The case for an Evaluator General: Why we need it and how we should configure it

Submission to the Thodey Review of public administration

November, 2018

> *Everyone is entitled to his own opinion, but not his own facts.*
>
> Daniel Patrick Moynihan

> *The first principle is that you must not fool yourself and you are the easiest person to fool.*
>
> Richard Feynman

Nicholas Gruen [1]

[1] Thanks to numerous others for comments on earlier drafts including Alexandra Ellinson, Keryn Hassall and Gene Tunny.

Foreword

I am presently on an extended tour of Australian cities as the National President of the Institute of Public Administration Australia (IPAA). It's an opportunity, in part, to discuss with officials from all jurisdictions how they imagine public service reform.

In a sense, the future is already with us. I am using the chance to highlight the many new and exciting approaches already being trialled - the utilisation of big data analytics, the embrace of behavioural psychology, the piloting of place-and community-based initiatives, enhanced cross-sectoral collaboration, the adoption of human-centred design, the collective creation of public impact, the introduction of digital democracy and the application of robotic process automation and cognitive technologies to complex but routine administrative tasks.

Change is afoot. Contrary to public perception, there is a great deal of innovation occurring at all levels and in many areas of public administration. Always, however, the audience at my talks end up debating why it is that so many of these creative improvements to our structures of democratic governance remain confined to its periphery. Too many demonstration projects, even when successful, fail to get scaled up. Pilot programs remain pilots. At the centre of public service, traditional approaches to policy implementation are rarely transformed. The existing state of affairs continues.

Perhaps, suggest some of the participants, that it reflects the inability of public servants to gain a positive authorising political environment. Perhaps, posit others, it is evidence of bureaucratic risk aversion. Perhaps there are vested interest in the maintenance of the status quo. Whatever the combination of factors, it is disheartening.

That is why this provocative discussion paper by Nicholas Gruen is so important. At the heart of the problem, he argues persuasively, is the lowly status accorded to the task of evaluation and to the perspectives and knowledge of those in the field and those they serve. That has detrimental consequences. We need to know where governments can invest public funds in order to best achieve the outcomes they seek.

That assessment needs to integrate systematic with contextual knowledge. In my own words, commitment to evidence-based advice needs to be informed by the pragmatic, real-world experience of front-line staff and by the citizens who access the services provided.

Gruen argues that the single best way to enhance the monitoring, appraisal and reporting of the delivery of government policy is to raise its profile, influence, authority and independence. To that end he contends that there would be significant advantage in establishing in Australian public services the role of an independent Evaluator-General alongside the existing Auditor-General. I think he’s right.

Gruen emphasises that his proposals would raise the status of evaluators and their work (and, I believe, would also enhance the

professional standing of the Project and Risk managers upon whom effective execution of government policy depends). Yet, to be successful, officers of the Evaluator-General embedded within agencies would also need to be ‘critical and expert friends’ of those delivering services, especially those in the field. They would not be top-down enforcers and regulators.

A critical aspect of Gruen’s proposal is that the Evaluator General would be in a position to independently compare the efficiency and effectiveness of different programs and approaches, whether undertaken by public servants or by contracted outside providers. This would help move public debate from its present preoccupation with government expenditure to a greater emphasis on measuring the financial and social returns on government investment.

Finally, Gruen proposes – in line with his long-standing advocacy for making publicly collected data available to the public – that, by making its independent monitoring and evaluation transparent to the public, an Evaluator-General could create a ‘knowledge commons’ of assessment methodologies and outcomes. Better practice would be shared across government agencies and service providers. Public understanding of performance-based outcomes (and the metrics necessary to measure them appropriately) would be enhanced. Critically, new programs and approaches can be assessed on a ‘level playing field’ with the incumbent ones that have proven so difficult to dislodge. The transparency of evaluation results would help build public support for difficult decisions where they are necessary.

I commend this paper by Nicholas Gruen. Its line of reasoning is convincing. Understanding that the quality of government policy can only be assessed by the manner in which it is delivered, it makes a bold but practical proposal on how to improve that process over time. I hope that his arguments are widely read, discussed … and implemented.

Professor Peter Shergold

Secretary to the Department of Prime Minister and Cabinet from 2003 to 2008

5th August 2018

Overview

The independent and the politically directed arms of the executive

Public sector services can be provided by either the 'politically directed executive' under the supervision of a political office-holder – in Australia's system a minister – or the 'independent executive' which typically reports to the legislature independently of political direction.

Agencies are often situated within the independent executive where they provide information or support the integrity of information and conduct more generally. Thus the Auditor General defends the integrity of government finance and wider systems. Others, such as the Surveyor General, bureaus of statistics or meteorology or the recently established Parliamentary Budget Office (PBO) form part of our informational infrastructure.

The establishment and expansion of Australia's PBO alongside the Federal Treasury illustrates the principles at work. It was established to provide parliamentarians with government funded fiscal policy expertise previously available only at the direction of the government through Treasury.

The current Opposition's proposal to move responsibility for the Government's economic forecasting from the Treasury to the PBO illustrates some principles by which functions should be allocated between the 'ministerially directed' and the 'independent' executive. As 'spin' engulfs political debate, and bureaucrats are increasingly drawn into assisting their political masters 'perform' government,[2] it seems sensible that this work be insulated from political direction or undue influence and for it to be seen to be so.

This paper takes these principles further in pursuit of evidence-based learning in policy and delivery. Just as, under the Opposition's policies, Treasury's advice to its Treasurer would be based on independent expertise about the *actual facts and most plausible futures* supplied to it by the independent PBO, similar principles apply more broadly to monitoring and evaluating the delivery of government funded services.

The Evaluator General

This paper proposes an Evaluator General,[3] an independent statutory agency having investigative and reporting powers similar to the Auditor General, though in the area of monitoring and evaluation rather than audit. Its existence would promote the profession of evaluation which,

---

[2] See Gruen, 2012, "How performing government is trumping transparency" http://clubtroppo.com.au/2012/03/12/secrecy-by-default-how-performing-government-is-trumping-transparency/

[3] These ideas were first set out in this two part essay. Gruen, N. 2016. "Why we accept travesties of 'evidence-based' policymaking", *The Mandarin, 9 May* at http://www.themandarin.com.au/64557-nicholas-gruen-evidence-based-policy-part-one/ and Gruen, N. 2016. "Why Australia needs an evaluator-general", *The Mandarin,* 9th May at http://www.themandarin.com.au/64566-nicholas-gruen-evaluator-general-part-two/

unfortunately enjoys far lower professional status and visibility than economics, accounting, audit or even public policy.

However, its role goes well beyond *sitting atop* government monitoring and evaluation systems. In this paper it is envisaged *as the institution through which a new demarcation would be operationalised* between program *delivery* on the one hand and resourcing expert knowledge on program performance on the other. Thus a line agency might deliver a program – or commission third parties to deliver it – but the Evaluator General would direct and provide substantial resources to the monitoring and evaluation system constituting the program's 'nervous system'.

Thus, monitoring and evaluation would be designed and operated *in the field by officers of the Evaluator-General*. For this to work well they and the delivery agency would need to collaborate closely. However the Evaluator General would have ultimate responsibility for monitoring and evaluation in the event of disagreement.

The Evaluator General would ensure that monitoring and evaluation outputs were available first and foremost to service deliverers to assist them optimise their performance. But subject to privacy safeguards, the Evaluator General would also make public the monitoring and evaluation system's outputs together with appropriate comment and analysis.

The objectives of the new arrangements

The finely disaggregated transparency of performance information made possible by this arrangement would support;

- the intrinsic motivation of most of those in the field to optimise their impact by building their own 'self-transparency' on their transparency to an impartial spectator and 'expert critical friend';
- public transparency to hold practitioners and delivery agencies to account;
- more expert and disinterested estimates of the long-run impact of programs to enable a long-run 'investment approach' to services; and
- a rich 'knowledge commons' in human services and local solutions that could tackle the 'siloing' of information and effort within agencies.

Further, by publicly identifying success as it emerged, an Evaluator General would place countervailing pressure on agencies to more fully embrace evidence based improvements even where this disturbed the web of acquired habits and vested interests that tend to entrench incumbency. Thus the tendency Peter Shergold laments for "too much innovation [to] remain at the margin",[4] might be ameliorated.

---

[4] Shergold, Peter, 2013. "My Hopes for a Public Service for the Future", *Australian Journal of Public Administration,* Vol 72, No. 1., pp. 7-13. 14 April https://doi.org/10.1111/1467-8500.12006

With journalism and political debate increasingly given over to spin, the public sector can strengthen its own independence from this process and help fill the gap left by the retreat of public interest journalism by strengthening the expertise, resourcing, independence and transparency of the evidence base on which it proceeds.

Implementation

As has been highlighted since at least the Moran Review, “Ahead of the Game” in 2010, governments need to increase substantially their investment in monitoring and evaluation. The establishment of the Evaluator General would be a good occasion on which to commit to this and would provide the appropriate institutional environment in which it should take place. Even without this deficit, it would take time to fully implement the vision set out here.

Accordingly, a substantial period should be set to move towards the vision. Five years would be a reasonable time. Priorities should be set to gain early experience of crucial aspects of the complete model being proposed. Thus in high priority areas experience should be gained in monitoring and evaluation of new and innovative programs and also with comparing their efficacy with incumbent systems. Capacity should be expanded to ensure the commitment to evidence-based policy is realised throughout the public service over a period of five years.

Introduction

In a wide range of areas of professional practice and service delivery much of the important knowledge and knowhow exists at the 'edge' of the system rather than at the centre. This is the case in most areas in which governments mandate the delivery of human services such as health, education, assistance in employment search, disability, aged care, corrective services, policing, child-protection and the list goes on.

The embrace of 'community based' approaches suggests some recognition of this, but learning from experience in the field is difficult and progress has been slow. The next section sets out the central policy problem; to improve the quality and flow of information, knowhow and learning between the bottom and the top of the system. We then explore the principles by which one would decide the ideal placement of public sector resources between ministerially directed and the independent executive. This then leads us to elaborate the functions of the Evaluator General.

But first a story.

Box 1: Decisions at the centre: learning at the edge

The Australian Centre for Social Innovation (TACSI) explored a previous government initiative to reunify children with families from whom they'd previously been removed because of abuse and neglect. This required identifying the best prospects and skillful support.

But inadequate regard for building and managing the necessary skills saw average success rates of 30 percent. In fact TACSI's investigations revealed that in one suburb, a team assembled the necessary skills to increase its success rate to 85 percent. However with such poor results being delivered across the state, the initiative was abandoned. The learning achieved by the successful team was overlooked by senior leadership and it was ultimately disbanded.[5]

Information within markets and centrally planned systems

As Friedrich Hayek put it, it is "more than a metaphor" to describe the price system within the market as a kind of telecommunications system

> Without an order being issued, without more than perhaps a handful of people knowing the cause, tens of thousands of people whose identity could not be ascertained by months of

[5] Weinstein, Lauren, 2016, Systematising the Steps from Evidence to Impact: Amplifying the Royal Commission's Recommendation for an Early Intervention Research Directorate, Sep 30. Previously at http://tacsi.org.au/systematising-steps-evidence-impact/ now available at https://web.archive.org/web/20170222035732/http://tacsi.org.au/systematising-steps-evidence-impact/. Declaration of interest, I was chairing TACSI at the time.

investigation, are made to use the material or its products more sparingly.

For profit seeking firms even where markets are quite imperfect and prices do not capture various externalities, they nevertheless offer considerable intelligence (transparency) about the economic value of specific goods or services and comparative costs between suppliers. And this helps keep managers of firms to the (usually) honest toil of satisfying their customers at minimum cost to themselves.

It's much harder to effectively govern many government funded human services, both conceptually and practically. First, in the human world, interactions are more complex and uniquely situated within a context than is typically the case in most markets. So it's much harder to know what works, how and why it works, how different things going on in different parts of the system affect each other, the difference between long and short-term impacts and so on.

Second, many human services aren't funded by paying consumers. This robs the resulting systems of all the transparency and feedback around cost, price and quality that emerges from the organic tension between buyers and sellers in a market. This also dulls cost-minimising incentives. Budget caps can be imposed from above, but it's a difficult business refining them appropriately down through the hierarchy.

Hayek's challenge to Soviet central planning has, thankfully, been won. Unfortunately however, a preoccupation with this issue diverted him and his followers from exploring the wider relevance of his ideas. The problem remains a live one for all firms – which are necessarily centrally planned,[6] though more urgently for larger ones.[7] By contrast even in Hayek's preferred world, and even more so the actual world we inhabit, governments remain large, centrally planned organisations largely bereft of disciplines or even guidance from market prices given that their revenue is predominantly through taxation and their expenditure is so heavily skewed towards public good and benevolent activities.

Centrally planned services must build the best system they can to generate high quality, granular information and incentives to optimise the value they deliver. In today's governments, system managers – often doing the bidding of politicians [8] – typically specify performance measures, often in a relatively unconsidered, even perfunctory way. However even if those at the top put more serious effort into designing measures with some diagnostic power for those seeking to optimise a

---

[6] By definition an 'organisation' is unitary and it is in this fact – this need to make singular decisions about what 'it' will do (whether to perform some activity or not, whether to hire or fire a particular person or a class of people) that it is centrally planned however autocratically or democratically it is managed.

[7] Note that even here they are disciplined by the market within which they operate and wherever smaller firms can compete with them.

[8] See eg. Mathis, W.J. & Trujillo, T.M., 2016. "Lessons from NCLB for the Every Student Succeeds Act" Boulder, Colorado: *National Education Policy Center*. Retrieved [23rd may 2018] from http://nepc.colorado.edu/publication/lessons-from-NCLB.

program's impact, they will often require intimate knowledge of the workings of the program in the field.

Box 2: How 'top-down' approaches to metrics can impede program learning

A foreign aid agency funded a program training agricultural extension workers to spread good agricultural practice to East Timorese farmers. However, most extension workers departed their agricultural extension work for higher paid jobs elsewhere. Given this, the contractor implementing the project suggested that the training should be delivered direct to farmers. The aid agency's representatives in East Timor agreed that this the new strategy was likely to work better.

At the agency's headquarters, the project appeared successful with the training of the government workers despite the fact it was failing to serve the original aim of helping farmers. Asking superiors to approve a revision of strategy would make it clear that the project was not serving its development goals. Failures had to be explained to senior agency officials and politicians to whom the agency reported. It meant a career risk for field agents wanting to improve the project. The existing strategy remained in place with many more trained extension workers and little discernible impact on farmers.

See Honig, 2018, pp. 3-4.[9]

For instance, an appealing measure of the performance of a job placement program would be how many job seeker clients found jobs. For a child protection program, it might be the number of children the system was forced to remove from struggling families following early intervention. But are these the right measures? Could too rapid job matching destroy value by foreclosing better matches, or by diverting valuable system resources to where they are redundant? And how does one weigh up the relative merits of child removal with poor home care?

Note that these observations apply to *the system* within which the program is delivered. In this sense the question of whether services should be delivered 'in-house' by lower levels of the hierarchy or by contractors is a separate question. In this sense the question of contracting out might reasonably be thought to be subordinate to the prior question of how well the system really understands its own performance. And yet somehow discussions of contracting out typically make little of this point,[10] as if theoretical attention to the nature of the

[9] Dan Honig, 2018. *Navigation by Judgment: Why and When Top Down Management of Foreign Aid Doesn't Work*, Oxford University Press, pp. 3-4.

[10] See Gruen, N and Vanstone, C, 2016. "Competition as a means to an end" parts one and two in The Mandarin, on 18th August at https://www.themandarin.com.au/69098-competition-markets-supply-chains-human-services/ and 19th August at https://www.themandarin.com.au/69140-competition-means-end-supply-chain-needs-brain/

market in which contracting will take place can somehow substitute for the system's practical knowledge of its impact.

Bureaucracies have a terrible habit of earnestly role-playing the tasks they've been given, while in reality going through the motions and responding to institutional imperatives. Thus as Lewis Hawke reports regarding performance budgeting in Australia:[11]

> both sides of the political spectrum have been concerned about strengthening the transparency, quality, and relevance of performance information to its intended purposes. Despite the unanimous agreement on its importance, this has proved to be an endless challenge.

In all this, performance measures imposed from the top sound like a mistake waiting to happen.

'Top down' and 'bottom up' information, knowhow and learning in human services

As early neoliberals like Michael Polanyi (from the left) and Friedrich Hayek (from the right) stressed, to function well, many systems must integrate both systematic knowledge – of the kind learnt from books and at university – with the local, practical, contextual and often tacit knowledge in the field.

[11] Donald Moynihan and Ivor Beazley, 2016, *Toward Next-Generation Performance Budgeting: Lessons from the Experiences of Seven Reforming Countries*, World Bank Group, https://openknowledge.worldbank.org/bitstream/handle/10986/25297/9781464809545.pdf?sequence=2&isAllowed=y, p. 50.

Box 3: Fred Chaney and Ian Marsh on indigenous programs [12]

Indigenous citizens must be engaged in the decisions that affect them, something that is impossible under existing highly centralised approaches to administration. But because taxpayers' money is involved, accountability also needs to be sustained. And efficiency requires attention to the effectiveness of programs and to learning, innovation and improvement. … An effective administrative architecture needs to square the circle between three not immediately compatible outcomes: local engagement, continuous improvement and central accountability. It must do this in a context in which responsibilities are shared between federal and state governments. …

There is no doubt evaluation is absolutely critical. But the question concerns its form. Think of the challenge of school attendance. This looks straightforward. But … the causes of non-attendance are many and varied: … mobility of families, deaths, funerals, sorry business, violence in the communities, sporting events, carnivals/shows, overcrowding, street parties at night … other cultural practices.… [C]ritical factors could be grouped … into family, community, school, governmental, cultural, economic, and other categories.

An approach that will work in Alice Springs is most unlikely to work in Redfern or Aurukun. Contextual factors will be critical. This is one reason local engagement is essential. The chances of there being anything even remotely resembling universal "best practice" are zero. This would throttle effective local service design. This is not to say that there will not be opportunities for learning and for the exchange of experience. But the learning that will apply across sites will be adaptive — not technical or codified.

In a context in which experience accumulates and circumstances change, learning will be continuous and dynamic. The surrounding system needs to enable these processes [and] be based on learning by doing.… Set objectives, yes, but all solutions are local and actions must be relevant to local circumstances. So the surrounding administrative system needs to invest in the capacity of local agents to learn from and improve their own efforts. This is the path to continuous improvement

As well as focusing on outcomes, the surrounding administrative design also must focus on the means used by different providers who are working towards broadly similar ends.… Monitoring should focus not only on outcomes but also on associated learning processes. These occur within and between providers as well as across agencies. The present box-ticking culture of compliance is wholly at odds with such relationships.

[12] Chaney, F, 2016. "Closing the gap between good intentions and results" The Australian, August 26. Fred Chaney indicated that Ian Marsh was a close collaborator.

Here we encounter a profound structural problem. Systematic knowledge is far more prestigious than the contextual knowledge in the field on which it's application so often depends. As Hayek put it in the late 1940s:

> Today it is almost heresy to suggest that scientific knowledge is not the sum of all knowledge. But a little reflection will show that there is . . . a body of very important but unorganized knowledge: . . . of the particular circumstances of time and place. [Here] practically every individual has some advantage over all others because he possesses unique information of which beneficial use might be made, but of which use can be made only if the decisions depending on it are left to him or are made with his active co-operation.[13]

Here Hayek bemoans the *cultural* supremacy of systematic knowledge over pragmatic knowledge but as he was well aware, this supremacy is reinforced by *power*. Those in the 'centre' of a centrally planned system are also formally at its commanding heights. As Donald J. Savoie put it recently:

> The ambitious know full well that the road to the top is through policy, generating ideas, managing the blame game, being visible in Ottawa circles, and central agencies, not through program management.[14]

The Secretary of the Australian Department of Prime Minister and Cabinet Peter Shergold suggested in 2005 that if there were "a single cultural predilection in the APS" he'd change, it would be "the unspoken belief that the development of government policy is a higher function – more prestigious, more influential, more exciting – than delivering results".[15] One upshot of this, as he observed eight years later, is that too much innovation remains at the margin of public administration, opportunities half-seized, new approaches remaining forever in 'pilot' mode".[16]

---

[13] Friedrich A. Hayek, 1945. "The Use of Knowledge in Society", *American Economic Review*. XXXV, No. 4. pp. 519-30.

[14] Donald J. Savoie, 2015, "What Is Government Good At? A Canadian Answer".

[15] http://www.apsc.gov.au/publications-and-media/current-publications/learning-from-failure/enhancing-project-management#_edn157

[16] Shergold, Peter, 2013. "My Hopes for a Public Service for the Future", *Australian Journal of Public Administration,* Vol 72, No. 1., pp. 7-13. 14 April https://doi.org/10.1111/1467-8500.12006

What functions should be independent?

Just as Adam Smith described economic growth as coincident with growth in the division of labour, so the growth of government since the early twentieth century has been coincident with the expansion of the number of separate government agencies. And many of these new agencies have been insulated from political direction (See Appendix). This section considers what criteria should determine what government functions should be part of this 'independent executive'?

Where an agency produces purely factual information, the quality of that work (including choices regarding what information is gathered and it is analysed) is probably enhanced by independence. Moreover there's a strong case for that information to be made publicly available in as rich and immediate form as possible.

As a matter of constitutional principle this provides a common factual or analytical basis on which public discussion can take place – including debate as to the quality of the information and analysis provided – which is less party-politicised. This in itself contributes to the transparency and accountability of government, particularly where the information and analysis provided by an agency relates to governments' success or otherwise in serving the public. Alongside this appreciation for the role of transparency in *political and institutional hygiene* has grown an appreciation of a corresponding *economic* case for openness.[17] Because information is indefinitely reusable, it is economically best to share it widely.[18]

As analysis is built on information provision, we can distinguish broadly between positive and normative analysis. The distinction is well illustrated in the Australian Labour Party (ALP) Opposition's proposal to move principal responsibility for economic forecasting from the Treasury to the PBO. Here it is intended that recommending the best policies remain with Treasury but that its advice rely on a common forecasting resource. It is hard to see any grounds on which such work should not be done independently of political direction and made available publicly. Indeed, there's a good case for going further than is proposed in the ALP policy.

---

[17] See Gruen, N. 2010. "Government 2.0 openness as micro-economic reform", at http://clubtroppo.com.au/2010/03/03/government-2-0-openness-as-micro-economic-reform/

[18] In economic jargon information is 'non-rival'. These considerations are captured in the objects of Australia's 2010 Freedom of Information Act, a comprehensive revision of Australia's initial FOI Act of 1982. In addition to recognising the constitutional value of government information in promoting "representative democracy by … increasing public participation, ... promoting better-informed decision-making and …. increasing scrutiny, discussion, comment and review of the Government's activities", the Act references the economic value of openness thus:

> The Parliament also intends, by these objects, to increase recognition that information held by the Government is to be managed for public purposes, and is a national resource.

Freedom of Information Act, 1982, Section 3, Objects--general at http://www5.austlii.edu.au/au/legis/cth/consol_act/foia1982222/s3.html

Box 3: Science, openness and open source economic modelling [19]

Economic and financial models by which government agencies forecast developments and seek to answer 'what if' questions offer an analytical framework within which normative questions can be asked such as "what policy would be most cost effective".

A range of benefits would flow from opening up such analysis to public scrutiny which are analogous to the benefits of openness in science and the open sourcing of software.

- Greater capacity utilisation and so greater efficiency from governments' sunk cost involved in building models.
- Wider public understanding of the public policy issues and constraints;
- Public critique of weaknesses of the modelling;
- Contributions from the community to improving the models.

In science, most governments already embrace the spirit of these principles by requiring researchers who receive government funding to openly publish the resulting research and data.[20]

Analysis of how to improve policy

Governments are in frequent need of independent but 'official' advice which they often obtain from independent sources either via some agency tasked with doing so – as in the case of the Australian Productivity Commission – or from some *ad hoc* body such as a royal commission or other commission of inquiry.

Further, agencies that are already independent may engage in normative analysis and arguably may wish to publish it. Indeed, if we are to take current practice as a guide, independent agencies may offer some public guidance on their own views of certain matters as for instance when a central bank governor may comment on some economic matter such as fiscal or trade policy which is outside of their direct, formal responsibility.

Despite all this however, governments also need access to expert advice and analysis from their officials from which they can take private counsel. This analysis is currently provided for them within departments of state subject to the constraints of freedom of information regulation and for eventual disclosure after some statutory archival period and these arrangements should continue.

---

[19] See also Gruen, N., and Kamper. N. 2016 "What if the crowd forecast the economy for Treasury?", *The Mandarin*, 8th April, available at https://www.themandarin.com.au/62739-gruen-kamper-treasury-forecasting-models-open-source/

[20] In fact this agenda is constantly frustrated by commercial interests in academic publishing.

By contrast, the delivery of services remains the province of line agencies subject to direction by ministers of the government of the day. While the government of the day or the legislature may seek to delegate some service delivery to independent agencies or otherwise contract it out, political accountability in our democracy is built on the ability of politicians to make promises to their constituents and to then obtain the assistance of the bureaucracy to deliver them.

An Evaluator General [21]

We can infer from the previous analysis that independent evaluation should be introduced for all government programs, and that such evaluations should be made public as a matter of course. Several authors recently published a paper "An Evaluator General for Canada" which would be to evaluation what an Auditor General is to audit in the public sector.[22] Where an Auditor General audits agencies' the quality, documentation and adherence to policies, procedures and systems of accountability, an Evaluator General in this mould would evaluate the social benefit of government activity. It would use wider sources ranging from agency documentation and data to the standard tools of evaluation such as surveys, direct observations, interviews and focus groups with a focus on.

Like the Auditor General, the Evaluator General proposed for Canada would have its influence from the top down:

> The Evaluator General, like the Auditor General, would function at a strategic level. Responsibility for the regular evaluation of departmental programs would continue to rest with the internal evaluation units that are currently within departments and agencies. Of course, if Parliament wanted the Evaluator General to attest to the quality of major departmentally-conducted evaluations (just as the Auditor General is required to attest to the annual financial reports of some agencies) then the Evaluator General would provide an opinion on the extent to which the evaluations met professional

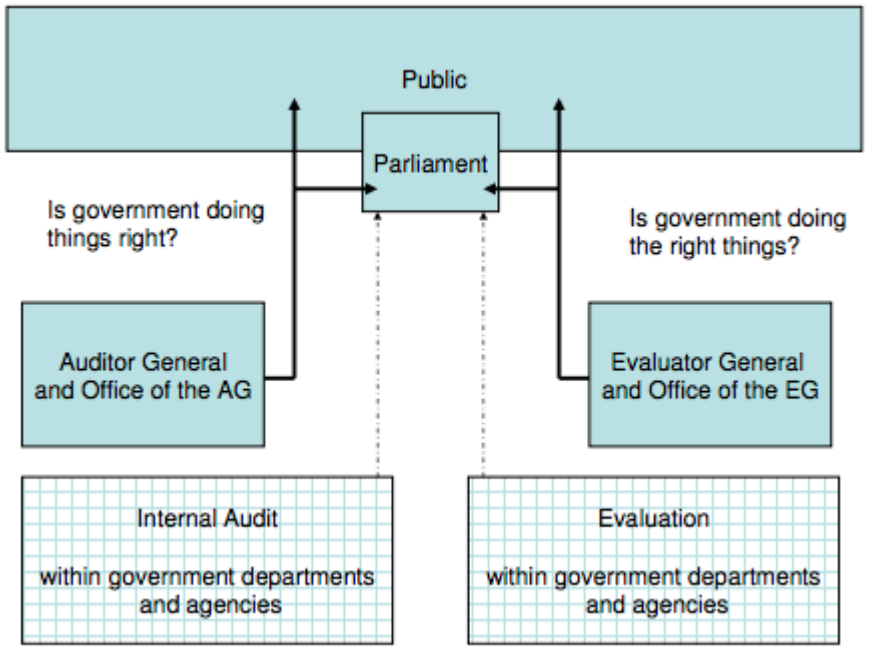
---

[21] See also earlier essays sketching out the case for, and the role and functions of an Evaluator General. See this two part essay "Why we accept travesties of 'evidence-based' policymaking" followed by "Why Australia needs an evaluator-general" in *The Mandarin,* 9th May 2016 at http://www.themandarin.com.au/64557-nicholas-gruen-evidence-based-policy-part-one/ and http://www.themandarin.com.au/64566-nicholas-gruen-evaluator-general-part-two/ respectively. These ideas were further elaborated in "Markets, supply chains, brains and human services" The Mandarin, *ClubTroppo*, August, 2016 at http://clubtroppo.com.au/2016/08/19/markets-supply-chains-brains-and-human-services/. A few passages in this paper are derived from passages in these articles.

[22] D'Aloisio, Guy *et al,* Undated, "An Evaluator General for Canada – A Solution for Filling the Accountability Void?" at https://evaluationcanada.ca/affichage/egc_e.pdf

standards. But overall, the responsibilities of departmental evaluation units would be unchanged.

Figure One: Proposed Evaluator General for Canada



Source: D'Aloisio, Guy et al*, 2007.*[23]

Imposing constraints and existing institutional imperatives: The spectre of unintended consequences

We should be alive to the possibility of unintended consequences when we seek to somehow discipline agencies from the outside. For the discipline imposes a new tension on the agency – between the institutional imperatives that presumably drove the problem we are seeking to address and the new constraint. Thus for example when freedom of information regulation is introduced, agencies usually adjust the candour of their record keeping accordingly.[24]

Further, the practice of evaluation is substantially more open to differences of opinion and approach than auditing. This suggests that the undesirable effects of existing institutional imperatives would be

[23] D'Aloisio, G. et al. "*Do we need an Evaluator General?",Canadian Government Executive,* Vol. 13, No. 7, *(September 2007), pp. 14-15*

[24] Likewise constraints are often placed on agencies in ways that can treat the symptoms, but not the causes of problems. For instance, to ensure the rights of families are respected, child protection workers will often be required to get a court order to remove a child from its family home for abuse and neglect. But if courts require child protection workers to exercise strong burdens of proof before removal, they will anticipate this in deciding which children to bring before the court. Child protection services may then be blamed for failing to apply for a court order when a child comes to harm, when in fact the wider system is at fault. The result is often learned helplessness within the agency.

more resilient to the introduction of a 'top down' constraint in auditing than they would be to the introduction of a 'top down' discipline on evaluation.

Were an Evaluator General to mirror the structure of the Auditor General, imposing some formal constraints on evaluation from the top down and from outside agencies, agencies might formally comply with some policy of meeting "professional standards" in evaluation whilst in fact managing both the commissioning of monitoring and evaluation and release of their outputs according to their current imperatives. Indeed it is relatively commonplace for operational business administration data and KPIs into which not a great deal of thought has gone to serve as 'metrics' for a program without their shedding any serious light on how well a program is working and why. In addition reporting might also become further compromised by 'messaging' considerations.

This pattern of following the letter rather than the spirit of some new policy imposed on agencies is familiar. For instance to ensure regulatory practice is guided by evidence of the relative costs and benefits of regulatory alternatives, 'regulation review' policies have constrained agencies to subject new regulation to formal regulatory impact analysis. Though, after over three decades in action the policy is *de rigueur*, its achievements appear very modest.[25] Bureaucracy and politics rewards 'can do' types, so regulatory impact analysis has become a box ticking exercise, obeyed in the letter, but not in spirit.[26]

We expect that introducing an Evaluator General even if it operated exclusively at the 'strategic' level as proposed in Canada would not fail as comprehensively as some argue regulation review has, though its success could well depend on its prominence and the strength of character of its senior management. It would offer some worthwhile benefits as outlined in the cited paper including raising the profile of evaluation, improving career pathways and doing more than is currently done now to hold governments to account for the outcomes of their programs.

---

[25] See for an example of regulatory review in action this column on regulation of Australian self managed super funds. http://clubtroppo.com.au/2012/12/13/regulation-review-superannuation-edition-the-column/.
As Amble and Chittenden put it over a decade ago in a report (pdf) commissioned by the British Chambers of Manufacturers, both sides of politics when in power "approach deregulation (removing existing laws) with enthusiasm, learn little or nothing from previous efforts, and have little if anything to show from each initiative". More recently two Australian scholars edited a recently published book arguing that "most Western governments have significant red-tape reduction programs but very few are successful". Chris Berg quoted in *The Australian*, 31st May 2018 launching *Australia's Red Tape Crisis.* https://www.theaustralian.com.au/national-affairs/governments-deregulation-agenda-an-abject-failure-says-new-book/news-story/45997d36dcfc540f4d6a5e8d5734cac7

[26] Interestingly the process did not work better when it was overseen by an independent agency such as Australia's Productivity Commission.

Nevertheless, wary of the extent to which an Evaluator General operating only at the 'strategic' level could really transform the culture and conduct of public agencies, we outline a more ambitious program to embed a learning and evaluation culture into government funded activity. Note this would generally include activity that was funded by government but contracted to be delivered by third parties.

Embedding independent evaluation in government

There is a clear tension between the use of evaluation by those running a program to optimise its effectiveness and efficiency, and its use to understand the relative merits of alternative programs and their relative funding. Certainly those engaged in the latter task should have independent evaluation to make evidence based decisions. But simply imposing an evaluation system designed from *their* perspective is likely to set off the kinds of unintended consequences critiqued in the previous section.

Further, evaluation cannot achieve its full potential if it is not integrated into delivery from the design stage. Nor can it do so unless it is assisting those at every level of delivering a program to understand *what* is being achieved, *how* and, accordingly the most promising prospects for improving effectiveness and efficiency. Thus, it must be integrated into service delivery in a way that is *collaborative with those delivering services*. If this can be done, this will offer the best way of holding people to account – after they've held *themselves* to account.

To succeed, the Evaluator General's must become the 'critical and expert friend' of agencies assisting the delivery agency and those in it become 'self-transparent' by going through the discipline of being transparent to another. The elements of such a relationship can be described as follows:

- The Evaluator General provides the resources and institutional support for a level of evaluation expertise to be cultivated within the public sector with clear career pathways through evaluation to the highest levels of the public service.
- Officers of the Evaluator General are 'embedded' in and work alongside officers of the delivery agency and have the lead role in designing and running monitoring and evaluation within the delivery agency. Though officers of the Evaluator General will seek to collaborate closely with line agencies, the Evaluator General determines any irresolvable disagreement between the two agencies regarding monitoring and evaluation.
- Data and analysis arising from monitoring and evaluation systems (including the specifications of any models) are published as soon as practicable.[27]

[27] This is subject to arrangements such as media embargoes to maximise the extent to which reporting is informed. We are not suggesting some PR role for the Evaluator General but rather practices like embargoes and media briefings to give media every

If such a mechanism is built around the felt need of the delivery agency and the practitioners within it to be 'self-transparent' concerning the quality of their work and the results it is achieving, this is the ideal foundation on which accountability can be built both up the chain of command in the agency and through to its political masters and more widely to the community in general.

These arrangements provide a general purpose interface between the 'brain' and the delivery of services, between knowing what one is doing and doing it. Accordingly they should apply not just internally to governments but also to circumstances where governments contract out sophisticated human services. Here a contractor would contract to deliver services but that would be subject to the Evaluator General being fully empowered and resourced to lead the monitoring and evaluation of that service delivery as it would if it were a line agency within government.

These arrangements also provide a means – a platform in common parlance – by which a community of practice might grow in its shared knowledge and knowhow and so drive learning in the entire system. They also provide an environment capable of nurturing the intrinsic motivation of those in the field, a crucial ingredient for the development and practice of high level professional and social skills and progressive learning from professional experience.[28]

It is intriguing to note that in a quite different context Toyota pioneered some of the ideas suggested here (See Box 4 below).

---

opportunity to become well informed. Experience suggests that on many matters the media will report constructively but that some may still report irresponsibly. There is no suggestion of any measures other than the Evaluator General's moral suasion and reputation for well informed truthfulness to prevent media irresponsibility.

[28] Economists and policy makers seem to pay surprisingly little heed to the extent to which employees' intrinsic motivation influences their effectiveness in performing their functions. As Herbert Simon puts it "pride in work and organizational loyalty are widespread phenomena in organizations. ... Willingness of employees at all levels to assume responsibility for producing results—not simply 'following rules'—is generally believed to be a major determinant of organizational success". Herbert A. Simon, 1991. "Hierarchy and Organisation", *The Journal of Economic Perspectives*, Vol. 5, No. 2 (Spring,), pp. 25-44 at p. 34-5

Box 4: Toyota builds a knowledge commons from the bottom up

Taking Henry Ford's ideas about eliminating waste beyond the point Ford had, Toyota revolutionised factory production. Under 'Fordism' essential knowledge work had been limited to engineers, technicians and senior managers, with middle managers' tasked with minimising the cost of following engineers' master plans. Unskilled factory workers were paid piece rates with supply going to the lowest bidding suppliers.

Toyota fostered distributed knowledge throughout its production system. Monitoring and evaluation was built from the get-go and from the bottom up with factory workers receiving ten times the training of Western competitors. Rather than being paced by the speed of production lines over which they had no control, teams of factory workers used the statistical control skills from their training to understand and endlessly optimise their own management of computerised numerical control machine tools.

Likewise Toyota gave its suppliers strong technical support, allowed them to keep part of the proceeds of innovation to encourage more of it but in return insisted that suppliers participate in a wider Toyota 'knowledge commons'. Toyota sponsored regular open days requiring suppliers to participate. This rapidly normalised the culture of sharing and collaboration thus increasing the rate at which successful innovations sped through the family of suppliers. In the upshot, Toyota plants often doubled their competitors' labour productivity while exceeding their production quality.

These arrangements provide the means by which the 'investment approach' to spending on various human services could be operationalised. The Evaluator General would uniquely be in a position to offer unbiased estimates of the impact of programs into the future. Further, with an eye to overcoming the curse of perpetual pilots to which Peter Shergold has referred, an important function of the Evaluator General would be to develop the monitoring and evaluation capability to compare the efficiency and effectiveness of new approaches with incumbent mechanisms of service delivery on as close to a 'level playing field' as is possible.

Conclusion

The Evaluator General is intended to address multiple issues simultaneously. It is analogous to the Auditor General in that can investigate and publish findings that have the effect of holding other agencies to account. But it also differs from the Auditor General in these ways:

- Rather than investigating others' compliance with some policy it is *itself* responsible for monitoring and evaluation within agencies and so its officers work within those agencies

designing, maintaining, reporting from and improving those systems.

- To do this well the Evaluator General and its officers must perform its duties in a way that is:
    - Closely collaborative with the agencies for whose monitoring and evaluation systems they are ultimately responsible;
    - Coincident with program design and delivery, not bolted on afterwards from outside;
    - Built from the bottom up. The Evaluator General's most fundamental task is to help those *throughout* agencies and certainly those delivering professional services to become 'self-transparent' – to measure, understand and so optimise the effectiveness of their services. Without this, measures of effectiveness do not reflect the knowledge of those in the field, are unlikely to be 'owned' by them and are at risk of being 'gamed' by them.
- In addition to being useful to the agency in optimising its effectiveness, outputs from the monitoring and evaluation system would also be released by the Evaluator General more generally which would:
    - Help build a knowledge commons around what works in the field and;
    - Have an accountability function.
- In particular, the Evaluator General would thereby be in a position to offer objective insights into:
    - the impact of programs into the future with a view to facilitating an approach to investment in human services based on optimising long run costs, benefits and outcomes as envisaged in the 'investment approach' to social investment.
    - the effectiveness of alternative ways of delivering services so that the cost effectiveness of new, innovative means of service provision could be compared – and so assessed on a 'level playing field' with – incumbent systems.

The ultimate design principle behind the Evaluator General is to follow a logic that offers a fine grained separation between the functions of government agencies to deliver or supervise the delivery of government funded services on the one hand – which is a function of the ministerially directed executive – and the delivery of objective analysis as to the efficacy of those services, which should be located within the independent executive.

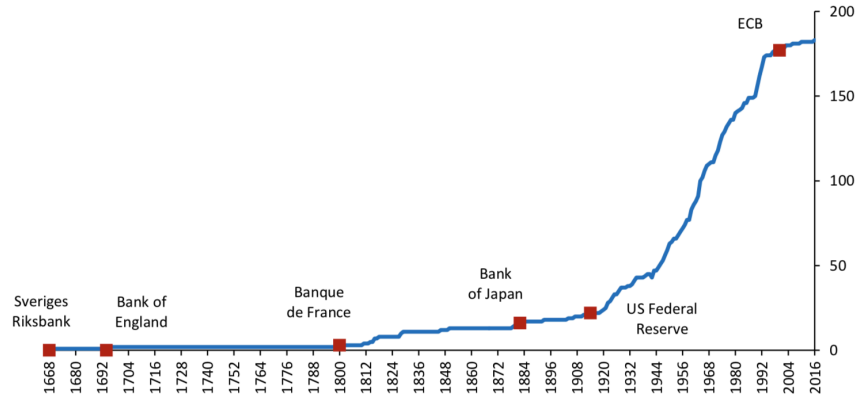
Appendix: The significance and growth of independent executive agencies

Until recently the bureaucracy predominantly comprised line departments of state directed by ministers from the governing party of the day. However agencies with greater independence from political direction go back well into previous centuries. Agencies of what we will call here the 'independent executive' are generally under the administrative supervision of a given ministerial portfolio, but their accountability is principally to the legislature.

One can distinguish several core functions of the independent executive. Some agencies such as the Auditor General,[29] defend the integrity of government systems, particularly financial systems. Others such as the Surveyor General, bureaus of statistics or meteorology form part of the society's informational infrastructure. Others administer regulatory regimes. Still others have been made independent of ministerial direction under legislation to ensure they operate more predictably according to specific mandates to operate within markets more commercially but often with regard to specified additional objectives as is the case with corporatised agencies such as Australia Post and all manner of other utilities.

Taking an international perspective, we can see many kinds of independent agencies being adopted according to the sigmoid 'adoption curves' that characterise technology adoption. Adoption responds slowly to the initial innovator but it gathers pace until it slackens off as saturation is approached. We can see this regarding central banks in the graph below.

Figure 1 Number of central banks



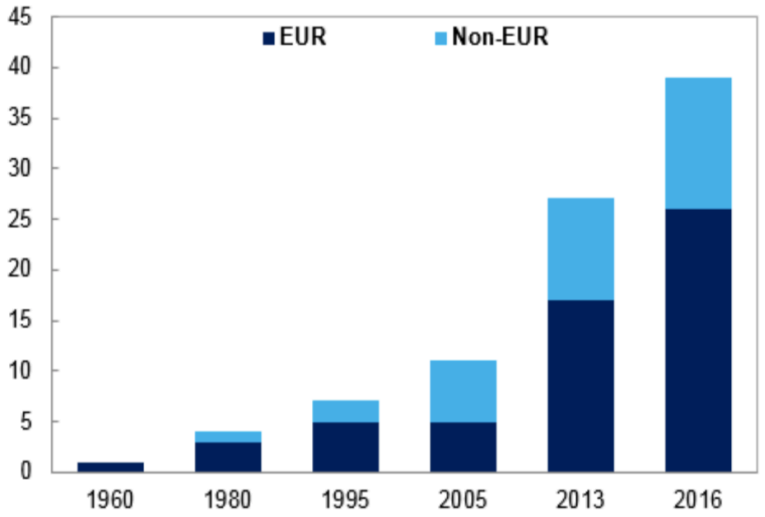
Source: Andrew G Haldane, 2017, "A Little More Conversation A Little Less Action" Speech, San Francisco, 31 March, p. 28

One can imagine similar adoption curves for many agencies from Freedom of Information agencies, Ombudsman's Offices,

[29] (and/or Comptroller General in the UK and and the US).

Environmental Protection Agencies, and utilities. Independent fiscal agencies appear to be following a similar path.

Figure 1. Number of Independent Fiscal Councils in the World



Source: IMF Fiscal Council Dataset. Note: "EUR" is European, "Non-EUR" is non-European.

Beetsma *et al*, 2018. "Independent Fiscal Councils: Recent Trends and Performance" IMF Working Paper, March.[30]

This trend has gathered pace since the 1960s. As Galadi put it in 2008 "During the past 25 years, independent regulatory agencies have become widespread institutions for regulatory governance".[31] For example freedom of Information regulation was pioneered in the USA in the 1960s and has been widely imitated since then. The Swedish office of the Ombudsman was established in the early 19th century and was increasingly imitated from the 1960s on. Ombudsman offices have also often grown in their scope, far beyond public administration complaints handling.[32]

In 1974 the US Congressional Budget Office was adapted from a Californian model (itself going back to 1941) and this has now spread to many other jurisdictions within and outside the USA. Insulating such agencies from political interference supports public and business confidence in the quality and predictability of their performance. Thus as Joyce puts it:

---

[30] Roel Beetsma *et al*, 2018. "Independent Fiscal Councils: Recent Trends and Performance" IMF Working Paper, March at http://www.imf.org/~/media/Files/Publications/WP/2018/wp1868.ashx

[31] Fabrizio Gilardi, 2008. *Delegation in the Regulatory State Independent Regulatory Agencies in Western Europe*, Edward Elgar. Note the quote comes from the publishers' 'blurb'.

[32] In New South Wales legislation has required there to be a Deputy Ombudsman overseeing and monitoring Aboriginal programs (The OCHRE Strategy). Rather than placing the role within a central agency, it was a powerful statement and endorsement of the importance of independence to place it within the Ombudsman's Office.

In the 40 years since its founding, the Congressional Budget Office (CBO) has become one of the most influential and well-regarded institutions in Washington, solidifying its place as the authoritative source of information on the budget and economy.[33]

[33] Joyce, Philip, 2015. “The Congressional Budget Office At Middle Age”, Brookings Institution, Hutchins Center, Working Paper No. 9.