Left-right party ideology and government policies:
A meta-analysis

LOUIS M. IMBEAU, FRANÇOIS PÉTRY & MOKTAR LAMARI
*Centre d'analyse des politiques publiques (CAPP), Département de science politique,
Université Laval, Canada*

Abstract. This paper summarizes how the partisan influence literature assesses the relationship between the left-right party composition of government and policy outputs through a meta-analysis of 693 parameter estimates of the party-policy relationship published in 43 empirical studies. Based on a simplified 'combined tests' meta-analytic technique, we show that the average correlation between the party composition of government and policy outputs is not significantly different from zero. A mutivariate logistic regression analysis examines how support for partisan theory is affected by a subset of mediating factors that can be applied to all the estimates under review. The analysis demonstrates that there are clearly identifiable conditions under which the probability of support for partisan theory can be substantially increased. We conclude that further research is needed on institutional and socio-economic determinants of public policy.

Introduction

The question of the impact of political parties on policy has been the object of much scrutiny and debate in the past decades. The main theoretical debate concerns the relative weight of socio-economic and political variables in determining policy outputs. Advocates of the so-called 'convergence' school of thought argue that industrialised societies of the twentieth century have become increasingly similar, facing the same kind of problems and applying the same kind of solutions. Consequently, the argument goes, political, institutional and cultural differences do not matter much when it comes to explaining variations in policy outputs (Skinner 1976; Thomas 1980). In reaction to the convergence argument, policy scholars have suggested that indeed politics does matter (Castles & McKinlay 1979b). Without denying the importance of economic factors, advocates of the 'politics matters' school of thought argue that there is a correlation between partisan variables and policy outputs.

   Much empirical research on the policy impact of political parties has been published in the form of comparative analyses based on precise quantitative definitions and measurements within the well-established scientific framework of partisan theory (Hibbs 1977, 1992). One key hypothesis of

partisan theory is that there exists a law-like tendency of policy to move in response to election outcomes, operationalized in terms of the left-right party composition of government. *Ceteris paribus*, changes in the left-right party composition of government are hypothesised to be related to changes in policy. This is expressed by a general equation of the form:

Policy outputs = *f* (left-right variations in the party composition of government, variations in controlling factors).

This general equation has been applied, in one form or another, in a large number of studies comparatively to measure partisan influence on government actions in a variety of policy domains across different countries. Researchers in such fields as finance, economics, sociology and political science, have reported their findings in North American and European journals. The results obtained have varied markedly, and they are sometimes contradictory even within a single policy domain. For example Solano (1983) and Swank (1988) both look at the relationship between total government spending as a percentage of GDP and the left-right party composition of government across roughly similar sets of OECD countries. Swank finds that changes in the party composition of government are associated with variations in spending but Solano finds no party effect on public spending. Pampel & Williamson (1988) and Hicks et al. (1989) examine the impact of parties on welfare expansion in advanced liberal democracies. But their conclusions disagree. Pampel & Williamson find no significant correlation between the party composition of government and welfare expansion; Hicks, Swank & Ambuhl find that left governments generate an increase in the rate of welfare expansion while governments of the right are associated with a decrease in the rate of welfare expansion.

The quantitative literature on partisan theory displays a high level of accumulation but it has not achieved a comparable level of scientific aggregation. By accumulation we mean that we have a large literature (with publications numbering in the hundreds) which does have a readily understandable theoretical point of reference and which is of high quality. Aggregation makes the additional requirements of scientific coherence and directionality within an expanding literature (McKinlay 1996). Why is it that the 'convergence' vs. 'politics matters' controversy at the core of the literature has not been resolved after several decades of quantitative research? Why does partisan theory generate such a wide range of empirical results across OECD countries? This is puzzling because qualitative studies often support the idea that the left-right partisan divide matters for policy outputs (Klingemann et al. 1994; Stokke 1989). One possibility is that parties matter in many different

ways which are more difficult to capture in quantitative studies (with few variables and many cases) than in qualitative studies (with in-depth examination of many variables across few cases). In other words, partisan effects would be too subtle to ensure sufficient robustness of cross-sectional statistical estimates. Another possibility is that some empirical studies may have tested the partisan influence hypothesis with the wrong methodology or with the wrong variables.

Our goal in this paper is to contribute to this debate by summarizing how the partisan influence literature assesses the relationship between the left-right party composition of government and changes in public policy through a meta-analysis of empirical tests of partisan theory in OECD countries. A meta-analysis is an analysis of analyses. It is the statistical analysis of a large collection of test results coming from individual studies for the purpose of integrating the findings. In meta-analysis, 'the findings of multiple studies should be regarded as a complex data set, no more comprehensible without statistical analysis than would be hundreds of data points in one study' (Glass et al. 1981: 12). The statistical methods used to accomplish this task are similar to the ones used in primary analyses of the original data. The major difference is that each case in a meta-analysis is an estimate obtained from the original data[1] whereas each case in a primary analysis is obtained directly from the original data.

Meta-analysis is useful not only to draw out general patterns from a wide variety of empirical sources but also because it enables researchers to resolve disputes in the literature, to determine which factors have contributed to systematic differences across studies, and to identify areas that have been neglected and warrant further investigation (Hunter & Schmidt 1990). One advantage of meta-analysis relative to traditional literature reviews is that it allows researchers jointly to assess the significance of multiple potential influences. In particular, meta-analysis helps identify both substantive influences on research findings and artefacts resulting from particular research methods and data analysis techniques. As one leading proponent pointed out, thanks to meta-analytic approaches, 'our summaries of research domains are likely to be more complete, more explicit, more quantitative, more powerful (in the sense of decreasing type II errors), and, for all these reasons, helpful to the process of cumulation' (Rosenthal 1998: 378).

The paper is divided into two parts. Part one discusses the meta-analytic research design, the data selection procedure and the variables included in the design. Part two presents the results of the meta-analysis. We conclude with a discussion of how our findings contribute to research in the field.

Data and methods

Although they share a common scholarly purpose, meta-analyses (MAs) and
literature reviews (LRs) are quite different from each other. LRs examine how
a particular phenomenon has been studied by scholars in published works
with the objective of justifying the need for further research. MAs deal with
specific empirical hypotheses and try to tell something about them with the
objective of assessing, on the basis of the empirical work so far published,
whether there is an impact of some independent variable upon some depend-
ent variable, and, if so, what the magnitude of this impact is. Therefore LRs
are about 'studies', MAs are about parameter estimates; LRs gather informa-
tion about various studies in order to describe and compare them; MAs extract
from these studies all the relevant parameter estimates together with inform-
ation about them. Furthermore LRs limit themselves to the most important
studies in a given field whereas MAs tend towards comprehensiveness.

There is a plethora of LRs in the partisan theory literature. Typically al-
most every partisan theory study, theoretical or empirical, contains a literature
review putting the issue in context. Some LRs however stand out as especially
illuminating, for example, those of Blais et al. (1993) and Caramani & Hug
(1998). On the other hand, MA is much less frequent. The only two meta-
analyses of partisan theory that we are aware of are Imbeau (1994) who
presented a preliminary analysis on a subset of our database, and Schmidt
(1996). The purpose of this essay is to expand on the work of Imbeau and
Schmidt by assessing systematically the predictive performance of empir-
ical tests of partisan impact reported in the cross-national partisan theory
literature since 1976. How often do these tests confirm the partisan theory
hypothesis? To what degree is the left-right partisan impact present in the
population? What (if any) are the main determinants of statistical support for
the left-right partisan influence hypothesis? More specifically, are partisan
cycle explanations more sensitive to substantive policy domains or to issues of
methods and measurement? How do our results contribute to the advancement
of knowledge in the field?

Like any other empirical research, MAs are vulnerable to various threats to
validity (Campbell & Stanley 1963; Wolf 1986). The main threat to validity in
this MA is heterogeneity among the studies incorporated in the analysis. We
tried to respond to this threat to validity; first, by establishing explicit criteria
for the selection of individual studies in order to ensure that the tests within
each study are comparable; and second, by coding relevant published inform-
ation in the form of variables most susceptible to influencing the predictive
performance of each test. The criteria for the selection of cases and the rules
for constructing the variables in the MA were organised in a codebook which
served as guidance for coders to identify individual parameter estimates and

their relevant attributes in each study. Routine and less routine controls were applied in order to ensure inter-coder reliability. Among other things special software with limited data-field size was developed to help data collection and minimise coding errors. At the end of the coding process, the principal investigator double-checked all coding and any coder-investigator disagreements were discussed and consensually resolved, though such disagreements were relatively rare and of minor importance.

Selection of cases

One concern in selecting individual cases for the meta-analysis was comprehensiveness. To have confidence in our measures, we needed to be sure that the selection process was representative of the available material. To achieve this goal, we tried to minimize discretion in selection by considering everything we could find in the published record. Accordingly, we searched the literature for studies containing statistical estimates of the partisan cycle hypothesis. Through a conventional bibliographic search starting with the most widely cited works in the field and applying cross-referencing techniques to identify additional relevant studies, and through an on-line search of the most important bibliographic search tools in the social sciences including Bartolini's impressive database of 11,500 publications on European parties and party systems since 1945 (Bartolini et al. 1998), we identified over 600 studies (mainly journal articles and book chapters) containing empirical results relevant to our problem.

Next, it was decided that the countries selected for analysis should all share similar democratic institutional arrangements and traditions. This is justified on theoretical grounds since partisan theory is explicitly applicable only to democratic systems (Hibbs 1992). Our choice appears sensible methodologically since sample heterogeneity in terms of democratic characteristics would certainly threaten the validity of our comparison of parameter estimates. We therefore limited our analysis to empirical test results in OECD countries.

Empirical tests of partisan theory come either in the form of longitudinal studies or in the form of cross-sectional studies. Longitudinal studies examine the variations in partisan impact over time within the same country. Cross-sectional studies, on the other hand, compare more than one country. In this MA we are interested in reporting cross-national comparisons and include results from cross-sectional studies and times-series cross-sectional studies only.

We sought to achieve a degree of sample homogeneity in the policy variables examined in individual studies (the *explanandum*). To be valid, a comparison of parameter estimates should be based on similar *explananda*.

Because of the wide variety of measures of the policy variable in the partisan theory literature, it is impossible to achieve perfect homogeneity. To ensure comparability of the policy variables, we decided that the parameter estimates in the individual studies under analysis should be related to a measure of policy outputs such as spending, taxation, legislation or regulation. We thus excluded statistical tests on macroeconomic outcomes such as growth, unemployment and inflation. This choice is justified on the theoretical ground that macroeconomic outcomes are much less susceptible to manipulation by governments than policy outputs.

The validity of the meta-analysis may also be threatened by the wide variety of measures of party ideology used by partisan theorists to explain policy outputs in individual studies (the *explanans*). To limit heterogeneity in party ideology variables, we consider only tests in which ideology is defined as either left or right. We therefore leave out estimates of the impact of centre (mostly Christian-democratic) and green parties on policy outputs. Following Downs (1957) virtually all studies in the partisan theory literature assume that party positioning can be ordered on a left-right scale corresponding to the desired degree of government intervention in the economy (the left supports a higher level of government intervention and the right supports a lower level). The issue of whether competition can be reduced to one dimension has been the subject of considerable theoretical debate (Sartori 1976; Laponce 1981). The left-right dichotomy has also been criticised recently on the ground that it ignores the distinctive character of Christian democracy (van Kersbergen 1995). However, it is clear that the distinction between left and right continues to dominate party competition and voter perceptions in advanced liberal democracies (Budge et al. 1987; Lewis-Beck 1988; Klingemann et al. 1994).

Another possible threat to validity in MA is the lack of sample homogeneity in the estimation methods. One of the objectives of this MA is to extract overall measures of the relationship between the party composition of government and policy outputs from single coefficients found in individual studies. Ideally, the individual coefficients to be compared should be highly similar. This is not the case here given the wide variety of statistical measures of association used by partisan theorists in their empirical tests. We decided to cast a wide net by including both bivariate and multivariate correlation and regression coefficients in the MA. Correlation coefficients, as well as standardized regression coefficients, are based on standardized values and therefore are directly comparable across studies. However, unstandardized regression coefficient estimates are not directly comparable, since their weights directly mirror the scale that is used to measure the variables in each individual study. Including both correlation and regression coefficients in the sample therefore prevents us from comparing the actual values of the

parameter estimates in the meta-analysis. We were also prevented from comparing significance levels of individual test results because of inconsistencies in reporting these statistics across individual studies.

Variables in the meta-analysis

The dependent variable in this MA is the predictive performance of individual tests of partisan theory, measured on the basis of parameter estimates of party impact. These estimates are operationalized in terms of three attributes: estimates that support the partisan theory hypothesis (the coefficient is significant and in the predicted direction) are recorded as 'successes'; estimates that fail to confirm the hypothesis (the coefficient fails the test of statistical significance) are reported as 'failures'; estimates that contradict the hypothesis (the coefficient is significant and in the wrong direction) are reported as 'anomalies'. Of course, comparing the relative predictive performance of individual tests cannot be assessed in a vacuum. To settle the question, one needs to look at possible sources of variation in predictive performance across cases.

Individual tests of the partisan theory hypothesis vary first and foremost in the definition and operationalization of policy outputs. The definition of policy outputs raises important theoretical debates with regard to whether partisan impact varies by policy domain, and whether policy outputs should be measured using financial data such as taxation and public expenditure or on the basis of non-financial data such as those relating to levels of regulation and public employment.

A large number of cross-national tests of partisan theory examine the impact of parties on the development of the welfare state. Sociological theories of cross-national convergence argue that industrialized countries experience similar welfare states as a direct result of their similar levels of economic wealth and technological development (Wilensky 1976). It is, therefore, no coincidence that partisan theorists have paid special attention to welfare policies (i.e., health, education and social security) in trying to refute the convergence thesis. Individual tests of partisan impact on welfare policies typically postulate that *ceteris paribus* welfare intervention by the state is positively correlated with the presence of leftist parties in government and negatively correlated with the presence of rightist parties in government (denoted L+, R−). Many partisan theorists have criticized the welfare approach because it fails to capture other dimensions of state intervention that may reflect partisan impact. One approach considers party impact on foreign policy. The approach postulates that spending on development assistance is positively correlated with the presence of the left in government and negatively correlated with the presence of the right in government (L+, R−). The hypothesis is reversed for military spending and military alliance membership:

military spending and state involvement in military alliances correlates positively with the presence of the right in government and negatively with the presence of the left in government (L−, R+). Another approach focuses on the impact of political factors on economic policy. One important variant of this approach is the political business-cycle hypothesis (Hibbs 1977, 1992; Tufte 1978). The hypothesis states that the standard relationship (L+, R−) holds for state intervention in macro-economic policy, public debt as a percentage of GDP, and progressivity of taxes. The reversed relationship (L−, R+) holds for privatisation, and budget balance as a percentage of GDP. A final approach examines how partisan factors affect the overall size of government. The approach postulates that *ceteris paribus*, government size correlates positively with the presence of the left in government and negatively with the presence of the right in government (L+, R−).

Individual tests of partisan theory also vary in terms of how they operationalize observed government activity. Most studies operationalize government activity in terms of public spending or revenues from taxes (often expressed as a percentage of GDP). This is justified on grounds of data availability (financial data are standardized in such a way as to make possible cross-national comparisons). However, the use of financial data has been criticized for failing to capture important dimensions of policy outputs. For example, it has been argued that, instead of focusing only on public expenditures as a measure of welfare effort, analysts should also consider various policy instruments, such as guaranteed basic income or conditional benefits, from which government can choose to affect income distribution in society (Esping-Andersen 1990). Others argue that, although financial data may constitute valid indicators of government priorities among capital intensive programs that compete for scarce resources, they provide less valid measures of government activity in other less capital intensive policy domains. Several empirical studies in the partisan theory literature have focused on state regulatory activity such as international agreements (Keman 1982), tariffs in the manufacturing sector (Blais 1986), or macro-economic policy (Pennings 1995). Other studies have been concerned with administrative aspects such as government employment (Cusack et al. 1989) or state privatization strategy (Boix 1997).

In order to assess the impact of variation in the definition of policy domains on our conclusions, we record whether individual parameter estimates apply to social welfare policies, the size of the state, economic policies, or foreign and defence policies. We also record whether the measure of government activity in each test is based on financial data (e.g., public spending) or non-financial data (e.g. regulation or administrative action).

Individual tests of the partisan theory hypothesis offer a wide variety of measures of the party composition of government (the main explanatory variable in these tests). Individual authors' definitions of the party composition of government vary depending on whether they report the presence of the left or the right in government, what parties (or families of parties) are included in the measure, and how the measure accounts for relative party strength.

Existing approaches to the left-right measure of the party composition of government fall into two basic categories. The most frequently used approach consists of reporting the presence of the left in government (Tufte 1978; Hibbs 1977, 1992; Cameron 1978). A theoretical justification for choosing this particular approach is given by Esping-Andersen (1990) who argues that the presence or absence of a leftist tendency is the most important political determinant of welfare expansion in capitalist democracies. The other approach reports the presence of the right in government. Castles (1982) argues that the presence or absence of the right makes the most important difference in public policy, especially social policy.

Individual authors often calculate the presence of the left or the right in government in the form of sophisticated indices that may or may not take into account the full partisan spectrum. There is some disagreement about the range of parties to be included in the measure. Most studies examined in this MA report the presence of the left and/or the right in government on the basis of either leftist parties, a left-right scale, conservative parties, or major parties of the right. These four measures are highly correlated with each other (Schmidt 1996: 161) so we can be confident that the choice of one or the other method of reporting what falls to the left and what falls to the right has little impact on estimates of partisan influence. However, a few studies also include centre parties in their measure of the left-right composition of government. This is a problem because left-right measures that include centre parties do not correlate well with other indicators of left-right party composition of government and might therefore make a difference to estimates of party influence on policy.

Individual measures of the party composition of government also differ in their definitions of relative party strength. The most frequently used measure is the party composition of the cabinet. This is justified on the grounds that in advanced liberal democracies, political power is basically exercised through the cabinet and that governmental policy is directly related to the allocation of ministerial portfolios among different parties (Austen-Smith & Banks 1990; Laver & Shepsle 1994). Another indicator of relative party strength is the percentage of parliamentary seats among the parties forming the cabinet. Blais et al. (1993) argue that since 'the basic decision-making unit is the cabinet, an assumption that underlies the whole approach, surely the most

logical indicator is the percentage of cabinet posts held by the parties of various stripes' (1993: 50). In any case, since cabinet seats tend to be directly proportional to parliamentary seats (Browne & Franklin 1973) we should expect little difference in the results achieved by each measure.

On the other hand, a case can be made in support of measuring the party composition of government in terms of the percentage of popular votes. A government's popular vote is a good measure of its electoral vulnerability: a government that enjoys little popular support is likely to be more vulnerable and, therefore, more inclined to retreat from its ideological ideal in order to court voters from an opposing coalition.

In order to control for the variation in individual definitions and operationalizations of the party composition of government in the MA, we collected and coded information about which party or family of parties (right or left) and which indicators of party strength (popular votes, seats in parliament, ministerial portfolios) were used as measures of the party composition of government in individual tests.

In spite of all the restrictive conditions that were applied to ensure sample homogeneity, there remains significant variation among the cases in the sample. This remaining variation still represents a threat to the validity of our meta-analysis. We address this problem by incorporating additional variables in the meta-analysis. First, we coded information about the period of analysis in each individual study so that any time effect on the results could be taken into account. This information was used to construct a historical period variable with two attributes, one for test results partly or entirely based on post-1973 annual observations, the other for test results entirely based on data up to (and including) the year 1973. The cut-off date of 1973 coincides with the onset of a major recession in industrialized countries, it is also used in studies that compare government growth over different periods (Swank, 1988). We also considered the effect of variation in sample size (N) on parameter estimates. Indeed since levels of significance are very sensitive to sample size, it is not uncommon to find significant correlations when N is very large. This may happen to an increasing extent in the body of research we are considering since researchers tend more and more to use time-series cross-section designs, thus sensibly inflating sample size. In order to protect our conclusions from variations in sample size, we coded each parameter estimate for the corresponding value of N and took that information into account in constructing a sample size variable.

Another factor that may affect the validity of our meta-analysis is the variation in the specification of the model in each individual study, i.e. the number and nature of explanatory or control variables that partisan theorists explicitly include in their datasets in addition to the party composition of

government variable. There is evidence that the statistical significance and the direction of estimates of partisan impact may be affected by institutional factors such as the division of power between parties in and out of government (Blais et al. 1993, 1996; Schmidt, 1996), the frequency of electoral competition (Cameron 1978), or the time a government has been in power (Swank 1988; Blais et al. 1993; Thérien & Noël 2000). Unfortunately, these (and other) potentially confounding factors have not been consistently examined in the literature. The small number of available estimates associated with each explanation precludes their inclusion in the MA.[2] On the other hand, we can still record the variation in model specifications with a dichotomous variable which identifies whether the estimation is bivariate or multivariate.

Findings

Out of over 600 studies identified in the bibliographic search, 43 proved to correspond to our selection criteria. The studies are listed in chronological order in Table 1 along with information about the period and the number of countries over which the tests were run, the relevant policy domain, and the left-right attribute used to define the party composition of government variable. Due to space limitation Table 1 reports only the summary information for each study. Individual tests within each study may involve different time periods, countries, policy domains and different measures of party ideology.

   The earliest study is the one by Wilensky (1976) on the impact (or rather the absence of impact) of parties on welfare state development. The latest studies examine partisan influence on the economy (Boix 1997) and fiscal policy (Carlsen 1997; Haan & Strum 1997). The frequency of publication has remained approximately constant over the years, with 23 studies published over the period 1976–1986 and 20 studies published over the period 1987–1997. Recent studies by Clayton & Pontusson (1998), Garrett (1998), Midtbø (1999), Swank (1998) and Thérien & Noël (2000) satisfy our selection criteria, but they were published after we started our analyses and are not included in this survey. Overall, the 43 studies include a total of 693 parameter estimates (a mean of 16 and a median of 8 estimates per study). Castles & McKinlay (1979a) and Friedland & Sanders (1986) provided each one single estimate, and Castles (1982) contained 95 relevant estimates. Castles is the most prolific scholar in the literature as he is associated with six studies containing 183 parameter estimates, 26 percent of our sample. The tests in the meta-analysis were all performed on data within the period 1950–1992.

Table 1. Characteristics of quantitative studies on the ideology-policy relationship

Author(s) & date	Period	No. of countries	Policy area	Ideology type	No. of tests	Failure	Success	Anomaly	Modal category	Effect size
Wilensky (1976)	1965–71	19	Welfare, fiscal	Left	15	15	0	0	Failure	0.0
Hewitt (1977)	1962–64	17	Social	Left	4	0	4	0	Success	1.0
Cameron (1978)	1960–75	18	Fiscal	Left	3	2	1	0	Failure	0.33
Castles & McKinlay (1979a)	1973–74	19–20	Social, education, welfare	Right	3	0	3	0	Success	1.0
Castles & McKinlay (1979b)	1958–72	19	Welfare	Right	1	1	0	0	Failure	0.0
Borg & Castles (1981)	1962–72	9–16	Social	Right, left	16	6	8	2	Success	0.37
Castles (1981)	1960–76	18	Social, health, education, total spending, welfare	Right	30	30	0	0	Failure	0.0
Wilensky (1981)	1965–71	19	Social, fiscal	Left, right	64	44	3	17	Failure	−0.22
Arnhem & Schotsman (1982)	1974	14	Fiscal	Right, left	5	5	0	0	Failure	0.0
Castles (1982)	1960–76	14–18	Social, education, health, total spending, welfare	Right, left	95	53	42	0	Failure	0.44
Keman (1982)	1956–75	18	Foreign	Left, right	32	32	0	0	Failure	0.0
Schmidt (1982)	1950–77	21	Fiscal	Left, right	33	20	11	2	Failure	0.27
Cameron (1983)	1965–81	20	Social, total spending	Left	3	3	0	0	Failure	0.0
Schmidt (1983)	1973–79	13–23	Social, administrative, total spending	Left	11	11	0	0	Failure	0.0
Solano (1983)	1968	18	Total spending, foreign, health, social, welfare	Left	16	16	0	0	Failure	0.0
Swank (1983)	1960–71	17	Social	Left	4	1	3	0	Success	0.75
Hicks & Swank (1984a)	1960–71	18	Social, welfare	Right, left	11	2	8	1	Success	0.64
Hicks & Swank (1984b)	1970	13	Fiscal	Left, right	4	1	0	3	Failure	−0.75
Blais (1986)	1980	17	Economic, social, fiscal	Left	4	3	1	0	Failure	0.25
Castles (1986)	1960–81	18	Education, social, health, welfare	Right	20	20	0	0	Failure	0.0

Table 1 (continued)

Author(s) & date	Period	No. of countries	Policy area	Ideology type	No. of tests	Failure	Success	Anomaly	Modal category	Effect size
Friedland & Sanders (1986)	1962–80	12	Social	Left	1	0	1	0	Success	1.0
Lane & Ersson (1986)	1975–77	24	Education, health, foreign	Left	4	4	0	0	Failure	0.0
Williamson & Pampel (1986)	1965–75	18	Social	Left, right	2	1	1	0	Success	0.50
Imbeau (1988)	1966–81	8–15	Foreign	Right, left	16	4	10	2	Success	0.50
Pampel & Williamson (1988)	1950–80	18	Social, health, welfare	Right, left	22	19	2	1	Failure	0.04
Swank (1988)	1960–80	18	Total spending	Center, right	5	1	2	2	Failure	0.0
Castles (1989)	1960–81	18	Education	Left	18	14	4	0	Failure	0.22
Cusack et al. (1989)	1960–83	13	Gov. employees	Right	3	0	3	0	Success	1.0
Griffin et al. (1989)	1959–80	22	Fiscal, welfare	Left	2	2	0	0	Failure	0.0
Hicks et al. (1989)	1957–82	16	Social	Left	4	0	4	0	Success	1.0
Imbeau (1989)	1966–81	8–15	Foreign	Right, left	94	83	9	2	Failure	0.07
Roubini & Sachs (1989)	1973–85	13	Total spending	Left	3	0	3	0	Success	1.0
Hicks & Swank (1992)	1960–82	18	Welfare, social	Left, right	12	4	3	5	Failure	−0.17
Blais et al. (1993)	1960–87	15	Total spending	Left	41	32	9	0	Failure	0.22
Huber & Stephens (1993)	1958–86	18	Social	Left, right	21	19	2	0	Failure	0.10
Haan & Sturm (1994)	1981–89	12	Total spending, fiscal	Left	9	1	8	0	Success	0.89
Noel & Thérien (1995)	1965–88	15–16	Foreign	Left	6	6	0	0	Failure	0.0
Pennings (1995)	1965–88	13	Economic	Left	8	8	0	0	Failure	0.0
Blais et al. (1996)	1962–91	15	Total spending	Left	4	0	4	0	Success	1.0
Hahm et al. (1996)	1958–90	9	Fiscal	Left	5	5	0	0	Failure	0.0
Boix (1997)	1979–92	18	Economic	Left, right	4	0	4	0	Success	1.0
Carlsen (1997)	1980–92	18	Fiscal	Left	32	20	1	11	Failure	−0.31
Haan & Sturm (1997)	1982–92	21	Fiscal	Left	3	3	0	0	Failure	0.0
Total (or average)					693	491	154	48	14 success 29 failures	0.28

Full bibliographic references are available in Imbeau 2001.

Overall left-right partisan effects

Table 1 also presents the number of relevant tests in each study and the frequencies with which these tests are reported as 'successes' (a coefficient associated with a right or a left partisan variable is significant ($p < 0.10$) and in the correct direction), 'anomalies' (a coefficient associated with the right or the left is significant and in the opposite direction to the one hypothesised) or 'failures' (a non-significant coefficient) in this study. Once the results of each individual test have been sorted into successes, failures and anomalies, we can use the simple 'vote-counting' meta-analytic procedure to assess the overall relationship between the left-right party composition of government (the *ex- planans*) and policy outputs (the *explanandum*). The procedure consists of counting individual *tests* in each category (successes, failures, anomalies) to see if a plurality of tests falls into one of the three categories with fewer falling into the other two categories, and to declare the modal category the winner. The modal categorization is then assumed to give the best estimate of the true relationship between the independent and dependent variables (Hedges & Olkin 1980; Light & Smith 1971).

Looking at the overall distribution of tests (bottom row of Table 1), we see that, out of the 693 tests, only 154 (22 percent) are reported as successes and 48 tests (7 percent) are reported as anomalies. The remaining 491 tests (71 percent) fail the statistical significance test (at 10 percent) and are reported as failures. The success rates appear remarkably stable over time: 86 out of 385 tests published in or before 1986 are reported as successes (22 percent); 68 out of 308 tests published after 1987 are reported as successes (22 percent). Clearly, these numbers cannot be interpreted as evidence in support of the left-right party impact hypothesis. Applying the vote-counting technique to the data suggests that, since the 'failure' category is the modal category, we cannot reject the null (convergence) hypothesis of no party impact based on the average predictive power of individual tests.

The vote-counting technique can also be applied to individual *studies* instead of individual tests. By counting studies instead of tests, we set the weight of each test in inverse proportion to the umber of tests reported in each study. This is recommended in samples – such as this one – where the distribution of the number of tests per study is highly skewed (Wolf 1986). A study is reported as a success if a plurality of tests contained in the study falls into the 'success' category; it is reported as a failure if a plurality of tests contained in the study falls into the 'failure' or the 'anomaly' category. The results are reported in the 'modal category' column of Table 1. We see at the bottom of the column that 14 studies are reported as successes (33 percent) and 29 studies are reported as failures (67 percent).[3] The 'failure' category is, once again, the modal category. Therefore, we cannot reject the

null hypothesis of no party impact based on the average predictive power of individual studies.

The vote-counting technique only gives a measure of the average success rate of tests of the party impact hypothesis. This is not an ideal measure for deciding whether or not we can reject the null hypothesis of no party impact. A better technique is the 'combined tests' technique (Wolf 1986). It consists of calculating the average effect of the *explanans* on the *explanandum* from the test results in each individual study. The technique normally involves the summation of actual test statistics (the individual coefficients and their standard errors) for the purpose of converting these summary statistics into a common metric, usually in the form of the Pearson Product Moment Correlation (Rosenthal 1991). This common metric is then used to estimate an 'average effect size' (the strength of the correlation coefficient) and to construct a confidence interval around the effect size in order to infer 'the degree to which a phenomenon is present in the population, or the degree to which the null hypothesis is false' (Cohen 1977: 10).

We cannot calculate a real average effect since the summation procedure in this MA does not involve actual test statistics. However, we can get around the problem by using a simplified version of the 'combined tests' procedure based on vote-counting results to calculate a proxy measure of average effect size. The technique consists first of calculating the approximate effect size r separately for each study using formula (1) which is constructed on the assumption that each success is assigned a weight of +1, each anomaly is assigned a weight of -1, and each failure is assigned a weight of zero.

$$r = (\text{successes} - \text{anomalies}) / (\text{number of tests}). \qquad (1)$$

Once the effect size for each individual study has been calculated, we compute the average of the effect sizes, r_{average}, using formula (2), to represent the estimate of effect size across all 43 studies.

$$r_{\text{average}} = \Sigma r_i / (\text{number of studies}). \qquad (2)$$

The proxy effect size for each individual study is reported in the right-hand column of Table 1. The numbers vary from a theoretical minimum of -1 and a theoretical maximum of +1. Therefore, they behave like correlation coefficients and so does the average size effect $r_{\text{average}} = 0.282$ which is reported at the bottom of the column. Based on our proxy measure for r_{average}, it is estimated that the left-right party composition of government affects policy outputs by approximately 0.28 standard deviation units. We then constructed a 95 percent confidence interval around the average effect size to examine whether it is statistically significant and found that it encompasses

zero (-0.15, $+0.71$). Applying a simplified 'combined tests' technique to the data therefore suggests that, since the proxy measure for average effect size across the 43 studies is not statistically significant, we cannot reject the null (convergence) hypothesis of no party impact.[4]

The finding that the null (convergence) hypothesis cannot be rejected may seem disappointing (at least to supporters of the 'politics matters' school of thought) given the operational creativity of various authors in the partisan theory literature. However, its significance should not be overestimated. First, the average success rate and the average effect size that emerge from this MA do not constitute an absolute yardstick. They are dependent on the severity of the criterion of success that we choose to apply to the estimates under analysis. Many estimates that we report as failures are in the correct direction and could be considered successes under a different criterion of statistical significance.[5] Second, the number of successes that we report must be considered in relation to the number of anomalies. Supporters of the 'politics matters' school of thought should take heart at our finding that 'successes' beat 'anomalies' in the ratio of 3 to 1. Last but not least, the results are potentially affected by mediating factors. Focusing only on overall effects and ignoring these mediating factors constitutes an oversimplification. In the next section, we will examine possible effects that might mediate the relationship between the left-right partisan composition of government and policy outputs.

However, before we do that, we need to address the question of whether our sample of studies is representative of the population. In spite of our efforts at collecting all the studies we could find, it is very likely that not all studies of partisan impact actually conducted are published and therefore available for review. It could be argued that published research studies tend to be biased toward positive findings because reports of non-significant findings are generally not publishable. This may enhance the likelihood of type I error (finding more positive significant results than is really the case were all studies to be included in the meta-analysis). There is no definitive answer to this question since we have no record of unpublished studies. However, we believe that the problem is not serious or pervasive in the partisan theory literature. The assumption that test results in support of the convergence (null) hypothesis are systematically buried away in file drawers does not square well with the high proportion of non-significant test results that we observe in the published literature. Furthermore, as can be seen from Table 1, non-significant results often coexist with positive significant results in the same study. Evidently, partisan theorists who publish their findings are not shy about reporting insignificant or anomalous test results. This can be explained in part by the rather common practice among partisan theorists of reporting insignificant or anomalous test results based on the 'wrong' approach to bolster *a contrario*

their own (obviously successful) approach. For example Blais et al. (1993) report a large number of unsuccessful (i.e., insignificant) cross-section test results in order to prove their point that statistically significant estimates are more easily achieved by way of time-series cross-section analyses. Other studies (e.g., Swank 1988) report insignificant (or even anomalous) estimates of left-right party impact in order to prove their point that statistically significant estimates exist in association with centre parties. Of course, we had no choice but to include all the tests that satisfy our selection criteria in the MA whatever the intention of individual authors in reporting those tests.

In any case, our research questions will not be settled by just looking at the average success rate or effect size of individual tests of partisan impact. We need to look instead for explanatory variables that reveal something theoretically interesting about possible sources of variation in the predictive performance of models of left-right partisan impact. The tests that are reported within a single study are often based on different combinations of explanatory variables: they vary in terms of policy domain, the measure of left-right party composition of government, the historical period, and the number of countries. This makes it impractical to use the studies as units of analysis when we assess the effect of these factors on success. The analysis of mediating factors that follows is, therefore, based on individual test results instead of individual studies.

Mediating factors

Table 2 presents a breakdown of individual tests by policy domain. We see from the table that 69 (24 percent) out of 288 tests in the welfare policy domain are reported as successes. The difference in predictive performance across specific areas within the welfare domain (from a low of 19 percent to a high of 27 percent) is not statistically significant. Next we see that 59 (37 percent) out of 161 tests of the size of the state are reported as successes. Note the perfect prediction score of tests of the public employment area within this domain. However, one cannot generalize from this because there are only three tests in the public employment area. As in the welfare domain, the deviations from the mean in the size of the state domain are not statistically significant. The overall success rates are much lower in the two remaining policy domains. There are 20 successes (13 percent) out of 148 tests in the foreign affairs and defence policy domain and only 6 successes (6 percent) out of 96 tests in the economic policy domain. The deviations from the mean success rate are statistically significant in both the foreign affairs and the economic policy domains. Tests of left-right party impact on the level of state intervention in the economy have a (relatively) very high degree of success

Table 2. Success rate of individual tests by policy domain

	Success rate	All tests
Development of the welfare state	0.24	288
Social security	0.25	164
Education	0.23	48
Health	0.19	31
Social security + Education + Health	0.27	45
Size of the state	0.37	161
Total spending	0.37	119
Revenue	0.31	39
Public employment	1.00	3
Foreign and defense policy*	0.13	148
Development allocations	0.16	117
Military expenditures	0	23
Alliance membership	0	8
Economic policy*	0.06	96
Privatization + state intervention	0.39	13
Public debt + budget balance	0.02	41
Progressivity of taxes	0	42
Overall	0.22	693

The entries indexed with an asterisk denote statistically significant differences ($p < 0.05$).

(39 percent). On the other hand, the success rate in the public budget and taxation areas is virtually zero.

From the data in Table 2, it appears that the choice of policy domain (and the choice of a specific area within a policy domain) significantly affects the rate of support for the partisan theory hypothesis. Interestingly, there is no correlation between the number of estimates and their success rate in each specific area ($r = -0.10$).

What other factors (if any) influence the variation in success rates? Table 3 provides some answers. The table displays the proportion of successes across the following variables: left-right party ideology, measures of party strength, the historical period, multivariate (as opposed to bivariate) model specification, sample size, and financial (as opposed to non-financial) measure of government activity.

We see from Table 3 that there are twice as many tests based on the presence of the left (472) as compared with tests based on the presence of the right (221) in our sample. Tests that report the presence of the right generate a higher rate of success on average (25 percent) than tests that report

Table 3. Success rate of individual tests across other variables

	Success rate	All tests
Party ideology		
Parties of the Right	0.25	221
Parties of the Left	0.21	472
Party strength		
% Portfolios	0.23	302
% Seats	0.26	125
% Popular votes	0.25	133
% Other	0.15	133
Historical period*		
Post-1973	0.27	339
Pre-1973	0.18	354
Model specification*		
Multivariate	0.28	279
Bivariate	0.18	414
Sample size*		
$N \leqslant 49$	0.20	544
$N > 49$	0.28	149
Government activity*		
Financial	0.23	633
Non financial	0.12	60

The entries indexed with an asterisk denote statistically signific-
ant differences ($p < 0.05$).

the presence of the left (21 percent). However, the paired difference is too small to be statistically significant. Looking at the party strength variable of Table 3, we see that using the percentage of cabinet portfolios, although the preferred method of calculating party strength, has no statistical impact on success when compared with alternative methods.

Next, we see that 79 (28 percent) out of 279 multivariate tests are reported as successes against 75 (18 percent) out of 414 bivariate tests. The paired difference for the model specification variable is statistically significant, as is the paired difference for sample size (the proportion of success is 20 percent for smaller samples against 28 percent for larger samples) and for the measure of government activity (the rate of success is 23 percent for financial data compared with 12 percent for non-financial data). The paired difference for the historical period variable is also statistically significant: 92 (27 percent) out of 339 tests based on the post-1973 period are reported as successes against 64 (18 percent) out of 354 tests based on the pre-1973 period. An interesting

Table 4. Success rate of individual tests by historical period and by policy domain

	Policy domains			
	Welfare success rate (N)	Size of state success rate (N)	Foreign success rate (N)	Economic success rate (N)
Historical period				
Post-1973	0.20 (106)	0.50* (98)	0.30* (71)	0.09* (64)
Pre-1973	0.27 (182)	0.16* (63)	0.07* (77)	0.00* (32)

The entries indexed with an asterisk denote paired differences with p values < 0.05.

pattern emerges when we crosstabulate the historical period variable with policy domains (Table 4). Post-1973 tests of party impact on the size of the state, foreign affairs and economic policy have significantly higher average success rates than pre-1973 tests of party impact in these policy domains. However, individual tests of party impact on welfare policy follow the opposite direction. The average success rate of estimates of party impact on welfare is smaller for the post-1973 period than for the pre-1973 period.

Table 5 presents a multivariate logistic regression model in which the dependent variable takes the value of 1 when a test is reported as a success and 0 otherwise. There were too few anomaly cases to include them as a distinct variable in this analysis. Three policy domain variables are included in the model to assess the predictive performance of tests of party impact on the size of the state (total public spending and taxation), foreign affairs, and the economy as compared with the predictive performance of tests of party impact on welfare (the reference policy domain). The next two variables in the model are dichotomous variables intended to assess whether the predictive performance of individual tests is affected by the use of non-financial policy data (as opposed to financial policy data) and by the inclusion of a left (instead of right) partisan variable. Next the model shows how indicators of party strength based on seats in the legislature, popular votes and cabinet portfolios compare with indicators based on other (mixed) measures of party strength (the excluded category). The model also assesses how the predictive performance of individual tests is affected by the historical period, the multivariate versus bivariate nature of the tests, and the number of observations. The numbers reported in each cell are the logistic regression estimates, the standard errors (in parentheses) and the odds ratios (in italics). We report only the odds ratios associated with significant regression coefficients.

From Table 5, we see that, *ceteris paribus*, the success rate of individual tests is not significantly affected by how partisan impact is operationalized.

Table 5. Multivariate logit analysis of individual tests' predictive performance

Dependent variable: Predictive performance			
Policy domains			
Size of the state	0.50**	(0.24)	*1.60*
Foreign affairs	−1.18***	(−0.34)	*0.30*
Economy	−1.82***	(−0.51)	*0.25*
Financial government activity	−0.36	(−0.51)	
Parties of the left	−0.21	(−0.22)	
Party strength			
Percent seats	0.13	(0.34)	
Percent portfolios	−0.08	(−0.31)	
Percent popular votes	0.67*	(0.35)	*2.00*
Multivariate test	0.56***	(0.22)	*1.80*
Sample size	−0.00	(−0.00)	
Post-1973	0.67**	(0.22)	*2.00*
Constant	−1.17**	(−0.57)	
Chi-square for covariates (df)	70.21	(11)	
% Correctly predicted	79.37		

The figures in each cell give the logistic regression estimate with standard error in parentheses. The italicized numbers are the odds ratios. The reference policy domain is Welfare. The reference attribute for party strength is the mixed measure. Details are provided in the text.
$*p < 0.10; **p < 0.05; ***p < 0.01$.

Whether individual tests use the left or the right to measure partisan impacts does not affect predictive performance, and neither does the definition of party strength (although measures of party strength based on the percent of popular votes perform better than measures based on cabinet portfolio allocation). The data also rule out sample size and the use of financial (as opposed to non-financial) data as causes of variation in predictive performance.

There is evidence that the predictive performance of individual tests is significantly affected by the policy domain variables. The estimate for the size of the state variable is significantly positive. The odds of success are multiplied by approximately one and a half in tests that use the size of the state variable compared with tests that use welfare spending (the reference category) as a dependent variable. In contrast, the estimates for foreign affairs and economic policy are significantly negative and the associated odds ratios are one-third and one-quarter of that associated with to the welfare policy domain, respectively. The data also indicate that predictive performance is significantly affected by model specification and the historical period. *Ceteris*

paribus, the odds of success for multivariate tests are approximately twice the odds of success for bivariate tests; the odds of success for post-1973 tests are also twice the odds of success for pre-1973 tests.

Thus, keeping the other variables in the model constant, the predicted logit of success for a multivariate, post-1973 test of the left-right division of the popular vote on the size of the state is $L = -1.17 + 0.56 + 0.67 + 0.50 + 0.67 = 1.23$. The predicted probability of success for such a test is therefore $P = 1/[1 + e^{-(1.23)}] = 0.77$. By comparison, the probability of success of a bivariate, pre-1973 test of partisan impact on welfare using a mixed measure of party strength is $P = 1/[1+e^{-(-1.17)}] = 0.24$. This demonstrates that there are clearly identifiable conditions under which the probability of support for the left-right party influence hypothesis can be substantially increased.

Various sensitivity tests were performed to check the robustness of our multivariate findings. We first ran multivariate regressions in which five studies that incorporate centre parties in their definition of the left-right partisan composition of government were omitted and found no significant change in the magnitude and sign of the main variables in the model. It is also possible that our multivariate findings are unduly influenced by the high number of estimates published by a handful of authors. To assess this possible bias, we ran multivariate regressions in which the estimates reported by Castles and collaborators (183 tests), Imbeau (110 tests) and Wilensky (79 tests) were in turn omitted. We found that the exclusion of the estimates published by Castles and collaborators affects the party strength variable (a significant increase of the impact of the percentage of parliamentary seats on success) and the policy domain variable (a significant increase of the impact of the size of the state on success). Excluding the estimates published by Imbeau affects the party strength variable (the coefficient for the percentage of popular votes becomes insignificant). The exclusion of the estimates reported by Wilensky decreases the statistical significance of both the post-1973 period variable and the size of the state variable. However, the coefficients for both variables remain positive and significant at 10 percent when the studies by Wilensky are omitted from the equation. None of these results contradict the main findings of the multivariate test of Table 5. It appears that they are sufficiently robust to allow a more general discussion.

Overview and discussion

We started this review of 693 published cross-section estimates of left-right party impact on policy with a series of questions that have not been addressed in a meta-analytic format before: how often do estimates support the left-right party impact hypothesis? What is the average magnitude of the effect size of

left-right party impact? What (if any) are the determinants of variations in statistical support? Are partisan cycle explanations more sensitive to substantive factors (e.g., the policy domain of government intervention; the historical period of analysis) or to issues of measurement and methodology? How do our results contribute to the advancement of knowledge in the field?

In response to the first two questions, we find that 154 (22 percent) estimates support the left-right party impact hypothesis, 48 (7 percent) estimates contradict the hypothesis, and 491 (71 percent) estimates fail to support the hypothesis. Applying the 'vote-counting' meta-analytic technique to these numbers leads us to conclude that the null hypothesis of no left-right party impact cannot be rejected. We reach a similar diagnosis when we use a revised 'combined tests' technique to generate an 'average effect size' of left-right partisan impact on policy. The 0.28 average correlational effect between party and policy indicates that approximately 8 percent of the variation in policy output is accounted for by the variation in the left-right party composition of government. A test of statistical significance of the average effect size shows that we cannot rule out the absence of a partisan impact in the population. These negative findings cannot be interpreted as evidence in support of convergence theory however. Given the extreme diversity of tests reviewed in this MA, judging solely on the basis of overall numbers of successes or effect size involves a real risk of committing a Type II error (accepting the null hypothesis while it is false). Indeed, the success rate raises from 0.22 to 0.50 for tests on the size of the state in the post-1973 period.

In response to the next two questions, we find from the multivariate logistic regression model of Table 5, that four variables positively influence the probability of success in empirical tests of left-right partisan theory. Success occurs more frequently with multivariate tests, with measures of the size of the state as opposed to measures based on specific policy domains (especially foreign affairs and the economy), with post-1973 data and with a measure of party strength based on popular vote (although this last result is not entirely robust). When other explanatory variables are kept constant, multivariate tests of the impact of the left-right division of the popular vote on the size of the state during the post-1973 period are approximately three times more likely to be reported as successes than bivariate tests of party impact on welfare during the pre-1973 period when party strength is assessed with a mixed measure (however the difference in the probability of success decreases when the tests in Wilensky (1976, 1981) are removed from the analysis).

On the other hand, the variables for financial government activity, left-right party ideology, and sample size all fail the statistical significance test in the multivariate analysis of Table 5. The absence of a statistical impact of the sample size variable on success appears to contradict the notion that left-

right partisan effects are virtually always present but that their magnitude is
so small that they can only be detected with large samples. The data also
indicate that whether partisan theorists report the presence of the left or the
right to measure the party composition of government does not affect suc-
cess in individual tests of partisan impact. However this conclusion must be
qualified in the light of one important caveat. We do not take centre parties
into account in our definition of the party composition of government. Centre
parties have been shown to influence policy in a distinctive and coherent way.
Christian-democratic parties in particular have proposed their own brand of
solutions with respect to social welfare (van Kersbergen 1995) and foreign
aid (Thérien & Noël 2000). Therefore, including centre parties in our design
would most likely alter some of our conclusions.

The relatively high success rate of post-1973 left-right partisan tests of
the size of the state suggests some degree of directionality in the field. How
did this come about? One possible explanation is that partisan theorists have
improved their methodology over time: post-1973 tests would be more suc-
cessful than pre-1973 tests because they are reported in studies that are more
recent and therefore more sophisticated methodologically. However, the data
do not support this interpretation. In order to test for a separate effect for
methodological improvement over time, we included the year of publication
of individual studies as an additional control and found that the coefficient
was not significant. The year of publication of an estimate does not affect its
predictive performance. Furthermore, the correlation between the post-1973
period variable and the year of publication of a study is only 0.53.

The statistical evidence of Table 4 suggests a better explanation. As
Table 4 illustrates, the left-right partisan composition of government predicts
welfare spending better than policy outputs in other domains in the pre-1973
period. In the post-1973 period, however, it is the size of the state that is
best predicted by the left-right partisan composition of government. The data
therefore suggest a strengthening of the left-right partisan divide over total
government spending (and a decrease in the left-right partisan divide over
welfare spending) in OECD countries during the period of analysis. This
appears compatible with the view of two historically divergent periods, each
characterized by a distinct partisan climate: the pre-1973 years coincide with
high economic growth and a broad left-right consensus over macroeconomic
implications of public policy. In this context, left-right partisan conflicts oc-
curred mainly over welfare spending as a means of redistributing a society's
abundant resources. By contrast, the post-1973 period is characterized by
lower rates of growth, higher inflation and unemployment and increased left-
right conflicts over the appropriate size of the public sector (Peters 1990;
Przeworski & Wallerstein 1982).[6] One might hypothesise that, as the sub-

stantive stakes of left-right party rivalry shifted away from social welfare and toward the overall size of the public sector, the frequency of success and the effect size of individual tests of partisan impact on the overall size of the state have tended to increase relative to tests of partisan impact on welfare.

How do our results contribute to the advancement of knowledge in the field? First, the norms developed in this meta-analysis help in interpreting individual studies or making comparison between studies. For example, we see from Table 2 that studies of left-right party impact on the size of the state report an average success rate of 0.37, whereas studies of left-right party impact on health report an average success rate of 0.19. Therefore, comparing tests of the size of the state with tests of health expenditures is likely to be misleading unless the 0.18 difference in average success rate is taken into account. Second, this MA brings into relief the areas that need further research. Policy outputs and the party composition of government – the two core variables of partisan theory – have been extensively researched and tested. As our theoretical discussion clearly points out, these variables have well-defined attributes and their definition no longer raises fundamental controversies. Our empirical findings further reveal that no important attributes of these variables has been neglected in the studies under review. Our MA therefore suggests that this area of the partisan impact literature meets the criteria of scientific accumulation and aggregation. By contrast, the study of institutional and socio-economic determinants of policy outputs appears to lack both coherence and direction. There is little agreement among partisan theorists on what these determinants are, let alone on their attributes and impacts on public policy. As a consequence, individual researchers have adopted a wide variety of definitions and methods of measurement of determinants of policy outputs other than the party composition of government (this is why we found it impractical to include these variables in this MA). Schmidt (1996: 177) stated that 'it would be valuable if direct effects and interaction effects of the party composition of government and state structures featured more prominently in future research on comparative public policy'. We fully agree with this statement. Existing methods of measuring party composition of government have contributed to knowledge about cross-national variation in policy outputs. However, these methods do have severe limitations. One is that they do not permit one to reject the null hypothesis of no party impact on policy. A related limitation is that they ignore important institutional and socio-economic factors affecting policy outputs.

Acknowledgments

Notes

1. As we will see below, studies may also be used as cases in a meta-analysis. Cases from studies that report more than one estimate must then be averaged across these estimates.
2. Other determinants of public policy that have been incorporated as control variables in multivariate tests of partisan theory include socio-economic variables such as unemployment (Blais et al. 1993) or the degree of openness of the national economy (Cameron 1978) as well as the distribution of power resources among social classes (Swank 1988). As with institutional factors, we found that it was impractical to include these variables in this meta-analysis. We are also aware that individual test results may be influenced by the country composition of each dataset. However the examination of a country effect proved impractical given the cross-sectional nature of the estimates that we review. Testing for a country effect must await a meta-analysis of longitudinal country specific studies.
3. More precisely, 28 studies fall into the 'failure' category and one study (Hicks & Swank 1992) falls into the 'anomaly' category.
4. If we compute the average effect size on the basis of individual estimates (N = 693) rather than of each study effect size (N = 43), we get a significant but trivial r of 0.15 ($p < 0.001$).
5. A weaker standard consists of reporting as success any slope coefficient in the proper direction and exceeding the value of its standard error. Applying this weaker criterion of statistical significance to a randomly selected sample of 192 estimates for which the information can be calculated, we find 85 (44%) successes, 19 (10%) anomalies and 88 (46%) failures.
6. The results of additional statistical tests involving the post-1973 period in interaction with other variables, not displayed here, confirm this interpretation. There is no evidence from the interaction models we performed that the variation in success rate across historical periods has depended on methodological factors: Higher success rates in the post-1973 period were not influenced by changes in model specification or sample size. Nothing significant is added by the variables for government activity, parties of the left, and party strength in interaction with the post-1973 period. These variables were not a factor in the higher success rate of estimates in the post-1973 period. In contrast, the variable for the size of the state in interaction with the post-1973 period is highly significant. It appears that the increase in predictive performance from the pre-1973 to the post-1973 period has primarily depended on the higher success of estimates of the size of the state.

References

Austen-Smith, D. & Banks, J. (1990). Stable governments and the allocation of policy portfolios, *American Political Science Review* 84: 873–891.

Bartolini, S., Caramani, D. & Hug, S. (1998). *Parties and party systems: A bibliographic guide to the literature on parties and party systems in Europe since 1945*, on CD-Rom. London: Sage.

Blais, A. (1986). The political economy of public subsidies, *Comparative Political Science* 19(2): 201–216.

Blais, A., Blake, D. & Dion, S. (1993). Do parties make a difference? Parties and the size of government in liberal democracies, *American Journal of Political Science* 37(1): 40–62.

Blais, A., Blake, D. & Dion, S. (1996). Do parties make a difference? A reappraisal, *American Journal of Political Science* 40(2): 514–520.

Boix, C. (1997). Privatizing the public business sector in the eighties: Economic performance, partisan responses and divided government, *British Journal of Political Science* 27: 473–496.

Browne, E. & Franklin, M. (1973). Aspects of coalition payoffs in parliamentary democracies, *American Political Science Review* 67: 453–469.

Budge, I., Robertson, D. & Hearl, D., eds (1987). *Ideology, strategy and party change*. Cambridge: Cambridge University Press.

Cameron, D.R. (1978). The expansion of the public economy: A comparative analysis, *American Journal of Political Science* 72: 1243–1261.

Campbell, D.T. & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Caramani, D. & Hug, S. (1998). The literature on European parties and party systems since 1945: A quantitative analysis, *European Journal of Political Research* 33: 497–524.

Carlsen, F. (1997). Counterfiscal policies and partisan politics: Evidence from industrialized countries, *Applied Economics* 29: 145–151.

Castles, F.G. (1982). The impact of parties on public expenditure, in F.G. Castles (ed.), *The impact of parties: Politics and policies in democratic capitalist states*. Beverly Hills, CA: Sage Publications.

Castles, F.G. & McKinlay, R.D. (1979a). Does politics matter: An analysis of the public welfare commitment in advanced democratic states, *European Journal of Political Research* 7: 169–186.

Castles, F.G. & McKinlay, R.D. (1979b). Public welfare provision, Scandinavia an the sheer futility of the sociological approach to politics, *British Journal of Political Science* 9(2): 157–171.

Clayton, R. & Pontusson, J. (1998). Welfare retrenchment revisited: Entitlement cuts, public sector restructuring, and inegalitarian trends in advanced capitalist societies, *World Politics* 51: 67–98.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. edn). New York: Academic Press.

Cusack, T., Notermans, T. & Rein, M. (1989). Political-economic aspects of public employment, *European Journal of Political Research* 17(4): 471–500.

Downs, A. (1957). *An economic theory of democracy*. New York: Harper.

Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Cambridge: Cambridge University Press.

Friedland, R. & Sanders, J. (1986). Private and social wage expansion in the advanced market economies, *Theory and Society* 15(1–2): 193–222.

Glass, G., McGaw, B. & Smith, M. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage Publications.

Haan, J. de & Sturm, J.-E. (1997). Political and economic determinants of OECD budget deficits and government expenditures: A reinvestigation, *European Journal of Political Economy* 13: 739–750.

Hedges L. & Olkin, I. (1980). Vote-counting methods in research synthesis, *Psychological Bulletin* 88: 359–369.

Hibbs, A. (1977). Political parties and macroeconomic policy, *American Political Science Review* 71: 1467–1487.

Hibbs, A. (1992). Partisan theory after fifteen years, *European Journal of Political Economy* 8: 361–373.

Hicks, A., Swank, D. & Ambuhl, M. (1989). Welfare expansion revisited: Policy routines and their mediation by party, class and crisis, *European Journal of Political Research* 17(4): 401–430.

Hunter, J. & Schmidt, F. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage Publications.

Imbeau, L.M. (1994). 'End of ideology' or 'Politics matters'? Two competing hypotheses in the comparative public policy literature, *History of European Idea* 19(4–6): 683–689.

Imbeau, L.M. (2001). *IDEOPOL: A database on the relationship between party ideology and public policy*, Québec: Centre d'analyse des politiques publiques (CAPP), http://www.capp.ulaval.ca/BD.HTM.

Keman, H. (1982). Securing the safety of the nation-state, in F.G. Castles (ed.), *The impact of parties: Politics and policies in democratic capitalist states*. Beverly Hills, CA: Sage Publications.

Klingemann, H.-D., Hofferbert, R., Budge, I. et al. (1994). *Parties, policies and democracy*. Boulder, CO: Westview Press.

Laponce, J. (1981). *Left and right: The topography of political perceptions*. Toronto: Toronto University Press.

Laver, M. & Shepsle, K., eds. (1994). *Cabinet ministers and parliamentary government*. Cambridge: Cambridge University Press.

Lewis-Beck, M. (1988). *Economic and elections: The major western democracies*. Ann Arbor, MI: University of Michigan Press.

Light, R. & Smith, P. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies, *Harvard Educational Review* 41: 429–471.

McKinlay, R.D. (1996). Accumulation, aggregation and eclecticism in political science: A case study of foreign policy analysis, in L.M. Imbeau & R.D. McKinlay (eds), *Comparing government activity*. London: Macmillan.

Midtbø, T. (1999). The impact of parties, economic growth and public sector expansion: A comparison of long-term dynamics in the Scandinavian and Anglo-American democracies, *European Journal of Political Research* 35: 199–223.

Pampel, F.C. & Williamson, J.B. (1988). Welfare spending in advanced industrial democracies: 1950–1980, *American Journal of Sociology* 93(6): 1424–1456.

Pennings P. (1995). The impact of parties and unions on welfare statism, *West European Politics* 18(4): 1–17.

Peters, B.G. (1990). *European politics reconsidered*. New York : Holmes and Meier Publishers.

Przeworski, A. & Wallerstein, M. (1982). Democracy at the crossroad, *Democracy* 2: 52–68.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage Publications.

Rosenthal, R. (1998). Meta-analysis: Concepts, corollaries and controversies, in J.G. Adair, D. Bélanger & K.L. Dion (eds), *Advances in psychological science: social, personal, and cultural aspects*. Hove: Psychology Press.

Sartori, G. (1976). *Parties and party systems: A framework for analysis*. Cambridge: Cambridge University Press.

Schmidt, M.G. (1996). When parties matter: A review of the possibilities and limits of partisan influence on public policy, *European Journal of Political Research* 30: 155–183.

Skinner, J.R. (1976). Technological determinism: A critique of convergence theory, *Comparative Studies in Society and History* 18: 1–29.

Solano, P.L. (1983). Institutional explanation of public expenditure among high income democracies, *Public Finance* 38: 440–458.

Stokke, O., ed. (1989). *Western middle powers and global poverty: The determinants of aid policies of Canada, Denmark, the Netherlands, Norway and Sweden*. Uppsala: Scandinavian Institute of African Studies.

Swank, D. (1988). The political economy of government domestic expenditure in the affluent democracies, *American Journal of Political Science* 32: 1120–1150.

Swank, D. (1998). Funding the welfare state: Globalization and the taxation of business in advanced market economies, *Political Studies* 46: 671–692.

Thérien, J.-P. & Noël, A. (2000). Political parties and foreign aids, *American Political Science Review* 94: 151–162.

Thomas, J.C. (1980). Policy convergence among political parties and societies in developed nations, *Western Political Quarterly* 23: 233–246.

Tufte, E. (1978). *Political control of the economy*. Princeton: Princeton University Press.

van Kersbergen, K. (1995). *Social capitalism: A study of christian democracy and the welfare state*. London: Routledge.

Wilensky, H.L. (1976). *The 'new corporatism': Centralization and the welfare state*. Beverly Hills, CA: Sage Publications.

Wolf, F.M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Berverly Hills, CA: Sage Publications.

Address for correspondence: Louis Imbeau, Centre d'analyse des politiques publiques (CAPP), Département de science politique, Université Laval, Québec, Qc, G1K 7P4, Canada
E-mail: Louis.Imbeau@pol.ulaval.ca