

BETTER PUBLIC SERVICES THROUGH EXPERIMENTAL GOVERNMENT

Jonathan Breckon
Foreword by Geoff Mulgan

March 2015

Thank you for the following for their help with this provocation paper:
Professor Jon Agar, Laura Baynton, Hasan Bakhshi, Albert Bravo-Biosca, Jo Casebourne, Stuart Deaton, Dan Jones, Charlotte Macken, Danni Mason, Geoff Mulgan, Oliver Quinlan, Jim Riccio, Ben Ritchie, Alex Sutherland, Professor Gerry Stoker, Professor Carole Torgerson, Stian Westlake. The views and any errors in the report remain the author's own. The author would welcome comments, which should be emailed to: Jonathan.Breckon@nesta.org.uk



The Alliance for Useful Evidence champions the use of evidence in social policy and practice. We are an open-access network of 2,000 individuals from across government, universities, charities, business and local authorities in the UK and internationally. The Alliance provides a focal point for advancing the evidence agenda, developing a collective voice, whilst aiding collaboration and knowledge sharing, through debate and discussion. We are funded by the Big Lottery Fund, the Economic and Social Research Council and Nesta. Membership is free. To sign up please visit: www.alliance4usefulevidence.org

Nesta...



BETTER PUBLIC SERVICES THROUGH EXPERIMENTAL GOVERNMENT

CONTENTS

FOREWORD	4
----------	---

1. INTRODUCTION - THE CASE FOR EXPERIMENTAL GOVERNMENT	6
2. CALLS FOR EXPERIMENTAL GOVERNMENT ARE NOT NEW	13
3. WHY EXPERIMENTAL GOVERNMENT MATTERS NOW	18
4. HOW TO DO EXPERIMENTAL GOVERNMENT	22
5. THE CHALLENGES OF EXPERIMENTAL GOVERNMENT AND HOW TO MEET THEM	27
6. CONCLUSION AND RECOMMENDATIONS	35

ENDNOTES	37
----------	----

FOREWORD

The idea that governments should experiment isn't new. China has a very old tradition of testing ideas out before taking them to scale. Nineteenth-century reformers in the West, inspired by the achievements of science, often advocated a deliberately experimental approach too.

But this idea has never been all that enthusiastically embraced. The public aren't keen to be guinea pigs. Politicians fear being crucified for the failures and given no credit for the successes. Others worry about the ethics of experimental methods like random assignment – if a new approach looks promising shouldn't everyone have a right to benefit from it?

Yet the benefits of experimentation are huge, as this report confirms. Far better to test ideas out on a small scale rather than on a whole nation. Far better to let promising ideas improve before they're imposed on everyone. And far better to find out quickly if apparently brilliant ideas that work well on paper don't work well in practice.

These arguments for experimentation help to explain why there is now much more use of formal and less formal experimental methods around the world, from welfare to education, health to criminal justice. But there is still uncertainty about what method to use and when, and whether, for example, the methods that work in clinical trials are always going to be appropriate for the work of government.

This paper attempts a remedy. It shows how experimental methods are being used, and how they could be used in practice. Its starting points are simple: the best way to find out if an idea is good is to test it, and anyway ideas are never born fully formed. They're always partial and imperfect and need to be refined and developed, and of course no plans survive their first encounters with reality intact.

There are many methods used for experiments and there is a craft to doing this well – for example, too much measurement too early can kill the most promising ideas (Google would never have survived a rigorous assessment in its first couple of years, and the same is probably true of everything from Surestart to the PFI). The report also covers the question of how to fit experiments into the daily life of public organisations. In some discussions of this issue there is an unhelpful polarisation between purist advocates of randomised controlled trials, who see them as the only truly valid source of knowledge about what works, and others who reject them outright. Both positions risk being equally simplistic, and equally blind to the often subtle evidence about evidence. RCTs can be immensely useful, particularly in fields where they are uncommon. But as medicine and healthcare have discovered they are far from being panaceas, and are often most useful in combination with other kinds of evidence. They are only one tool; they are not always the best tool; and they are not well-suited to testing many types of policy. That's why this paper advocates attention to the quality of evidence and the appropriateness of the tools used to generate it.

The report is also clear about the limits as well as the potential for experimentalism. No amount of experiments will persuade people to change their minds on issues like same-sex marriage, or the UK's membership of the European Union, or whether bankers bonuses should be taxed. These, and many other choices, have more to do with values. Other areas are just too messy and systemic, like wholesale banking reforms, to be easily testable.

Then there is the challenge of time. Governments often do slowly what should be done fast, and fast what should be done slowly. Ill thought out reforms are rushed into implementation at great cost. The experimental method offers a reasonable compromise – fast action, but on a small scale, leading to phased adoption at a larger scale. That gives politicians plenty of examples to point to, but at less risk. But that may not always be possible, and there is an unavoidable tension between the demands of the public, as represented by politics, and the ideal of cool and calm experimentation. The public understandably want problems to be solved, and are frustrated by inaction. Telling them that a useful pilot is underway that will provide lessons in five years time is unlikely to be very satisfactory.

I hope this paper will be useful. There is growing interest in introducing more experimentation to more areas of public action. We all stand to benefit in the long run, from better policies and less wasted money. Here is both useful ammunition, and useful attention to the subtleties and nuance that will be needed alongside the rigour if experimentation is to become a habit.

Geoff Mulgan,
Chief Executive, Nesta

1. INTRODUCTION – THE CASE FOR EXPERIMENTAL GOVERNMENT

If we want effective public services, we need an experimental, learning government – robustly and systematically testing things out, measuring them and growing what works.

Without experimentation government stagnates. And we can't afford business as usual in the face of the immense challenges to adapt to ageing populations, climate change and fiscal pressures. In business, experimentation is commonplace. Google claims to have run 12,000 randomised experiments in just one year, with about 10 per cent of these leading to business changes.¹ In a recent survey by Nesta of UK internet economy businesses, controlled trials and experiments were the fastest growing analytical techniques used.² Experimentation is also mainstream in science, medicine and international development. But it is not routinely used in social policy interventions in the UK.³

Breakthroughs in transforming public services only come from a willingness to push at boundaries, to take risks, and, sometimes, to fail. If the government is to play a role in innovating, it's essential that it does so in the expectation that some of what it does may well fail, and that when things are not working they will be stopped.

A learning, adaptive government

Experimental government is not about trying things out in a haphazard way. It needs to be done in a way where we can genuinely learn from those experiments and adapt. Indeed, governments experiment all the time – rolling out new policies on the populace, but without really learning if they are doing any good. The Russian Soviet 'experiment', for example, was not really an experiment at all. The government could not really evolve, test different approaches and set up feedback loops. The reality was that Soviet central planners had a 'pathological inability to experiment', creating some disastrous command-and-control industrial policies.⁴ In the book *The Blunders of our Governments*, academics Anthony King and Ivor Crewe catalogue a range of big, failed projects over the last three decades in the UK that were not properly tested out, such as Individual Learning Accounts and the Child Support Agency.⁵

But when pilots or more rigorous experiments are used, they can be ignored. In the face of political pressures and the urgency to deliver results within short timetables, experimentation can be sidelined. That is why we are calling for a genuine ethos of experimental government. It requires a degree of humility and openness to risk and failure. We usually do not know exactly which policies will work. Experimental government is a way of finding out.

An experimental approach is also vital in moving beyond a static 'evidence-based policy'. Evidence can sometimes only reflect what worked in the past – or in other countries. Focusing on what seemed to work in the past reinforce the *status quo*.⁶ We need to constantly strive for new alternatives and new knowledge to meet current challenges.

Confronting complexity and risk

When dealing with any open complex system about which there is missing information and substantial uncertainty, the best way to discover is to probe and experiment.⁷

In the area of business innovation policy, for instance, experimentation is more effective than old-school industrial policy of command-and-control – where government knows best, picks winners, and rolls out business-support policies with little idea if they will succeed or not.⁸ The superior approach is to embrace uncertainty, and for government to have an innovation policy which actively experiments with business to create valuable (open) data and intelligence. As Hasan Bakhshi, Alan Freeman and Jason Potts put it, “Innovation policy... can turn up knowledge of opportunities faster, it can spread this knowledge faster, and it can dramatically reduce the social cost of repetitive failure.”⁹

Experimental government means avoiding an approach where everything is well and truly fixed from the outset. As the author of a UK government report on piloting, the late Professor Sir Roger Jowell said, we need a “spirit of experimentation, unburdened by promises of success”.¹⁰

This requires a grown-up approach to government that supports risk-taking and is comfortable with failure and learning from mistakes. This will mean a shift in attitude, at Westminster, Holyrood, Cardiff and Stormont, and also at city government levels. It includes accepting that we won’t always have ready-made solutions to social and economic problems. As Geoff Mulgan, CEO of Nesta puts it:

“Ideas are never born fully formed. They’re always partial and imperfect and need to be refined and developed... before long the idea needs to be tested out in the real world. No plans survive their first encounters with reality intact, and often the fastest way to develop an idea is to put it into practice, usually on a small scale so that the risks are contained.”¹¹

Such an approach to running the country is hard in practice. We as voters and readers of media unrealistically demand perfect 20:20 policy vision from our political masters. Government is not allowed to fail, or do U-turns, or to test out new ideas. This makes it hard for adjustments, reversals and tweaks when, say, developing new approaches to teaching children in schools, or tackling crime on our streets. But we can – and do – already experiment in these areas. Even in sensitive areas experimentation is possible. The police have trialled different ways of stopping and searching potential terrorist threats under Schedule 7 of the Terrorism Act 2000 in the real-life setting of an international airport.¹²

Small is (mostly) beautiful

Many experiments need to start small, so as to limit the political and economic cost to a government of a trial that fails.

The advantage of an experimental approach is that you can be ‘innovative and cautious at the same time...try things out in an overly tentative manner’, according to a US position statement on social policy experimentation funded by the MacArthur Foundation.¹³ You can test your new idea and, if it fails, it’s only on a limited scale, limiting damage.

But if the innovation succeeds, you can extend it to the wider world. This allows policies to be tested, rigorously researched and adjusted, before being rolled out to scale. For example, trials by HMRC, the UK's tax authority, and the Behavioural Insights Team on ways to encourage people to submit their tax returns, resulted in over £200 million in additional tax take.¹⁴ The educational benefits of free fresh fruit have trialled in some schools to see if it might work nationally.¹⁵

Experimental government should happen locally and regionally, not just at the national level.* Sub-national, local or city government offers a ripe opportunity to rigorously test out different approaches to policies, programmes and practice.¹⁶ Testing out new policies by state governments is commonplace in the US. Their federal structure enables individual states to mount their own fairly large-scale experiments to test the likely impact of a proposed new policy or delivery mechanism or both. One overview in 1997 found 240 ongoing experiments in US states across all social policy areas including welfare, schooling and addiction.¹⁷

While the UK doesn't have the scale of the US, we could do more to test out different innovations between the constituent parts of the UK. The National Assembly of Wales was the first in the UK to vote to ban smoking in public places (although not to implement it), and the first to raise a levy on plastic bags.¹⁸ Although these weren't formal trials, it shows the Welsh government's appetite to try new things out and do things differently within the UK, an appetite that that could lead to more robust experimentation.

But scale is relative. 'Small' for US and China may look very large to others. Despite China's heavily centralised, bureaucratic and hierarchical government, it has been able to experiment in a relatively bottom-up way but on major economic policies (see Box 1).

Box 1. The politicised experimentation of China's economic rise.

Several scholars of China's economic rise have stressed the important role of experimentation.¹⁹ What's been so surprising is that so much of this flexibility has taken place in the extremely inflexible institutions of the Communist party-state.²⁰

The approach to experimentation in China's party-state is very different from other democratic countries. It does not use the small-scale pilot projects seen in Europe and the US. China has used bolder experiments that have impacted China-wide national policies, such as more positive approaches to private business after bottom-up experimentation.²¹ But these experiments were highly politicised. They were often championed by ambitious government officials. And there was little public evidence, data or social science to share from these experiments. Throughout the 1980s and 1990s there were thousands of regulations that were formally titled as 'experimental or provisional', according to the sinologist Professor Sebastian Heilmann, Director of the Mercator Institute for China Studies in Berlin.²² China also set up a high-level Bureau for Comprehensive Planning and Experimental Points in 1988 to coordinate local level economic experiments. These 'experimental points' focused on targeted policy areas or economic sectors, dealing with rural policies, financial regulation, social security, healthcare, education.²³

China also used 'experimental zones', the best known of which are the special economic zones, such as Shenzhen Special Economic Zone near Hong Kong. Ideas were tested in this 'laboratory for legislation' that had a strong influence on central government, including testing out ideas borrowed from East Asian neighbours, including land auctions or wholly-owned foreign-owned companies.²⁴

*An 'experimentalist' approach is also happening beyond national boundaries – at transnational level. See for example, this account for 'experimentalist governance' relating to regulation across national boundaries, such as the European Union's Water Directive, where transnational institutions can frame problems in a relatively open-ended way, and subjected to periodic revisions in the light of knowledge learnt through experimentation. Sabel, C.F. and Zeitlin, J. (2012) 'Experimentalist Governance.' In Levi-Faur (2012) eds., 'The Oxford Handbook of Governance.' Oxford: Oxford University Press. See: [http://www2.law.columbia.edu/sabel/papers/Sabel%20and%20Zeitlin%20handbook%20chapter%20final%20\(with%20abstract\).pdf](http://www2.law.columbia.edu/sabel/papers/Sabel%20and%20Zeitlin%20handbook%20chapter%20final%20(with%20abstract).pdf)

The use of the internet for experiments also means we can have a different approach to scale. Governments have the opportunity to test things with large and statistically meaningful samples. No longer are we constrained by the practical challenge of experimenting with people 'out in the field' in small geographical areas. We can experiment with small changes to government websites at very low cost and reach large numbers of people. The Government Digital Service and the Behavioural Insights Team have experimented to improve everything from encouraging organ donations, to filling in HMRC tax forms (see Box 2). In these cases they started 'small', which meant relatively low cost both to the public purse, and in terms of the potential political and reputation damage of a failed experiment.

A duty to experiment?

There is an ethical case for governments to do experiments.

Too often new public policies are rolled out nationally with little trialling or evaluation. In effect, governments experiment on the whole population at once but without learning what works.²⁵ In the UK, £66 billion worth of government projects have no plans to evaluate their impact, according to a National Audit Office report in 2013.²⁶ It is unethical to experiment on us in this arbitrary way, where we will not learn if policies are doing more harm than good.²⁷

But knowing if we are failing only comes by robustly testing government innovations. Much of what we are currently doing may indeed be failing to deliver. It may even be doing harm. In the 1970s, as part of the America's 'war on drugs', it seemed like a brilliant idea to steer teenagers away from crime by getting them to visit prisons and meet convicted criminals. Shock the kids and scare them off a life of crime. Despite support from US police forces and the White House, programmes like Scared Straight backfired.²⁸ Rather than scaring kids, it made crime look like it *did* pay and offending increased in some areas. But it was only after robust evaluations that we were able to find out that it wasn't having the desired effect.

Box 2. Why robust experiments matter – the case of synthetic phonics and how children learn to read.

The question of how best to teach a child to read is a taxing one for any parent – as well as for government policymakers. The current approach is for synthetic phonics, by which children learn to match sounds to letters and groups of letters. But this is still effectively a large-scale but weak ‘experiment’ on our children as robust evidence is lacking. A team commissioned by the Department for Education and Skills and led by Professor Carole Torgerson, then at York University, was asked to review the evidence on synthetic phonics. Although there was some promising evidence from the US and a small-scale study in Clackmannanshire, Scotland, that this approach worked, the evidence was relatively weak. Their government review found only a dozen small trials, the biggest of which involved 120 children.²⁹ They urged caution in making national policy. Torgerson recommended in her report that the government should roll it out gradually, with the first areas to benefit to be chosen at random in a trial. But this advice was ignored, according to Torgerson, and ‘it just became policy’:

“As a result, we still don’t know whether or not phonics works as the main teaching strategy for all children. Some of the recent evaluation work has demonstrated synthetic phonics may not be having the impact that was hoped for. If we’d done a randomised trial we would have known before the policy went national.”³⁰

Setting up government programmes based on partial evidence from overseas or results from small sample sizes is not good enough. There is a responsibility amongst those in power to not over-claim that they know best, but to have the humility and integrity to know that we will need to try things out in a robust way to learn if our new – and old – ideas work.

A fog of names... and our working definition

We are aware that there is a great deal of confusion in some of the language of the *methods* employed for experimental government. Interviews with UK civil servants and ministers back in 2003 discovered a ‘fog in relation to nomenclature’ between the different policy-testing mechanisms: ‘trailblazers’, ‘policy trials’, ‘pioneers’, ‘prototypes’ or ‘benchmarks’, ‘randomised controlled trials’ and other terms were bandied around:

“Not only do departments across UK administrations differ in their use of labels, but so too do divisions of the same department or administration, compounding the confusion. Fanciful terms for early evaluations of policy have multiplied – to the extent that one of the Ministers we interviewed reported having been given the option to choose the tag that he or she liked best for a pilot from a range of competing but equally inappropriate options.”³¹

So how do we get a clearer approach? We were reminded by our economist Nesta colleagues, Hasan Bakhshi and Albert Bravo-Biosca, that it's worth starting with a dictionary definition of experimentation: 'A test or an investigation, especially one planned to provide evidence for or against a hypothesis' (*Collins English Dictionary Third Edition*). We have adapted this to government.

Our definition is:

experimental, learning government is one that robustly and systematically tests things out, measures them and grows what works.

We are not interested here in 'seat-of-the-pants innovations in public policy';³² trying out new policies and programmes with no meaningful way of evaluating their impact. Our definition is not equivalent to 'freewheeling trial and error or spontaneous policy diffusion'.³³ It's best to see experimentation as a continuum, ranging from the messiness of all attempts by government to try something new, right through to experiments using the best available research, such as randomised controlled trials or other more appropriate techniques. In this paper we advocate moving to the right of this continuum.

A continuum of experimental government:



This provocation paper does not travel down the well-trodden path of accounts for and against different experimental research methods *per se*. But we do touch on the methods that can be used by governments that take up the challenge of being more experimental. It's important that we use robust methods to evaluate the impact of experiments such as – if appropriate – randomised controlled trials or quasi-experimental research methods.* It is, of course, always the case that politicians and governments will need to balance evidence with other political pressures in their decision making. But strong evaluation techniques means that governments can see if the experiment worked or not, rather than having to rely on gut instinct.

*For some non-specialist guides for policymakers on experimental research techniques see: Behavioural Insights Team (2012) 'Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials.' London: Behavioural Insights Team. Or this Nesta and Ben Goldacre D.I.Y website. For more on quasi-experimental methods see (2014) 'Guide to Testing Social Policy Innovation.' London: European Commission. See: <http://www.alliance4usefulevidence.org/publication/spark-guide-to-testing-social-policy-innovation/#sthash.SbrjWEoC.dpuf>

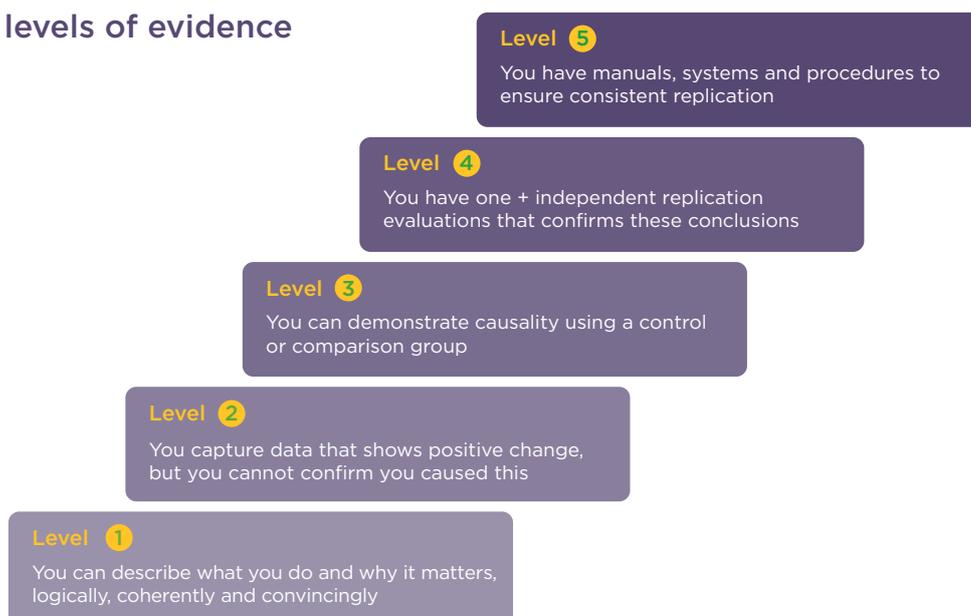
Using the right evidence – the dangers of a ‘methods-led’ approach

But experimental research methods, such as RCTs, can be the wrong way to measure impact. We **do not** advocate using a ‘methods-led’ approach, where you only start off with one method, such as randomised controlled trials, and then apply it to understand the impact of your initiative. It is much better to think about the appropriateness of your methods. Having a research design of brilliant technical quality is not enough. You could have a sound controlled trial, but one that tells you nothing about whether the same intervention might work elsewhere. As the founding President of the UK Evaluation Society, Professor Elliot Stern, put it: ‘We need to start with what we want to know about programmes, rather than a particular toolkit, however fashionable it might be’.³⁴ Other designs, such as those using statistical, qualitative, theory-based and other approaches, might be much more useful in understanding impact.³⁵

Some early innovations may not be suitable for experimentation, as there is no clear hypothesis or coherent activity. We must avoid the danger of having too high a threshold of evidence for new activities, compared to more established government work. In some cases, it’s just too early to say yet what needs to be tested, and we need to let it grow before designing a robust test.

One way to avoid blanket-approaches to evaluating experiments, is to use standards of evidence, a method used by Nesta, Project Oracle, Pearson Education and the What Works centres, such as the Early Intervention Foundation.³⁶ These standards of evidence mean starting from a basic level where you can describe what you do and why it matters, logically, coherently and convincingly. You then move up the scale to areas towards Level 3, where you may require experimental research, and greater confidence on causality, and then beyond to evidence of replication and scaling. Your level of evidence is appropriate for where you are in developing an intervention.

Nesta levels of evidence



One final point about the scope of our argument. Our main interest is on government experimenting in social policy, although we discuss other areas, such as innovation policy, research and development (including defence), medicine and health.

2. CALLS FOR EXPERIMENTAL GOVERNMENT ARE NOT NEW

Experimental government has been thought about for many years. Four hundred years ago Sir Francis Bacon challenged conclusions based on everyday experience as fraught with potential for bias.³⁷ He called for the greater use of experiments to minimise such bias. Below is a brief recent history of experimentation and government, with a particular focus on the US experience.

Box 3. Sir Francis Bacon's 'Salomon's house'

Sir Francis Bacon was no ivory tower philosopher. He was a politician who served as an MP, Lord Chancellor and Attorney-General to King James I. In his book *New Atlantis* published in 1624, he proposed a utopian state that he called Salomon's House, a proto-government backed centre for science and experimentation. But he didn't propose this in a work of philosophy. His book *New Atlantis* was a work of narrative fiction, where he describes some shipwrecked sailors who come across a perfect island state. The aim of this novel was to influence King James I.

It didn't work. His attempts to get James I to back his plan to set up a Salomon House-type institute fell on deaf ears. But much later, after he died, the plans outlined in *New Atlantis* inspired the Royal Society, founded in 1660. Bacon's ideas influenced experimentation in many areas, from science to government. Indeed, some have regarded him to be **the** philosophical influence behind the dawning of the industrial age. His philosophy set out a vision for an experimental and inventive state – that eventually led to the mechanical inventions that made possible the industrial revolutions in the subsequent centuries after he died.³⁸

The term 'social experiment' was widely used amongst some 19th century European³⁹ thinkers, according to a history of the 'experimenting society'.⁴⁰ Thinkers such as Auguste Comte, John Stuart Mill and George Cornewall Lewis discussed it. But it held a very different meaning in those days. It was merely a metaphor, borrowed from natural science, for describing events disturbing social order. The metaphor indicated that something can be learned about normal social life if events disrupting it are carefully observed. But the disruptions were not deliberate and planned, but accidental 'Acts of God' such as famines or floods. Indeed, these early social scientists didn't believe that deliberate experiments – controlled by researchers or others – were ethical.⁴¹

It wasn't until the 20th century that government experimentation was really given more traction. The rise of experimentation in the early 20th century arose from a combination of increasing interest in applying Bacon's logic of science to social phenomena, and the development of the welfare state by Western governments. Centralised government grew. More social policies were introduced, from slum clearance projects to new work programmes for the unemployed. Government agencies were set up to run these activities. And their administrators increasingly turned to social scientists to deliver 'hard proof' of the efficacy of government interventions.⁴² The proof requested had to be as objective and standardised as possible. Increasingly scientific experimental research designs were seen as the answer.

More recent influential scholars such as Friedrich Hayek, Douglass North and others have pointed to policy experimentation as an important tool in economic policy.⁴³ It is, they argue, an effective mechanism to find out ‘what works on the ground and ...produce innovations that are conducive to entrepreneurialism, investment and economic growth’.⁴⁴ However, some of these approaches to experimentation have been ‘surprisingly ill-defined and vague’.⁴⁵ Our definition of experimentation focuses more on the purposeful and deliberate testing of an idea. Not the incremental policymaking of tinkering and ‘muddling through’ to use Lindblom’s well-worn description of government.

The bold experimentation of FDR’s New Deal

The first big wave of experimentation occurred during the 1930’s New Deal years in the US. In 1932, Franklin D Roosevelt called for ‘bold, persistent experimentation’ and he surrounded himself with social scientists as his personal advisers.

“The country needs and, unless I mistake its temper, the country demands bold, persistent experimentation. It is common sense to take a method and try it: If it fails, admit it frankly and try another. But above all, try something. The millions who are in want will not stand by silently forever while the things to satisfy their needs are within easy reach.

(Franklin D Roosevelt address at Oglethopre University, 22 May 1932)

By 1940, 8,000 social scientists were working for the federal government. FDR called one of the leading proponents of controlled trials for government programmes, Professor Charles Meriam, his ‘uncle Charley’.⁴⁶

A reforming zeal – the 1960s

However, interest waned in experimentation with successive governments after FDR. Interest in the US did not return until a quarter of a century later, with the reforming agendas of John F Kennedy’s ‘New Frontier’ and Lyndon Johnson’s ‘Great Society’ in the 1960s.⁴⁷ There was increased interest in experimental designs of new programmes – not just collecting data, but comparing different approaches to control groups. The 1960s battles against racial segregation, the decay of inner cities and unemployment created political bases for mobilising scientific experimental approaches:⁴⁸

“The contemporary version of experimentation grew principally out of the domestic social amelioration emphasis of the 1960s when the federal government launched dozens of major programs aimed at eliminating poverty, raising the condition of minorities, and solving urban problems. Some social program planners realized that information was not sufficient or appropriate to guide the design of certain intervention programs. Accordingly, they opted for controlled trials of miniature versions of plausible alternative programs, rather than staking everything on a single intervention mode that was merely consistent with existing inadequate data.”⁴⁹

Lyndon Johnson's 'Planning, Programming, Budgeting System' was set up to find the most effective and least costly alternatives to achieve social progress: vast experiments with welfare allowances were instigated to establish whether financial aid reduced people's motivation to work. The New Jersey Negative Income Tax Experiment, for instance, was set up to understand the benefits of the government giving extra pay to the poorest workers, instead of paying taxes.⁵⁰ This experiment inspired several comprehensive US federal plans proposed during the 1970s: President Nixon's Family Assistance Plan, Senator George McGovern's Universal Demogrant Proposal, and President Carter's Program for Better Jobs and Income.⁵¹

Another of the social scientists who got involved was Donald Campbell. His seminal paper in *American Psychologist*, 'Reforms as Experiments', reads like a manifesto for experimental government. He called for modern nations to be ready for an experimental approach to social reform.⁵² In subsequent years he provided one of the classic texts on 'quasi-experimentation': how to remain rigorous even under the limitations of practical research and evaluation.⁵³ His name inspired the international Campbell Collaboration to encourage greater use of experimental research and systematic reviews for social policy.⁵⁴

However, experimental government did not take off in the way in which Campbell had envisaged in his call to action. There were a number of 'growing pains' with some of the prominent experiments in the US in the 1960s and 1970s.⁵⁵ Researchers overpromised on what they could deliver. The ability of experiments in themselves to affect change may also have been oversold in the 'golden age' of the 1960s. Policymakers were impatient with the length of time it took for experiments to provide answers. This led to a decline in experimentation.⁵⁶ Many of the earlier experiments were not matched to the right issue and 'addressed questions that in retrospect were not easily answered within an experimental context. We are more humble now; we have a better sense about when to experiment'.⁵⁷ These growing pains led to a re-evaluation of the merits of experimentation in the 1980s and 1990s.

The return of the experimenters in the UK

The co-founders of LSE and the Fabian Society, Sidney and Beatrice Webb, were early advocates of social experimentation (though seemingly unaware of the rise of social experimentation in the US). They pointed out that there is far more experimentation going on in 'the world social laboratory in which we live', but as most of this is 'wrapped in secrecy' it yields nothing to science and called for more 'scientific' social policy.⁵⁸

But, Peter John, Professor of Political Science and Public Policy at University College London, suggests experimentation for UK social policies is a relatively recent phenomenon.⁵⁹ He argues that it started in the late 1990s, exemplified by the publication in 2003 of HM Treasury's *Magenta Book*, a guide for government research, which set experimental designs as the gold standard of evaluation. There have since been some successful Government-backed experiments, such as the trial of the effectiveness of new services to improve job retention and advancement prospects for low-wage workers in the Employment, Retention and Advancement Demonstration.⁶⁰

Box 4. CASE STUDY: Helping people progress out of low pay and no pay

The most systematic attempt to help people out of poverty towards better paid jobs was through the Employment, Retention and Advancement demonstration project mentioned above (ERA) – which sought to test approaches in an experimental way between 2003 and 2007, in sites in the UK.

The project was aimed at long-term unemployed jobseekers, lone parents on Income Support, and lone parents receiving Working Tax Credit. Under the programme, participants were supported in three ways, through: face-to-face adviser support; money for training; and cash bonuses for staying in work.

The project ran as a randomised controlled trial – where customers were referred randomly either into the programme or into a ‘control’ group that did not receive that support, with the differences between the two being compared.

The evaluation found that ERA did produce short-term improved retention and earnings for all groups, and that these held over the longer term for the long-term unemployed group.⁶¹

The ERA work flags up an important point about experiments: average effects can mask important differences in effects on subgroups. Experiments can help determine whether interventions work better for some types of people than others. For example, ERA had substantial earnings impacts on lone parents with A-level qualifications, and on long-term unemployed living in social housing.

Aside from the policy test it conducted, ERA influenced the experimentation movement in the UK in a few ways. It was an unusual pilot in that it tested a new policy before government committed to the policy on a national level, whereas most ‘pilots’ at the time were tests of early rollouts of policies that were already adopted by government. ERA also helped establish the feasibility of very large-scale randomisation within the UK. This was not to be taken for granted, as opposition to randomisation was considerable at the time. The fact that 16,000 people were successfully randomised in six regions of Britain in 58 local Jobcentre Plus offices with no major controversies was an important achievement. It was hard to argue afterward that ‘it couldn’t be done.’⁶² ERA was also an important model as it included a comprehensive and careful benefit-cost analysis, which could set standard for other experiments.

However, experimentation in the UK has only really taken off since 2010, according to Professor John, with a series of ‘nudge’ experiments conducted by the Cabinet Office’s Behavioural insights Team. In the next section we will argue why experimentation matters to government now.

Box 5. Learning from the past: the Slough Experiment and Labour's Ministry of Technology

This failed but visionary experiment in 1966 shows a scale of ambition that existed in the past and a lesson about not attempting too much at once.

The idea for an experiment came from Sir Ieuan Maddock, a deputy controller of the Ministry of Technology – MinTech – the department for Wilson's White Heat policies. This department, said Maddock, 'should launch a concentrated 'industrial experiment'... an area which contains a wide variety of industrial activities should be selected to serve as a Microcosm of the Nation at large'.⁶³ Computers and every other modern telecommunication innovation should be concentrated in one area: Slough.

'We are dreaming up a so-called 'Slough Experiment', wrote an official in May 1966, 'in which the Slough area would be exposed to every technological trick on the books of circumambient Ministry of Technology agencies'.⁶⁴ Prime Minister Wilson and HM Treasury were consulted and there was ministerial support for the experiment. A feasibility study was completed by consultants paid for by the atomic weapons side of UK Atomic Energy Authority (UKEA). But according to the historian Professor Jon Agar who has studied the archives on this little-known episode, it was killed off by critics worried that 'It was in the wrong location, ...a subsidy for the prosperous South East, and the Atomic Energy Authority was the wrong agency'.⁶⁵

Although it never saw the light of day, the Slough Experiment provides lessons current to today's proponents of experiments in government. If the experiment had been launched, it would have failed to tell us much as they attempted to do too many things at the same time (so it would have been far to the left of our continuum on page 11). There were no controls or comparison groups, and it would have been impossible to tell what, of all the interventions, worked. Another lesson is that 'normal politics will disrupt intentions', such as the antipathy to hosting it in the prosperous south of England. According to Professor Jon Agar, the Slough Experiment also reminds us that in-house government teams can be highly innovative: 'The idea (of an experiment at all), and the innovations (computing, specific expertises) were largely grown in-house, successfully'.

3. WHY EXPERIMENTAL GOVERNMENT MATTERS NOW

The current policy climate is ripe for experimentation. In a time of increased pressures on public services to meet the demands of many ‘wicked’ and intractable social issues, such as family violence or multi-generational poverty, we need to find what works by ‘fast testing in real world situations rather than paper development followed by national implementation’.⁶⁶

Less ideological-based policy

Impatience with belief-based government means the time is right for experimentation. Author and commentator Caitlin Moran captured this impatience in her new year’s *Times* column in 2014 when she said that the public are bored of ‘huffy editions of *Question Time*’, left vs right, Keynesian economics vs pro-market, based mostly on what was ‘inherited’ views. She argues that what we need to do is:

“...stop talking about our political feelings and beliefs and find out, once and for all, which one actually works... Run some experiments. Work out some figures. Do randomised, controlled trials in, say, local government. Perhaps neither method works. Great. Now we can abandon both and search for an original solution.’

Caitlin Moran, *The Times* 28 December 2013 ‘2013 and the end of belief’

That is not to say that ideology and conviction-politics should never be a part of government. Values and beliefs should always be core to government.* Judging acceptable levels of income equality, or laws on same-sex marriage, will and should be dominated by our beliefs. But there is always room to experiment.⁶⁷ This can be within the context of belief systems, such as testing-out *delivery*, rather than the policy itself. The Mexican government under President Ernesto Zedillo did have a strong conviction to help the poorest of the poor (see Box 6). But Yale-educated economist President Zedillo backed up his convictions with experiments to test out if the new approach of Conditional Cash Transfers would work, or not.

*At our fringe events on this topic at the Labour, Conservative and Liberal Democrat party conferences in September and October 2014, the audience disputed that ideology is less important in government. One MP also suggested to us that the issue was not ideology but ‘prejudice-based government’ that ignored evidence. Also, NatCen’s British Social Attitudes survey or the Hansard Society’s Audit of Political engagement show that values and beliefs are still important to many. It was suggested that the issue was a public aversion to mainstream Westminster political parties.

Box 6. An enduring anti-poverty programme experiment in Mexico.

The Mexican social assistance programme Progresa is a prime example of how experimentation can be done on an ambitious national scale – and survive different political regimes, transferring from Presidents Zedillo to Fox, and onwards to the present day.

All too often, new government policy innovations can be dropped by incoming governments which don't want to be associated with the past. But the strong experimental evidence of Progresa, built-in from day-one, survived changes in government. Evidence of success didn't come from political rhetoric – politicians saying it was a success – but randomised controlled trials.

Progresa has evolved and rebranded (first to Oportunidades, and is now as Prospera). It involved a powerful mix of an ambitious national policy commitment – namely reducing poverty – with a pragmatic approach that tested out different methods on the ground to see if they worked.⁶⁸ Chief amongst these was a method called Conditional Cash Transfers (CCT), offering payments to poor families if they encouraged their children to regularly attend school, or go to health clinics.

It was subject to a rigorous evaluation: 320 communities were randomly assigned to a treatment group, with 186 communities assigned to a control group. Between 1997 and 2005, it covered five million families, practically all of the Mexican households living in extreme poverty.⁶⁹

Randomisation took place at community level as household level randomisation was deemed to be unfeasible. Programme officials were particularly concerned about the ethical implications of withholding benefits for the purpose of evaluation – however, it was argued that budget limitations meant that not all eligible families could get the benefits from the beginning of the programme – so there was always going to be some withholding of benefit.

The positive benefits meant that a number of Latin American countries adopting the same programme.⁷⁰ A similar intervention was also trialled in the US by Mayor Bloomberg: Opportunity NYC, funded by private sources. Would the positive effects found in rural areas of a developing country translate to an urban area in the richest nation in the world? In this case, the CCT policy has not been adopted yet in New York, because the jury is still out on whether the new version of the model is effective; the evidence on the initial model, while having a number of important and positive effects, did not live up to its promise enough to justify adoption as policy. The mixed early results of NYC Conditional Cash Transfer experiments led to a redesign of the model and a new test, which is still underway.

And in a further twist, the Mexican officials from Oportunidades (the successor to Progresa) are in conversations with the New York team about the next stages in the evolution of their Mexican CCT model. Government staff are interested in learning from New York and other US experiments to design and test strategies to promote 'productive exits' (in their terminology) from the programme.

A policy environment open to experiment

The current climate also offers practical opportunities for experimentation. Budget constraints that result in the rationing of resources on either a geographical or demographic basis, and incremental roll out of major new policies and programmes (e.g. Universal Credit) could provide fertile ground for experiments (see section 4: How to do experimental government).⁷¹

Decentralisation also provides more chances to experiment.⁷² Devolving power away from Whitehall to schools, local authorities and frontline professionals provides opportunities to try out different approaches. However, one politician⁷³ told us that it is still important to coordinate local experiments, and suggested that a body like the Local Government Association could have an oversight role; facilitating parallel experiments in multiple locations, to compare different local approaches, such as in the four ‘Whole Place’ pilot projects in Essex, Greater Manchester, West London and West Cheshire.⁷⁴ Central Governments should also have a role in developing more coherent programmes of experiments that build on and complement each other, rather than commissioning *ad hoc* experiments on single initiatives.⁷⁵

The increasing use of the internet and social media in public services allows more experimentation – large scale, low cost and fast, such as the randomised controlled trial on voter mobilisation that went to 61 million users in the 2010 US Congressional elections.⁷⁶ The experiments to improve organ donation by the Behavioural Insights Team (BIT) involved a million UK citizens.⁷⁷

Box 7. UK Government Digital Service ‘nudge’ experiments.

Over 350,000 people have now registered for organ donation thanks to a clever and low-cost internet-based experiment run by the Government Digital Service (GDS).

The simple idea was based on the knowledge that people are generally more open to trying out new things after completing a successful online transaction. So, people who bought their tax disc via the GOV.UK. website were encouraged to join the NHS organ donation register, right after the ‘Thank You’ page on the Government website. But they weren’t sure how to encourage people, so different pages were trialled.

Working with the Cabinet Office Behavioural Insights team, DVLA and NHS Organ Register team, the GDS tested eight different calls to action with a total of over 1,000,000 visitors to the tax disc ‘Thank You’ page.

The eight variants appeared at random, and they worked closely with the NHS Organ Donation team to see which one generated the most completed registrations. This is known as A/B testing. The most successful variant introduced concepts of ‘reciprocity’ and fairness by asking people: ‘If you needed an organ transplant, would you have one? If so please help others.’

Since the DVLA tax disc online service is used by around two million people every month, the impact of such optimisation is significant. GDS predicts that some 96,000 more people will register compared with the original call to action. In total over 350,000 people have registered for organ donation via this one link on GOV.UK since they added the original call to action.⁷⁸

We also now have institutional support for experimental government, through the ‘what works’ centres, co-funded by the Economic and Social Research Council and the UK, Welsh and Scottish governments, covering everything from local economic growth to crime reduction. One of these centres, the Education Endowment Foundation has funded 93 projects in total since it was set up in 2011, working in over 4,000 schools with over 600,000 pupils. Of these, 77 are set up as RCTs. They are experimenting⁷⁹ on what works best in the classroom in areas such as testing different types of digital education⁸⁰ or school-wide approach to improving speech, language and communication support.⁸¹

There are other institutions to help experiment. Some of them have been around for many years and provide vast knowledge on advanced experimental social science, such as Dartington’s Social Research Unit, University of London’s EPPI Centre, and the York Trials Unit. The Behavioural Insights Team, once in Cabinet Office, now with a free reign to work with governments anywhere, can help officials to run experiments.⁸² Nesta and the Kauffman Foundation launched the Innovation Growth Lab in October 2014, a new global centre to encourage experimentation on ideas for innovation and growth policy in partnership with several governments around the world. We will also soon see the launch of policy lab set up by Nesta in Wales, partly inspired by the J-PAL Poverty Lab in US (who have run over 400 randomised controlled trials around the world). These institutional players will help smooth the road to more experimentation by providing expertise and research capacity.

4. HOW TO DO EXPERIMENTAL GOVERNMENT

We recognise that in the real world of politics and the day-to-day demands of government that experimentation is challenging. So how best to go about it? We need good leadership, the right incentives and the right skills to create more experimentation.⁸³

Embrace risk (and make failure OK)

Putting experimentation at the heart of government means challenging negative attitudes to risk.⁸⁴ To be fair, there can be good reasons why government has a cautious attitude: there can be some harsh political consequences of getting experimentation wrong. Testing out ideas on children at risk of sexual abuse or live experiments on air traffic control could, for instance, could have dire results. Public sector innovators face the political consequences of failure and have to contend with social equity issues when experimenting in a way that private sector innovators do not.⁸⁵ There is ‘loss aversion’ in the public sector, as the personal cost of being associated with failure is more severe and certain than the potential benefits of being association with success. This bias towards a fear of loss – rather than embracing success – means it is safest to stick with the *status quo*.⁸⁶

But a more benign attitude to risk and failure is central to fostering experimentation. As Professor Sir Roger Jowell, author of a key government guide on pilots said:

“*A pilot that reveals a policy to be flawed or ineffective should be viewed as a success rather than a failure, having potentially helped to avert a potentially larger political and/or financial embarrassment.*”⁸⁷

So how can we engender such an attitude to risk? Research has hinted at a range of ways that this can be done. Suggestions⁸⁸ for how to change the culture include: leaders in government – and their political masters – taking responsibility for failure so that employees feel safe to experiment and take risks; we also need to praise and recognise staff successes, even if a project fails; and we need a positive narrative and training on the benefits of experimentation and risk-management.

Prizing success

It’s an obvious point but worth stressing: if there is no motivation to transform public services, then there is no motivation to experiment. The *status quo* must feel like it is not good enough. We need to strive for public services to be better. One piece of research has even shown a link between the desire to experiment and the desire to improve: US school administrators who had little interest in educational achievement working on the Tennessee STAR (Student/Teacher Achievement Ratio) experiment, were also the ones who had low demand for experiments.⁸⁹

One practical measure is to set up positive reward structures to celebrate experimentation to harness motivation, perhaps through challenge prizes or awards. For example, the innovation awards for civil servants run by Queensland Public Service Commission, or PS21

in Singapore.⁹⁰ *Financial Times* economist Tim Harford has also called for an annual award for politicians who admitted they got something wrong, learned from their mistakes, and completed a policy ‘U-turn’.⁹¹

But a word of caution. There is a flip side to craving success. Whether consciously or unconsciously, experiments can suffer from the ‘Hawthorne effect’ whereby ‘public officials try to make the intervention look especially good because they are under the watchful eye of an experimenter’.⁹² Making interventions ‘blind’ to those implementing it, the ideal of ‘gold standard’ medical research trials, may not always be possible, but we must be aware of the potential for government officials to inadvertently or advertently bias experiments.

Create ‘Skunkworks’ institutions

New or improved institutional structures within government can also help with experimentation. These can be at the centre – or local and city level. The Behavioural Insights Team (BIT) operates a classic ‘skunkworks’⁹³ model; teams tasked with innovative projects work semi-detached from government, acting with a high degree of autonomy and unimpeded by day-to-day bureaucracy. BIT has worked with government departments on specific commissions, but ‘deliberately maintain(ed) a low profile as it developed and honed its method’.⁹⁴ The UK Ministry of Justice Data Lab,⁹⁵ working closely with New Philanthropy Capital, has opened up re-offending data to charities and voluntary organisations, allowing for experimental research (particularly quasi-experimental designs) so that charities can see if their work on reoffenders is having any impact against a control group. Other teams that are independent of (but close to) government include the UK Education Endowment Foundation or the French public policy laboratory Fonds d’Expérimentation pour la Jeunesse, which funds experimental youth programmes and evaluations, with government and private money.⁹⁶

These institutions have the resources and expertise to run experiments, and the space to make things happen and impact government thinking:

“They can ‘unfreeze’ organisational embedded practices, operating as neutral spaces dedicated to problem-solving in a highly experimental environment, bringing people together with a diversity of skillsets, creating opportunities for collaboration, and using methods such as prizes, design, ethnography, co-creation, rapid prototyping, experimentation, sharing evidence and supporting adoption and diffusion.

Casebourne, J. and Puttick, R. (forthcoming) ‘Setting up effective institutional arrangements for supporting innovation.’ Chapter 4, Public sector innovation report. Paris. OECD.

As well as institutions,* we need more strategic innovation funds – perhaps from private and charitable funds – to foster experimentation. In the US, funding for experiments has often come from charitable foundations. Indeed, this may have helped sustain the long-term viability of experiments, so that they outlived White House incumbents (see Box 7).

* For further ideas on institutional models to encourage experimentation relating to innovation policy – such as testbed environments and business ‘forensics laboratories’. See Bakhshi, et al., (2011) ‘State of Uncertainty; Innovation policy through experimentation.’ London: NESTA. p. 13–15.

The US not-for-profit Coalition for Evidence-Based Policy has an open competition for 'low-cost RCTs' in social policy three times a year, up to a value of £100,000.⁹⁷ But the money could come directly from government itself. The US Justice Department's National Institute of Justice has launched a 'Randomized Controlled Trial Challenge', a monetary prize to encourage criminal justice agencies to use low-cost experiments as a 'standard and straightforward approach to answering their questions and conducting their day-to-day business operations'.⁹⁸

A Nesta report on experimentation, *State of Uncertainty*,⁹⁹ recommended a public strategic competitive fund for experimentation around innovation policy. It could be created in several ways, perhaps from establishing an independent endowment (such as from Lottery revenues for example). Such funds could be used to directly support collaborative business experiments which test propositions of wider business and/or social interest.

Listen to the front-line and the 'base of the pyramid'

Experimental government also requires bottom-up approaches, listening to service-users and front-line staff who know the realities of delivering policies in practice. Those 'on the ground' can also feed in new ideas for an experiment.

Box 8. Schools experiment on the best methods to teach disadvantaged pupils.

Teachers have been invited by the UK government to suggest innovative approaches to teaching disadvantaged pupils.¹⁰⁰ Schools and teachers can apply to a £4 million research scheme called 'Closing the gap: test and learn' and be part of a controlled trial, run through the government-backed National College for Teaching and Leadership. The best approaches come 'bottom up' from teachers. Around 700 schools are taking part in the trial. The ones that look effective from the experiment will be replicated across the country.

A report by the OECD and World Bank recommends more extensive 'base of the pyramid' experimentation in government approaches to improve the lives of the poor in developing countries.¹⁰¹ Their starting premise is that innovation policies characterised by top-down government-led policies are not the right approach to promoting development, and call for a 'more appropriate approach to innovation... policy (that) involves search, experimentation, monitoring, learning and adaptation...' (p.15).

The authors of the OECD and World Bank report argue that it is important to embed evaluation in programmes and policies from the outset, and call for a shift away from a 'design globally, execute locally' approach to 'design locally, execute globally'. This involves the co-production of innovative solutions that draws on local tacit knowledge and social capital. To support this shift towards a bottom-up approach, the authors argue that new practices will need to be developed to ensure that promising innovations are shared and scaled. And poorly-performing policies and programmes need to be revised or eliminated.¹⁰²

Box 9. General David Petraeus and 'trial and error' in the Iraq war.

Deaths of Iraqis and Americans were mounting on a weekly basis during 2004. Radical thinking was needed to reverse the decline. The top-down command-and-control approach of Donald Rumsfeld, the Pentagon and State Department was failing in the post-Saddam Iraq counter-insurgency. But soldiers such as General David Petraeus and Major HR McMaster had other plans. They allowed experimentation with troops on the ground.

Away from Washington, more junior US army personnel would come up with a hypothesis for tackling the conflict in areas such as Mosul or Tal Afar. Test it out, drop it if it failed, or expand it if it worked. It was dramatically more effective when the troops on the ground had to devise their own strategies and tactics to limiting the carnage. These 'bottom-up' approaches eventually led to wider policies to contain the chaos, and improved counterinsurgency doctrine. The trial-and-error approach created a dramatic turn-around by 2008 and 2009 in limiting deaths and destruction. Of course this was no panacea, conflict continues in this region and the Petraeus strategy did not end problems conclusively, but the more pragmatic approach did at least dramatically reverse a downward trend of chaos, by learning from mistakes and making the best of a bad situation.¹⁰³

As well as dispersing power and listening to other voices, we must also learn from past experiments and wider research evidence. As with all policy development, experiments should be preceded by a systematic review of the evidence from the UK and overseas.¹⁰⁴

Apply a 'learning as you go' philosophy

Experimental government needs an ethos of flexibility; to adapt to events and unexpected consequences. Drawing on the New Zealand experiences of complex policy implementation, the NZ Institute of Policy Studies has set out an approach to experimental government: intelligent policymaking needs a highly pragmatic approach, 'learning as you go', based on accommodating every day complexities, testing policies in practice, evaluating rigorously, and applying what is learned to future policy thinking and decisions.¹⁰⁵

This approach means departing from purist experimental research methods – such as RCTs – towards something more iterative; learning from experience as you go along and adapting the programme.¹⁰⁶ Some academic experimenters also suggest providing early findings to decision makers through pragmatic trials and 'design experiments' as precursors to full field experiments.¹⁰⁷

This messier approach to experimentation may not be popular with researchers who want methodologically cleaner and valid experimental research, such as through RCTs. But the proponents of this 'pragmatic' approach argue that some sort of compromise often has to be struck between 'knowing everything' and 'knowing something'.¹⁰⁸

There is value in evaluating your progress as you develop – through what's called formative evaluations – that feedback into the design of a programme. Experimentation should draw on mixed methods approaches, including more qualitative techniques, to answer questions about theories of change and 'pry open the black box' to find out what's really going on in an experiment.¹⁰⁹

Evaluations within the experiment can be vital in developing critical thinking, as they ‘provide an opportunity for organisations to examine accepted truths, question the justification for specific claims and call attention to unstated assumptions’.¹¹⁰

Ultimately, those involved in policy implementation (policymakers, practitioners and evaluators) need to be flexible if they are to adopt an experimental approach:

“They need to be able to deal with ambiguity and changing situations without losing sight of strategic goals. They also need to be sufficiently flexible and nimble in their thinking and actions to take advantage of the serendipity that will arise during implementation.”¹¹¹

Experimenters must not be politically naïve

Researchers who want to experiment need to empathise with their government partners. According to one seasoned experimental researcher, Professor Gerry Stoker at the University of Southampton, academics need to sharpen their political skills as they often have a ‘naïve’ approach to influencing policy; an unrealistic rational decision-making model that fails to take account of the messiness of the real world.¹¹² Professor Stoker recommends that experimenters need to develop their ‘people’ skills – sensitivity of context, contingency and subjects that is more closely associated with qualitative research. They also need to be more aware of the political and policy context within which the experiments are taking place.

But the closer relationship between government and researchers has to be reciprocated. Officials in government would do well to link-up – at an early stage – with experienced evaluators and researchers who know the best way to design experiments. A recent history of MDRC in the US, a think-tank that has run RCTs on government social programmes for four decades, stressed the importance of creating a ‘learning community’:

“the central factor that sustained forty-five years of random assignment was (the) creation of a community – or researchers, welfare administrators and other public officials, funders, advocates, and state and legislative staff – that recognised and valued the distinctive credibility of such studies.”¹¹³

The main ‘heroes’ of their history of MDRC were the local and state administrators who ‘displayed the courage, determination and political skill needed to put their programs on the line’.¹¹⁴

Seize opportunities to experiment

Experimenters working with government need to be opportunists. Ready to seize windows of opportunity to experiment (see page 20 on a policy environment open to experiment).¹¹⁵ Social researchers often do not make the most of natural chances to try out different approaches.

For example, researchers could take advantage of budget constraints, with access to new interventions initially restricted to a random selection of an eligible population.

5. THE CHALLENGES OF EXPERIMENTAL GOVERNMENT AND HOW TO MEET THEM

There are areas of policy where this approach is just not the right one. Some decisions need to be taken swiftly by government and based on political instincts. Experimentation may just not be realistic, feasible, ethical or politically acceptable in cases such as legislating on same-sex marriage or going to war against an invading army. But there are plenty of other challenges with experimental government that are set out below.

Public fears of governments experimenting on people

There is a horrific history of government-sanctioned scientific experiments on human beings: medical experiments and torture conducted on Jews by Aribert Heim, 'Dr Death', at the Mauthausen concentration camp or Josef Mengele in Auschwitz; the infamous Tuskegee experiments on syphilitic poor black US citizens by the government's Public Health Authority, right up to the early 1970s; the Ministry of Defence experiments between 1940 and 1979 on a million unwitting people on the South Coast of Britain to see the impact of germ warfare.¹¹⁶

With such a history, it's no wonder there might be an immediate suspicious public response to the phrase 'experimental government'.

Box 10. CASE STUDY: The notorious Tuskegee syphilis study by the US Public Health Service

The US Government's Public Health Service was involved in an infamous unethical research experiment that ensured that 399 African-American men with syphilis went untreated. When the experiment ended in 1972, many men were dead and children with congenital syphilis had been born.

The purpose of the Alabama-based study entitled the 'Tuskegee Study of Untreated Syphilis in the Negro Male' was to examine the natural progression of syphilis in poor black men who received free health care from the government. The reason it felt justified in the 1930s was that no treatment worked very well for syphilis. So the medical researchers – supported by the government – believed it was acceptable to observe the progression of syphilis to help find a cure.

But what became a notorious medical decision was that the study continued even when a treatment was found. By 1947 penicillin was the standard of care for treating syphilis. It is hard to imagine such callousness now, but researchers prevented their subjects from participating in the national campaigns to treat syphilis with penicillin. In essence, the researchers denied an effective treatment to their subjects.

By the end of the study in 1972, of the original 399 men in the study, 28 had died of syphilis; 100 were dead of related complications; 40 wives had been infected with syphilis; and 19 children had been born with congenital syphilis.¹¹⁷

Public suspicions about experimentation are not unique to governments. Business, charities and scientific bodies can also engender distrust. There was a widespread backlash in newspapers and social media against Facebook in the summer of 2014 when it came to light that they had been doing experiments with 689,000 of their web-users. Researchers had re-jigged the Facebook News Feeds so that some were more positive than others. The idea was to see if Facebook users would write more positive posts through a ‘contagion’ of reading good news stories. It seemed innocuous enough but there was widespread criticism and a public apology from COO Sheryl Sandberg.¹¹⁸

The hostile public response flagged up two things: firstly, our innate distrust of being manipulated by those in power – whether business or government. Particularly if the perceived benefits to the individuals themselves seem small.

Secondly, the Facebook story raises the thorny issue of consent. Just because we clicked ‘yes’ and agreed an electronic Facebook terms of service form, doesn’t constitute *informed* consent. It’s an ethical issue that dogs medical and scientific trials – people agreeing to take part in experiments may be unaware of exactly what the research is about. They could be signing-up to be experimental subjects without fully grasping what it means – such as the fact that they may just be getting a ‘sugared pill’ of an inert placebo, rather than the pill that actually works. One survey in the UK found that 47 per cent of physicians thought their patients were *never* aware that they were about to participate in an experiment when signing a consent form.¹¹⁹ We return to how to handle the issue of ethics and public perception below on page 31.

Expensive knowledge

People sometimes worry that trials are expensive and complicated.¹²⁰ This does not have to be the case. Experiments to encourage organ donation by the Government Digital Service and Behavioural Insights Team (Box 7 above) involved an estimated cost of £20,000.¹²¹ This was because the digital experiments didn’t involve setting up expensive new interventions – just simple new web pages.

It’s also important to think about the relative costs of a trial. The cost of running a trial can be merited if compared to the cost of simply pushing ahead with an untested idea that may fail, at great expense and even harm to the public, as seen in the Scared Straight and D.A.R.E programmes (see page 9). The Nobel Laureate economist Robert Solow praised the cost of the US ‘Supported Work’ experiment to get former prisoners and drug addicts back to work because of the benefits – accrued to government:

“That is expensive knowledge, but I’m convinced the price is worth paying. Governments must be made to realize that even such large sums pale into triviality next to the much larger amounts that get poured down rat-holes in the belief that it is more important, or at least better politics, to respond hastily and visibly to social needs on the basis of no tested knowledge at all.”¹²²

Experimental research is also not necessarily more expensive than other types of evaluations, such as interviews, surveys and observational studies. Those studies can be very resource-intensive. We can also limit costs by using routine data that is already

being collected, such as figures from government tax collection, click-through rates on government websites, or exam results from schools. Nesta has also teamed up with Ben Goldacre to create a D.I.Y website <http://www.randomiseme.org/> to help any government official – or member of the public – design their own experiment. At no cost.

Getting the timing right

Experimentation can take time to get the results you need. This can be frustrating for those in power. As one senior civil servant put it, politicians are not going to wait for evidence that ‘answers yesterday’s question tomorrow’.¹²³ The Justice Secretary, Chris Grayling, made his frustration of pilots clear in 2013:

“*The last Government were obsessed with pilots. Sometimes those in government just have to believe in something and do it, but the last Government set out a pilot timetable under which it would have taken about eight years to get from the beginning of the process to the point of evaluation and then beyond. Sometimes we just have to believe something is right and do it, and I assure Members that if they went to Peterborough [the location of a criminal rehabilitation policy of payment-by-results pilot study] to see what is being done there, they would think it was the right thing to do.*¹²⁴

But rushing to conclusions on experimentation can be a dangerous approach. As Roger Jowell said in his government guide to pilots:

“*A policy pilot should be seen above all as a ‘test run’ the results of which will help to influence the shape and delivery of the final policy. It follows that a policy pilot must be allowed to run its course and produce its findings before the policy is rolled out. Too often, this has not been the case.*¹²⁵

Early interim results from an experiment may give some helpful pointers. But they may give a ‘misleading picture of long-term policy outcomes’.¹²⁶ We should not rush to conclusions. At our series of seminars with the Institute of Government, Nesta’s Hasan Bakhshi illustrated why it was important for experiments to be maintained in place for some time and longitudinal data collected. The initial results of Nesta’s Creative Credits experiment¹²⁷ in Manchester showed a highly positive effect: SMEs which benefited from innovation support from creative service businesses compared with the control group. But it would have been premature to have generalised policy lessons from those early findings:

“*Stopping the experiment at that stage would have concluded that this was a programme worth rolling out just as it had been designed. But those benefits decayed rapidly thereafter, such that as the programme ran on it did not offer significant benefits. Too often the political temptation would be to seize on earlier positive results and scale up a programme without waiting for the long-term results to come through.*¹²⁸

Even when the time for formal experiments – such as pilots – finishes, it should not be the end of the learning process. It should continue as the policy is scaled-up, with good evaluation and monitoring systems put in place.

However, there may be an urgency for government to deliver results. There may not be time to wait years to find out if an experiment is working or not. One MP at our fringe¹²⁹ event at the Liberal Democrat 2014 party conference in Glasgow said that most ministers think they have two years, maximum, to make a difference, before being moved on. Some experiments such as campaigns to reduce obesity, or changes in the school curriculum, can take years before we see any measurable impact. Producing interim findings may help if urgent results are needed. But as the example of Creative Credits above shows, they need to be treated with caution. Unless the trial is long enough to detect certain impacts, it can create a false impression of long-term policy success.

What works here, might not work there

It can be hard to generalise from a single experiment to the wider world. The unique context of one time and place means that it can be difficult to learn policy lessons to roll out nationally. For instance, the pilots using ‘payments by results’ in Doncaster and Peterborough are looking at new ways to reduce re-offending rates for those just out of prison.¹³⁰ It may be rolled out across the UK. But how do we know if the effects are due to the local context and human geography of Doncaster and Peterborough? Would it really apply in rest of UK – or even internationally? In their book *Evidence-Based Policy: A Practical Guide to Doing It Better*, philosopher Professor Nancy Cartwright and ex-businessman Jeremy Hardie have argued that this is a fundamental flaw in things like RCTs.¹³¹ It is very hard in practice to move from knowing that something that was effective there, will work here.

The best way to respond to this is to replicate any trial in multiple locations (and even the same location to see that not a flash in the pan result). Replication is an important part of experimental science and social science.¹³² It should be an important part of experimental government.

For example, replication is central to a ground breaking scheme for deprived first-time young mothers and their babies. Although there is 30 years of evidence on the life-long benefits of home visits by specially trained family nurses for vulnerable mothers, most of this evidence is from the US.¹³³ So in Scotland they are running a trial to see if the same benefits can happen there.¹³⁴ We need to be sure that the US results on intensive nurse visits are relevant outside the US. Scotland has also piloted the Drug Treatment and Testing Orders, even though they had been thoroughly piloted in England¹³⁵ (see also Box 6 on the efforts to test how well the Mexican Progresá anti-poverty measures would work in New York City.)

Another response to the limitations of context is to do much larger experiments.¹³⁶ The BIT trials to encourage organ donation came from a sample of one million people using the GOV.UK website. Not one specific UK location, but using the country at large. Businesses such as Facebook, Capital One and Google carry out experiments that are global in nature.

Local context will always be a critical factor in learning from experiments. But it’s also an issue for any form of evaluation or any roll out of a policy (a policy that is effective in West Lothian may be totally inappropriate in, say, the West Midlands). Experimentation is also a

never-ending process. We should always keep testing and trying policies out – even when rolled out nationally. The Family Nurse Partnership is an example of this attitude – despite three decades of evidence to support it, we still need to keep testing it, adapting and learning.

Too many policies ‘set in stone’

A fundamental challenge to an experimental state-of-mind is a political system that doesn't allow for change. Manifesto commitments, trailed in the media, debated in Parliament, may have too much political capital invested in them to really allow for experimentation. You have to stick with the original high-profile plan. As the 2003 Cabinet Office review of pilots put it, in such cases the only thing that matters is smooth implementation, ‘unencumbered by unwelcome news of the sort that suggests that the policy may incorporate flaws after all’.¹³⁷ The problem may be worse in the UK than in the US:

*“The British legislative process is, in practice, not very conducive to genuine piloting. By the time a policy has reached the statute books, its content (and often its methods of delivery too) have run the gauntlet of parliamentary debate, media examination, pressure from lobbies and scrutiny by committees. Emerging from this process, the final version of a policy may well incorporate numerous carefully worked compromises which are by then far too complex to be re-opened.”*¹³⁸

But it's not impossible to build in some room for manoeuvre at the outset. The UK's Department for Work and Pension's Employment Retention and Advancement (ERA) project was ‘explicitly designed to influence the existence and shape of legislation’.¹³⁹ The policy allowed for amendments and learning as it was rolled out, and wasn't forced to follow every letter of parliamentary law.

Allowing policy to be more ‘project-based’ could encourage such room for manoeuvre, according to a previous Nesta report on experimentation.¹⁴⁰ Policies could be more time-limited. It would offer policymakers space to try new approaches out, without committing to a blanket open-ended policies. For example, sunset-clauses at the completion of project – a sort of policy expiry-date – could foster experimentation by allowing new ideas to be tried out for a fixed period. According to the authors of the Nesta report, this would mean a significant institutional shift away from ‘endeavours to make policy into law... but instead to internalise the notion that the policy ends when the learning ends’.¹⁴¹

Ethics and (un)informed consent

A common criticism that we heard at our Conservative, Labour and Liberal Democrat party conference fringe events in Autumn 2014 is that experimentation is unethical. The usual argument is that a robust trial involves creating a comparison group, a control, and it may mean withholding an effective ‘treatment’, such as new policing tools, or financial incentives to reduce poverty. It is thus unfair, so the argument goes, that only some of the population gets what might be an effective treatment.

But on examination, what people really mean is not ethical issues, but concerns about reputation and politics. The fact that those missing out will eventually get the treatment (as long as it works) may not hold much sway with voters and citizens; they demand from their politicians and officials the same treatment that they heard their neighbours were getting.

The dilemma can be even worse if attempts are made to experiment based on randomisation. There is an instinctive antipathy to something that feels so arbitrary. There is a strong view amongst some that it is unfair to give treatments based on chance, rather than who deserves it most.¹⁴² Political distrust of randomisation assignment was a major issue for the Mexican government in the Progresa antipoverty programme (see Box 6). There was a great deal of resistance in political circles for randomisation of key services to help the poor. It didn't feel fair, and could ruin your political capital if your constituents weren't getting help – but neighbours were.

Maybe one approach to tackle these concerns is to appeal to the public's better nature? Persuade people that findings from experimentation may benefit society at large. But this is unlikely to help. Research evidence from medicine – where trials are commonplace – shows that participants have little idea what trials are about. 'Informed consent' for all patients may be an 'unobtainable ideal'.¹⁴³ Even if patients do understand, they sign-up for trials out of self-interest, hoping to get the latest innovative treatment, rather than out of public spiritedness and to help the population at large.¹⁴⁴ If self-interest dominates medical patients, we may have the same problem – perhaps worse – in areas of social policy like schools or policing. How can we persuade parents of children in schools or worried crime-ridden neighbourhoods that we need to experiment on them?

There are also practical issues with consent. Dean Karlan is professor of economics at Yale, who has been involved in many development projects using experimental research says that telling participants too much about the trial destroys the validity of the results by changing everyone's behaviour.¹⁴⁵ Camilla Nevill of the Education Endowment Foundation says that trials are often agreed to and conducted by schools, but not parents. Trying to persuade every parent to agree explicitly to the trial 'decimates' the number of participants.¹⁴⁶

Box 11. Moral philosophy and 'equipoise'.

Does government have a moral duty to experiment for the greater good? Moral philosophers in the Benthamite utilitarian tradition would argue that one's ultimate duty is to maximise utility by the greatest happiness of the greatest number of people. On the other hand, Kantians dispute this and say that you should always treat people with respect – never treat people merely as the means to other people's ends. The tension, as so often, is balancing the duties of individual and society.¹⁴⁷

For medical experiments, this tension has been tackled by a concept bioethicists call 'clinical equipoise'; trials can only be justified if there is genuine uncertainty in the expert medical community about the preferred treatment. A physician must have an equal state of uncertainty – or 'equipoise' – between the available options. The idea was first articulated by the philosopher Charles Fried in 1970s and, although it has plenty of critics,¹⁴⁸ it has become mainstream in medical thinking. Equipoise sought to answer the fundamental conflict that arises from the Hippocratic Oath where 'patients may not be used merely as the means to an end, however admirable that end might seem'.¹⁴⁹

The question arises if 'equipoise' could work as a justification for experimentation in social policy. If government has a duty to protect the interests of citizens, could we also use the ethical principle of equal uncertainty to justify policy experiments? The issue for social policy is that we **DO** sometimes have a pretty good idea of what works. So you can't argue

that there is a balance of uncertainty. In China, one experiment tested whether academic performance could be improved by providing glasses to those with poor vision.¹⁵⁰ Surely it was obvious to expect that academic test scores improved from those myopic Chinese rural students who did get glasses? Yes, it may seem obvious but this misses another side of experimentation. Sometimes the ‘obvious’ answer, based on common sense, proves to be incorrect (the Scared Straight example was based on a reasonable assumption that exposing children to criminals might scare them off, but the evidence showed a different story). We also need experiments to test out if innovations are cost-effective and workable for government. The cost-benefit analysis in China suggested very high economic returns to wearing glasses that justified state funding for free glasses. From the standpoint of social policy, the salient question is not ‘whether an intervention produces any benefit, but how much benefit and delivered at what cost’.¹⁵¹ Experimentation can give government valuable policy insights, even when the answers might seem obvious. One World Bank economist has argued for us to look at ‘policy equipoise’, and not clinical equipoise, where those wider considerations of cost and benefits are included.¹⁵²

The democratic deficit: why we need to make the case to the wider public

If randomisation is key to your experiment (and we are at pains to say that this paper is not about RCTs alone – other methods may be more appropriate), then one recommendation is to exploit opportunities where decisions were going to be arbitrary anyway. A team working in Mexico on a project related to Progresa recommended that:

“Because some decisions are always made below the level of political radar, we offer the generalization that randomization is always acceptable at one level below that at which politicians care.”¹⁵³

Once officials in Mexico understood this point, the team of researchers could work with them to find the best way to design an experiment – including randomisation – within the political and administrative constraints.

There are, however, some dubious democratic assumptions in such opportunistic experiments as seen in Mexican Progresa antipoverty case. Is it not inherently undemocratic to work ‘below the political radar’? While this might be pragmatic, surely it would be better to have a grown-up debate with your electorate about the necessity for experimentation?

One option to meet the challenge of public resistance, is to try and change the language. Academic experimenter Peter John at UCL has recommended a linguistic shift.¹⁵⁴ He argues that the term ‘experimentation’, rather than simply ‘experiments’, captures the idea that experimentation can become a ‘way of life’ that encourages curiosity and a willingness to take calculated risks to find out new knowledge.

Others have suggested a dedicated public understanding programme to explain the importance of experimentation.¹⁵⁵ This would be a bit like the ‘public understanding of science’ campaigns run by The Royal Society, Wellcome Trust and the science community on controversial issues such as stem cell research or genetic modification. There are, for instance, deliberative councils to embed the public in the complex ethical and political decisions, such as those run by Involve and Sciencewise.¹⁵⁶ The Citizen Council at the National Institute for Health and Care Excellence, uses lay audiences to help with its health guidance, so that their recommendations take into account social values and social equity.¹⁵⁷

Could we extend these public science programmes to experimentation in social policy?¹⁵⁸ Such a drive could build on the work of others. The work of US Coalition for Evidence Based Policy has attempted to educate Washington about the benefits of RCTs for improved government. In the UK, the Alliance for Useful Evidence is an open-network of 2,000 individuals in government, charities and academia who want to see a bigger demand for those in power to learn from evidence. A new public campaign by Sense about Science, Evidence Matters, will also champion the experimental cause. Mark Henderson's book *Geek Manifesto* that advocates RCTs in social policy has been sent to every single Westminster MP.¹⁵⁹ Ben Goldacre has championed experimental research in social policy across the media and government.¹⁶⁰ So there are institutions and individuals catalysing change. We need to keep building on this to encourage government to experiment.

Ethics by committee?

One approach to taking tough ethical decisions on experiments is to refer to a research ethics committee. These are commonly placed in universities and hospitals. A criticism of the Facebook experiment discussed above is that they worked with researchers at Cornell University who had not received assent from the research ethics committee. There are also other learned bodies producing ethical guidelines that occasionally censor unethical experiments, such as the Indian Council of Medical Research.

We could set up a similar body for government-related experiments outside of science. But this would still rely on the 'great and the good' and not engage with wider public resistance. A public campaign is still a necessary part of promoting experimentation. Regardless of where you think the government and states' legitimacy and social 'contract' with citizens lies, it is surely the right thing to publically make the honest case for the benefits – and potential pitfalls – of experimental government. The debate cannot be left to expert 'ethics' committees in academia or the civil service, but must spread out into wider public and political domains. The politics of experimentation is challenging as it means a degree of humility. Launching an experiment is an admission of ignorance.¹⁶¹ But we can't duck the need for making the political case for why experimentation is needed – even if it's hard to sell to voters in practice.

6. CONCLUSION AND RECOMMENDATIONS

The time is now for experimental government. The wicked societal challenges facing us means that ‘business as usual’ is unlikely to be good enough to meet 21st century challenges. According to an earlier call for action in the 1980s for more experimentation, it’s an approach that is realistic for those in power as it ‘permits adventurousness tempered with caution’.¹⁶² Rather than recklessly pursuing a new paper-based policy that ends in political trouble, there is a method to test it out.

FDR’s ‘bold experimentation’ showed a path of ambitious and fast-paced legislative work – sending 15 major legislative proposals to Congress in his first 100 days – mixed with an openness to try things out through robust experimental research. We should be inspired by this, and build on the long and impressive history of experimentation.

Recommendations:

1. Government officials and researchers should set up more **ambitious and bold experiments** on nationally important issues seen in the scale of FDR’s vision or the highly politicised experimentation in China (Box 1), not just experimentation on issues of implementation and delivery. We also need to replicate experiments – particularly from overseas – to see if they will work here (see page 6), and to coordinate local level experiments to ensure widespread benefits (see page 8).
2. Experiments need the **most robust available evaluation methods such as, if appropriate, experimental research designs** to share learning about what worked or failed, and move to the right of our continuum of experimentation (see page 11 on definition). We also need formative evaluations to learn about why and how things are working (see page 25), and use qualitative research and a mixture of methods to pry open the ‘black box’ (see page 25). Governments need to ensure their officials have the expertise to work with researchers using these methods. We support the training and skills development on research-use, currently being developed by the What Works Network, Civil Service Learning, BIT, and the Alliance for Useful Evidence. We also support the formation of a trial advice panel comprising government and external trialling experts to support the use of experimentation.
3. Leaders in government need to **embrace risk and reward success**, by taking responsibility for failure, praising staff successes, and having a positive narrative and more training on risk management (see page 22). Governments must recognise that some experiments fail, but – if they are properly designed – the knowledge acquired through failure can be valuable. One practical measure is to set up challenge prizes or awards to incentivise experimentation in government (see page 22).
4. Create the right **institutional support for** experimentation, such as a public strategic competitive fund (see page 24), more semi-detached skunkworks-type institutions with on-tap expertise on experimentation (see page 23) or bodies like the Ministry of Justice Data Lab (see page 23), and sunset-clauses for government policies to create the space to learn and test out new ideas (see page 31).

5. Researchers and government officials need to **work closely together to co-produce experiments**. Officials should seek out researchers with technical expertise in experimentation. Researchers need to improve their political skills (see page 26), co-producing with government officials and being opportunistic with political and policy cycles (see page 26).
6. Work with the grain of current policy environment to devise experiments. **Seize opportunities to experiment**, such as working with decentralised public services and front-line professionals (see page 26), or the staged roll out of services due to budget constraints (see page 26).
7. Policymakers at national or local level need to **start policies that are not set in stone**, as seen in the UK Employment Retention Advancement Demonstration (see page 16) and allow some flexibility and opportunities to ‘learn as you go’.
8. We need to grow the **public debate about the importance of experimentation** (see page 33), not just relying on expert, ethics committees or ‘working below the political radar’ (see page 33), and promote public engagement activities and replicate bodies like NICE’s Citizen Council (see page 33).

Potential areas of UK social policy for experimentation

Below are four specific areas where more experimentation is merited. They have been chosen as they are politically feasible. These are areas of social policy where minds are open, and innovation possible. Where ideology and political belief has not completely straight jacketed decision making. Some of these ideas are not new, but are about growing or replicating emerging ideas. Many more areas are possible, but this list gives a flavour of what might be possible. We welcome suggestions of other subjects that are ripe for experimentation.

1. **Labour market and jobs schemes.** For example, replicating the US experiment with ‘Jobs Plus’ to increase employment and income for residents of social housing. This uses a ‘saturation approach’ – working with all residents in targeted, low-income communities – underpinned by specialist advisers, community mentors and incentives to help people stay in work. Any experiment in the UK would need to be adapted to the UK context.
2. **Probation and community justice innovations.** Although Transforming Rehabilitation is politically contested, there is a broad range of less contentious areas, such as involving former probation service users in delivery of rehabilitation to others, schemes targeting prolific offenders such as shoplifters, and improving accessibility of public services for those undertaking probation. Experimenters should collaborate with the Probation Institute in delivering trials that are useful to probationers.
3. **Social media and computers in schools.** Significant numbers of schools are investing in either tablets or laptops for students, or allowing them to bring their own to school. Vast sums of money are being spent with very little evidence of the impact this may have. Also, we need experiments with teachers using social media and new tools to collaborate, share and indulge in reflection and professional development.
4. **Improving business-university links/knowledge transfer.** What methods are more effective to exploit knowledge created by universities and to help firms address their technological challenges? Do activities like Knowledge Transfer Partnerships, mentoring, face-to-face engagement deliver?

ENDNOTES

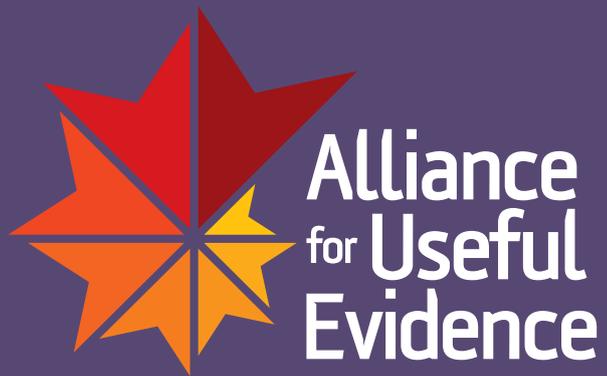
1. Manzi, J. (2012) 'Uncontrolled; The Surprising Payoff of Trial-and-Error for Business, Politics and Society.' London: Basic Books.
2. Bakhshi, H. and Mateos-Garcia, J. (2012) 'Rise of the Datavores: How UK businesses analyse and use online data.' London: Nesta. See: http://www.nesta.org.uk/sites/default/files/rise_of_the_datavores.pdf
3. This assertion is made in a variety of reports relating to Randomised Controlled Trials, such as 'Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials.' (Cabinet Office, Behavioural Insights Team, 2012). We do not, however, have an estimate of the quantity of experimentation in UK government. Internationally, there is a register of randomised evaluations from all across the world run by the American Economics Association: <https://www.socialscienceregistry.org/> But it does not include many past trials so it is hard to reliably estimate how many trials have been done in the UK.
4. Hartford, T. (2011) 'Adapt: Why Success Always Starts with Failure.' London: Abacus.
5. Crewe, I. and King, A. (2013) 'The Blunders of Our Governments.' London: OneWorld Publications.
6. Puttick, R. (2012) 'We don't need "what works". We need to know what is working.' London: Nesta. See: <http://www.nesta.org.uk/blog/we-dont-need-what-works-we-need-know-what-working#sthash.e3EVPoIJ.dpuf>
7. Kastle, T. and Potts, J. (2010) Public sector innovation research: What's next? 'Innovation: Management, Policy and Practice.' Vol. 12, number: 2, pp. 122-37.
8. Bakhshi, H., Freeman, A. and Potts, J. (2011) 'State of Uncertainty; Innovation policy through experimentation.' London:Nesta. See: <http://www.nesta.org.uk/publications/state-uncertainty>
9. Bakhshi, H., Freeman, A. and Potts, J. (2011) 'State of Uncertainty; Innovation policy through experimentation.' London:Nesta. See: <http://www.nesta.org.uk/publications/state-uncertainty>
10. Cabinet Office (2003) 'Trying it out; The Role of 'Pilots' in Policy-Making. Report of a Review of Government Pilots.' London: Cabinet Office.
11. Mulgan, G. (2014) 'Innovation in the Public Sector; How can public organisations better create, improve and adapt?' London: Nesta See: http://www.nesta.org.uk/sites/default/files/innovation_in_the_public_sector-_how_can_public_organisations_better_create_improve_and_adapt.pdf
12. Langley, B. (2014) 'A randomised control trial comparing the effects of procedural justice to experienced utility theories in airport security stops'. Dissertation for Master's Degree in Applied Criminology and Police Management, University of Cambridge.
13. Berk R.A., Boruch R.F., Chambers D.L., Rossi P.H. and White A.D. (1985) Social Policy Experimentation: A Position Paper. 'Evaluation Review.' Vol. 9, Number 4, pp. 387-429. See: <http://www.crim.cam.ac.uk/alumni/theses/Langley,%20B.%20A%20randomised%20control%20trial%20comparing%20the%20effects%20of%20procedural%20justice%20to%20experienced%20utility%20theories%20in%20airport%20security%20stops.pdf>
14. Baeck, P., Colligan, P. and Puttick, R. (2014) 'I-teams The teams and funds making innovation happen in governments around the world.' London: Nesta and Bloomberg Philanthropies.
15. The Big Lottery Fund provided £42 million for 'School Fruit Pilots'. The aim is to test out if there are benefits from a free piece of fruit each school day, given to children aged four to six: Big Lottery Fund Research Issue 27 (2007) '5 A Day Programme and School Fruit and Vegetable Scheme: findings from the evaluations.' London: The Big Lottery Fund. See: file:///C:/Users/jbreckon/Downloads/er_eval_5aday_findings.pdf
16. See for example the experiments on attitude and behaviour change in Manchester: Cotterill, S. and Richardson, L. (2010) Expanding the Use of Experiments on Civic Behavior: Experiments with Local Government as a Research Partner. 'Annals of the American Academy of Political and Social Science.' Vol. 628, pp. 148-164. However, Sarah Cotterill and Liz Richardson discuss the difficulties of getting support from local government. In their interviews, they heard that local government workers thought they were a luxury 'extra' which local resources should not be responsible for covering. In discussions, some practitioners wanted to benefit from experimental evidence, but did not feel they should be the ones to pay.
17. Greenberg, D. and Shroder, M. (1997) 'The Digest of Social Experiments.' Third Edition. London: Urban Institute Press.
18. Gatehouse, M. and Price, A. (2013) 'State of Innovation; Welsh Public Services and the Challenge of Change.' London: Nesta.
19. For example: Rawski, T.G. (1995) Implications of China's reform experience. 'The China Quarterly.' (144):1150-73; Roland, G. (2000) 'Transition and economics: politics, markets, and firms.' Cambridge: MIT; Cao, Y., Qian, Y. and Weingast, B. (1999) From federalism, Chinese style to privatization, Chinese style. 'Economics of Transition.' 1999;7(1):103-31.
20. Heilmann, S. (2008) Policy Experimentation in China's Economic Rise. 'Studies in Comparative International Development.' Vol. 43, Number 1, p. 26.
21. Ibid. p. 16.
22. Ibid. p. 6.
23. Ibid. p. 7.
24. Ibid. p. 8.
25. Mulgan, G. and Puttick, R. (2013) 'Making Evidence Useful: The case for new institutions.' London: Nesta.
26. National Audit Office (2013) 'Evaluation in Government.' UK: London: National Audit Office.
27. See for example: Chalmers, I. (2003) Trying to do more good than harm in policy and practice: the role of rigorous, transparent, up-to-date evaluations. 'Annals of the American Academy of Political and Social Science.' Vol. 589, pp. 22-40.
28. See for example this systematic review of evidence on the 'Scared Straight' and other juvenile awareness programmes for preventing juvenile delinquency. Cochrane Collaboration (2012), 'Scared Straight' and other juvenile awareness programmes for preventing juvenile delinquency. See: <http://www.cochrane.org/features/scared-straight-and-other-juvenile-awareness-programmes-preventing-juvenile-delinquency>
29. Brooks, G., Hall, J. and Torgerson, C.J. (2006) 'A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling.' London: Department for Education and Skills.
30. Henderson, M. (2013) 'The Geek Manifesto: Why science matters.' London: Corgi. p. 17.
31. Cabinet Office (2003) 'Trying it out; The Role of 'Pilots' in Policy-Making. Report of a Review of Government Pilots.' London: Cabinet Office.

32. Berk, R.A., Boruch, R.F., Chambers, D.L., Rossi, P.H. and White, A.D. (1985) Social Policy Experimentation: A Position Paper. 'Evaluation Review.' Vol. 9, Number 4, pp. 390.
33. Heilmann, S. (2008) Policy Experimentation in China's Economic Rise. 'Studies in Comparative International Development.' Vol. 43, Number 1, pp. 1-26, p. 3.
34. Stern, E. (forthcoming) 'Impact Evaluation; A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.' Prepared for Bond, Comic Relief, Big Lottery Fund and DFID. p. 4.
35. For an excellent overview of different approaches see Medical Research Council (2008) 'Developing and evaluating complex interventions: new guidance.' London: Medical Research Council. See: <http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/>
36. Mulgan, G. (2014) 'Innovation in the Public Sector; How can public organisations better create, improve and adapt?' London: Nesta. p. 11.
37. Sherman, L.W. (2003) Misleading Evidence and Evidence-Led Policy: Making Social Science More Experimental. 'Annals of the American Academy of Political and Social Science.' Vol. 589, pp. 6-19.
38. Farrington, B. (1951) 'Francis Bacon, philosopher of industrial science.' New York NY: Henry Schuman Inc., Musson, A.E. and Robinson, E. (1969) 'Science and technology in the Industrial Revolution.' Manchester: Manchester University Press. p. 16.
39. It is not just unique to Western government. Chinese Emperors talked about it openly - and Deng Xiaoping's commitment to facts, what works, and crossing the river by walking on the stones deliberately echoed that tradition. Mulgan, G. (2013) 'Experimental government.' London: Nesta. See: <http://www.nesta.org.uk/blog/experimental-government>
40. Dehue, T. (2001) Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design. 'American Journal of Psychology.' Vol. 114, Number 2, p. 288.
41. Dehue, T. (2001) Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design. 'American Journal of Psychology.' Vol. 114, Number 2.
42. Dehue, T. (2001) Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design. 'American Journal of Psychology', Vol. 114, Number 2, p. 288.
43. For other views on experimentation relating to economic policy see: Hayek, F.A. (1978) 'The constitution of liberty.' Chicago: University of Chicago Press; North, D.C. (1990) 'Institutions, institutional change and economic performance.' Cambridge: Cambridge University Press; Mukand, S.W. and Rodrik, D. (2005) In search of the Holy Grail: policy convergence, experimentation, and economic performance. 'American Economic Review.' 2005;95(1):374-83.
44. Heilmann, S. (2008) Policy Experimentation in China's Economic Rise. 'Studies in Comparative International Development.' Vol. 43, pp. 1-26, p. 3.
45. Heilmann, S. (2008) Policy Experimentation in China's Economic Rise. 'Studies in Comparative International Development.' Vol. 43, pp. 1-26, p. 3.
46. Ross, D. (1991) 'The Origins of American Social Science.' New York NY: Cambridge University Press. p. 455.
47. Boruch, R.F. and Riecken, H.W. (1978) Social Experiments. 'Annual Review of Sociology.' Vol. 4, pp. 511-532.
48. Gross, M. and Krohn, W. (2005) Society as experiments: sociological foundations for a self-experimenting society. 'History of the Human Sciences.' Vol. 18, No. 2.
49. Boruch, R.F. and Riecken, H.W. (1978) Social Experiments. 'Annual Review of Sociology.' Vol. 4, p.512.
50. Dehue, T. (2001) Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design. 'American Journal of Psychology.' Vol. 114, Number 2.
51. Institute for Research on Poverty 'Negative Income Tax.' See: <http://www.irp.wisc.edu/research/nit.htm>
52. Campbell, D. (1969) Reforms as Experiments.' 'American Psychologist.' Vol. 24(4), pp. 409-429.
53. Campbell, D.T. and Cook, T.D. (1979) 'Quasi-experimentation: Design and Analysis Issues for Field Settings.' Boston MA: Houghton Mifflin Harcourt.
54. Interestingly, Donald Campbell was also a strong advocate of qualitative research, not just experimental and quasi-experimental designs, and argued for a 'humanistic, validity-seeking, case study method.' (See Campbell, D. T., Foreword in Yin, R.K. (1994) 'Case Study Research: Design and Methods.' London: SAGE.)
55. Berk, R.A., Boruch R.F., Chambers D.L., Rossi P.H. and White A.D. (1985) Social Policy Experimentation: A Position Paper. 'Evaluation Review.' Vol. 9, Number 4, pp. 387-429, p. 389.
56. Oakley, 1998.
57. Berk, R.A., Boruch R.F., Chambers D.L., Rossi P.H. and White A.D. (1985) Social Policy Experimentation: A Position Paper. 'Evaluation Review.' Vol. 9, Number 4, pp. 387-429, p. 396.
58. Webb, B. and Webb, S.(1932) 'Methods of Social Study.' pp. 224-5.
59. John, P. (2013) Experimentation, Behaviour Change and Public Policy. 'The Political Quarterly.' Vol. 84, No 2, pp. 238-246.
60. Blundell, R., Green, H., Greenberg, D., Lissenburg, S., Morris, S.T., Mittra, B., Riccio, J., (2003) 'The United Kingdom Employment Retention and Advancement Demonstration Design Phase: An Evaluation Design.' GCSRO Occasional Papers Series No.1. London: Cabinet Office.
61. Hendra, R. et al., (2011) 'Breaking the low-pay, no-pay cycle: Final evidence from the UK Employment Retention and Advancement (ERA) demonstration.' London: Department for Work and Pensions.
62. Email correspondence with authors Jim Riccio, MDRC, 17 December 2014.
63. Agar, J. (2014) 'Learning from the past: Slough, MinTech and experimental government.' See: <http://www.alliance4usefulevidence.org/learning-from-the-past-slough-mintech-and-experimental-government/>
64. Ibid.
65. Ibid.
66. Mulgan, G. (2014) 'Rewiring the brain: a rough blueprint for reforming centres of government.' London: Nesta. See: http://www.nesta.org.uk/sites/default/files/reforming_centres_of_govts_-_geoff_mulgan.pdf
67. Mulgan, G. (2013) 'Experimental government.' London: Nesta. See: <http://www.nesta.org.uk/blog/experimental-government>
68. Levy, S. (2006) 'Progress Against Poverty Sustaining Mexico's Progresa-Oportunidades Program.' Washington, DC: Brookings Institution Press.
69. Levy, S. (2006) 'Progress Against Poverty Sustaining Mexico's Progresa-Oportunidades Program', Washington, DC: Brookings Institution Press.
70. Parker, S.W. and Teruel, G.M. (2005) Randomization and Social Program Evaluation: The Case of Progresa. 'Annals of the American Academy of Political and Social Science.' Vol. 599, pp. 199-219.
71. Stoker, G. (2010) Translating Experiments into Policy. 'Annals of the American Academy of Political and Social Science.' Vol. 628, pp. 47-58.
72. Stoker, G. (2010) Translating Experiments into Policy. 'Annals of the American Academy of Political and Social Science.' Vol. 628, p. 48.
73. This was from the Alliance for Useful Evidence party fringe events in Glasgow, 7 October 2014. See: <http://www.alliance4usefulevidence.org/event/liberal-democrat-2014-party-conference/>

74. Local Government Association (2013) 'Local Public Service Transformation; A Guide to Whole Place Community Budgets.' London: Local Government Association. See: <http://communitybudgets.org.uk/wp-content/uploads/2013/03/Guide-to-Whole-Place-Community-Budgets.pdf>
75. Martel Garcia, F. and Wantchekon, L. (2010) Theory, External Validity, and Experimental Inference: Some Conjectures. 'Annals of the American Academy of Political and Social Science.' Vol. 628, pp. 132-147.
76. Bond, R. et al., (2012) A 61-million-person experiment in social influence and political mobilization. 'Nature.' Vol. 489, No 7415, pp.295-298.
77. Cabinet Office Behavioural Insights Team (2012) 'Applying behavioural insights to reduce fraud, error and debt.' London: Cabinet Office. See: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60539/BIT_FraudErrorDebt_accessible.pdf
78. Cabinet Office Behavioural Insights Team et al., (2013) 'Applying Behavioural Insights to Organ Donation: preliminary results from a randomised controlled trial.' London: Cabinet Office. See: http://www.behaviouralinsights.co.uk/sites/default/files/Applying_Behavioural_Insights_to_Organ_Donation_report.pdf
79. Although RCTs are an important part of their research, they are also funding earlier stage projects: they are only doing RCTs where the scale and development level is appropriate to do so.
80. Nesta (2014) 'As part of our Digital Education programme, we're looking at opportunities presented by one-to-one remote tuition.' London: Nesta. See: <http://www.nesta.org.uk/project/digital-education/remoted-tutoring>
81. The Communication Trust (2014) 'Randomised Controlled Trials.' London: The Communication Trust. See: <https://www.thecomunicationtrust.org.uk/projects/talk-of-the-town/randomised-controlled-trial/>
82. Bloomberg Philanthropie and Nesta (2014) 'i-teams: The teams and funds making innovation happen in governments around the world.' London: Nesta. See: <http://www.nesta.org.uk/publications/i-teams-teams-and-funds-making-innovation-happen-governments-around-world>
83. Caseborne, J. (2014) 'Why Motivation Matters in Public Sector Innovation.' London: Nesta.
84. Ibid.
85. Potts J. and Kastle T. (2010) Public sector innovation research: What's next? 'Innovation: Management, Policy and Practice.' Vol. 12, Number 2, pp. 122-37.
86. Townsend, W. (2013) Innovation and the perception of risk in the public sector. 'International Journal of Organizational Innovation.' Vol. 5, Number 3, pp. 21-34.
87. Cabinet Office (2003) 'Trying it out; The Role of 'Pilots' in Policy-Making. Report of a Review of Government Pilots.' London: Cabinet Office. p. 7.
88. See for example Potts J. and Kastle T. (2010) Public sector innovation research: What's next? 'Innovation: Management, Policy and Practice.' Vol. 12, Number 2, pp.122-37; Queensland Public Service Commission (2009) 'Submission on Advancing Public Sector Innovation.' Australia: Queensland Government; Townsend, W. (2013) Innovation and the perception of risk in the public sector. 'International Journal of Organizational Innovation.' Vol. 5, number 3, pp. 21-34.
89. Jacob, B. and Ludwig, J. (2005) Can the Federal Government Improve Education Research? 'Brookings Papers on Education Policy.' Vol. 8, pp. 47-87.
90. Bloomberg Philanthropies and Nesta (2014) 'i-teams: The teams and funds making innovation happen in governments around the world.' London: Nesta. See: <http://www.nesta.org.uk/publications/i-teams-teams-and-funds-making-innovation-happen-governments-around-world>
91. Harford's book 'Adapt' sets out a powerful case for how governments need to evolve and adapt by testing out new ideas, whether it be counter-terrorism, climate change or dealing with financial crises. Harford, T. (2014) The random risks of randomised trials. 'The Financial Times.' 25 April 2014.
92. Stoker, G. (2010) Translating Experiments into Policy. 'Annals of the American Academy of Political and Social Science.' Vol. 628, p. 52.
93. Inspired by the successful aeronautical Lockheed Martin's Advanced Development Programs 'skunk works' programme in California that created such designs as the famous U-2 bomber, it's now common in engineering and design.
94. Bloomberg Philanthropie and Nesta (2014) 'i-teams: The teams and funds making innovation happen in governments around the world.' London: Nesta. See: <http://www.nesta.org.uk/publications/i-teams-teams-and-funds-making-innovation-happen-governments-around-world>
95. Ministry of Justice (2014) 'Accessing the Justice Data Lab Service.' London: Ministry of Justice. <https://www.gov.uk/government/publications/justice-data-lab>
96. Bloomberg Philanthropies and Nesta (2014) 'i-teams: The teams and funds making innovation happen in governments around the world.' London: Nesta. p. 18. See: <http://www.nesta.org.uk/publications/i-teams-teams-and-funds-making-innovation-happen-governments-around-world>
97. See: <http://coalition4evidence.org/wp-content/uploads/2014/02/Low-cost-RCT-competition-December-2013.pdf>
98. National Institute of Justice (2014) 'The National Institute of Justice Randomized Controlled Trial Challenge in Criminal Justice Agencies.' Washington DC: National Institute of Justice. See: <http://nij.gov/funding/Pages/rct-challenge.aspx>
99. Bakhshi, H. et al., (2011) 'State of Uncertainty; Innovation policy through experimentation.' London: NESTA. See: <http://www.nesta.org.uk/publications/state-uncertainty>
100. 'New randomised controlled trials will drive forward evidence-based research.' Press release. London: Department for Education. 3 May 2013. See: <https://www.gov.uk/government/news/new-randomised-controlled-trials-will-drive-forward-evidence-based-research>
101. Dutz, M.A., Kuznetsov, Y., Lasagabaster, E. and Pilat, D. (eds.) (2014) 'Making Innovation Policy Work: Learning from Experimentation.' Washington DC: OECD.
102. Bunt, L. and Leadbeater, C. (2012) 'The Art of Exit: In search of creative decommissioning.' London: Nesta.
103. Hartford, T. (2011) 'Adapt: Why Success Always Starts with Failure.' Little, Brown.
104. This was also a recommendation in Cabinet Office (2003) 'Trying it out; The Role of 'Pilots' in Policy-Making. Report of a Review of Government Pilots.' London: Cabinet Office. For a guide on how to do a systematic review see: Gough, D., Oliver, S. and Thomas, J. (2013) 'Learning from Research: Systematic reviews for informing policy decisions.' London: Alliance for Useful Evidence. See: <http://www.alliance4usefulevidence.org/assets/Alliance-final-report-08141.pdf>
105. Sanderson, I. (2009). Intelligent policy making for a complex world: Pragmatism, evidence and Learning. 'Political Studies.' Vol. 57, pp. 699-719.
106. Oxman, A. D. et al., (2009) Why we will remain pragmatists: Four problems with the impractical mechanistic framework and a better solution. 'Journal of Clinical Epidemiology.' Vol. 62, Number 5, pp. 485-488.
107. Stoker, G. and John, P. (2009) Design experiments: engaging policy makers in the search for evidence about what works. 'Political Studies.' Vol. 57, pp. 356-373.
108. Stoker, G. (2010) Translating Experiments into Policy. 'Annals of the American Academy of Political and Social Science.' Vol. 628, p.53.

109. Riccio, J.A. and Bloom, H.S. (2002) Extending the Reach of Randomized Social Experiments: New Directions in Evaluations of American Welfare-to-Work and Employment Initiatives. 'Journal of the Royal Statistical Society.' Vol.165, Number 1, pp.13-30; Ludwig, J., Kling, J.R. and Mullainathan S. (2011) Mechanism Experiments and Policy Evaluations. 'The Journal of Economic Perspectives.' Vol. 25 Number 3, pp.17-38; Card, D., DellaVigna, S. and Malmendier, U. (2011) The Role of Theory in Field Experiments. 'The Journal of Economic Perspectives.' Vol. 25, Number 3, pp. 39-62.
110. Oldsman, E. (2014) Making evaluations count: Toward more informed policy. In Dutz, M.A. et al., (eds.) (2014) 'Making Innovation Policy Work: Learning from Experimentation.' Washington DC: OECD. p. 245.
111. Eppel, E., Turner, D. and Wolf, A. (2011) 'Experimentation and Learning in Policy Implementation: Implications for Public Management.' Institute of Policy Studies Working Paper. Wellington NZ: Victoria University of Wellington. p. 12
112. Stoker, G. (2010) Translating Experiments into Policy. 'Annals of the American Academy of Political and Social Science.' Vol. 628, pp. 47-58.
113. Gueron, J.M. and Rolston, H. (2013) 'Fighting for Reliable Evidence.' New York NY: Russell Sage Foundation.
114. Ibid. p. 428.
115. Stoker, G. (2010) Translating Experiments into Policy. 'Annals of the American Academy of Political and Social Science.' Vol. 628, pp. 47-58.
116. See: <http://www.theguardian.com/politics/2002/apr/21/uk.medicalscience>
117. Adapted from the Tuskegee University Bioethics Centre website. See: http://www.tuskegee.edu/about_us/centers_of_excellence/bioethics_center.aspx
118. Facebook apologises for psychological experiments on users. 'The Guardian.' 2 July 2014. See: <http://www.theguardian.com/technology/2014/jul/02/facebook-apologises-psychological-experiments-on-users>
119. Taylor, K. (1987) Interpreting physician participation in randomized control clinical trials: the physician orientation profile. 'Journal of Health and Social Behaviour.' 1987b;28:389-400.
120. This criticism that experiments are too expensive if succinctly rebutted in this government-sponsored paper by Goldacre, B. (2013) 'Building Evidence into Education.' London: Department for Education. See: <https://www.gov.uk/government/news/building-evidence-into-education>
121. Email correspondence with Government Digital Service, 27 September 2014.
122. Quoted in Gueron, J.M. and Rolston, H. (2013) 'Fighting for Reliable Evidence.' New York NY: Russell Sage Foundation.
123. Rutter, J. (2012) 'Evidence and Evaluation in Policymaking; A problem of supply or demand?' London: Institute for Government. p. 10.
124. Hansard 09 Jan 2013: Column 315. 'Transforming Rehabilitation.'
125. Jowell, R. (2003) 'Trying It Out - The Role of 'Pilots' in Policy-Making.' London: Cabinet Office.
126. Ibid. p. 9.
127. Bakhshi, H. et al., (2013) 'Creative Credits: A Randomised Controlled Industrial Policy Experiment.' London: Nesta. See: http://www.nesta.org.uk/sites/default/files/creative_credits.pdf
128. Rutter, J. (2012) 'Evidence and Evaluation in Policymaking; A problem of supply or demand?' London: Institute for Government. p. 12.
129. We held fringe events on experimental government at the Labour, Conservative and Liberal Democrat fringe in September and October 2014.
130. Ministry of Justice (undated) 'Statistical Notice: Interim re-conviction figures for the Peterborough and Doncaster Payment by Results pilots.' London: Ministry of Justice. See: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206686/re-conviction-results.pdf
131. Cartwright, N. (2012) 'Evidence-Based Policy: A Practical Guide to Doing It Better.' New York NY: OUP.
132. Sutherland, A. (2013) 'Social science and replication.' Alliance for Useful Evidence: London. See: <http://www.alliance4usefulevidence.org/replication/>
133. The outcomes are extraordinary and wide ranging: from improved early language development, to reductions in children's injuries and abuse; from mums being much more likely to return to education or employment when their children are old enough, to reductions in smoking while pregnant and higher use of breastfeeding. It's not just the mums and their babies that benefit. Estimates of savings to taxpayers in the US are around US\$ 3-5 return for every dollar invested. For further information see: 'Family Nurse Partnerships.' Dartington: Social Research Unit. See: <http://dartington.org.uk/projects/family-nurse-partnership-2/>
134. The Scottish Government (2014) 'Evaluation of the Family Nurse Partnership Programme in NHS Lothian, Scotland.' See: <http://www.scotland.gov.uk/Publications/2014/02/6499/2>
135. Eley, S. et al., (2002) 'Evaluation of the Scottish Pilots Drug Treatment and Testing Orders.' Scotland: the Scottish Government. See: <http://www.scotland.gov.uk/Publications/2002/10/15537/11660>
136. Moehler, D.C. (2010) Democracy, Governance, and Randomized Development Assistance. 'Annals of the American Academy of Political and Social Science.' Vol 628, pp. 30-46
137. Jowell, R. (2003) 'Trying It Out - The Role of 'Pilots' in Policy-Making', London: Cabinet Office. p. 10.
138. Jowell, R. (2003) 'Trying It Out - The Role of 'Pilots' in Policy-Making.' London: Cabinet Office. p. 9.
139. Ibid. p. 10
140. Bakhshi, H., Freeman, A. and Potts, J. (2011) 'State of Uncertainty; Innovation policy through experimentation.' London: NESTA. See: <http://www.nesta.org.uk/publications/state-uncertainty>
141. Bakhshi, H. et al., (2011) 'State of Uncertainty; Innovation policy through experimentation.' London: NESTA. p.13 See: <http://www.nesta.org.uk/publications/state-uncertainty>
142. For further discussion on ethical fears relating to RCTs, see Puttick, R. (2011) 'Day 3 Debunking the myths about Randomised Control Trials (RCTs).' London: NESTA. See: http://www.nesta.org.uk/sites/default/files/ten_steps_to_transform_the_use_of_evidence_blogs.pdf

143. Edwards, S.J. et al., (1998) Ethical issues in the design and conduct of randomised controlled trials. 'Health Technology Assessment.' 1998 Dec;2(15):i-vi, 1-132.
144. Ibid.
145. Harford, T. (2014) The random risks of randomised trials. 'The Financial Times.' 25 April.
146. Ibid.
147. There are of course many other philosophical traditions relevant here such as those working on social contract theories, or the revival of interest in virtues. For an excellent and accessible review of the ethics of medical RCTs, including an overview of research on 'informed consent', see Ethical Issues in the design and conduct of randomised controlled trials, NHS R&D HTA Programme. 'Health Technology Assessment 1998.' Vol. 2; No. 15.
148. See for example, London, A.J. (2009) Clinical Equipoise: Foundational Requirement or Fundamental Error. In Steinbock, B. (Eds.) 'The Oxford Handbook of Bioethics.' Oxford: Oxford University Press.
149. Edwards, S. et al., (1998) Ethical issues in the design and conduct of randomised controlled trials. 'Health Technology Assessment.' 1998 Dec;2(15):i-vi, 1-132.
150. Glewwe, P., Park, A. and Zhao, M. (2012) 'Visualizing Development: Eyeglasses and Academic Performance in Rural Primary Schools in China.' Center for International Food and Agricultural Policy Working Paper: WP12-2. See: <http://ageconsearch.umn.edu/bitstream/120032/2/WP12-2.pdf>
151. Friedman, J. (2014) 'Should impact evaluation be justified by clinical equipoise or policy equipoise?' Blog: World Bank Group. See: <http://blogs.worldbank.org/impacetevaluations/africacan/should-impact-evaluation-be-justified-clinical-equipoise-or-policy-equipoise>
152. Friedman, J. (2014) 'Should impact evaluation be justified by clinical equipoise or policy equipoise?' Blog: World Bank Group 04/17/2014. See: <http://blogs.worldbank.org/impacetevaluations/africacan/should-impact-evaluation-be-justified-clinical-equipoise-or-policy-equipoise>
153. King G., et al., (2007) A "Politically Robust" Experimental Design for Public Policy Evaluation, with Application to the Mexican Universal Health Insurance Program. 'Journal of Policy Analysis and Management.' Vol. 26, Number 3, pp. 487.
154. John, P. (2012) 'Experimentation, Behaviour Change and Public Policy.' Inaugural lecture to the Department of Political Science, University College London, 20 Nov 2012.
155. Sherman, L.W. (2003) Misleading Evidence and Evidence-Led Policy: Making Social Science More Experimental. 'Annals of the American Academy of Political and Social Science.' Vol. 589, pp. 6-19.
156. See: <http://www.involve.org.uk/>
157. Breckon, J. and Ruiz, F. (2014) 'The Nice Way; Lessons for Social Policy and Practice from the National Institute for Health and Care Excellence.' London: Nesta. See: http://www.alliance4usefulevidence.org/assets/NestaAlliance_and_NICE_paper_v5.pdf
158. For one well known attempt to show the ethical case for more experimental research see Chalmers, I. (2003) Trying to Do More Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-Date Evaluations. 'Annals of the American Academy of Political and Social Science.' Vol. 589, Number 1, pp.22-40.
159. Henderson, M. (2013) 'The Geek Manifesto: Why science matters.' London: Corgi.
160. For example, Goldacre, B. (2013) 'Building Evidence into Education.' London: Department for Education. See: <https://www.gov.uk/government/news/building-evidence-into-education>
161. Berk, R.A., Boruch R.F., Chambers, D.L., Rossi, P.H. and White, A.D. (1985) Social Policy Experimentation: A Position Paper. 'Evaluation Review.' Vol. 9, Number 4, p.396.
162. Ibid.



Alliance for Useful Evidence

1 Plough Place
London EC4A 1DE

www.alliance4usefulevidence.org

March 2015

