

- Guinaudeau I. and Perisco S. (2014) What is Issue Competition? Conflict, Consensus and Issue Ownership in Party Competition. *Journal of Elections, Public Opinion and Parties* 24(3): 312–333.
- Hobolt S., Klemmensen R. and Pickup M. (2009) The Dynamics of Issue Divergence in Party Rhetoric, working paper, Oxford University, Oxford, UK.
- Kedar O. (2009) *Voting for Policy, Not Parties: How Voters Compensate for Power Sharing*. New York: Cambridge University Press.
- Lin T.-M., Enelow J. and Dorussen H. (1999) Equilibrium in Multicandidate Probabilistic Spatial Voting. *Public Choice* 98: 59–82.
- Mauerer I., Thurner P. W. and Debus M. Forthcoming. Under Which Conditions do Parties Attract Voters' Issue Reactions? Party Varying Issue Voting in German Election 1987–2013. *West European Politics*.
- Merrill S. III and Adams J. (2002) Centrifugal Incentives in Multicandidate Elections. *Journal of Theoretical Politics* 14(2): 273–300.
- Neundorf A. and Adams J. (2014) The Micro-Foundations of Party Competition and Issue Ownership, working paper.
- Petrocik J. (1996) Issue Ownership in Presidential Elections. *American Journal of Political Science* 54(8): 825–850.
- Schofield N. (2007) The Mean Voter Theorem: Necessary and Sufficient Conditions for Convergent Equilibrium. *The Review of Economic Studies* 74(3): 965–980.
- Spoon J.-J. and Klüver H. (2014) Do Parties Respond? How Electoral Context Influences Party Responsiveness. *Electoral Studies* 35(1): 48–62.
- Volkens A., Bara J., Budge I., McDonald M. D., Klingemann H.-D. (eds.) (2013) *Mapping Policy Preferences from Texts III. Statistical Solutions for Manifesto Analysts*. Oxford: Oxford University Press.
- Wagner M. and Meyer T. M. (2014) Which Issues Do Parties Emphasize? Salience Strategies and Party Organization in Multiparty Systems. *West European Politics* 37(5): 1019–1045.

doi:10.1017/S0143814X15000161

## The Comparative Policy Agendas Projects as measurement systems: response to Dowding, Hindmoor and Martin

BRYAN D. JONES

Department of Government, University of Texas at Austin, USA  
E-mail: bdjones@austin.utexas.edu

Dowding, Hindmoor and Martin (hereafter DHM) have contributed a critique of the Comparative Agendas Project (CAP) enterprise, and the editors of the *Journal of Public Policy* have asked me to respond.<sup>11</sup> I am happy

<sup>11</sup> The author appreciates comments on an earlier draft from Frank Baumgartner, Shaun Bevan, Christian Breunig, Laura Chaques-Bonafant, Emiliano Grossman, Christoffer

to have the opportunity, but I shall do so in a somewhat roundabout way. The reason for this indirection is that I intend this essay to be an independent contribution to our understanding of how measurement works in social science – a woefully under-addressed topic, but one that DHM’s essay requires.

Here is how I proceed. First, I develop the general distinction between measurement systems, which maintain time series reliability, and other types of computerised information systems. Then, I present an extended discussion of the queen of measurement systems in the social sciences – the National Income and Product Accounts (NIPAs) – and show why that is a measurement system and how the issues it faces as dynamic economies change is similar to any other measurement system, particularly the Comparative Policy Agendas Projects (PAPs). Then, I turn to DHM’s explicit critiques, showing how they are better understood through the notion of measurement systems.

Many of the points made by DHM are reminiscent of how the Obama Administration itself characterises its foreign policy: “Don’t do stupid things” (although the statement from the White House actually was a little more colourful). On this, I am not only in agreement with DHM, but I also applaud their emphasis here. Some remarks, although basically correct, require a great deal of context to understand their implications. In particular, they imply serious tradeoffs in research designs that cannot be easily managed. All of these tradeoffs stem from the issue of maintaining the PAPs as a measurement system, as I will make clear below. Finally, some of their comments I do not agree with, especially what I see as an unfortunate conflation between CAP and punctuated equilibrium. Before I address any of these, I ask the reader’s indulgence as I provide an extended discussion of what I term “measurement systems”.

## Measurement systems

We are generally familiar with the basics of measurement theory as it is applied to assessments of a single variable. For example, in political science, the concept of party identification has provided a mainstay in the study of mass political behaviour for more than half a century, with vigorous discussions about validity, reliability, random error, bias and the like. With the advent of the ability to assemble information from a very large number of documents and trace this information through time, a new and more comprehensive form of measurement activity has emerged. *Measurement systems* are designed to produce reliable time series information on indicators and address them in a comprehensive way. I explain this below.

Green-Pederson, Will Jennings, Peter Mortensen, Brian Roberts, Stefaan Walgrave, John Wilkerson, Chris Wlezien and Sam Workman.

Obviously, the archived data available for the study of government and public policy have vastly expanded since the advent of digital computing. *Information systems* are systems that allow a user access to a document base through the capacity to “search and retrieve”. The library card catalogue served as the fundamental information system for the access of archived information for 150 years until computerised search engines replaced it. The fundamental goal in computerised information systems is to make it easy for the user to find relevant information. This can be done in two ways: search strings (such as a Google search) and keywords. Searching a large document through search strings is enormously inefficient; keyword search limits the search to what is relevant, but relies on a tight connection between the keywords and the material.

An excellent modern example of an information system whose primary goal is search and retrieve is the US Library of Congress’ Legislative Indexing Vocabulary (LIV). As we noted more than a decade ago: “This indexing system was developed to enable congressional staff and other researchers to identify legislative actions that are relevant to their interests ... LIV currently includes more than seven thousand subject terms, and a given bill can be coded as relevant to several dozen of these terms. While such an approach is desirable for information retrieval, it is practically useless for studying policy trends” (Baumgartner et al. 2002, 41).

The Congressional Research Service (CRS), responsible for LIV, provides a set of keywords that are modified as the nature of Congressional action changes. A team of human coders codes each document (bill, hearing, etc.) according to the keywords, often using multiple keywords for a single document. In adding new terms, the CRS applies it to new material going forward. It does not change the keyword system as it was applied to documents generated in the past.

LIV is not a measurement system for two reasons. First, and most importantly, there is no commitment to the principle of *backward compatibility* – that is, if the search terms are modified or some are added, CRS does not make the new terms consistent with the old ones (Baumgartner et al. 2002, 37). The new term is carried forward, with new documents coded using the new term, but older documents are not re-coded to reflect it. As a consequence, one cannot be certain that any keyword has the same meaning across time. Second, the categories are not *mutually exclusive* because any document can appear in more than one category. This means any category (keyword) that one might want to trace through time can refer to different aspects of a document at different times, and because different numbers of keywords can be used in categorising documents, the proportionality of documents associated with any keyword cannot be calculated. Owing to these two facets of LIV, it cannot be used to generate reliable time series information on

legislative activity. The LIV system does incorporate a third criterion for coding documents – it is exhaustive. That is, every incidence of a hearing, for example, is placed in at least one category.

A major challenge for the use of archived non-quantitative records as the basis of systematic research is transforming information systems into measurement systems – that is, transforming an unreliable set of indicators into a reliable set. It is worth reiterating that reliability here refers only to inter-temporal reliability, that the measure has the same meaning throughout the period of measurement. Even when a document shifts meaning over time, or when different individuals have different readings of the document, the indicators based on that document must have the same meaning. One might have other forms of reliability (such as intercoder reliability) that do not ensure time series measures, because coders may all agree on the (changed) meaning, and therefore undermine intertemporal reliability.

To summarise, a measurement system is a set of indicators derived from documents in which each indicator derived from the documents satisfies the criteria of *backward compatibility* and *mutual exclusivity*. A third criterion, *exhaustiveness*, is not, strictly speaking, necessary to establish a quality measurement system; however, like with any measure, too much missing data will cause reliability problems. For example, the failure to categorise some documents in the system because they are unavailable or for other reasons may not do much harm if these documents are approximately randomly spaced through time. However, if they are concentrated in one period, they can cause potential problems with backward compatibility.

## NIPAs

The gold standard for transforming documents into quantitative reliable time series measures is the NIPAs (Bureau of Economic Analysis 2014). The NIPAs provide “a detailed picture of economic activity at a given time, as well as a consistently defined series of measures over time” (Bureau of Economic Analysis, US Department of Commerce 2007a, 1). NIPAs are “a set of economic accounts that provide detailed measures of the value and composition of national output” (Bureau of Economic Analysis, US Department of Commerce 2007a, 2), granting “a detailed snapshot of the myriad transactions that make up the economy—buying and selling goods and services, hiring of labor, investing, renting property, paying taxes, and the like” (Bureau of Economic Analysis, US Department of Commerce 2007a, 2).

The queen measure of the system is the gross domestic product (GDP), which sums the value of all measured economic transactions in an economy. The ability of economists to measure the size of economies through the analysis of such documents as business regulatory filings and tax returns

and to get the system adopted across the world is a remarkable accomplishment—one of the great milestones in the development of the scientific measurement of social phenomena. NIPA does more: it provides the assessment of the value of economic transactions within industries and, because these measures are reliable across time, allows researchers to trace changes in the sizes of economic transactions by industry. It also allows researchers to trace the flow of activity from one economic sector to others, and to estimate the value added by each industry or sector. Where countries use the same rules for assigning transactions to the industry codes, comparison between countries is also possible.

Underlying this complex system for reducing various data and document sources to indicators is a commitment to backward compatibility – keeping the same transactions in the same categories across time. What happens when economic activity emerges in areas that are not directly assessed by the system? Before 1997, the Federal Government used a set of industry codes, called the Standard Industrial Classification System (SIC), to classify business establishments in order to collect and analyse statistical data relating to businesses. In the late 1990s, a surge of economic growth came from newly developing industries involved in the production of information and communications, areas that were not well-assessed by the SIC. Using existing SIC codes distorted the contribution of this emerging sector. That is, the validity of the measuring categories came into question. The response was to produce a new set of industry codes, the North American Industry Classification (NAICS). Employing these codes in the future, however, would make time series comparisons invalid; therefore the Bureau of Economic Analysis developed a system for converting historical SIC-based data to NAICS (Yuskavage 2007).

A few other aspects of NIPAs are important to this discussion. First, the system is designed to measure economic activity broadly and comprehensively, and yields a large number of variables and indicators. Second, it was not designed to test any particular economic theory, but it has been used to test several. It does rely on a theoretical foundation: input-output analysis, in which various sectors are linked by flows of economic activity. Third, the system does not try to measure what one might cite as the primary function of an economic system. “While GDP is used as an indicator of economic progress, it is not a measure of well-being (for example, it does not account for rates of poverty, crime, or literacy)” (Bureau of Economic Analysis, US Department of Commerce 2007b, 2). Moreover, although it assesses some facets of the non-market economy, such as defence expenditures, it misses wide swaths of non-market activity, such as the care of one’s children at home or black market activity (Bureau of Economic Analysis, US Department of Commerce 2007b, 2). These are vast and productive enterprises; thus, in an

important respect, NIPAs do not measure economic productivity very well. Even worse, if, for example, black market activity increased as a proportion of an economy, GDP would measure different proportions of productivity at different times. What is measured, however, is well-defined, internally consistent and reliable across time.

This brief overview of some aspects of NIPAs points to the key aspects and major challenges in developing and maintaining a measurement system. It also makes clear that, although NIPA and the Policy Agendas system are vastly different, in a major respect they share a common logic – the goal to provide reliable historical time series data. The same issues arise in tracing other activities relevant to policy studies, and I want to highlight these briefly.

### Government budgets

The major issue in tracing public budgets and expenditures (which are not the same thing; see Wlezien and Soroka 2003) is the tendency of public officials to move expenditure items from one category to another, hence violating the principle of backward compatibility. This and other failures to maintain budget data as measurement systems can lead to systematic error, which further leads to major errors in model estimation (Soroka et al. 2006). This usually does not affect aggregate spending, but it can affect breakdowns that would seem to be simple, such as domestic versus defence expenditures. In a study of historical expenditure patterns in the United States (US), my colleagues and I found substantial adjustments were necessary to create a consistent time series for domestic and defence spending (Jones et al. 2014).

In the US, the Office of Management and Budget provides figures for budget authority (appropriations) that are consistent back to Fiscal Year 1977. The US PAP calculated estimates for consistent categories back to 1946 (using Office of Management and Budget categories, not Policy Agendas categories, which was impossible), so that we have a backward compatible tabulation since the Second World War. Although other country projects also provide budget figures, the difficulty in making proper adjustments means that these series may be of varying quality.

Tabulated budgets do not capture elements such as benefits provided by lowering the tax rate, typically referred to as tax expenditures. As is the case for the correspondence between GDP and economic progress, budget experts do not try to estimate the “impact” or “real value” of government expenditures. Note that if one wants to be closer to impacts, one would want to study expenditures, but if one were interested in connecting decision-making processes to budgets one would want to focus on appropriations (or in the US, budget authority). What is measured depends on one’s objective.

## The Comparative PAP

The Comparative PAP consists of 15 country projects, a European Union Project and the US State of Pennsylvania project; others are in process. Each country project tabulates occurrences of various activities, such as US Congressional hearings and roll-call votes, parliamentary questions, executive speeches and laws enacted, within a set of policy content categories and subcategories, arranged hierarchically. That is, each major category is built up from the subcategories below it. Each incidence is assigned to one – and only one – subcategory. Subcategories do not cross major categories – that is, the system is fully hierarchical (Bevan 2014).

The subcategories, and thus the categories, are *mutually exclusive* (each occurrence is placed within one and only one category) and *exhaustive* (all occurrences are placed in a category). In addition, the system adheres to the rule of *backward compatibility*, as discussed above. Owing to these three characteristics, each of the country PAPs constitute a measurement system. In many cases, most particularly the US State of Pennsylvania project, the system is also a retrieval system, but that is not a uniform requirement. Most projects also have a budget component, but those categories are different from the others in two respects. First, they rely on systems of categorisation adopted by governments, not imposed by the research team, and hence do not use a content category system that allows direct comparisons of budget activity with those in other areas of activity. Second, budgets are tabulated in monetary units, not occurrences.

It should be clear that the failure to ensure any of the three primary requirements of a measurement system (mutual exclusivity, exhaustiveness and backward compatibility) will invariably lead to deteriorated or even nonsensical time series, as we have noted before (Baumgartner et al. 2002) and DHM emphasise as well. What that means is a commitment to temporal reliability over other desirable aspects of measurement. A research team may feel that a policy “really” should be assigned to two or more policy content categories, but if the team does so tracing change within the category is no longer possible.<sup>12</sup> Exhaustiveness is less problematic, but it causes the measured series to fail to incorporate all elements of a well-defined measure. Most often, problems in exhaustiveness come from defects in the source material. For example, the US PAP encountered some difficulties with Congressional hearings not released for publication in the public record. Luckily, years later, these were dumped into the system,

<sup>12</sup> There is a possible fix for this, but it is enormously expensive: find some way to weight the occurrence within each of the categories used, and maintain a consistent weight system throughout the series.

requiring us to match the hearing to the year of occurrence, but maintaining exhaustiveness.

Backward compatibility failures occur when a research team detects a new category or subcategory of policy that has emerged with more visibility. This is a validity issue – the categories are not faithfully picking up the full range of policy-making activity in the category. The research team has three choices: maintain the existing category system and, therefore, reliability; change the category system to address validity and destroy the system as a measurement system; or change the system to address validity and do the hard work of restoring backward compatibility (which can involve substantial recoding).

### **Modularity and retrieval**

It should be obvious that this system will not serve all policy researchers' needs, nor is it meant to do so. The PAP is designed to provide consistent time series data on the incidences of various policy-making activities. An investigator using this data may find that the categories do not fit his or her research aim. A researcher may have a broader definition of labour policy, for example, than the projects provide. One option is to take advantage of the system's modularity and combine subcategories from more than one category. Alternately, one might not use all the subcategories from a category. Modularity provides such flexibility with no loss of reliability in the time series structure.

The only alternative if this does not suffice is to recode the source material. In assessing the necessity of recoding documents for a project, the researcher may well be aided. This is because many of CAP data sets allow for direct document retrieval, as policy topic codes are linked to the original document. Therefore, if there are debates over framing or impact, researchers can trace the data back to their source, be it a line of a speech or the title of the law. In these cases, CAP data sets serve both as measurement and retrieval systems.

### **Other measurement systems in political science**

Poole and Rosenthal (2000) have analysed millions of roll-call votes in the US Congress and subjected them to a scaling algorithm, interpreting the results as indicators of ideology. Their clever technique called DW Nominate provides what I have termed above a measurement system, as it provides consistent time series information on the scaled measures. The coding of these votes (at least back to the Second World War) provides an indication of the policy substance of those debates. Whether or not a legislature is ideologically organised may or may not affect the mix of issues decided;



both are important facets of legislative decisionmaking. Another system that would seem to me to qualify as a measurement system is the Party Manifestos Project, which provides estimates of the ideological positions of political parties based on their manifestos. These estimates are intended to be consistent across time and between countries. The new website allows retrieval from the source material. These measurement systems have different but complementary objectives from CAP.

### DHM's critique

Although our detour has been long, the payoff will be dual: it will make clear just how important some of DHM's points are, while putting others in a context that will allow the reader to judge them more completely.

The first issue that DHM raise, appropriately enough, involves what is measured. They write "PAP/CAP generally measures policy attention – what is being discussed in various forums – rather than what the government is actually doing". This is not really correct. I have noted above that the system tabulates incidences of policy activity by content categories. This activity can certainly involve talking about something – as is the case for speeches by the executive. In other cases, it is incorrect, as is the case for counts of laws (which nearly all CAP projects have provided) or regulations (which will soon be available for the US project, thanks to the work of Sam Workman of the University of Oklahoma). In the case of parliamentary questions or Congressional hearings, the measure is certainly "what government is doing", although it is also a measure of what topics the government is paying attention to. In the case of a measure like Congressional roll-call votes, the counts may be viewed as the mix of substantive topics that decisions are made on.

In one important sense, however, DHM have a point: a major research question in CAP has certainly been the role of attention in policy choice. Moreover, they note, "Certainly, the substance of policy will often flow from policy attention but the relationship is far from a perfect one". I agree. They note that attention is different from impact, and attention does not assess distributional effects. I also agree. Nor does the system assess ideology. Right, but systems assessing ideology, such as the Poole-Rosenthal system, do not tabulate policy substance. Yet, whether the issue to be decided is funding infrastructure or financing a new health initiative matters beyond where these issues fall on an ideological continuum, if they do. *In providing a measurement system, one must decide exactly what is to be measured and stick with it.* Studies of the distributional impacts of policies are important, and if they could be assessed across time, this would be even more important. However, PAP/CAP cannot and should not provide that, any more than the NIPAs

should offer measures of social well-being or other assessments of economic impacts. Indeed, the system is fundamentally focussed on decision-making processes rather than other aspects of policymaking. As a consequence, DHM's distinction between the *subject* or content and the *substance* is a useful one, as is the unstated but implied distinction between policy decision and impact, ones that researchers always ought to keep in mind.

DHM offer something they term *implementation style* as a complicating factor for CAP assessments of policy. I prefer to think of this as the policy solution used to address an issue, but the broader point is correct and important. Different political systems may use different solutions (implementation styles) to address the same problem. A comparison of US health policy with European policies offers this issue in stark relief. However, the *subject* of the policy is not in question; hence, providing consistency and comparability in CAP/PAP time series is not difficult. This is generally true, but not always. Is, for example, the US Earned Income Tax Credit both tax policy and welfare policy? It could justifiably be double-coded, but that destroys time series consistency, as I have noted above.

DHM note that the issue of what goals policymakers have in mind can complicate matters; they write that “substantially different policies with the same impacts might be coded differently”. Everything here centres on the term “the same”. DHM write “Two governments might see poverty relief as an important part of their agenda, for example. One might address poverty by means-tested direct welfare benefits. The other might tackle it through a full-employment policy, trying to ensure that everyone has access to work”. Indeed, these two policies would be coded in different categories. Moreover, in all nations, political leaders distinguish between welfare policies and employment policies; therefore, there is little fear that we will miscode them in our system; however, are these different solutions to poverty “the same” in impact? Not at all – at least if we adopt a broader view of impact that includes more than the income levels of participants. The policies empower different bureaucracies, mobilise different constituencies and have different side consequences. For example, although welfare policies may have as a consequence reduction in work efforts, clearly full-employment policies do not.

As I mentioned before, CAP projects were designed to monitor decision-making, including agenda processes, not impact. Again, we return to the key issue in a measurement system: are the indicators of interest reliable across time? There are of course good reasons for policy scholars to study impact, and even more reasons to study the relationship between policy processes and impact, but CAP will help only with the monitoring of policy processes.

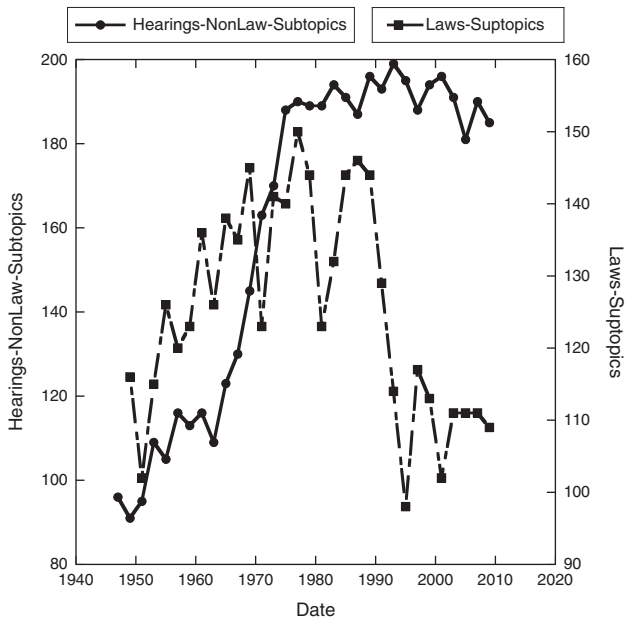
DHM raise the issue on which we are in most agreement when they note that “Issues relating to the boundaries between coding categories become

important in considering framing effects”. As multiple codes for a single occurrence will destroy time series consistency, and yet multiple codes more faithfully represent the policy topic, we face the tradeoff between reliability and validity seriously here. DHM’s suggestion that PAP/CAP codes be supplemented with unsupervised machine coding algorithms is not only correct, but should open up new avenues of research into framing and more: for example, PAP/CAP is best at tabulating established issues, but not as useful in assessing the politics of emerging issues when definitions are still fluid, but unsupervised systems are capable of that. However, of course, they are not useful in providing consistent time series. Another possibility is to code documents at a lower level. As laws are more and more likely to include multiple topics, the US PAP is in the process of coding laws at the title level (titles are subcategories of statute laws).

DHM return to the theme of care in interpreting what is measured again and again, which is quite proper. One interpretation of some of the activities coded, such as parliamentary questions, is attention to one subject rather than others, indicating prioritisation, at least of attention; however, attention may not assess importance, nor is attention necessarily related to what happens later. These are correct cautions, but I would note that PAP/CAP is uniquely suited to study these questions empirically.

To give the reader an indication of just how this might work, I offer the following observations. First, the correlation between the total number of hearings that the US Congress holds and the number of laws enacted is a robust 0.8. Discussion in this case clearly leads to action. Second, examine Figure 1. This is a tabulation of the number of Policy Agendas subtopics in a given Congressional session, with at least one hearing in which a law was not considered and the number of subtopics with at least one law enacted. This is an important measure because it assesses the *scope* of government activity – and over time indicates the displacement of civil society by government activity (Baumgartner and Jones 2015). The scope of lawmaking increased throughout the 1950s, 1960s and 1970s, peaking in 1978 and declining afterwards. This is, doubtless, a result of the conservative countermobilisation to the policy activism “Long Great Society” (Grossman 2014). Hearings not considering new laws can focus on problems that might be addressed in the future and are what are termed oversight matters – overseeing the implementation of laws. As the graph shows, the scope of hearing activity increased as well, but instead of declining, it continues, year after year, at a very high level. Monitoring the vast new programmes created by liberal activism in policy area after policy area requires continual legislative vigilance. Conservatives could stanch the flow of laws, but could not undo what had been done, nor was it productive for them to do so.

This example demonstrates the importance of the allocation of attention, as measured in Congressional hearings. Attention leads to laws. It demonstrates



**Figure 1** The number of policy agendas subtopics with at least one hearing for laws and for hearings not involving a law.

the utility of categorising policy subjects. The more the subjects are addressed over time, the more the scope or reach of the government. It indicates the vast superiority of consistent time series measuring incidences of policy activity by topic, because it allows the tracing of trends across time. Finally, it shows the permanent impact of policy discussions – they not only lead to laws, but to subsequent bureaucratic growth and legislative oversight.

I would note that many other studies of the effects of attention using CAP data might also be cited. Indeed, the authors of the critique are themselves developers of the Australian PAP (which is another reason to take their critiques very seriously), and they have focussed on examining the connection between attention allocation and other aspects of the policy process (Dowding et al. 2010, 2013). Therefore, not only is the connection between attention allocation and outputs an empirical question, it has been vigorously studied using CAP data.

### Theory and measurement systems

Thus far, the reader may have had the impression that I have only trivial disagreements at best with DHM, and that would not be far off the mark.

However, on the question of the connections between PAP/CAP and policy theory, we do have more substantial disagreements.

The relationship between theory and measurement systems is an interesting topic, which unfortunately DHM do not explore; rather, they veer off into a discussion of punctuated equilibrium and bounded rationality, an unfortunate deviation given their generally sophisticated and productive discussion of measurement issues in PAP/CAP and the essential irrelevance of the later discussion to the earlier.

Any measurement system requires some theory, but it does not necessarily require much. Returning to NIPA, economists have tested all sorts of theories of economic growth using this data, because it is amenable to doing so. However, the system is based on input-output analysis, which economist Christ (1955, 137) described as capable of being “regarded as a vast collection of data describing our economic system, and/or as an analytical technique for explaining and predicting the behavior of our economic system”. The technique is based on an input-output table with one row and one column for each sector of the economy, and “shows, for each pair of sectors, the amount or value of goods and services that flowed directly between them in each direction during a stated period” (Christ 1955, 137). Clearly, one must have a theory of economic transactions among sectors, and the NIPA system must categorise an economic transaction such that it can at most be associated with one sectorial interaction.

Similarly, the Party Manifestos system relies on the theory that political parties have preferences or ideological positions, but it does not require that governments put into practice the preferences of the parties of the governing coalition, nor that democratic accountability is established through party ideologies and elections. These are empirical issues. Finally, the PAPs do in fact rely on the idea that policy content matters in the conduct of public affairs – that the choices of governments to allocate scarce resources (from attention to money) to building roads and bridges instead of towards funding armies or welfare – is important. Moreover, policy agendas scholars have emphasised that attention allocation can serve as a unifying approach (not theory) to understanding multiple data sets (Green-Pederson and Walgrave 2014). Just how that works, and whether the process works incrementally or not, can be tested using the data, but the data system itself does not require us to find punctuations or anything else in particular.

As there is no link between PAP/CAP in either the intentions of the founders or in the requirements of the measurement systems, the second part of the DHM essay is both disconnected from the earlier part and unfortunately less useful to the user of the system. First, by no means was the US PAP established to test punctuated equilibrium. It was explicitly established to build the data and measurement infrastructure of policy

studies. Nor is the entire CAP Project some sort of grand attempt to test punctuated equilibrium. We did indeed find evidence of punctuated behaviour in our budget studies, and this allowed us to examine the friction or resistance in a political system as opposed to the static concept of gridlock, but these questions actually became accessible because of the data sets. This may be a fine distinction, but I do not think it serves any purpose to confuse the research agenda of Jones or Baumgartner with the PAP.

When I circulated DHM's article and a draft of this response article to a number of policy agendas colleagues, they all objected to DHM's conflation of punctuated equilibrium theory (PET) with CAP. My colleagues highlighted that most work from the CAP measurement system has had little or nothing to do with PET, focussing instead on such topics as media framing, responsiveness to public opinion, implementation issues, the role of political parties in raising issues to the agenda, the effects of agenda changes on policy outcomes and many others.

It is important that we distinguish carefully between policy punctuations and an increase or decrease in attention. Indeed, the whole notion of agenda-setting implies that attention shifts are not enough. Baumgartner and I postulated that changes in attention were necessary but not sufficient to bring about policy punctuations. While I suppose big budget changes can be associated with attention, more importantly they are big changes in the government's commitment. However, as DHM highlight, every big budget punctuation need not be associated with a big policy change, as one would want to examine the possibility of reversals and the connections between budgetary commitments and major laws. Although the former of these has been examined, if too lightly, the latter has not been. Therefore, I am in complete agreement with DHM's call for qualitative process tracing as a parallel approach to quantitative policy changes documented by the Policy Agendas data.

A final point that DHM make on which I am in disagreement is the idea that somehow the PAP needs an "overarching theory". Indeed, I have absolutely no idea what this is. I do agree that the new study by Bertelli and John conceiving the distribution of issue attention as a portfolio is potentially important for future research – as much for its insistence on treating policy attention as a distribution across all issues as much as its basic conception that such attention is an investment that leads to future returns. However, I cannot see the overarching part – stick-slip dynamics involve the study of institutional and other mechanisms associated with resistance, leading to an explicit prediction of leptokurtic outcome distributions, whereas the Bertelli-John approach leads us to study the calculations of individual political leaders (or maybe political parties). They do different things, and neither was intended to be "overarching"; however, in any case, the Bertelli-John study highlights

what I have been emphasising all along: a quality measurement system allows the exploration of many different ideas grounded in different approaches, theories or empirically derived hypotheses; it definitely does not require some sort of straitjacketing “overarching theory”.

### Concluding comments

DHM have done the Comparative Policy Agendas community a favour by raising explicitly both the potential and pitfalls of using the now-vast data sets being assembled, coded and analysed by an increasing number of country teams of scholars interested in the comparative potential of the data sets. I have tried, here, to put their observations in a broader context by developing the concept of a measurement system and showing how the demands of maintaining such a system open up the vast potential of systematic time series analysis, but simultaneously limit the scope of the questions that can be addressed by using the data sets. Many projects should and have supplemented the Policy Agendas data with other information, quantitative or qualitative.

I have further shown that a good measurement system requires some theory, but not much. Obviously, one has to have some notion of what is relevant to the system, but there is no need for the imposition of some sort of theoretical orthodoxy to make sense of the data.

### References

- Baumgartner F. R. and Jones B. D. (2015) *The Politics of Information*. Chicago, IL: University of Chicago Press.
- Baumgartner F. R., Jones B. D. and Wilkerson J. D. (2002) Studying Policy Dynamics. In Baumgartner F. R. and Jones B. D. (eds.), *Policy Dynamics*. Chicago, IL: University of Chicago Press, 29–46.
- Bevan S. (2014) *Gone Fishing: The Creation of the Comparative Agendas Project Master Codebook*.
- Bureau of Economic Analysis, US Department of Commerce (2007a) *An Introduction to the National Income and Product Accounts*. Washington, DC: Bureau of Economic Analysis, US Department of Commerce.
- (2007b) *Measuring the Economy: A Primer on GDP and the National Income and Product Accounts*. Washington, DC: Bureau of Economic Analysis, US Department of Commerce.
- (2014) *NIPA Handbook: Concepts and Methods of the U.S. National Income and Product Accounts*. Washington, DC: Bureau of Economic Analysis, US Department of Commerce. <http://www.bea.gov/national/pdf/chapters1-4.pdf>.
- Christ C. (1955) A Review of Input-Output Analysis. In Goldsmith R. W. (ed.), *Input-Output Analysis: An Appraisal*. National Bureau of Economic Research Princeton, NJ: Princeton University Press, 137–182.
- Dowding K., Hindmoor A., Iles R. and John P. (2010) Policy Agendas in Australian Politics: The Governor-General Speeches, 1945–2008. *Australian Journal of Political Science* 45: 533–557.

- Dowding K., Hindmoor A. and Martin A. (2013) Attention, Content, and Style. *Australian Journal of Public Administration* 72: 82–88.
- Green-Pederson C. and Walgrave S. (eds.) (2014) *Agenda-Setting, Politics, and Political Systems*. Chicago, IL: University of Chicago Press.
- Grossman M. (2014) *Artists of the Possible*. Oxford: Oxford University Press.
- Jones B. D., Zalányi L. and Érdi P. (2014) An Integrated Theory of Budgetary Politics and Some Empirical Tests: The US National Budget, 1791–2010. *American Journal of Political Science* 58: 1–18.
- Poole K. T. and Rosenthal H. (2000) *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.
- Soroka S., Wlezien C. and McLean I. (2006) Public Expenditure in the UK: How Measures Matter. *Journal of the Royal Statistical Society A* 169(Part 2): 255–271.
- Wlezien C. and Soroka S. (2003) Measures and Models of Budgetary Policy. *Policy Studies Journal* 31: 273–286.
- Yuskavage R. (2007) Converting Historical Industry Time Series Data from SIC to NAICS. Paper presented to the 2007 Research Conference of the Federal Committee on Statistical Methodology, Arlington, VA, 5–6 November.

doi:10.1017/S0143814X15000173

## Attention, content and measurement: rejoinder to Adams and Jones

KEITH DOWDING

*Department of Political Science, Australian National University, Australia*  
E-mail: keith.dowding@anu.edu.au

ANDREW HINDMOOR

*Department of Politics, University of Sheffield, UK*  
E-mail: a.hindmoor@sheffield.ac.uk

AARON MARTIN

*Department of Political Science, University of Melbourne, Australia*  
E-mail: aaron.martin@unimelb.edu.au

It is a pleasure to respond to the thoughtful comments of Jones (2015) and Adams (2015). We first want to explain the motivation behind our review article. As we began a project on policy agendas in Australia, we set our intentions to audiences at conferences in Australia, Europe and the United States and ran a workshop with policy specialists in Australia (which resulted in a special issue of the *Australian Journal of Public Administration* 2013). What we discovered was that many of the political scientists we were talking to were extremely sceptical about the value of the Policy Agendas