

Revisiting the study of policy failures

Mark Bovens & Paul 't Hart

To cite this article: Mark Bovens & Paul 't Hart (2016) Revisiting the study of policy failures, Journal of European Public Policy, 23:5, 653-666, DOI: [10.1080/13501763.2015.1127273](https://doi.org/10.1080/13501763.2015.1127273)

To link to this article: <https://doi.org/10.1080/13501763.2015.1127273>



© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 22 Feb 2016.



Submit your article to this journal [↗](#)



Article views: 7320



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 41 View citing articles [↗](#)

Revisiting the study of policy failures

Mark Bovens and Paul 't Hart

ABSTRACT The analysis of policy failures is, by definition, not a neutral endeavour, since policy fiascos are not neutral events. Moreover, they are often, usually implicitly, but sometimes explicitly, permeated with prosecutorial narratives, blame games and a search for culprits. Fiascos do not just 'happen'. They are constructed, declared, and argued over in labelling processes that are not necessarily 'evidence-based'. This presents a challenge for any academic endeavour to identify, analyse and explain policy fiascos. Against this backdrop, we assess the study of policy failure as it stands today, and offer some reflections for its further development.

KEY WORDS Policy evaluation; policy failure; policy fiasco; policy success.

POLICY FAILURES: CONTESTED CONSTRUCTS

In 2012, the Australian historian Christopher Clark published *The Sleepwalkers*, a 700-page-thick account of how the European great powers stumbled into the First World War. The book gave a very detailed description of the years, months, days and even hours that preceded the outbreak of war in August 1914. Soon after its publication, the book became the subject of controversy. Clarke's conclusion was that 'the outbreak of war was a tragedy, not a crime' (Clarke 2012: 561). The crisis that resulted in one of the world's most atrocious wars was 'the fruit of a shared political culture', according to Clarke, and could not be blamed solely on Germany and Austria, the axis powers. The other European superpowers, but also the Serbian nationalistic government, bore part of the blame, too. In Germany, with its contemporary history of soul-searching and introspection and its 'battle of the historians', Clarke was portrayed as an apologist. In Serbia, on the other hand, he was accused of anti-Serbian bias.

The reception of *The Sleepwalkers* shows how, even one century after the fact, international crises and diplomatic failures can be the topic of vehement academic debate and political controversy. The analysis of policy failures is, by definition, not a neutral endeavour, since policy fiascos are not neutral events. Moreover, they are often, usually implicitly, but sometimes explicitly, permeated with prosecutorial narratives, blame games and a search for culprits (see Oppermann and Spencer 2016). Starting 20 years ago, we published several books and papers on policy failures, successes and the politics of policy evaluation, including cross-national and cross-sectoral case comparisons (Bovens and 't Hart 1995, 1996; Bovens *et al.* 1999, 2001, 2006; Gray and 't Hart

1998). We argued in these studies that ‘success’ and ‘failure’ are not inherent attributes of policy, but rather labels applied by stakeholders and observers. Fiascos do not just ‘happen’. The same goes for ‘successes’ (McConnell 2011). They are constructed, declared and argued over. Clearly, these labelling processes are not necessarily ‘evidence-based’. Formal evaluations or benchmarking exercises are not always conducted, and when they do their results may not necessarily be widely known or accepted.

Particularly in cases of complex and contested programmes, stakeholders’ views may have been shaped long before. Moreover, evidence will be met by counter-evidence. In other words, the verdict about a public policy, programme or project is shaped in ongoing ‘framing contests’ between its advocates and shapers on the one hand, and its critics and victims on the other (Boin *et al.* 2009). In this contribution, we re-examine this argument, review the study of policy failure as it stands today, and offer some reflections for the analysis of foreign policy failures.

IDENTIFYING FAILURE: TWO LOGICS OF EVALUATION

Governments plan, decide, do, deliver, adjust, reverse and terminate many things, all the time. Three decades of rhetoric about shrinking the scope of the state notwithstanding, their interventions cover almost every conceivable aspect of our lives and our societies. Even deregulation, privatization and more community-based public service provision require proactive and alert government, we have since discovered (Moran 2001). Only a part of this myriad of ambitions and activities unfolds as hoped, expected and planned for by policy-makers. Another part throws up surprises, complications, delays, disappointments and unintended consequences. Though to our knowledge no one has ever attempted to make a comprehensive estimate of the proportion, four decades of implementation research have taught us that the share of the latter group is non-trivial at best: it has yielded a litany of ‘great expectations dashed’ and ‘ruined hopes’ (Pressman and Wildavsky 1984; see also Schuck 2014).

So, if we regard instances of delivery-and-results-as-planned as ‘successes’, what are we to make of the other category? Are they all, by definition, ‘failures’? That seems overly harsh, and hardly helpful. But how much deviation from original plans – in terms of coverage, timing, production costs, impacts, stakeholder attitudes and behaviours – can a policy stand before its reputation begins to suffer? There is no Archimedean point, no self-evident yardstick. The exercise of discretionary judgment is inevitable, and that is precisely what goes on both in the political and the academic arenas in which policy reputations are being shaped.

But this exercising of judgment in how people perceive the performance of public policies is an imperfect business. Nor does it take place in a vacuum. There are many sources of inadvertent or explicit bias in assessing policies and their outcomes (Bovens and ‘t Hart 1996). Values, positions, interests,

time and culture come into play as much as the observable attributes (the 'facts') of a policy. Evaluation is, after all, both a normative exercise, in that it presumes standards against which performance will be assessed, and a political exercise, in that attaching certain labels to a programme or project can have significant consequences for those involved in and affected by it.

Of course, there are instances where few observers will hesitate to attach the label 'policy failure' to government operations that evidently went wrong. Think of spectacular, unambiguous and highly consequential mishaps such as the United States-led invasion of Iraq, the regulation and oversight of the banking and financial services industry, the acute collapse of public infrastructures and other man-made disasters, conspicuous public procurement and IT-innovation failures, or cases of entrenched institutional corruption. However, some analysts seem to think that this suffices, and don't bother to define what they mean when they talk about 'blunders' or 'fiascos' (e.g., Crewe and King 2013).

But more often it is fundamentally ambivalent. Think of the Australian government's suite of 'Keynesian' macroeconomic policy measures (worth A\$42 billion) taken at the onset of the global financial crisis and designed to prevent it causing an economic slump in Australia by putting money in the hands of local businesses and consumers as soon as possible. It was a marked success overall: Australia was the only Western country to avoid recession at that time. But at the same time, one of the measures – generous subsidies for home insulation – generated not just a massive uptake but also appeared to have contributed to the electrocution of four young apprentices working for several of the many electricians that scrambled to perform the work, as well as to a marked upsurge in house fires owing to improper installation of insulation materials. How to assess this latter programme? The four fatalities that had occurred across the 200,000 installations that had been performed during this period actually constituted an improvement upon the industry's average safety record. But all four incidents occurred within a few weeks, and thus became a story rather than a statistic. Massive and critical attention was trained on the programme by media and the parliamentary opposition. The government tried to counter with 'facts'. A framing contest ensued, and the image of a 'mess' was the one that stuck. It cost the responsible minister his job, the programme was shut down and the public service division that administered it was moved into a different department. Some policy analysts argued that the programme's woes were a product of avoidable failures in policy design (Lewis 2011), but others claimed that had the government conducted its impression management a little more astutely, the chips could well have fallen the other way and the story could have crumbled. What was now a major 'fiasco' could well have ended up as an incident (Tiffen 2010).

This example is the rule rather than the exception. There are many shades of grey in how we perceive public policies, programmes and projects. The study of policy success and failure is therefore one of the dynamics of *reputation* as much

as it is one of *performance*. Bovens and 't Hart (1996) acknowledged this by making a distinction between the 'programmatic' and the 'political' dimensions of policy evaluation. *Programmatic* evaluation pertains to the world of facts and social balance sheets: observable costs and benefits, original intentions and eventual outcomes. Its currency is 'performance as measured'. Mark Moore's (2013) public value scorecard is just the latest in a whole series of attempts to provide policy analysts with standardized tools for programmatic assessment. Clearly, though sometimes presented as 'objective' and 'evidence-based', there is ample debate over the extent to which such technical assessment is free of political judgement/evaluation. For instance, the European Commission has introduced impact assessments for all major policy initiatives, designed to ensure that all economic, social and ecological consequences of alternative policy options are described, estimated and evaluated. However, such impact assessments have been characterized as justifications for certain pre-determined policy narratives, rather than representing an objective instrument for evidence-based policy (Radaelli *et al.* 2013).

Political evaluation, on the other hand, squarely pertains to the world of impressions: lived experiences, stories, frames, counter-frames, heroes and villains. These are constructed in the way policies are being perceived and debated among their stakeholders, in the media and in the forums where policy-makers are held to account, such as citizen and institutional watchdogs, legislatures, courts. The currency of political evaluation is 'reputation as conferred' (by those accountability forums). Performance statistics and professional judgments are only a few among many reputation-shaping forces in the politics of evaluation. In the political arena, critical, well-publicized incidents, however unrepresentative for the policy at large they might be, can make for more compelling narratives than any programmatic evaluation exercise, however richly documented and methodologically robust.¹

That these two logics of evaluation are not necessarily highly correlated was demonstrated empirically in our six-country, four-sector comparative empirical investigation of success and failure in public governance (Bovens *et al.* 2001; see also Gaskarth 2016; Kruck 2016). Some significant programmatic successes – such as health care reforms seeking to regulate doctor's pay or government efforts to provide soft landings to sectors and regions afflicted by terminal industrial decline – were not recognized as such in the political arena. Tax collection is also a classic case. In quite a few countries, governments manage, year in, year out, to peacefully extract very significant parts of wages, profits and asset values from citizens and corporate bodies, with tax compliance rates well above 90 per cent. But tax laws and tax authorities are rarely getting any credit and remain nearly always politically controversial, loathed by the corporate sector and met with at best quiet desperation by the public. They can never be popular, regardless of their performance on the relevant performance indicators; at best, they are taken for granted and accepted as a necessary evil. Much of government policy is not about conferring benefits and delivering services, but

about restricting freedoms, administering pain and eliciting compliance. To make those types of policies politically popular is a big ask.

Likewise, marked performance failures were not followed by a collapse of a programme's reputation or of its chief architects and implementers. Most of the six countries in our study experienced major banking and/or currency crises as unintended consequences of deregulation policies pursued during the 1980s and early 1990s. In some instances, major monetary and social costs were incurred by sloppy or toothless regulatory oversight of the now much more autonomous financial institutions. But in the closed, technocratic, old-boys-network world of finance, political pressures for transparency and accountability about what had happened remained limited. Twenty years on, the scenario appeared to have changed only somewhat. Programmatically, very similar causes lay at the root of the spectacular regulatory failure that allowed financial institutions to embark on the excessively complex products and risky investment strategies that triggered the 2008–9 global financial meltdown. But this time, politically, virtually all incumbent governments responsible for the lax regulatory regime were punished by the voters, particularly in hard-hit countries ('t Hart and Tindall 2009).

When programmatic and political evaluation logics are out of kilter, strange incongruities may occur. Take for example the handling of HIV-infected blood samples during the early days of the AIDS epidemic. Countries all struggled to get to grips with this new and ill-understood challenge. Technically, most adopted more or less similar measures at roughly the same time. In France, however, the limited number of HIV infections that occurred resulted in one of the most vehement post-World War 2 political scandals, with several ministers having to step down and former prime minister Fabius ending up in a criminal court. At the same time in Spain where, seen from a programmatic perspective, medical authorities were far less effective, resulting in many avoidable HIV contaminations, little or no political 'fallout' occurred (Bovens *et al.* 2001).

Table 1 Two logics of evaluation

<i>Reputation:</i> Political assessment	++	--
<i>Performance:</i> Programmatic assessment		
++	Success	Tragedy
--	Farce	Fiasco

Table 1 visualizes the four ideal typical outcomes that result from acknowledging these two logics of evaluation. The table in effect creates a two-dimensional assessment space. Two symmetrical outcomes ('success' and 'failure') are contrasted with two incongruent, asymmetrical outcomes. One is the

'tragedy' box, consisting of policies that deliver admirable outcomes but yield little or no public or political credit. Wildavsky's (1977) political paradox of health policy – 'doing better, feeling worse' – captures this. Intelligence and counterterrorism programmes, whose accomplishments can often not be publicized, may suffer from a similar lack of community and political appreciation, particularly when salient enemies and threats seem to be lacking.

Conversely, there is also the 'farce' of a policy failing to deliver on its promises but suffering no real dent in its reputation or that of its makers. During the 1970s and '80s, industry policies seeking to resuscitate dying but publicly treasured, regionally pivotal and politically sacralized sectors like shipbuilding, car manufacturing and steel production triggered truly massive but strategically futile public expenditure. But in the short run, jobs were 'saved' and pain postponed, and so these policies for a long time enjoyed solid support across the political spectrum. It was a classic case of 'too much invested to quit', with all the cognitive dissonance reduction it entailed. Likewise, foreign policy fixtures like 'deterrence' (and its close connection to 'domino theory' thinking, however ill-grounded in misperceptions of adversaries' will power, resources and rationality), and 'economic sanctions' appear to enjoy continued popularity among policy-makers in any states, despite their chequered performance records and the complicated conditions under which they are said to be working at all (Blanchard and Ripsman 2013; Payne 2015).

Table 1 highlights what related fields of research – on agency reputation (Maor 2014), public accountability (Bovens *et al.* 2014) and the social dynamics of credit and blame (Tilly 2008) – have also found: there is no 'just world' of policy assessment in which reputation naturally reflects performance. The nexus between the two is constructed, negotiated and therefore contingent, and often variable over time. The dynamic (dis)equilibriums between performance and reputation that may result are always the result of two different and only weakly related evaluation processes: the deterioration or improvement of its programmatic accomplishments as measured in technical assessment exercises; and the political waxing and waning of the coalitions of actors lending it support or criticizing it.

Over time, a policy may thus evolve an assessment trajectory across the two-dimensional space of Table 1. One example would be NASA's space shuttle programme (1981–2011), which was the initially popular and seemingly well-performing follow up to its iconic Apollo programme. But five years after its inception, budgetary difficulties and growing vulnerabilities in the management of the programme undermined its performance. This became dramatically visible, and politically challenging, with the January 1986 Challenger explosion, which caused a major slump in both the programme's and the agency's reputation and autonomy. Reputation recovered somewhat and ostensibly performance improved as 'lessons were learned'. But the reputations of programme and organization were shattered again in 2003, when the Columbia shuttle was lost upon its re-entry to earth and the subsequent investigation revealed that the agency had learned less well than it thought it had (Starbuck and Farjoun 2005).

McConnell (2011) has since made three additions to the Bovens and 't Hart's framework. One, he has turned the original focus on fiascos upside down by concentrating on the other end of the spectrum, policy success. Second, he devised a continuum of assessment outcomes that comprises two intermediate positions between the two opposites of complete (in his terminology: 'durable') success and failure: conflicted and precarious success. *Conflicted* success comprises instances where, notwithstanding some demonstrable achievements, there is significant political contestation about the merits of a policy. *Precarious* success denotes instances where 'policies operate at the edge of failure', as there are major shifts or deviations from original goals and 'high-profile and bitter conflicts' among stakeholders and in the political arena (McConnell 2011: 61). Third, he supplemented the programmatic–political dichotomy with a third dimension: process assessment (see also McConnell 2016). McConnell (2011: 64):

Process basically “gets the job done” for government in the sense that it puts in place the quality that government wants, bringing constitutional legitimacy to the outcome, with a strong coalition of interests behind it.

In our view, this way of defining process comes perilously close to what has been defined as political assessment. Each of the three criteria mentioned by McConnell essentially refers to reputational concerns, namely (ways of) getting and maintaining support and legitimacy for a policy. The technocratic side of 'process' – methodological rigour and procedural correctness in policy design, professional management of implementation and oversight – the quality of 'puzzling' to counterbalance the 'powering', to use Hecló's (1974) classic phrase – is noticeably absent.

This is not the place to debate the merits of various incarnations of the success/failure assessment framework as such (see Bovens 2010; Marsh and McConnell 2010). Suffice to say that the entire line of research discussed here makes it clear that researchers of alleged failures, fiascos, blunders and messes of governments' labour have an obligation to be methodical, transparent and clever in how they position themselves *vis-à-vis* these quintessentially political acts of applying synthetic labels to often complex, ambiguous, changeable realities of performance and reputation.

Howlett (2012) has made several important recent additions to the analytical repertoire of students of policy failure that can help in this endeavour. He observes, first of all, that one should distinguish between policy failures born out of ill will and malevolence, such as studied in the literature on state crimes (De Haven-Smith 2006), and those that arise inadvertently. Secondly, focusing on the latter class of events, Howlett (2012: 545) rightly argues that:

[I]n many contemporary studies . . . different aspects and types of failure are often poorly specified and incorrectly juxtaposed, and this conceptual confusion has stood in the way of cumulative theory-building into the causes and consequences of policy success and failure.

Table 2 Howlett's typology of policy failure

		Magnitude (extent and duration)	
		High	Low
Saliency (intensity and visibility)	High	Type I: major failure e.g., climate change (international treaty) policy failure	Type II: focused failure e.g., sports crowd control (riots) policy failure
	Low	Type III: diffuse failure e.g., anti-poverty policy failure	Type IV: minor failure e.g., policy service contract bid failure

To remedy this, he presents a typology (see Table 2) that employs two dimensions: highly public and widely agreed-upon versus more opaque and more ambivalent failures; and system-wide and long-lasting versus limited scope and ephemeral failures. Howlett is right. If the analytical ambition is to not only systematically identify but also explain policy failures, lumping together systemic and protracted failures (such as the creeping regulatory failure in financial sector supervision in the decade leading up to the 2008–9 global financial crisis) with a local, one-off events like, for example, the crowd disaster at the 2010 Love Parade electronic dance festival in the German city of Duisburg (a case of crowd turbulence which caused the death of 21 people by suffocation), is not a particularly helpful case selection strategy. Aiming for a grand theory covering all forms of policy failure identified in the typology is fruitless. It makes more sense to develop several mid-range theories covering particular clusters of policy failures, and Howlett's typology is one plausible way to get this process underway.

EXPLAINING POLICY FAILURE: IN SEARCH OF MID-RANGE THEORIES

With these distinctions in mind, let us explore this ambition of explaining policy failures a bit further. There are at least two forms that ambition can take. The first is to *explain specific instances of policy failure*. The literature is replete with such attempts. Some of the most impressive efforts in this vein have not only been made by academic researchers such as Vaughan(1996), but by 'post-mortem-style' public inquiries. Notable examples include the 9/11 Commission study of the roots of the US intelligence failure that failed to forestall the Al Qaeda surprise attacks on the Twin Towers and the Pentagon; and the Dutch Institute for War Studies report on the failure of the United Nations (UN) peacekeeping operation to prevent the 1995 massacre of Bosnian men by Bosnian Serb forces at Srebrenica (Netherlands Institute for War Studies [NIOD] 2002). Contemporary high-saliency, high-magnitude

cases, such as the failure of the US-led invasion of Iraq, now routinely elicit a multi-disciplinary array of academics seeking to explain their occurrence.

Much of this oft-admirable work is either at best implicitly using existing theory (e.g., Byman 2008 on Iraq). Much of it privileges a single explanation at a single level of analysis – for example, cognitive failure because of excessive reliance on misleading historical analogies (Khong 1992), ‘risk-taking’ and ‘hubris’ on the part of ambitious but unreflective government leaders (Boetcher 2005; Owen 2012), ‘groupthink’ in the inner circle elite groups in which foreign policy decisions get made (Badie 2010), or ‘bureaucratic politics’ marring interorganizational information exchange and co-ordination (Gabriel 1986). Piling up single case studies using single theories of failure remains useful to an extent. But it should be clear that it does not advance the field as much as do multi-perspectivist and integrative approaches (e.g., Houghton 2001; Kam 1988; Mintz and DeRouen 2010; Mitchell and Masoud 2009; Parker and Stern 2005; Walker and Malici 2011; Yetiv 2011) to analysing single instances of policy failure. Equally useful are medium-n comparative case designs, particularly ones that compare the dynamics of policy failures with those of non-failures (Patashnik 2008; Shafer and Crichlow 2010).

The second form of explanatory analysis, as we have seen, is to *develop theories explaining a range of policy failures of a certain type*. Peter Hall’s (1981) classic study of planning disasters in the United Kingdom led the way. Moran’s (2001) study of policy disasters of the regulatory state, Bovens *et al.*’s (2001) structured-focused comparison of successes and failures of governments dealing with four different types of governance challenges, and Flyvbjerg *et al.*’s (2003) study of failures of ‘megaprojects’ are all examples of controlled comparisons that enable the development of mid-range theories. Contemporary advances in causal process-tracing and fuzzy-set methodology allow for more explanatory rigour and stronger theory development than can be found in any of these examples (Blatter and Haverland 2012; Ragin 2008). The present collection’s emphasis on explaining failures in foreign policy stands in a long, US-based tradition that was given momentum by Irving Janis’s *Victims of Groupthink* (Janis 1972) and has reached an impressive new peak with Schafer and Crichlow (2010). It treats foreign affairs as a distinct policy endeavour and develops purpose-built models explaining instances where foreign policy initiatives are seen to have failed.

The challenges of such attempts are manifold. They entail issues of scope (does, for example, ‘foreign policy’ still not cover much too broad a spectrum of interventions to be a suitable target for mid-range theorizing?), typologizing (are, for instance, ‘intelligence failures’ or ‘megaprojects’ coherent enough categories to be treated as types in a comparative design?) and case selection (what can we learn from a sample of ‘foreign policy fiascos’ in country X that includes a military misadventure, a grand development aid scheme coming unstuck and a deterrence failure?). Fortunately there are now great examples of mid-range theory construction based on rigorous comparison of (foreign) policy-making process that we can emulate and adapt for the purposes of

explaining particular types of policy failure (George and Smoke 1974; Haney 2002; Kaarbo 2012).

MOVING FORWARD: LEARNING FROM POLICY FAILURES

In closing this contribution, it might be useful to reflect on the underlying purpose of studying policy failure. Most analysts who analyse policy failures probably are driven by a desire to create knowledge that will help prevent the recurrence of failure. By documenting what went wrong and explaining why, policy analysts create a knowledge base that should enable future policy-makers to do better. One cannot help but wonder if we have gotten any closer to achieving that objective. To get complex organizations and policy networks to actually *learn* from feedback, rather than make symbolic, opportunistic or minimal impact changes in response to it, (March and Olsen 1975) about their past performance is hard enough – that much we know from the spate of studies on policy change and policy learning across a wide range of sectors and jurisdictions. To induce policy (as opposed to political) learning from highly public, politically charged forms of feedback such as that produced by evaluators and public inquiries in the context of policy fiascos is proving even more difficult (Birkland 2006; Boin *et al.* 2008; Hansén 2007; Howlett 2012; May 1992; Walsh 2006).

In short: big policy failures can be, but all too seldom are, a trigger for big policy learning that reduces the likelihood of their reoccurrence. This should not deter us from trying, but it also means that in doing so we need to move ‘from scientific demonstration and verification to the giving of reasons and the assessment of practical arguments’ (Fischer 1995: 215). Articulating their research-based ideas about institutional and policy design, decision-making and implementation and injecting them into the right places at the right time is a challenge that students of policy failure (and success) should take to heart just as much as getting the next study underway and the next article published. Policy analysts must be prepared to make a difference in the ‘real world’ if they are serious about preventing fiascos (Hallsworth 2011).

This is not just frivolous sniping from the sidelines. It is about taking our own findings seriously. They tell us that the best way to avoid fiascos is to ‘open up’ policy-making processes to genuine contests of ideas. This is truer the more complex and interdependent the world is getting, and the bigger the stakes of avoiding the destructive potential of flawed government. Christopher Clark, the author of *The Sleepwalkers* ends on a positive note. One of the major differences between the years before World War 1 and the decades after World War 2 is that:

decisionmakers and the general public alike grasped in a visceral way the meaning of nuclear war ... As a consequence, the greatest arms race in

human history never culminated in nuclear war between the superpowers. (Clark 2012: 562)

At the same time, minute analysis of several crises and nuclear alerts during the Cold War standoff between the US and the Soviet Union suggests that this was sometimes more owing to coincidence and level-headed individuals lower down the hierarchy exercising sound professional judgments than it was a product of failure-proof weapons systems, military doctrines or far-sighted foreign and defence policy-makers (Jervis et al. 1985; Lebow and Stein 1995; Sagan 1993).

One of the founding fathers of modern policy analysis, Yehezkel Dror, has demonstrated acute awareness of the rising stakes of governance failures in the contemporary world. In his latest book, *The Avant-Garde Politician* (Dror 2014), he makes a compelling case that bad governance can now compromise the very survival of species and planet. His solution is to advocate what, in inimitable fashion, he has earlier called a 'novo-Platonic' approach to public policy-making: a circle of strong, enlightened, dedicated, disciplined, well-educated 'high-quality' rulers (Dror 2001). Only constrained by a global constitutional court, they should offset democracy's penchant for short-termism and opportunism in dealing with our future world's increasingly complex and often dangerous policy predicaments. To some this may be a tempting idea. But if the yield of several decades of studies of policy failure is anything to go on, it is likely to generate terrible unintended consequences.

The biggest fiascos are not caused by division, ceaseless debate, all too powerful checks and balances and institutional paralysis, but by the closing up of policy-making processes: concentrating authority in too few hands; constraining the scope and duration of deliberation; and shutting down diversity and dissent (*cf.* Schuck 2014). Another founding father of the field, Charles Lindblom, was absolutely right when he observed that the very 'messiness' of democratic policy-making – active citizens, wide participation, dispersed decision making, multi-stakeholder involvement, contending bodies of evidence and argument, persuasion and negotiation, institutional checks and balances during policy design as well as implementation – with all the suboptimization, incrementalism and prevarication it entails, is still a much better safeguard against spectacular failures than a top-down, monolithic, tightly held, 'evidence-based', linear and lean process dominated by a single small 'team' within the policy arena (Lindblom 1965). We forget this number one lesson for preventing policy failure at our own peril.

Biographical notes: Mark Bovens and Paul 't Hart are professors of public administration at the Utrecht University School of Governance.

Addresses for correspondence: Mark Bovens, Bijlhouwerstraat 6, 3511 ZC Utrecht, Netherlands. email: m.bovens@uu.nl / Paul 't Hart, Bijlhouwerstraat 6, 3511 ZC Utrecht, Netherlands. email: p.thart@uu.nl

NOTE

1 To complicate things further, in the evaluation practice technical assessments will often not be free of political judgements. For instance, the European Commission has introduced impact assessments for all major policy initiatives, designed to ensure that all economic, social and ecological consequences of alternative policy options are described, estimated and evaluated. However, such impact assessments have been characterized as justifications for certain pre-determined policy narratives, rather than representing an objective instrument for evidence-based policy (Radaelli *et al.*, 2013).

REFERENCES

- Badie, D. (2010) 'Groupthink, Iraq, and the War on Terror: explaining US policy shift toward Iraq', *Foreign Policy Analysis* 6(4): 277–96.
- Birkland, T. (2006) *Lessons of Disaster*, Washington, DC: Georgetown University Press.
- Blanchard, J.F. and Ripsman, N.M. (2013) *Economic Statecraft and Foreign Policy*, London: Routledge.
- Blatter, J. and Haverland, M. (2012) *Designing Case Studies: Explanatory Approaches in Small-N Research*, Basingstoke: Palgrave.
- Boetcher, W. (2005) *Presidential Risk Behavior in Foreign Policy: Prudence or Peril?* New York: Palgrave MacMillan.
- Boin, A., 't Hart, P. and McConnell, A. (eds) (2008) *Governing After Crisis: The Politics of Investigation, Accountability and Learning*, Cambridge: Cambridge University Press.
- Boin, A., McConnell, A. and 't Hart, P. (2009) 'Crisis exploitation: political and policy impacts of framing contests', *Journal of European Public Policy* 16(1): 81–106.
- Bovens, M. (2010) 'A comment on Marsh and McConnell: towards a framework for establishing policy success', *Public Administration* 88(2): 584–85.
- Bovens, M. and 't Hart, P. (1995) 'Fiascos in the public sector: towards a conceptual frame for comparative policy analysis', in J.J. Hesse and Th.A.J. Toonen (eds), *The European Yearbook of Comparative Government and Public Administration*, Baden-Baden: Nomos, pp. 577–608.
- Bovens, M. and 't Hart, P. (1996) *Understanding Policy Fiascoes*, New Brunswick, NJ: Transaction.
- Bovens, M., 't Hart, P., Dekker, S. and Verheuver, G. (1999) 'The politics of blame avoidance: defensive tactics in a Dutch crime-fighting fiasco', in H.K. Anheier (ed.), *When Things Go Wrong: Failures and Breakdowns in Organizational Settings*, London: Sage, pp. 123–47.
- Bovens, M., 't Hart, P. and Peters, B.G. (eds) (2001) *Success and Failure in Public Governance: A Comparative Analysis*, Cheltenham: Edward Elgar.
- Bovens, M., 't Hart, P. and Kuipers, S.L. (2006) 'The politics of policy evaluation', in M. Moran, B. Goodin and M. Rein (eds), *Oxford Handbook of Public Policy*, Oxford: Oxford University Press, pp. 317–33.
- Bovens, M., Schillemans, T. and Goodin, R. (eds) (2014) *Oxford Handbook of Public Accountability*, Oxford: Oxford University Press.
- Byman, D. (2008) 'An autopsy of the Iraq debacle: policy failure or bridge too far?', *Security Studies* 17(4): 599–643.
- Clark, C. (2012) *The Sleepwalkers: How Europe Went to War in 1914*, London: Penguin.
- Crewe, I. and King, A. (2013) *The Blunders of Our Governments*, London: Oneworld.
- De Haven-Smith, L. (2006) 'When political crimes are inside jobs: detecting state crimes against democracy', *Administrative Theory & Praxis* 28(3): 330–55.

- Dror, Y. (2001) *The Capacity to Govern: A Report to the Club of Rome*, London: Frank Cass.
- Dror, Y. (2014) *Avant-Garde Politician: Leaders for a New Epoch*, Washington, DC: Westphalia Press.
- Fischer, F. (1995) *Evaluating Public Policy*, Chicago: Nelson-Hall.
- Flybjerg, B., Bruzelius, N. and Rothengatter, W. (2003) *Megaprojects and Risk: An Anatomy of Ambition*, Cambridge: Cambridge University Press.
- Gabriel, R. (1986) *Military Incompetence*, New York: Hill and Wang.
- Gaskarth, J. (2016) 'The fiasco of the two Syria votes: decline and denial in British foreign policy', *Journal of European Public Policy*, doi:10.1080/13501763.2015.1127279.
- George, A.L. and Smoke, R. (1974) *Deterrence in American Foreign Policy*, New York: Columbia University Press.
- Gray, P. and 't Hart, P. (eds) (1998) *Public Policy Disasters in Europe*, London: Routledge.
- Hall, P. (1981) *Great Planning Disasters*, London: Penguin.
- Hallsworth, M. (2011) *Policymaking in the Real World: Evidence and Analysis*, London: Institute for Government.
- Haney, P.J. (2002) *Organizing for Foreign Policy Crises*, Ann Arbor, MI: University of Michigan Press.
- Hansén, D. (2007) *Crisis and Perspectives on Policy Change: Swedish Counter-terrorism Policymaking*, Stockholm: Swedish National Defence College.
- Heclo, H. (1974) *Modern Social Politics in Britain and Sweden: From Relief to Income Maintenance*, New Haven, CT: Yale University Press.
- Houghton, D.P. (2001) *US Foreign Policy and the Iran Hostage Crisis*, New York: Cambridge University Press.
- Howlett, M. (2012) 'The lessons of failure: learning and blame avoidance in public policy', *International Political Science Review* 33(5): 539–55.
- Janis, I.L. (1972) *Victims of Groupthink*, Boston: Houghton Mifflin.
- Jervis, R., Lebow, R.N. and Stein, J.G. (1985) *Psychology and Deterrence*, Baltimore, MD: Johns Hopkins University Press.
- Kaarbo, J. (2012) *Coalition Politics and Cabinet Decision Making: A Comparative Analysis of Foreign Policy Choices*, Ann Arbor, MI: University of Michigan Press.
- Kam, E. (1988) *Surprise Attack*, Cambridge, MA: Harvard University Press.
- Khong, Y.F. (1992) *Analogies at War*, Princeton, NJ: Princeton University Press.
- Kruck, A. (2016) 'Resilient blunderers: credit rating fiascos and rating agencies' institutionalized status as private authorities', *Journal of European Public Policy*, doi:10.1080/13501763.2015.1127274.
- Lebow, R.N. and Stein, J.G. (1995) *We All Lost the Cold War*, Princeton, NJ: Princeton University Press.
- Lewis, C. (2011) 'A recent scandal: the home insulation program', in K. Dowding and C. Lewis (eds), *Ministerial Careers and Accountability in the Commonwealth Government*, Canberra: ANU E Press, pp. 153–76.
- Lindblom, C.E. (1965) *The Intelligence of Democracy*, New York: Free Press.
- Maor, M. (2014) 'Theorizing bureaucratic reputation', in W. Arild and M. Maor (eds), *Organizational Reputation in the Public Sector*, London: Routledge, pp. 17–36.
- March, J. and Olsen, J.P. (1975) 'Organizational learning under ambiguity', *European Journal of Political Research* 3(2): 147–71.
- Marsh, D. and McConnell, A. (2010) 'Towards a framework for establishing policy success: a reply to Bovens', *Public Administration* 88(2): 586–7.
- May, P.J. (1992) 'Policy learning and failure', *Journal of Public Policy* 12(4): 331–54.
- McConnell, A. (2011) *Understanding Policy Success*, Basingstoke: Palgrave.

- McConnell, A. (2016) 'A public policy approach to understanding the nature and causes of foreign policy failure', *Journal of European Public Policy*, doi:10.1080/13501763.2015.1127278.
- Mintz, A. and DeRouen, K. (2010) *Understanding Foreign Policy Decision Making*, Cambridge: Cambridge University Press.
- Mitchell, D. and Massoud, T.G. (2009) 'Anatomy of failure: Bush's decision-making process and the Iraq War', *Foreign Policy Analysis* 5(2): 265–86.
- Moore, M. (2013) *Recognizing Public Value*, Cambridge, MA: Harvard University Press.
- Moran, M. (2001) 'Not steering but drowning: policy catastrophes and the regulatory state', *Political Quarterly* 72(3): 414–27.
- Netherlands Institute for War Studies (NIOD) (2002) *Srebrenica, een 'Veilig' Gebied: Reconstructie, Achtergronden, Gevolgen en Analyses van de Val van een Safe Area*, Amsterdam: Boom.
- Oppermann, K. and Spencer, A. (2016) 'Telling stories of failure: narrative constructions of foreign policy fiascos', *Journal of European Public Policy*, doi:10.1080/13501763.2015.1127272.
- Owen, D. (2012) *The Hubris Syndrome*, London: Methuen.
- Parker, C. and Stern, E.K. (2005) 'Bolt from the blue or avoidable failure? Revisiting September 11 and the origins of surprise', *Foreign Policy Analysis* 1(2): 301–33.
- Patashnik, E. (2008) *Reforms at Risk: What Happens After Major Policy Changes Are Enacted*, Princeton, NJ: Princeton University Press.
- Payne, K.B. (2015) *Deterrence in the Second Nuclear Age*, Lexington, KY: University of Kentucky Press.
- Pressman, J., and Wildavsky, A. (1984) *Implementation*, Berkeley, CA: University of California Press.
- Radaelli, C., Dunlop, C.A. and Fritsch, O. (2013) 'Narrating impact assessment in the European Union', *European Political Science* 12(4): 500–21.
- Ragin, C.C. (2008) *Redesigning Social Inquiry: Fuzzy Sets and Beyond*, Chicago, IL: University of Chicago Press.
- Sagan, S.D. (1993) *The Limits of Safety*, Princeton, NJ: Princeton University Press.
- Schuck, P.H. (2014) *Why Government Fails so Often, and How it can do Better*, Princeton, NJ: Princeton University Press.
- Schafer, M. and Crichlow, S. (2010) *Groupthink Versus High-Quality Decision Making in International Relations*, New York: Columbia University Press.
- Starbuck, W.H. and Farjoun, M. (eds) (2005) *Organization at the Limit: Lessons from the Columbia Disaster*, New York: Wiley Blackwell.
- 't Hart, P. and Tindall, K. (eds) (2009) *Framing the Global Meltdown*, Canberra: ANU Press.
- Tiffin, R. (2010) 'A mess? A shambles? A disaster?', *Inside Story*, 26 October 2010, <http://insidestory.org.au/a-mess-a-shambles-a-disaster> (accessed 1 February 2015).
- Tilly, C. (2008) *Credit and Blame*, Princeton, NJ: Princeton University Press.
- Vaughan, D. (1996) *The Challenger Launch Decision*, Chicago, IL: University of Chicago Press.
- Walker, S.P. and Malici, A. (2011) *US Presidents and Foreign Policy Mistakes*, Stanford, CA: Stanford University Press.
- Walsh, J.I. (2006) 'Policy failure and policy change', *Comparative Political Studies* 39(4): 490–518.
- Wildavsky, A. (1977) 'Doing better and feeling worse: the political pathology of health policy', *Daedalus* 106(1): 105–23.
- Yetiv, S. (2011) *Explaining Foreign Policy: Explaining US Foreign Policy in the Gulf Wars*, Baltimore, MD: Johns Hopkins University Press.