

# HEAD OR HEART?

## THE CHALLENGE TO EVIDENCE-BASED POLICY-MAKING

E

The last 20 years have seen far more effort to base policy-making on “the use of evidence” to help governments understand their policy problems and options to solve them. This practice is not without limitations or even always adhered to, but is more widespread than in the past. More recently there have been signs that the use of analysis to support policy is losing momentum. Indeed, politicians have become acutely aware that pronouncements based on evidence seem less effective than those that touch the public’s emotions. The lesson is that evidence-based policy is more likely to succeed in its objectives, but that the evidence does not speak for itself: it needs to be communicated well to compete with emotions-based policy-making.

A

D

V

I

S

The rationale for evidence-based policy-making is simple enough: all policy decisions involve trade-offs, and good analysis is needed to make those decisions efficiently and effectively. This might seem particularly obvious when spending is involved: government resources are limited and evidence is needed to measure the trade-offs between one use and another. It would not make sense to expend all resources on improving the nation's health; some are needed to educate people and try to keep them safe. The question that can surely only be answered by analysis is: how much should go in each pot?

In the first quarter of the 20th century, the Cambridge economist Arthur Pigou elegantly described the rational way to approach this question, with:

*...the postulate that resources should be so distributed among different uses that the marginal return of satisfaction is the same for all of them... Expenditure should be distributed between battleships and poor relief in such wise that the last shilling devoted to each of them yields the same real return.*

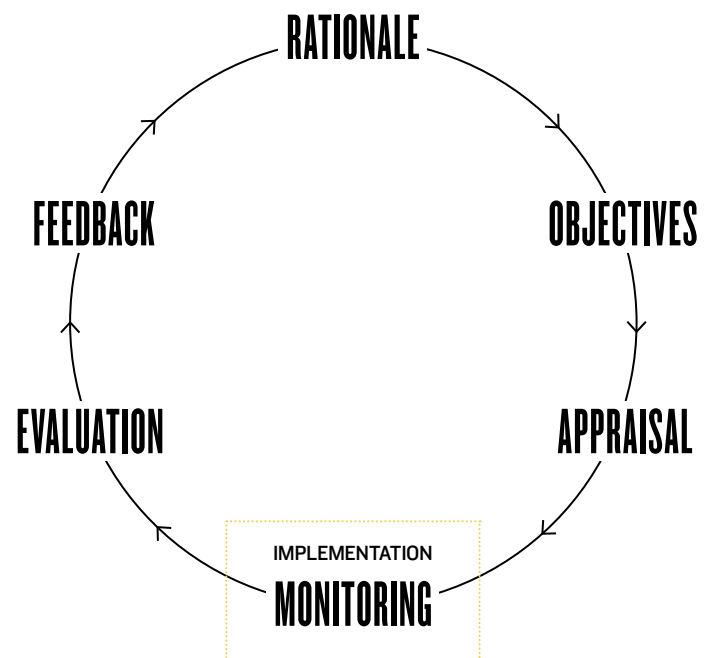
Over the following century, governments and officials have made repeated efforts to articulate this “postulate”, or principle, in practical terms. As the Permanent Secretary to the UK Treasury, Sir Tom Scholar, reminds us in the 2018 version, the Treasury's *Green Book* – the official how-to guide to the appraisal and evaluation of proposals for government policy – has been in publication for over 40 years. In the UK and many other countries, legislative proposals have to be accompanied by “impact assessments”. Cost-benefit analyses are the common accompaniment to policy-making in most jurisdictions.

The *Green Book* describes in detail the approach to be followed with respect to proposals for legislation or public spending in the UK. It articulates the different elements of the “policy cycle”, beginning with the identification of the rationale for intervention, and concludes with the need for a feedback mechanism, as shown in [Figure 10.1](#).

The aim of appraisal, according to the *Green Book*, is to help “decision makers to understand the potential effects, trade-offs and overall impact of options by providing an objective evidence base for decision making”. But if that has long been the purpose of policy appraisal, the phrase “evidence-based policy-making” did not begin to trip off political tongues until the 1990s.

**“THE QUESTION THAT CAN SURELY ONLY BE ANSWERED BY ANALYSIS IS: HOW MUCH SHOULD GO IN EACH POT?”**

FIGURE 10.1:  
POLICY CYCLE



The desire for robust evidence to justify the use of public resources has come from a number of different political angles. Governments in the 1970s, when the rise in public spending was beginning to cause unease in a number of countries, were occasionally attracted by ideas of “zero-based budgeting” – making every department argue for every penny – but the inertia in any public spending system easily overcame such ambitions.

In the 1980s, President Ronald Reagan struck a chord with US voters when he argued that “government is not the solution to our problem; government is the problem”, and the British prime minister, Margaret Thatcher, began a long programme of returning government enterprises to the private sector. In *Green Book* terms, the first link in the policy cycle – establishing a rationale for intervention – became less of a formality: it had to meet the test of a new political scepticism about the role of the state.

By the turn of the millennium, however, the pendulum had swung back towards a vision of government as enabler and facilitator of societal improvement. But if the “rationale” had again become easier to establish, the demand for evidence to support policy had become even stronger, was articulated publicly, and had a new, avowedly practical focus on well-managed delivery.

The British prime minister, Tony Blair, argued that the job of government was to decide on the right policies and then “deliver them effectively...this means learning from mistakes, seeing what works best”. And in 2009, in his inaugural address, President Barack Obama declared that:

*...The question we ask today is not whether our government is too big or too small, but whether it works – whether it helps families find jobs at a decent wage, care they can afford, a retirement that is dignified. Where the answer is yes, we intend to move forward. Where the answer is no, programs will end. And those of us who manage the public's dollars will be held to account – to spend wisely, reform bad habits, and do our business in the light of day – because only then can we restore the vital trust between a people and their government.*

Nor was this only an Anglo-American preoccupation. In Germany in 2003, a series of active labour market programmes were introduced to try to bring down the country's high unemployment rate. This depended on the findings of an expert commission chaired by Peter Hartz, the Personnel Director of Volkswagen (see Box 1).

## HARTZ AND MINDS

**Some useful lessons about the making of complex policies can be drawn from a comparison of two recent social programmes in Germany and the UK: the Hartz reforms and the introduction of Universal Credit. Both had as one of their key objectives an increase in work incentives, although this element in the Hartz reforms was part of a wider package of labour market measures, while the proposal for Universal Credit was an ambitious plan to swallow up the full range of means-tested benefits and disgorge them as a single payment.**

**Germany's reforms were launched in 2003, after a period in which painfully slow economic growth and a sclerotic labour market had driven the country's unemployment rate up into double figures – on one measure, to over 13%. The chancellor, Gerhard Schroeder, had appointed an expert commission chaired by Peter Hartz, the Personnel Director of Volkswagen, and moved swiftly to implement its findings.**

**There were no pre-implementation policy trials, but the reforms were introduced over three years, and post-implementation expert reviews were used to weed out unsuccessful elements (of which the make-work schemes were the most notable). These reviews helped to demonstrate that the reforms had succeeded in reducing Germany's underlying unemployment rate by three percentage points, and Germany's more flexible labour market proved remarkably resilient to the subsequent financial crisis.**

**The institutional and benefits changes were, in technical terms, executed smoothly. But with the final phase (Hartz IV), which sharply reduced payments to the long-term unemployed, income inequality in Germany increased and the centre-left government began to lose popular support. The Social Democrats suffered for this in the 2005 federal election, a lesson on the extent to which even “successful” policies need to be constantly recalibrated and resold to the electorate.**

**By comparison, in the UK, Universal Credit has got off to a very rocky start. It was the personal project of a departmental minister, or secretary of state, formerly (and briefly) leader of the Conservative Party, Iain Duncan Smith, and was passed into law in 2012 with the intention that it should be fully rolled out by 2017.**

**A dramatic simplification of the welfare system may have been a laudable ambition, but it was an extremely expensive one, in conflict with the government's efforts to cut the welfare budget. The secretary of state failed to secure Treasury buy-in as the costs of Universal Credit mounted and criticisms of design flaws went unheeded. He resigned in 2016, claiming the welfare budget had been short-changed.**

**Criticisms of the Universal Credit scheme multiplied, both with respect to costs (the UK's National Audit Office concluded it would be more expensive to administer than the six benefits schemes it replaced) and to design, particularly the long delay before those moved onto the new benefit received their first payment.**

**Subsequent ministers secured extra funding from the Treasury and began redesigning the scheme. Fortunately it was being "trialled" with selected groups and in selected areas, and the period for such trials was soon extended. In 2018 a further £4.5 billion was pumped into the scheme and its conditions softened. By 2019 the latest secretary of state, Amber Rudd, had announced that another "careful" trial migrating 10,000 people to Universal Credit would be started the following July, and there would be another parliamentary vote before the decision was taken to roll the scheme out in full.**

**It is, of course, far too soon to reach conclusions on the effectiveness of this reform. But unlike the Hartz reforms, Universal Credit does not seem to be cutting welfare costs, even though it is dogged by stories of hardship caused to losers, and the use of food banks is reported to have increased in areas where it was being trialled. Unemployment has continued to fall to levels not seen in the UK since the 1970s, but it is far too soon to say whether anticipation of Universal Credit has had any part to play in this shift.**

**The scheme certainly illustrated the benefit of pre-implementation trials, but also their limitations: trials are not without cost when real people's lives are affected. Accurate costing and full government buy-in before policies are launched remain essential. ■**

In France, social policy experimentation has also grown in popularity. Several large-scale randomised control trials have been implemented, including a quasi-experimental trial of the new French Minimum Income Scheme in 2009. All these examples demonstrated an appetite for "evidence-based policy-making". But what precisely was meant by this catchphrase? Was it just a fashionable label to attach to different policy approaches? Or did it have a hard core of must-have elements?

## DEVICES AND DESIRES

A broad definition of evidence-based policy-making was provided in a UK Cabinet Office paper in 1999 – it is an approach that:

*...helps people make well informed decisions about policies, programmes and projects by putting the best available evidence from research at the heart of policy development and implementation.*

Beyond such generalities, a powerful influence on the understanding of what was needed to fulfil such an agenda was the American psychologist Donald Campbell, who had refined his vision of *The Experimenting Society* between the 1970s and the 1990s:

*...The US and other modern nations should be ready for an experimental approach to social reform, an approach in which we try out new programs designed to cure specific problems, in which we learn whether or not these programs are effective, and in which we retain, imitate or discard them on the basis of their apparent effectiveness on the multiple imperfect criteria available.*

Thus the key elements in the approach were seen to be the collection of relevant evidence, and its refreshment through experimentation to discover "what works".

In what was possibly the high point of faith in experts, the UK Cabinet Office attempted to define what could be counted as "evidence" in its 1999 white paper *Modernising Government*. The list runs as follows:

*...Expert knowledge; published research; existing research; stakeholder consultations; previous policy evaluations; the Internet; outcomes from consultations; costings of policy options; output from economic and statistical modelling.*

Evidence can play a role at various stages in policy-making, as the table below illustrates:

### THE USE OF EVIDENCE IN POLICY-MAKING

Policy process	Data and methods
Identification of a problem	Surveys, administrative data, statistical analysis
Cause of the problem	Econometric models, experimental designs
Rationale for intervention	Theoretical analysis, distributional analysis
Appraisal of options	Administrative data, cost-benefit analysis
Monitoring and evaluation	Surveys, randomised control trials, independent audit

# “THE PENDULUM HAD SWUNG BACK TOWARDS A VISION OF GOVERNMENT AS ENABLER AND FACILITATOR OF SOCIETAL IMPROVEMENT”

The final category takes the list into the second key element – experimentation. The most powerful influence on this part of the approach to evidence-based policy-making has been the development by, among others, the US criminologist Lawrence Sherman of the five-level Maryland Scientific Methods Scale (SMS).

The screening of policy evaluations carried out by What Works Centres in the UK, which were launched in 2013, makes use of this approach. The version of the scale used by one of these centres is summarised below:

## The Maryland SMS (What Works Centre for Local Economic Growth)

### Level 1:

Either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group. *No use of control variables in statistical analysis to adjust for differences between treated and untreated groups or periods.*

### Level 2:

Use of adequate control variables and either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group. *In (a), control variables or matching techniques used to account for cross-sectional differences between treated and control groups. In (b), control variables are used to account for before-and-after changes in macro-level factors.*

### Level 3:

Comparison of outcomes in treated group after an intervention, with outcomes in the treated group before the intervention, and a comparison group used to provide a counterfactual (e.g. difference in difference). Justification given to choice of comparator group that is argued to be similar to the treatment group. Evidence presented on comparability of treatment and control groups. Techniques such as regression and propensity score matching may be used to adjust for difference between treated and untreated groups, but there are likely to be important unobserved differences remaining.

### Level 4:

Quasi-randomness in treatment is exploited, so that it can be credibly held that treatment and control groups differ only in their exposure to the random allocation of treatment. This often entails the use of an instrument or discontinuity in treatment, the suitability of which should be adequately demonstrated and defended.

### Level 5:

Reserved for research designs that involve explicit randomisation into treatment and control groups, *with Randomised Control Trials (RCTs) providing the definitive example.* Extensive evidence provided on comparability of treatment and control groups, showing no significant differences in terms of levels or trends. Control variables may be used to adjust for treatment and control group differences, but this adjustment should not have a large impact on the main results. Attention paid to problems of selective attrition from randomly assigned groups, which is shown to be of negligible importance. There should be limited or, ideally, no occurrence of “contamination” of the control group with the treatment.

The What Works Centres, focusing on education, crime prevention, local growth initiatives, children’s social care and early years and ageing, are gaining a strong reputation for their ability to filter vast quantities of research for the nuggets of evidence useful to policy-makers. They had an impressive forerunner in the National Institute for Clinical Excellence, which was created in 1999. This has shown that it is possible to detach difficult decision-making from the political process and rest it on expert evidence and careful trialling – for at least part of the time.

## ON THE PLUS SIDE...

The growing role of evidence in the policy-making process has stimulated the collection and storage of relevant data by researchers, while the rapid growth in computing capabilities, information and communication technologies, and statistical sciences has led to the better use of such data. Randomised control trials, deemed by such frameworks as the Maryland SMS to set the highest standards for policy evaluations, are far from universal. But a large number of quasi-experimental methods as well as qualitative research methods have become the norm in evidence-based policy-making.

Similar bodies to the What Works Centres have been set up in many other countries. The US now has the What Works Cities network funded by Bloomberg; Canada is setting up its own centre; and countries from France to Japan are developing their own initiatives to embed evidence, sometimes drawing on the UK experience. There are many ways in which these institutions can evolve: making better use of vastly greater flows of data; connecting to neighbouring fields like impact measurement; and sharing global experience.

Such approaches have clearly had an impact. Professor Sherman's development of evidence-based policing in the US is an obvious example. His randomised control trials of such innovations as the wearing of body cameras by police officers, or of "hot spots" policing, together with his development of the "Triple T" framework – targeting, tracking and trialling – have had worldwide impact on the use of police resources. In the UK, notable evidenced-based policies in other fields include the Sure Start programme, the Educational Maintenance Allowance and the Employment and Retention Advancement (ERA) Demonstration Project, among others.

But the UK is a long way from a model of government where evidence would have primacy in terms of influence on policy decisions, and a cursory glance at the news flow in other countries, democratic or otherwise, demonstrates the same influences of emotion or opinion on major decisions. It is time to weigh up these factors, and ask which way the balance of policy-making is leaning at the start of the 2020s.

## THINKING FAST, NOT SLOW

John Maynard Keynes put the other side of the story of policy-making to the elegant "postulate" of his contemporary, Arthur Pigou:

*...There is nothing a government hates more than to be well-informed; for it makes the process of arriving at decisions much more complicated and difficult.*

Policy-making is complicated by the pressure of different objectives: not only political imperatives (parties' manifestos, opinion polls) but also the differing agendas of civil servants and their departments, as well as the deep inertia in resource allocation that constrains dramatic shifts.

There is, moreover, a further set of pressures that distracts from the use of evidence, or rather which confuses evidence of inputs with evidence of outputs. Success, in the eyes of ministers and their departments, lobby groups or political parties, is often articulated in terms of the amount of money spent, or the rate at which it is increasing – X billion more on defence, welfare, health or education. The nonsensical extreme to which such an approach can lead is illustrated by the story of the UK's "National Programme for IT" for its National Health Service.

In the first decade of the new millennium, something of the order of £20 billion was proudly spent on this project – a sum that would have paid for perhaps 30 new hospitals. Eventually, the scheme had to be abandoned, with little or no cost recovery; one of the most spectacular failures in a long history of disastrous government IT projects.

This was also an example of a common weakness of governments for engaging in grand legacy projects that prove to be spectacularly undercosted or whose costs are poorly controlled (see Box 2). In the UK, the Treasury specifically warns against "optimism bias" in proposal costing, but probably only succeeds in scraping the tip of the iceberg.

In a 2018 article for the *Journal of European Political Research*, Messrs Jennings, Lodge and Ryan analysed 23 policy "blunders" committed in a variety of different countries that fall into this category, including such well-known extravaganzas as the Millennium Dome and the Hamburg Concert Hall. A common feature of these projects is that their reversal costs are high, both financially and politically.

Few politicians relish the U-turn. Searching for other common elements or causes, the authors highlight "over-excited politics" and a lack of administrative capacity, sometimes exacerbated by the choice of the wrong instrument.

Policy-making is, in short, beset by biases in favour of over-spending, intended or unintended, rather than output-focused analysis. It was, after all, the arch-proponent of evidence-based policy among British prime ministers, Tony Blair, who gave the final go-ahead to the Millennium Dome.

Even where lip service, at least, is paid to the need to base proposals on evidence, it is not difficult to find examples of the following common weaknesses:

- The evidence that has been developed is not well matched to the particular policy being proposed. It may be outdated or collected in relation to a different environment, geography, cohort or community.
- The resources that have been put aside for a robust evaluation are insufficient. Only some of the critical policy questions have been answered.
- The objectives are intangible (e.g. the “Big Society”). This does not make them worthless, but it does make it hard to appraise the value of policies proposed to deliver them.
- The production of evidence is an afterthought, produced just in time to meet statutory requirements (e.g. regulatory impact assessments).

In short, even within evidence-based frameworks, objective evidence may not have proved decisive against the demands of policy-makers and/or their political masters. And this may as often have been caused by “cognitive biases” as by the pursuit of conflicting objectives.

## CALLING THE SCORE

So how do we weigh up the “evidence” on the direction of policy-making, as we approach the third decade of a new millennium? Until recently, it seemed possible to argue that, despite frequent and inevitable dips in the road, we were on a long upward journey, on which each new stage in the development of evidence-based policy-making had led to further improvements in the available techniques, and further acceptance of its importance. However, the recent trends in global and national politics have brought this optimism into question.

Three worrying phenomena are much in evidence. In some ways, these could be seen as problems of success; but they also look like kinks in the road.

- A wide range of evidence is available to politicians, from competing research organisations/think tanks to internet posts. It is not difficult to find a study that can support a particular position, allowing policy-makers to cherry-pick. They also tend to weigh more highly, or come to rely on, those parts of the media that share their prejudices – an obvious example of “confirmation bias”. However, the public’s difficulty in sifting the evidence may be compounded by:

- A misguided attempt by “impartial” media to demonstrate balance by giving equal weight to unequal sources of information. In their attempts to appear even-handed, public service broadcasters frequently give equal prominence to minority voices whose research may be far less thorough than mainstream conclusions. News programmes have long relied on argument between opposing “talking heads” and are only gradually waking up to the need to give listeners and viewers “fact checks”.
- A resentment of experts, fed by populists, which has made voters keener to listen to non-expert voices that resonate with their own concerns.

Meanwhile, neuroscientists and psychologists have been building up our knowledge about human responses to evidence. The nub of it is that human beings don’t really like changing their minds, and are certainly not very willing to have their minds changed by data.

A range of psychological and neuroscientific studies suggest that people are not much influenced by facts, figures or data. According to conclusions reached in 2018 in *The Influential Mind: What the Brain Reveals About our Power to Change Others* by Professor Sharot, an accomplished neuroscientist at University College London:

*...the problem with an approach that prioritises information and logic is that it ignores what makes you and me human: our motives, our fears, our hopes and desires...data has only a limited capacity to alter the strong opinion of others.*

The reluctance of political leaders to change their minds or abandon their projects has already been noted. In fairness, it must be acknowledged that they often got to where they are by battling against the “evidence” – that is to say, contrary to the expectations of the prevailing consensus – with strong instincts that they then naturally tend to rely on in preference to evidence.

The election of Donald Trump, the economic policies of Margaret Thatcher, arguably even Winston Churchill’s refusal to make terms with the Nazi regime in 1940, all took place in defiance of the judgement of “wiser heads”. But the best leaders know how to weigh up the evidence before deciding to ignore it: Churchill had been avid for information about German rearmament throughout the pre-war period, and it was only when Mrs Thatcher became cavalier about evidence that she embarked on the policy that cost her the premiership (see Box 2).

Photo: Dave Sinclair – poll tax protests.



# POLLIS



Photo: Dave Sinclair – poll tax protests.



## POLLS APART

Britain's brief and disastrous experiment with local government finance in the late 1980s provides a vivid illustration of the clash between emotions and evidence in policy-making. The change was introduced after a surge of political emotion against a local tax loosely based on property rental values called the "rates". Many of the Thatcher Government's supporters saw "rates" as unfair to elderly single occupants, compared with households which contained multiple income-earning (and local service-using) residents but paid no more tax. A flat-rate "Community Charge" would instead be levied on every resident.

Arguments that such a tax would prove hard to collect, put by the Treasury, were ignored by the prime minister, although a "trial" was undertaken by introducing the charge first in Scotland. This added to the political disaster by leading the Scots to complain their country had been used as a guinea pig by English Tories. And as what was rapidly rechristened the "poll tax" was rolled out in England too, opposition erupted nationwide and its weaknesses became evident.

Property, which can't move, is not difficult to tax: people (especially the young) proved much harder, disappearing rapidly off the register. Exceptions had to be made for those unable to pay, particularly as the shrinking tax base drove up the level of charge that local authorities had to levy, in a vicious spiral of policy failure. The anger MPs experienced in their constituencies was a more powerful factor in the downfall of Mrs Thatcher than any disagreement with her over Europe. Local revenue collapsed.

The new prime minister, John Major, moved quickly to plug the hole in government finances with a 2.5% increase in VAT, and announced a review of alternative systems of local government finance. A local income tax was seen to be impractical in a densely populated country, where many people live in one local authority area and work in another. A local VAT rate would have suffered from the same border difficulty. But although a property base looked the best option, emotions were still running high against any risk of a "return to the rates".

So the department responsible for local government came up with a scheme for a tax made up of two different elements – one based on property values, the other, "personal", element being levied on each of the residents in the property. Analysis by the Treasury and No. 10 Policy Unit showed this to be extremely expensive to administer. They also conducted demographic analysis to demonstrate that the number of households containing more than two adults who could be expected to pay a personal tax of any size – that is to say, adults who were not students, non-earning and/or on benefits – was small. Too small, that is, to be worth the construction and maintenance of a complex personal register to try to capture as many of them as possible.

This paved the way for the introduction of a straightforward property tax with a discount for lone occupants – a discount which had to be claimed by the occupant, thus obviating the need for a full register to be maintained. The tax would be levied at rates covering broad bands of property values, avoiding the tax spikes which characterised "the rates". The result – the council tax – was introduced in 1992 and remains in place to this day. ■

# APART

What, however, these neuroscientific and psychological studies demonstrate is the equal resistance of those outside the public eye to data-based arguments intended to make them change their minds. It follows that politicians are more likely to succeed by appealing to emotions than by relying purely on reason.

A vivid example of this, cited by Professor Sharot, is given by the responses of two aspirants for the Republican candidacy for the US presidential elections in 2016: the paediatric neurosurgeon Ben Carson and the successful candidate Donald Trump. In one of the primary debates, the discussion turned to the link between childhood vaccines and autism. Mr Carson, the expert, pointed out that:

*...the fact of the matter is that we have extremely well-documented proof there's no autism associated with vaccinations.*

Mr Trump's response was:

*...Autism has become an epidemic...out of control...you take this little beautiful baby and you pump – I mean it looks like it's meant for a horse not a child...just the other day a beautiful child went to have the vaccine and came back, a week later got a tremendous fever...now is autistic.*

As we have all learnt since, it is that appeal to the emotions, however incoherent, that gets through.

## CONCLUSION

What does this mean for evidence-based policy-making? While the use of analysis to support policy has increased immeasurably in the past 20 years, there is reason to be concerned that, for the most important policies, this approach is losing momentum. In areas such as trade, for example, there seems little appetite for evidence to support policy. Instead, it appears that politicians have become increasingly aware that pronouncements based on evidence are less effective than those that touch public emotions.

The all-important lesson is that evidence does not always speak for itself. It remains fundamentally important to build frameworks within which as much public policy as possible is grounded in robust evidence, and constantly retested on the journey from proposal to implementation. Where high or heated politics are not involved, that may be all that is needed. But for all policies, and especially for those that are contentious, policy-makers need to be constantly aware of the need to persuade and retain support.

**“WHEN THE  
INFORMATION  
CHANGES, I CHANGE  
MY CONCLUSIONS.  
WHAT DO YOU DO, SIR?”**

Dry explanations of the research evidence will not meet the need. Another dimension is required, and that is to understand how the public thinks, and how to deliver research in the most compelling way imaginable. Effective communication of objectives, engagement with the emotions of those whose support is needed, a continuous taking of the public temperature and anticipation of likely reactions may be as critical to the completion of a successful “policy cycle” as the analysis of the academic data and statistical results of control trials.

Both appraisal and communication have one thing in common: the need to remain sensitive, flexible and responsive. Evidence-based policy-makers, in short, are the one group who cannot afford to be resistant to new information. Whether apocryphal or not, there is much to be said for the story of Keynes's response to a criticism of inconsistency:

When the information changes, I change my conclusions. What do you do, Sir?