Analysing Talk, Text and Interaction* (London: Sage, 1993), pp. 13–14.

[30] Daniel Kahneman and Amos Tversky, 'Choices, Values and Frames', in Daniel Kahneman and Amos Tversky, eds, *Choices, Values and Frames* (Cambridge: Cambridge University Press, 2000), pp. 1–16; 'The Framing of Decisions and the Psychology of Choice', *Science*, 211 (1981), 453–8; 'Judgement under Uncertainty: Heuristics and Biases', in Daniel Kahneman, Paul Slovic and Amos Tversky, eds, *Judgement Under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982), pp. 3–22.

[31] See, for example, the discussion in B. Fischoff, 'Value Elicitation: Is There Anything in There?' *Amercan Psychologist*, 46 (1991), 835–47.

[32] MSR, p. 465.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

a situation. For these reasons, interview material may not be sufficient fully to reveal all the processes at work.

Economists, in particular, have tended to be sceptical about the stated-preference approach. For them, talk is cheap, and preferences revealed through action are all that can be trusted. The 'revealed preference' approach has received many criticisms over the years, though most of these involve the claim that the revealed preference approach means we can do away with 'intentions' altogether and just study behaviour, raising the question of whether or not we may examine revealed preferences with just the 'internal consistency' conditions of rational choice theory.[33] However, for the purposes of evaluating the bureau-shaping model, the conventions of much empirical political science and the approach followed by Dunleavy and James suggest that stated-preference should not be relied on exclusively and should be cross-checked with revealed-preference to check for *conflict* and possible misrepresentation. After all, how many people would continue to trust someone who constantly *stated* undying friendship but seemed to continually undermine one through their *actions*? Such a situation would require further investigation to see why he or she was behaving in such a manner, rather than simply taking the individual's statements as the end of the story.

It is important that MSR should have followed this approach because they claim to be using the bureau-shaping model and this methodological approach is intrinsic to the model. But even if they do not claim to be adhering to the methods associated with the model this point still raises a problem for their whole evaluation because it does seem that politicians and officials' accounts cannot simply be taken at face value. You do not need to be a rational choice theorist to think that the quote from an official in the Efficiency Unit about civil servants becoming more interested in management needs to be viewed in the context of this unit's mission to introduce such practices.[34] Officials in the unit were under pressure to present their activities as a success.[35] MSR's use of such a narrow form of evidence is disappointing because other data are available.

MSR characterize the evidence preferred by rational choice theorists as aggregate budget data.[36] However, the validity of budget information could be compromised by strategic misrepresentation by actors, for example to inflate the cost of delivering a programme to get more resources in future budget rounds. But, in the context of evaluating the bureau-shaping model, the data form an unobtrusive indicator in 'bureaumetric' terms.[37] Researchers do not have to interact with officials in order to have data collected specifically for the purpose, reducing the possibility of strategic misrepresentation. MSR acknowledge that 'such [aggregate] data are relevant' and appear to have no methodological or theoretical objections to using such material.[38] But they do not use this form of evidence. In contrast, a study that used a mix of interview and alternative forms of evidence found considerable support for the models.[39] There is evidence after the reform that substantial budgets and routine work tasks were passed on to agencies, that most senior officials (of Grade 5 and above in the old grading system) were located in departments rather than in agencies, and that the Efficiency Unit was sensitive to the opinions of departmental officials in drawing up plans for a reform that would meet with broad approval from senior officials.[40]

In their conclusion, MSR advance a general critique of the applicability of rational choice in this area, the problems of which the model is taken to reflect. However, this critique appears only indirectly related to the bulk of the article, which is an attempt to engage with the bureau-shaping model rather than castigate the fundamental assumptions of rational choice. They comment,

[33] See Keith Dowding, 'Revealed Preference and External Reference', *Rationality and Society*, 14 (2002), 257–82, for a defence of revealed preference and rational choice in this regard.
[34] MSR, p. 468.
[35] James, *Executive Agency Revolution in Whitehall*.
[36] MSR, p. 465.
[37] Christopher Hood and Andrew Dunsire, *Bureaumetrics* (Farnborough, Hants.: Gower, 1983).
[38] MSR, p. 465, though this seems to contradict their 'social construction' of reality position.
[39] James, *Executive Agency Revolution in Whitehall*.
[40] James, *Executive Agency Revolution in Whitehall*.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

The emphasis is on intentional explanation and preferences are assumed not explained. No role is given to structure or culture in explaining outcomes. Indeed, what is assumed, or ignored, is much of the legitimate subject matter of political science and sociology. In essence, the key problem is that rational choice theory cannot explain the dynamics of change.[41]

One of us has commented elsewhere on this and will not repeat that response here.[42] But we may note that they support this with a selective quotation from Hugh Ward, who in the same piece referred to by MSR, comments: 'I do not believe that practising rational choice theorists typically exhibit much desire to squeeze out structure; rather, they often seek to illuminate how choices are made within structures, the agenda sometimes stretching to the consideration of how rational choices reproduce or transform structures.'[43]

Their conclusions lead us to question why MSR bothered to engage with the bureau-shaping model. They very briefly sketch an alternative perspective to overcome the problems they identify, arguing: 'To do so requires a dialectical approach to the relationship between structure and agency and a recognition that the meaning individuals ascribe to structure, cultures and actions affects their behaviour.'[44] But the reader should note that asserting the greater usefulness of such an approach does not demonstrate its usefulness in answering research questions. Their preferred 'dialectical approach' can only be superior if it generates some predictions *at variance* with the bureau-shaping model which their empirical evidence then corroborates, or is able to explain *everything* the bureau-shaping model does and some more.[45] Those predictions are seriously lacking. We may suggest that even if the bureau-shaping model does not correctly explain the entire Next Steps reforms at least it offers a way of pursuing key issues of interest to academics and practitioners. MSR appear to accept at least some of these arguments in their reasons for looking at the model, which are in large part based on how influential it has been,[46] and we suggest that the model is worthy of further exploration as a way of understanding administrative reform.

[41] MSR, p. 481.

[42] Keith Dowding, 'There Must be End to Confusion: Policy Networks, Intellectual Fatigue and the Need for Political Science Methods Courses in British Universities', *Political Studies*, 49 (2001), 89–105.

[43] Hugh Ward, 'Rational Choice Theory', in David Marsh and Gerry Stoker, eds, *Theory and Methods in Political Science* (London: Macmillan, 1995), pp. 76–93, at p. 84.

[44] MSR, p. 465.

[45] We may also note that Hugh Ward has criticized the 'British Institutionalist Rational Choice' approach for being too informal and not producing enough hard predictions and empirical corroboration in 'The Fetishisation of Falsification: The Debate on Rational Choice', *New Political Economy*, 1 (1996), 283–96, pp. 292–4.

[46] MSR, p. 461.

Understanding and Explaining Civil Service Reform: A Reply to Dowding and James

DAVID MARSH, DAVID RICHARDS AND MARTIN J. SMITH*

Dowding and James's response to our critique of the bureau-shaping model makes some important criticisms that we welcome.[1] However, we suggest that they ignore the empirical evidence we

* Department of Political Science and International Relations, University of Birmingham; School of Politics and Communication Studies, University of Liverpool; and Department of Politics, University of Sheffield, respectively.

[1] D. Marsh, D. Richards and M. J. Smith, 'Bureaucrats, Politicians and Reform in Whitehall: Analysing the Bureau-shaping Model', *British Journal of Political Science*, 30 (2000), 461–82.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

presented in our original article (because of their narrow definition of acceptable data) and we believe that they misunderstand our arguments because of the positivist paradigm within which they operate. Our response covers three main issues – epistemology, methods and evidence.

EPISTEMOLOGY

Dowding and James criticize or praise us for being 'interpretative sociologists'. If labels matter, then we regard ourselves as critical realists. As such, we are interested in explanation, as are positivists like Dowding and James. However, we have a different understanding of explanation than they have, and, in particular, we have a different view of the role of theory in explanation. To a positivist, deductive theory is used to generate hypotheses that are then tested. If a hypothesis is falsified, then the theory needs to be revised to account for that observation. Our position differs in three key respects. First, in our view, not all relationships between social phenomena are observable; so, hypotheses cannot just be tested against direct observation. Secondly, to us, theory is not something that is used to generate hypotheses, rather, it is something that is used to interpret what is observable, and indeed what cannot be observed (because it is a deep structure). Thirdly, we would argue that the way in which structures (deep or otherwise) affect outcomes is mediated by the agent's understanding and interpretation of his or her structured environment. What this means is that our view of social science, and thus of doing research, differs from that of Dowding and James. Our research focused on two approaches that are usually absent in rational choice research: the use of interview data that deal with the agent's understanding of both his/her situation and the actions of others; and a historical analysis.

ON METHODOLOGY

Dowding and James are critical about the use of interview material, because it can often be 'unreliable' or 'soft' data; here, the usual criticisms are that respondents may lie, may selectively recall or may tell the interviewer what they think s/he should know. We recognize the problems of this type of data.[2] Equally, such biases can be countered or reduced by triangulation – comparing the views of respondents, interviewing a variety of politicians and civil servants, comparing with existing primary and secondary literature. At no point is our evidence based *solely* on memoirs or on interviews.

There are three points important here. First, Whitehall is a predominantly closed world and, as such, it is relatively difficult to generate other forms of data (indeed, it is the reliance on crude aggregate data that leads to the problems of the bureau-shaping model). Secondly, Dowding and James presume that there are some 'good', unflawed, data somewhere, when in fact all data have flaws. Thirdly, and most importantly in our view, most politicians and civil servants stressed the role of politicians in the reforms of the civil service. This is what they understood had happened and that understanding must surely play an important part in any fuller explanation of the changes that took place in the 1980s. It seems astounding that Dowding and James believe that actions are a better indicator of preferences than interview responses. This is a voluntarist, ahistorical and astructural approach. People often do things that they do not want to do.

Elsewhere, we emphasized that Next Steps cannot be explained independently of the historical context in which it occurred. So, we argue first that the Next Steps reforms have to be viewed against the background of the Rayner Scrutinies[3] and the Financial Management Initiative and not as an isolated change. In terms that a historical institutionalist would use, there is a considerable degree of path dependency involved here. This is something that rational choice theorists often neglect, in their desire to attain rigorous and, hence, parsimonious explanation. So, in our original article, we suggested that the political context was important. The Thatcher government's reform strategy

[2] See D. Richards and M. J. Smith, 'Interpreting the World of Political Elites: Some Methodological and Epistemological Problems' (paper presented to the PSA Annual Conference, University of Aberdeen, 2002).

[3] The Rayner Scrutinies were a set of reviews of the efficiency of government which took place in the 1980s.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

was driven by a belief that Whitehall was an institution committed to consensus politics, big government and defending the status quo. In resisting this 'conservative' force, the government was also emphasizing its image of governing competence, what Bulpitt calls its Conservative statecraft.[4] The point here is that rational choice approaches tend to downplay such historical contextual factors because they assume preferences; certainly, Dunleavy's approach appears ahistorical. The rational choice response is of course that these are external factors and it was senior civil servants' reactions to these factors that led to bureau-shaping. However, our point is that these changes occurred despite, not because of, the civil service, and in most measures of official preferences this is not the type of world that civil servants want. Dunleavy may have been right in predicting the form of some of the changes that occurred, but his explanation that it was a consequence of utility maximization by officials is *wrong*.

ON EXPLAINING CHANGE IN THE BRITISH CIVIL SERVICE

Our article refuted the core hypothesis of the bureau-shaping model in four ways:

(1) Our interviews suggest, whatever the acknowledged problems of elite interviewing, that it is difficult to categorize the sole preference of civil servants, in terms of the pursuit of policy work. Whilst some clearly prefer this type of work, others prefer management (indeed this must be an empirical question and not an assumption of the model). There is not one simple, single, preference.
(2) A clear hypothesis of the bureau-shaping model is that core budgets will be protected and programme budgets cut. The opposite in fact had occurred by the mid-1990s with the Fundamental Expenditure Review and Comprehensive Spending Review under Labour producing cuts in core budgets.
(3) Using a range of policies, government tried to make senior officials more managerial – even though James clearly states that the aim of the agency programme was to stop this occurring. In particular, the Senior Management Review reduced the number of senior civil servants (something the bureau-shaping model would not have predicted) and clearly demarcated Grade 5s upwards as managerial positions.[5] Consequently, those involved in making policy are increasingly Grades 7 and Higher Executive Officers. Under the Labour government, it also appears that the policy-making role is shifting to consultants, political advisers and special advisers. As a result officials are losing their monopoly of policy advice. The clearly stated bureau-shaping hypothesis – that the pattern of reform is a consequence of the desire of officials to retain control of policy – has been strongly undermined by the reforms that have occurred since 1997.
(4) Despite their claims to be 'empirical political scientists', Dowding and James's argument seems to be based on a counterfactual – that had the reforms not been undertaken, officials would be undertaking more management than they are now. This is non-falsifiable.

There is no doubt that rational choice has offered new ways of thinking about the motivation of agents and, more generally, the processes of reform in Whitehall. However, we believe that alternative methodologies can underline some of the simplifications that occur as a consequence of rational choice assumptions. We believe that, even in its own terms, rational choice fails to demonstrate what the preferences of officials are (it just assumes them) or to accept that its hypotheses are refuted. Moreover, its ahistoricism leads to the false assumption that reform is a consequence of the preferences of particular agents. Whatever the elegance of the bureau-shaping model, the lack of empirical support suggests that it cannot explain public sector reform.

Dowding and James claim that the 'dialectical approach' we advocate can only be 'superior' (*sic*), presumably to their own public choice accounts, 'if it generates some predictions at variance with the bureau-shaping model which their empirical evidence then corroborates'. They do not

[4] J. Bulpitt, *Territory and Power in the United Kingdom* (Manchester: Manchester University Press, 1983).
[5] Here, we use the pre-1995 Senior Management Review grading system in order to assist in understanding.

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press

acknowledge that, unlike their rather lofty view of what is achievable by political scientists, we never set out in our article to create a 'better' predictive model. Instead, our aim was to provide a critique of Dunleavy, as own research indicated his analysis was wrong, and to provide what we saw as a more convincing explanatory account of the process of Next Steps reform. As critical realists, we see our goal as trying to provide as full an explanation as possible of political phenomena. We are sceptical of positivist approaches to political science that aspire to generate predictive models. Our epistemological position leads us to conclude that this is more often than not a futile task, strewn with methodological problems. Our interest in critiquing the bureau-shaping model was never to generate a 'better' predictive model, particularly as the original model did not work in the first place!

https://doi.org/10.1017/S0007123403210395 Published online by Cambridge University Press