

**BUDGET
PERSPECTIVES
2019
PAPER 2
July 2018**

USING BEHAVIOURAL EXPERIMENTS TO PRE-TEST POLICY

PETE LUNN AND DEIRDRE ROBERTSON



USING BEHAVIOURAL EXPERIMENTS TO PRE-TEST POLICY

Pete Lunn

Deirdre Robertson

July 2018

BUDGET PERSPECTIVES 2019

PAPER 2

Available to download from www.esri.ie

DOI: <https://doi.org/10.26504/bp201902>

© 2018 The Economic and Social Research Institute
Whitaker Square, Sir John Rogerson's Quay, Dublin 2

ABOUT THE ESRI

The mission of the Economic and Social Research Institute is to advance evidence-based policymaking that supports economic sustainability and social progress in Ireland. ESRI researchers apply the highest standards of academic excellence to challenges facing policymakers, focusing on 12 areas of critical importance to 21st-century Ireland.

The Institute was founded in 1960 by a group of senior civil servants led by Dr T.K. Whitaker, who identified the need for independent and in-depth research analysis to provide a robust evidence base for policymaking in Ireland.

Since then, the Institute has remained committed to independent research and its work is free of any expressed ideology or political position. The Institute publishes all research reaching the appropriate academic standard, irrespective of its findings or who funds the research.

The quality of its research output is guaranteed by a rigorous peer review process. ESRI researchers are experts in their fields and are committed to producing work that meets the highest academic standards and practices.

The work of the Institute is disseminated widely in books, journal articles and reports. ESRI publications are available to download, free of charge, from its website. Additionally, ESRI staff communicate research findings at regular conferences and seminars.

The ESRI is a company limited by guarantee, answerable to its members and governed by a Council, comprising 14 members who represent a cross-section of ESRI members from academia, civil services, state agencies, businesses and civil society. The Institute receives an annual grant-in-aid from the Department of Public Expenditure and Reform to support the scientific and public interest elements of the Institute's activities; the grant accounted for an average of 30 per cent of the Institute's income over the lifetime of the last Research Strategy. The remaining funding comes from research programmes supported by government departments and agencies, public bodies and competitive research programmes.

Further information is available at www.esri.ie

THE AUTHORS

Pete Lunn is with the Economic and Social Research Institute (ESRI) and the Department of Economics, Trinity College Dublin. Deirdre Robertson is with the ESRI and the Department of Psychology, Trinity College Dublin.

ACKNOWLEDGEMENTS

Thank you to the two anonymous reviewers and Professor Emer Smyth who reviewed this paper.

This paper has been accepted for publication by the Institute, which does not itself take institutional policy positions. The paper has been peer reviewed prior to publication. The authors are solely responsible for the content and the views expressed.

TABLE OF CONTENTS

ABSTRACT.....	1
1. INTRODUCTION.....	1
2. SOME ECONOMIC CONSIDERATIONS	3
2.1 Proof of concept.....	4
2.2 Generalisability and scaling up.....	4
2.3 Opportunity cost	7
2.4 Spillovers	8
3. INTERNATIONAL EXAMPLES OF EXPERIMENTAL PRE-TESTS	8
3.1 Overview of international applications of behavioural science	8
3.1 Pre-testing using a field trial.....	10
3.2 Pre-testing in the laboratory	12
3.3 Pre-tests with multiple methods.....	12
4. PROGRESS IN IRELAND	14
4.1 Revenue.....	14
4.2 The ESRI’s Behavioural Research Unit.....	15
4.3 Other examples in Ireland	18
5. EVOLVING METHODS: PROCESS TRACING	19
5.1 Pictorial warning labels on tobacco products in the US.....	20
5.2 Country-of-origin labelling of meat in the EU	20
5.3 Consumer protection in Colombia’s communications market	21
5.4 Calorie posting in Ireland	21
6. CONCLUSIONS	22
REFERENCES	25

LIST OF FIGURES

Figure 1	Methods used in 159 international applications of ‘behavioural insights’	9
Figure 2	Where behavioural insights are being used in the policy cycle internationally.....	10
Figure 3	Meta-analysis of 20 RCTs conducted by Revenue.....	15
Figure 4	Pre-test of EAB intervention for residential electricity packages	17

USING BEHAVIOURAL EXPERIMENTS TO PRE-TEST POLICY

Experience never errs; it is only your judgments that err by promising themselves effects such as are not caused by your experiments.

(From the notebooks of Leonardo da Vinci)

ABSTRACT

Good policy development requires the conviction and courage to know when to push through and when to admit uncertainty. This paper argues that policy can be improved when uncertainty is admitted and paired with rigorous scientific methodology. We use international and Irish examples to show how experimental behavioural research is a powerful but underused tool that policymakers can use to reduce uncertainty and add a scientific foundation to the policymaking process.

1. INTRODUCTION

History does not relate whether Leonardo was referring in the above quote to experiments with only physical objects. Given his intricate studies of anatomy, his experimental subject matter may well have included the human body. But it is unlikely that he had behavioural experiments in mind. Leonardo died 360 years before the establishment of the first experimental psychology laboratory, which is usually credited to Wilhelm Wundt in 1879. Nevertheless, whether we consider physical, biological or behavioural science, what unites experimentalists is embodied in the quote. To see the need for experimentation, you have to understand something of what you don't know. You have to live with uncertainty about your own judgement; you have to doubt, and to doubt openly.

This observation is one reason why the recent breakthrough of behavioural science into policymaking is remarkable. Doubt can be difficult territory for many policymakers, by which we mean not only government ministers and parliaments, but senior civil servants and officials in state agencies, all of whom help to decide and implement policy. Openly admitting to uncertainty about whether a policy intervention is good or bad can, if not carefully phrased, seem to signal lack of expertise or competence. Changes in policy often have to be 'driven through', requiring persuasion, support, and the conversion of practitioners to the cause. Openly stating that one does not know the best policy is an admission of weakness, albeit one that wiser heads know to be almost always true of everyone. People like strong leaders; we like certainty. So the increased application to public policy of behavioural science, which relies strongly on experimentation, is remarkable.

Behavioural scientists themselves did not see it coming. Just over 10 years ago a group of prominent psychologists and behavioural economists lamented the lack of influence of behavioural science on policy, despite its clear relevance for diagnosing policy problems and understanding citizens' responses to policy interventions (Amir et al., 2005). But since the publication of *Nudge* (Thaler and Sunstein, 2008) and the establishment by the UK government in 2010 of the Behavioural Insights Team (BIT), the use of behavioural science in the development of public policy has spread rapidly. Such has been the success of behavioural science that Executive Order 13707, 'Using Behavioural Science Insights to Better Serve the American People', was signed by President Obama in September 2015, directing the Federal Government to develop its policies and programmes using empirical findings from behavioural science research. The spread of behavioural science to public policy has largely occurred not through the application of scientifically grounded theories of behaviour to policy problems, although this has happened, but through the integration of experimental evidence into the policy development process (Sunstein, 2011; Lunn, 2014). This is telling, because it only makes sense to conduct experiments if the outcome is uncertain. Experiments are a reasoned response to doubt.

This is not to say that the application of behavioural science to policy always involves such open investigation. The Joint Research Centre at the European Commission has published a useful classification of behavioural policy initiatives (Sousa Lourenço et al., 2016), which groups them into three categories. 'Behaviourally aligned' policies are those that were not developed with any input from behavioural science but turned out nevertheless to be aligned with behavioural evidence. An example would be where a regulator has put time and effort into simplifying a compliance process, in the belief that increased convenience may have a disproportionate effect on compliance, despite penalties for non-compliance. In general terms, behavioural research in multiple domains shows that simplification and convenience can have such disproportionate effects (Sunstein, 2013). Hence the policy, although not *based* on behavioural research findings, is nevertheless *aligned* with them. 'Behaviourally informed' policies are those that are designed at least in part on the basis of behavioural evidence.

For instance, the 2014 EU Consumer Rights Directive banned the use of pre-ticked boxes that default online consumers into purchasing additional products (e.g. insurance, deluxe features, gift wrapping) unless they untick the box. This policy was directly informed by evidence that default settings have a powerful influence on choices in multiple domains, much of which was experimental evidence (e.g. McKenzie et al., 2006). However, no experimental study was undertaken to test the likely impact of the ban before it was introduced. 'Behaviourally tested' policies constitute the final category of behavioural policy initiative and refer to instances where an explicit behavioural test of the policy itself has been undertaken. Policies can be experimentally tested after implementation as part of a process of

evaluation, or they can be experimentally pre-tested prior to being rolled out. This last category of experimental pre-tests is the focus of this paper.

At one level the argument for pre-testing is straightforward and obvious. The history of public policy development is littered with expensive mistakes, which for understandable reasons often receive more public attention than expensive success stories. In principle, pre-testing has the capacity to reduce the likelihood of expensive mistakes, thereby contributing to the efficient use of public spending and resources. Perhaps it makes sense always to ask if measures to be introduced in government budgets, and at other times, can potentially be pre-tested. If the policymakers responsible for a proposed initiative are open to the possibility that it doesn't work, and if it can be relatively cheaply and quickly pre-tested in an experiment, it would seem to make sense to do so. Why not announce the intention to pre-test the potential intervention, with a view to funding it fully if the test proves successful, rather than announce the funding and cross the fingers?

How good is this argument? What is the scope for using experiments to pre-test policy interventions? This is the main question addressed here. The paper sets the scene by considering the different sorts of policies that might be pre-tested. Then, we document the use of experimental pre-tests internationally, in the context of the rapidly expanding application of behavioural science to policy. We consider specific and, we hope, instructive examples where either laboratory or field experiments, including randomised controlled trials (RCTs),¹ have been deployed to pre-test policy interventions. This is not intended to be an exhaustive review, which would be beyond the scope of the present paper. The aim is instead to highlight what is possible and give insight into how multiple methods can be used to undertake experimental pre-tests. We then consider how the growing application of behavioural science to policy is leading to an increase in pre-testing also in Ireland. We highlight some recent advances in the study of decision-making processes that may also be of use to policymakers. The final section pulls together the material and, based on experience thus far, looks to draw some conclusions regarding the potential for experimental pre-testing and relevant lessons concerning how best to conduct pre-tests.

2. SOME ECONOMIC CONSIDERATIONS

Before describing specific examples, we outline a framework within which studies to date might be considered in the context of government budgets and administration. It makes sense to assume that resources available for experimentation are finite. Although behavioural approaches to policy are

¹ An RCT works by randomly allocating participants either to a group receiving the intervention under investigation (the 'treatment' group) or to group that receives no intervention (the 'control' group). The logic of an RCT is that randomisation minimises selection bias, allowing researchers to determine the effects of the intervention compared with no intervention while all other factors are held constant.

increasingly taught in universities and as part of continued professional development, there is a limited supply of trained researchers and technical staff. Ideally, therefore, these resources would be directed where they can be most effective from a cost–benefit perspective. This requires consideration of issues beyond the success or otherwise of a specific pre-test.

2.1 Proof of concept

Many of the early behavioural interventions undertaken by BIT involved experimental trials of interventions designed to improve the efficiency of service delivery or to increase rates of regulatory compliance. In some cases, the return to the UK Exchequer from testing simple and cheap changes to communications, such as deploying text reminders or behaviourally informed messages in letters, was large in comparison to the associated costs. For instance, the scaling up of experimentally tested changes to communications regarding tax was estimated to have increased revenue in one tax year by over £200 million. In Ireland, Revenue has undertaken similar trials of behavioural interventions, which are discussed further in Section 4.1. Where pre-testing of interventions improves the efficiency of the system that raises taxes, the benefit for the public finances is directly measurable and, it seems, quite substantial.

Most of the early interventions trialled by BIT took place at the policy coalface, where local public bodies communicate directly with citizens. Among many others, these included testing alternative text messages to increase payment of court fines, testing messages given to hospital outpatients to reduce missed appointments, testing webpages to increase sign-up to the organ donor register, and testing friendlier communications to increase the numbers of black and ethnic minority candidates to the police force. In these and other cases, researchers recorded statistically significant improvements in policy outcomes. Such initiatives have the potential to generate incremental or even substantial improvements in targeted policy outcomes, through relatively cheap and simple interventions.

BIT has published its own manual for how to go about applying behavioural science in this way (Haynes et al., 2012), advocating the systematic and iterative use of RCTs. Where an RCT in a local area or sector suggests that an intervention works well and there are good reasons to believe that the result will generalise beyond the specific context of the trial, it can be rolled out on a wider scale, in the expectation that the effect can be scaled up.

2.2 Generalisability and scaling up

A first issue that needs to be considered is whether the measured effect can indeed simply be scaled up. The strength of the experimental approach lies in its ability to separate out one hypothesised effect from other factors that have been controlled for. While this can demonstrate whether an effect is present and often identify the mechanism behind it, there is no guarantee that it will operate in the same way

when scaled up to policy level where other factors are present. Banerjee et al. (2017) identify six challenges that can impact on the success of scaling up an intervention from proof of concept to policy. These are: (1) market equilibrium effects; (2) spillover effects; (3) political reactions; (4) context dependence; (5) randomisation or site-selection bias; and (6) piloting bias/implementation challenges.

The first two of these involve the potential for interactions between individuals targeted by the intervention and those left alone. A straightforward example of a market equilibrium effect is a job placement assistant programme for individuals that increased the likelihood of obtaining employment for those in the treatment group but reduced it for those in the control group. As there were a limited number of vacancies, the intervention changed only which individuals were in employment, not the total number (Crepon et al., 2013). Spillover effects refer to more direct interaction via contamination. For example, Duflo and Saez (2003) recorded higher retirement savings among individuals who were exposed to an intervention but equally large effects among those who were not, but who worked in close proximity to those who were. In this case the spillover was positive; it can be negative. The third challenge is the simpler observation that successful pre-tests do not ensure lack of resistance from interested parties when scaled up, which can alter the format of the intervention to an extent that also alters the intended effect.

The other three challenges surround the validity of making the inference that whatever effect has been observed in the experimental study will operate as strongly when the policy is widespread. This challenge is perhaps more obvious when looking at the results of a laboratory study, but it applies to field studies including RCTs too. While many argue that RCTs are the best method for policy evaluation (Haynes et al., 2012; van Bavel et al., 2013), there are important methodological arguments regarding when RCTs can and cannot be generalised beyond the particular context in which they take place (Deaton, 2010; Cartwright and Hardie, 2012) and when RCTs of specific policy interventions are and are not the best method for providing evidence for policy (Ludwig et al., 2011; Lunn and Ní Choisdealbha, 2018). The final three challenges listed capture most of these arguments. Even a well-designed and successful RCT demonstrates only that an intervention works in the specific geographic location of the trial, with the trial sample, at the time the trial took place, when the policy is implemented by the people who implemented it in the trial. There may be good reasons to suppose it will work beyond this context, but there is no guarantee that the results will extend to different settings when scaled up. The effect could be context-dependent.

For instance, if experiments are tested on a homogeneous group of individuals then the same effect may not work for a more heterogeneous population. Sometimes the location of trials is not random, or randomisation into treatment and control groups is imperfect, biasing results. Lastly, scale-up can be affected not

only by the intervention being tested but by challenges in the implementation. This is a particular problem if the trial was conducted by people keen on the policy; implementation may be less successful when demanded of officials who are less keen.

Generally, however, the effect can go either way, as two examples from development studies illustrate. Trials in Kenya found that a primary school intervention was successful at reducing class sizes when implemented by parent–teacher associations, but unsuccessful when implemented by the government (Bold et al., 2015). By contrast, a programme in Indonesia that distributed identity cards to allow collection of rice subsidies managed to reach only 30% of targeted participants when trialled but got close to 100% when the programme was scaled up, meaning that it turned out to be more successful than the trial had implied (Banerjee et al., 2018).

Consideration of such challenges is important when scaling up from pre-tests to policy, but they are not insurmountable and, arguably, tackling them makes the experimental process stronger. Market equilibrium effects can be factored into either the design or the analysis stage depending on the topic. Spillover effects can be explicitly measured by varying the level of exposure of different groups. Political reactions are sometimes unforeseen, but working alongside policymakers and consulting stakeholders while designing a pre-test helps to ensure that views are taken into account, that what is being tested is feasible, and that there is widespread agreement that the study constitutes a fair test.

The size of context dependence effects varies with the topic at hand, but implies benefits to pre-testing policy in the culture in which it is going to be implemented, or conducting tests in more than one location. Site-selection biases and implementation challenges can also be factored in to pre-tests. There are experimental techniques that can account for individual differences in the population, meaning that, with larger sample sizes, moderators of the effect can be accounted for. This underlines the importance of carrying out tests on samples of the general population rather than student samples. Finally, implementation challenges illustrate again the importance of the policymaker–researcher relationship at all stages of experimental design. Pre-tests are designed to specifically inform one policy and, as such, implementation difficulties should be taken into consideration in the design stages. The ongoing relationship and conversation between researchers and policymakers during the scale-up can also help to attenuate unintended effects caused by changes during implementation.

The aim of a pre-test is to scale up the relevant intervention, and thus the above factors are important to consider from the earliest stages of the design. They also highlight the benefits, where possible, of engaging in an iterative process of experimentation when pre-testing policy, with stages of testing and re-testing to fine-tune and bolster the effects of a policy before attempting to scale it up. Where

multiple methods in the behavioural toolbox can be applied to the main research questions, pre-testing is likely to be stronger and more persuasive.

2.3 Opportunity cost

Notwithstanding issues associated with scaling policies up, the successes of BIT and others have stimulated international interest in the application of behavioural science to policy. Early pre-tests were largely motivated by the desire to provide ‘proof of concept’, to show that the application of behavioural science and the experimental method directly to policy problems can work – indeed often does. Armed with this knowledge, it is not unreasonable to ask whether the policy problems and associated interventions being selected for pre-testing are the right ones to target, given limited resources of expertise. Typically, success or failure is evaluated by comparing the cost of the intervention to the benefit of the policy outcome, first measured in a pre-test at a local level, then estimated for the scaled-up policy. Before considering more examples, however, there are at least two other economic considerations that one might want to take into account.

The first and most straightforward one is opportunity cost. If research resources are being directed to one specific policy problem, then those same resources are not being directed to another. As described above, most early behavioural pre-tests of policy interventions have taken place at the administrative coalface and have involved the experimental manipulation of communications to citizens. The goal is to improve the efficiency of the communications and the cost of failure is not high. There are, potentially, many other applications of behavioural science to policy decisions where the cost of failure might be very high and so the argument for pre-testing should be proportionally stronger. For instance, in areas such as financial regulation, employment law, environmental regulation and public health, policies that are intended to produce changes in behaviour, or perhaps to constrain certain behaviours, are frequently manifested in primary legislation, statutory instruments or national regulations. In such cases, the cost of an intervention not working may be great.

Rules designed to have beneficial effects on behaviour routinely impose substantial economy-wide costs on businesses, such as requirements to disclose product information, to comply with certain human resources practices, or to undertake data protection measures. In addition to the costs imposed on the object of the regulation, the time and effort of the public servants and organisations involved in the development of the legislation and rules can amount to a hefty public cost, as can the time and effort taken to reform or refine an ineffective policy subsequently. These considerations of cost are important when considering the scope for applying pre-tests and making good use of the limited resources available for designing and undertaking experiments to inform policy. While experimentally pre-testing routine administrative communications is beneficial, the greater prize may be to be deploy the same method and resources

to pre-test more central and far-reaching government decisions, to try to avoid costly failures. Of course, in an ideal world, a pre-test will confirm that the policy in question produces the desired behavioural effect sufficiently to justify the cost.

2.4 Spillovers

The second consideration, over and above the direct costs and benefits of a specific intervention, is more subtle. The power of the experimental method derives from its ability to identify causal effects. Where outcomes are compared under two conditions that differ by a single factor, we can be confident that any difference observed is due to that single factor. This means that experiments are a good way to identify reliable behavioural mechanisms.

Most behaviours that policymakers might seek to influence involve explicit decisions: whether to take on more debt, which mode of transport to take, whether to drink alcohol knowing that you need to drive home, whether to take exercise, and so on. Such decisions have common individual factors, such as tolerance for risk and uncertainty, or preferences for outcomes now versus outcomes in the future. They also vary in the extent to which they share contextual cues, such as to what extent many others visibly engage in the same behaviour, or exposure to official advice. Because decisions and contexts have these commonalities, results from one domain of behaviour can be instructive for results in another. Consequently, the value of an applied experiment often extends beyond the immediate policy context, spilling over into other domains and potentially informing policy elsewhere. The influence over decisions of default options offers an instructive example. Experiments designed to test the influence of default options on pension participation (Madrian and Shea, 2001) and organ donation (Johnson and Goldstein, 2003) have had effects on multiple other policy areas (Sunstein, 2013).

Overall, therefore, it is important when considering the costs and benefits of pre-testing a potential policy not to consider costs and benefits too narrowly. Efficient use of pre-tests does not depend only on measuring the cost of implementation of a given policy against an experimental measure of the effect size it produces. Consideration needs to be given to whether larger decisions involving greater potential costs might be tested instead, as well as to the potential benefits of an experiment for other policy domains.

3. INTERNATIONAL EXAMPLES OF EXPERIMENTAL PRE-TESTS

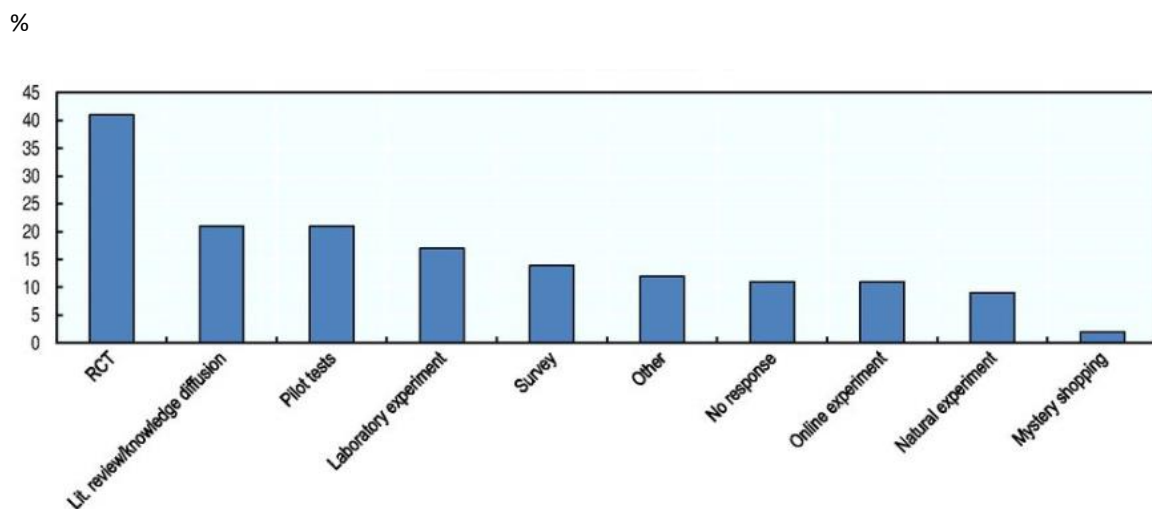
3.1 Overview of international applications of behavioural science

As the application of behavioural science to policy has spread internationally, it has become increasingly difficult to document in a comprehensive fashion. Perhaps the most complete analysis is contained in a recent report from the Organisation for

Economic Co-operation and Development (OECD, 2017a) on the use of ‘behavioural insights’, which was produced through direct contact with relevant governmental and regulatory bodies. The report presents findings from a survey of over 23 countries, with 60 institutions involved in 159 case studies of applying behavioural science to policy decisions. The majority of the institutions were central governmental departments, or regulatory and tax authorities. In the 129 case studies for which detailed information was available, most were in the policy area of consumer financial regulation, although a large number of other policy areas were represented including health and safety, labour markets, energy, public service delivery, environment, tax, education, telecommunications and consumer policy more broadly.

The OECD’s report contains data on the scientific methods typically employed, a breakdown of which is provided in Figure 1. Across the case studies, the most commonly deployed method was an RCT, which made up 25% of the represented methods, followed by literature reviews (or similar forms of knowledge diffusion), pilot tests and laboratory experiments. However, an important point made by the OECD bears repeating, which is that there is no methodological ‘one size fits all’ – the method chosen for behavioural studies should match the particular policy problem at hand (OECD, 2017a). The need to be more careful in matching the research question to the method also underpins a recent analysis that questions whether laboratory experiments are being under-used relative to RCTs and other field trials (Lunn and Ní Choisdealbha, 2018). The issue depends on the trade-off between the downside of studying behaviour in an artificial setting and the upside of the increased experimental control, flexibility and replicability that accompany laboratory investigation. Below we give examples of deploying both kinds of methods to pre-test policy.

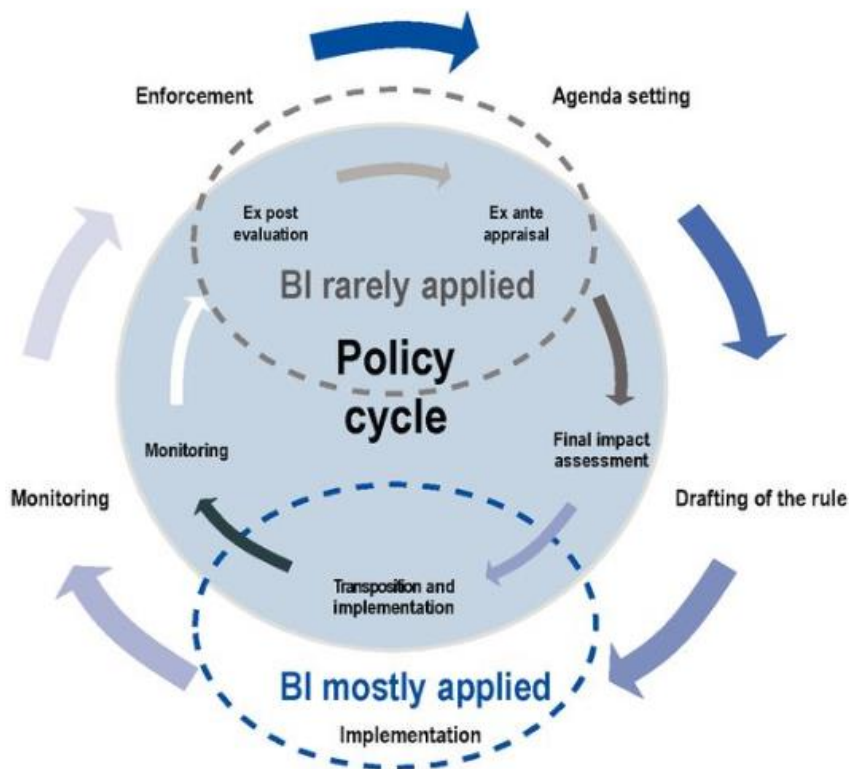
FIGURE 1 METHODS USED IN 159 INTERNATIONAL APPLICATIONS OF ‘BEHAVIOURAL INSIGHTS’ AS RECORDED IN A 2017 OECD SURVEY



Source: OECD (2017a).

The OECD also gathered data on when behavioural insights are being used in the policy cycle. Using three sequential stages of policy decision-making – research/diagnosis, design of decisions/interventions, implementations – the survey found that behavioural insights seem to be used primarily at the third rather than at the first or second stages of policymaking. While this is perhaps understandable given the success of the approach in tailoring communication to improve the implementation of policies, it does show that the use of behavioural science to pre-test the design of policy accounts for a fairly small minority of instances. As the OECD notes, there is the potential to use behavioural science both at the beginning of the policy cycle to design policy and at the end to monitor and adapt it (Figure 2).

FIGURE 2 WHERE BEHAVIOURAL INSIGHTS ARE BEING USED IN THE POLICY CYCLE INTERNATIONALLY



Source: OECD (2017a).

Despite the fact that they represent a minority of applications, pre-tests of policy via behavioural experiments are becoming more common. We turn now to illustrative examples.

3.1 Pre-testing using a field trial

The roll-out of energy smart meters in the UK provides an example of how a field

trial can provide a pre-test of policy. Smart meters track real-time energy usage and automatically send readings to energy suppliers. The UK government committed to rolling them out across the country as standard by the end of 2020. Smart meters offer the promise of reducing energy usage, giving consumers control over their usage levels, and supporting time-of-use tariffs that can help to spread the demand for electricity more evenly throughout the day. In this context, the Energy Demand Research Project was set up in 2006 as a means of understanding how consumers react to information about their energy consumption. Although initially the trials were not specifically designed to inform the smart meter policy, which was announced later, there was a significant focus on smart meters in the trials that were conducted and thus the findings have been communicated as a pre-test of the smart meter policy (AECOM, 2011; OECD, 2017a).

Following a call for tenders, four energy suppliers tested a series of interventions either individually or in combination. These included: energy efficiency advice, providing historic energy consumption information, benchmarking the household's consumption against comparable households, engaging customers using targets for reduced consumption, smart electricity and gas meters, real-time display devices showing energy use, control of heating and water with a real-time display, and financial incentives to reduce or shift consumption away from peak periods. The trials found that the most effective interventions combined smart meters with the installation of real-time information displays. All but two of the interventions that did not use smart meters showed no demonstrable reduction in energy usage. The two that showed a small effect (energy savings of approximately 1 per cent) were real-time displays and benchmarking against comparable households. Interventions using smart meters showed marked reductions in energy usage. The finding of particular importance for the smart metering policy is that coupling smart meter interventions with real-time displays led to energy savings that were 2–4 per cent greater than with smart meters alone (AECOM, 2011; OECD, 2017).

An important aspect of this pre-test was that it provided evidence about what did not work as well as evidence about what did. With reference to the argument of the previous section in relation to opportunity costs and avoiding costly mistakes, pre-tests like this are a way to avoid interventions that are well motivated and appear to be sensible, but impose widespread costs on businesses and turn out to be ineffective in altering outcomes. Some related, although somewhat different, findings have arisen from field trials of smart meters in Ireland, which tested more specific effects of feedback on energy consumption (Carroll et al., 2014).

Such studies are good examples of the potential benefits of pre-testing and the matching of research questions to methods. The roll-out of smart meters is a policy that will ultimately affect all households. Residential energy usage is a major contributor to climate change, and the extent of behaviour change associated with

the installation of smart meters is hence an important policy outcome. In this case, where the behaviour of interest represents the culmination of multiple decisions taken on a daily basis within the household, the use of field trials to investigate the impact is appropriate. However, other aspects of the roll-out of smart meters, such as the tariffs households choose, may be more suited to pre-testing by other means (see Section 4.2).

3.2 Pre-testing in the laboratory

The Joint Research Centre at the European Commission undertook laboratory pre-tests of regulatory measures designed to protect online gamblers. Behavioural evidence suggests that gamblers are often prey to time-inconsistent decision-making, whereby they set an initial limit on the amount of money they are willing to gamble in a session, but increase that amount in response to encountering losses. The study tested a series of potential regulatory interventions designed to counteract this tendency, comparing warnings and other messages delivered prior to a gambling session with those delivered within a session. The study was conducted both in a laboratory setting and online. Participants engaged in online gambling tasks for real money, using virtual roulette wheels and slot machines. The results clearly showed that interventions delivered prior to a session were far less effective than those delivered during a session, especially where these were combined with self-commitment strategies to stick to an expenditure limit.

As well as illustrating how a well-designed laboratory study can be used to pre-test policy, this study is a useful example of the benefits of pre-testing for a number of reasons. Firstly, it again showed the ineffectiveness of some potential interventions that policymakers might reasonably have expected to work. By pre-testing many alternative messages and warnings and finding most of them to be ineffective, good evidence was supplied to avert potentially costly regulatory policies. Secondly, it is a case study of a problem routinely encountered by regulators, namely that the firms they are trying to regulate often have far better data on the behaviour of the individuals regulators seek to protect. The study helped to correct that imbalance. Lastly, as data were collected in multiple experiments on different, realistically designed platforms, the evidence supplied could be regarded as fairly strong.

3.3 Pre-tests with multiple methods

This last example brings us to another aspect of pre-testing that is worth illustrating. Where possible, the deployment of multiple methods can strengthen the evidence generated. When multiple methods are applied to the same research question, this is referred to as 'triangulation' of methods. Increasingly, behavioural scientists look to apply multiple approaches that use traditional data analysis to supplement the experimental method, or multiple experimental methods.

An example of pre-testing via multiple methods was undertaken by the UK

telecommunications regulator, Ofcom, in relation to encouraging consumer switching. The first of two studies investigated the impact of automatically renewable contracts (ARCs) on consumer switching behaviour (OECD, 2017a). The second compared consumer behaviour when presented with gaining provider led (GPL) switching processes (where the consumer only contacts the new provider when they want to switch) compared to losing provider led (LPL) switching processes (where the consumer makes up to three contacts during switching and can be given a counter-offer by the losing provider) (Huck and Wallace, 2010).

ARCs, contracts that are automatically renewed after the minimum contract period, were introduced by the provider BT in 2008. Concerned about the effect on switching behaviour, Ofcom carried out an initial econometric data analysis of the frequency of switching of BT customers on rollover contracts compared to comparable customers on standard contracts. They found that customers on ARCs switched significantly less than those on standard contracts. This analysis was a key part in informing the decision to prohibit ARCs (OECD, 2017a). The second study investigated whether GPL processes led to more switching behaviour than LPL processes. The study used laboratory tasks designed to mimic the telecommunications market, including different levels of demand, minimum term contracts with penalties for early departure and search costs for switching. The findings suggested the GPLs were better for consumers and resulted in better switching behaviour, but only if verification was first provided. The benefit of GPLs disappeared when GPL processes were carried out without verification (referred to as 'slamming') from the consumer. Furthermore, early termination charge warnings were not found to be helpful in either GPL or LPL processes.

There are two factors to note. The first is the different methodologies used to pre-test policy decisions, with the first study using traditional econometric analyses of behavioural outcomes and the second using a laboratory experiment. The second is that the behavioural studies can pre-test the efficacy not only of the policy itself but also of the factors that may drive its success or failure, such as the inclusion of additional information (e.g. real-time displays for smart meters) or the exclusion of other behaviours (e.g. loss of benefits for GPLs when slamming is a part of the process).

These three case studies are just examples of some of the ways that field trials, laboratory experiments and mixed-method studies have been used to pre-test policies. Other examples for which there is not space to include in full include the Behavioural Insights Team RCT field trials to measure the efficacy of back-to-work schemes run by the Department for Work and Pensions in the UK (Haynes et al., 2012), laboratory experiments run by the Financial Conduct Authority on regulation around disclosure of information for pensions (OECD, 2017a), and European Commission mixed-methods research on environmental car labels (Codagnone et al., 2013). Nevertheless, the trials that pre-test policy are a minority in the wide breadth of behavioural science research for policy, most of which has

to date focused on informing policy but stopped short of pre-testing it.

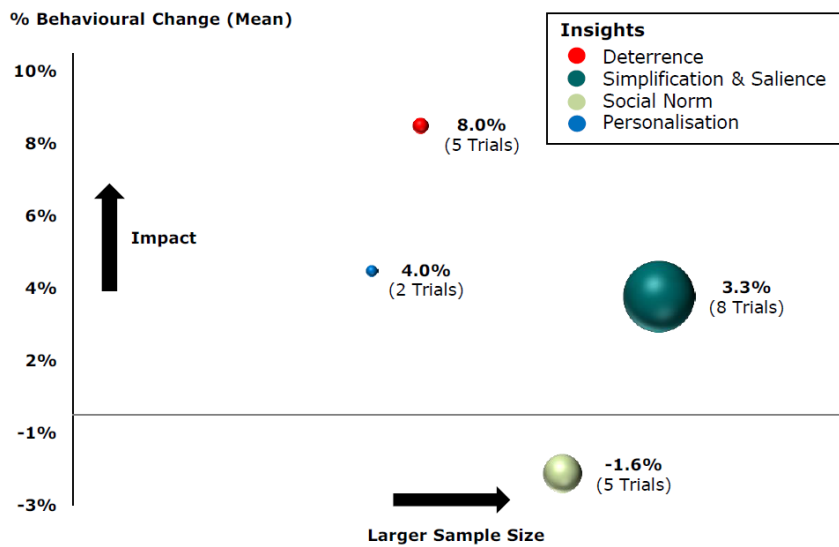
4 PROGRESS IN IRELAND

While the USA and, particularly, the UK have led the charge in applying behavioural science to policy, Ireland is among a group of countries not far behind. Dedicated teams of researchers applying behavioural science to policy problems now operate within Revenue, the ESRI and the Sustainable Energy Authority of Ireland (SEAI). Within the university sector, a new behavioural science group at the UCD Geary Institute for Public Policy has a strong applied focus, while individual behavioural scientists conduct some research for policy in most of Ireland's universities. The Irish Government Economic and Evaluation Service (IGEES) is also developing capability in the application of behavioural economics. The Irish Behavioural Science and Policy Network acts as a forum where members of these teams interact with multiple policymakers, academics and interested people from the private sector. With all this activity, the potential benefits of pre-testing policy interventions are becoming better known within the civil and broader public service and a substantial number of relevant studies have been conducted.

4.1 Revenue

Revenue was the first state body in Ireland to implement RCTs to test the application of behavioural research. It has been conducting trials of communications within the Irish tax administration for the past seven years. Overwhelmingly, these trials have involved the manipulation of written communication. Twenty RCTs are summarised in Kennedy et al. (2017), which conducted a meta-analysis. The behavioural levers tested broadly fell into four categories: (i) making a deterrent salient; (ii) simplifying information or making key information more salient; (iii) communicating a social norm (e.g. stating in the communication that the majority of the target group files tax returns on time); (iv) personalising the message (including a handwritten component, using individuals' names, etc.). Figure 3 summarises the results of this meta-analysis by plotting the mean effect size measured across the RCTs by type of behavioural intervention, expressed as the percentage-point difference in the main outcome variable (treatment group minus control group), with the sample size on the horizontal axis. This analysis suggests that, on average, highlighting a deterrent was the most successful behavioural lever. However, these averages mask some differences between individual studies. For instance, one of the personalisation manipulations involved affixing hand-written 'post-it' notes to letters sent to small and medium enterprises encouraging them to complete and return a survey. This generated a particularly large effect size, almost doubling initial response rates.

FIGURE 3 META-ANALYSIS OF 20 RCTS CONDUCTED BY REVENUE, DISTINGUISHING FOUR TYPES OF BEHAVIOURAL INTERVENTION, OR 'INSIGHT'



Source: Revenue (Kennedy et al., 2017).

Although the measures trialled in these studies reflect relatively small administrative changes rather than pre-tests of substantial policy interventions, the strategy of conducting multiple trials is of wider benefit, as it permits useful inferences to be made about the likely effectiveness of these behavioural levers in other domains. Three of the four levers were generally effective, but it is notable that the communication of social norms was not. This stands in contrast to international results (Coleman, 2007; Behavioural Insights Team, 2012). Although more research is required on this, it is possible that people in Ireland are less receptive to the idea of compliance with social norms, with implications for the design of interventions in other domains.

4.2 The ESRI's Behavioural Research Unit

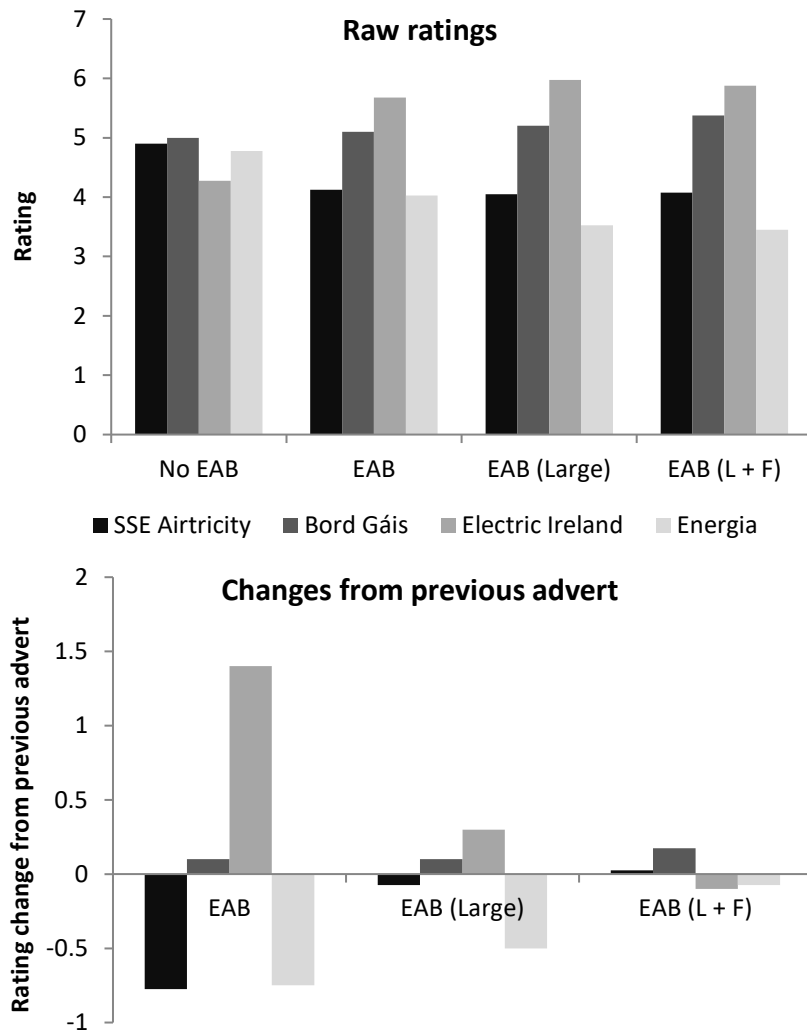
Some pre-testing of more substantial policy changes (as opposed to changes in administrative practice) is now being undertaken in Ireland. The Behavioural Research Unit (BRU) at the ESRI recently conducted a laboratory pre-test of new regulations on price transparency in the residential energy market (Lunn and Bohacek, 2017). The study followed a previous experiment indicating that the marketing practice of expressing prices as discounts from standard unit rates, which vary between providers, makes it substantially harder for consumers looking for cheaper electricity to choose better-value offerings. The Commission for the

Regulation of Utilities (CRU)² proposed and consulted on a regulatory requirement to include an estimated annual bill (EAB) in all advertising and marketing materials. The EAB is calculated for a consumer of average usage, as set by regulation, such that it is a directly comparable price across providers – similar to an APR on credit products. The experimental pre-test set out to determine whether it would influence how consumers perceived the value of offerings and whether they found it easier to choose cheaper electricity tariffs when the EAB was present.

Figure 4 shows the results of one section of the study, in which a sample of consumers rated advertisements for value. The adverts corresponded to the offerings in the market from the four largest providers in Ireland at the time of the study. Four conditions were tested: (i) a control condition (No EAB) consisting of typical adverts prior to the regulation; (ii) a condition (EAB) in which the EAB was legibly displayed alongside other price information; (iii) a condition (EAB Large) in which the EAB was displayed with the same font size as other price information; (iv) a condition (EAB L + F) in which the EAB was displayed with the same font size as other price information and an explanatory footnote was shown. The providers are listed as A to D in decreasing order of their unit rates at the time. The results revealed that showing the EAB produced a large and statistically significant swing in favour of cheaper providers, which strengthened when the EAB was shown with the same font size as other price information.

² The study preceded the recent name change from the Commission for Energy Regulation (CER).

FIGURE 4 PRE-TEST OF EAB INTERVENTION FOR RESIDENTIAL ELECTRICITY PACKAGES. ADVERTISEMENT RATINGS WERE SYSTEMATICALLY ALTERED IN FAVOUR OF THE MORE COMPETITIVE PROVIDERS WHEN THE EAB WAS DISPLAYED WITH THE SAME FONT SIZE AS OTHER PRICE INFORMATION ('EAB LARGE')



Source: Lunn and Bohacek (2017).

Further tests within the same study showed that displaying the EAB increased the likelihood that consumers would choose the cheaper offering and improved consumers' ability to trade off price information against other product attributes accurately. Following this pre-test and the consultation period, CRU introduced the requirement for providers to provide the EAB in all marketing material.

This study was the first to pre-test a new regulation experimentally in Ireland. More laboratory pre-tests are currently being designed and undertaken in the BRU in relation to communication of information about telecommunications products, pensions, car finance, other credit products, calories on restaurant menus, and

smart meters. Studies vary from pre-testing the contents of consumer advice webpages to testing the detail of regulations ahead of new legislation (the study on calorie posting, which is described further below). In addition to these laboratory pre-tests, the BRU is designing field trials and RCTs of interventions that aim to reduce nitrate pollution on Irish farms, to encourage action to remove any lead fixtures in domestic water piping, and to increase levels of physical activity among the socially disadvantaged.

4.3 Other examples in Ireland

Applications involving the use of behavioural science to pre-test policy are also under way on a smaller scale elsewhere within the public service. In 2017, SEAI established a behavioural economics unit with the intention of engaging in pre-tests of interventions. Work under way includes laboratory pre-tests designed to increase the effectiveness of the Building Energy Rating (BER) certificate, trials of an online calculator designed to assist consumers' understanding of electrical vehicles, and pre-tests of alternative webpages that aim to encourage the take-up of grants for energy efficiency upgrades. Some of the behavioural researchers at Revenue, whose work is described above, are members of IGEES. Other IGEES staff in central government departments are involved in various trials designed to improve the efficiency of administrative practice. Most of this work is at an earlier stage of development than the research undertaken in Revenue. It includes trials of behaviourally informed communications in employment centres and of letters to outpatients designed to improve the management of hospital waiting lists. This type of pre-testing is broadly similar to that undertaken by BIT in the UK, in terms of both scientific method and the sort of policy research questions addressed. The work is summarised in an IGEES paper (Purcell, 2016).

Overall, it appears that the understanding of the potential benefits of pre-testing policy interventions is spreading within Irish policymaking. As in other countries, most work is designed to pre-test behaviourally informed improvements in the effectiveness of administrative communications. In the process, the behavioural science community is growing and lessons regarding behavioural levers that potentially work differently in Ireland to elsewhere are being learned. Some pre-tests are now being undertaken of larger policy interventions where behavioural experiments can be deployed. One notable feature, however, is that in Ireland, unlike most other countries, there is little central direction to this expansion of work. Experimental pre-tests are essentially being undertaken by departments and agencies within which individual officers and executives have become aware of the possibilities and have had the wherewithal to engage with this alternative approach to policy development.

5 EVOLVING METHODS: PROCESS TRACING

The methods used to pre-test policy interventions most frequently fall under the categories of RCTs, laboratory experiments or other field trials. Within each of these categories lies a range of techniques and methodologies that can be adapted to suit the research question and proposed design. These can include analysing consumers' preferences, testing the quality of individual decision-making, or recording the extent of desirable changes in behaviours. The type of study undertaken is dictated by the policy to be tested and the specifics of the research question. For example, a policymaker aiming to regulate marketing material in a specific domain may be interested in how consumers' preferences differ depending on the format of the information they are exposed to. Alternatively, a policymaker may consider mandating the inclusion of a warning label alongside marketing material and may therefore pre-test whether inclusion of this label increases the consistency of (and hence presumably reduces the confusion within) consumers' decisions. If behaviour change is the target of a policy then a pre-test to determine whether implementation of the policy really does change behaviour – as opposed to the intention or motivation to change – in either a field trial or a laboratory setting will be the most useful technique.

While these are the most commonly used methods in pre-tests, they are not the only ones. There has been an increasing interest in recent years in 'process tracing', which refers to analyses of not just what decision individuals make but how they make it.

Process tracing methodologies include: verbal protocol analyses, in which experimental participants are asked to verbalise their thoughts as they make a decision; hand movement analyses, in which decision-makers' movements of a computer mouse are recorded while they make the decision; and eye tracking, in which decision-makers' eye movements are recorded. While all have been used in decision-making research, there is the possibility that if a technique involves awareness of the measurement or effort during the decision-making process it can change the decision itself (Glaholt and Reingold, 2011). This is of particular relevance to methodologies that place additional demands on participants while they engage in a study. This may in part explain why there is a growing interest in the use of eye tracking as a measure in behavioural pre-tests. Eye tracking offers a non-invasive and unobtrusive means of assessing what consumers are looking at and, at least in part, attending to (Glaholt and Reingold, 2011).

Modern eye-tracking equipment uses a combination of a near-infrared illuminator and a high-resolution camera to track eye movements, most often to assess where someone is looking on a screen. The illuminator shines near-infrared light into the centre of the eye, which causes a reflection on the cornea. The camera can then track the position of this reflection to estimate where a person is looking. Advances in the technology mean that modern eye-tracking equipment can take over 1,000

samples per second, so the estimations of gaze location are updated and precise. The data available from eye tracking include fixations, which are pauses in movement and thus show what someone has looked at and how many times they have looked at it, and saccades, which show movements themselves and thus can show the order in which someone looked at different pieces of information.

Eye tracking is becoming a widely used method in behavioural studies to assess whether information is attended to. It is predictive of choices and has been used to investigate when consumers look at what information. We briefly outline three international examples and one from Ireland.

5.1 Pictorial warning labels on tobacco products in the US

Recent research funded by the Food and Drug Administration (FDA) in the United States used eye tracking to assess (i) whether smokers attend to graphic pictorial warning labels (PWLs) on cigarette packaging and (ii) which of the FDA's proposed PWLs were most effective at capturing attention and memory (Lochbuehler et al., 2017). Using eye tracking, it was shown that smokers' attention was drawn to the images more quickly than to text and that they spent longer looking at the images than the text. In a follow-up survey the research demonstrated that the warning messages from FDA PWLs that had a congruent text and pictorial warning were more likely to be remembered than the PWLs that had an incongruent text and pictorial warning. This research is to be used as support for the FDA policy in a lawsuit taken by tobacco companies against PWLs.

5.2 Country-of-origin labelling of meat in the EU

In 2015 European Union legislation required mandatory country-of-origin labelling within the EU for beef, pigs, sheep, goats, poultry, fruit and vegetables, olive oil, wine, eggs, honey and hops (Fraser et al., 2015). As the list of products within the remit of this legislation grew, the Department for Environment, Food and Rural Affairs (DEFRA) in the UK noted a growing consumer demand for country-of-origin labelling on other products, as well as voluntary labelling by multiple retailers (Fraser et al., 2015). DEFRA also noted that consumers expected such labelling to provide correct and not misleading information.

In light of this, and with the expectation that the country-of-origin labelling legislation would expand in future, DEFRA carried out research to identify and understand UK consumer preferences for labelling on a range of meat products, to ascertain values for different labelling requirements, and to check how attention to country-of-origin information is influenced by other information on packaging using eye tracking (Fraser et al., 2015). Choice experiments conducted online were validated via a face-to-face eye-tracking study that measured consumers' attention to different aspects of labelling combined with their willingness to pay for products with the labelling. Results for the online and eye-tracking samples were consistent: UK country-of-origin labelling was valued positively, particularly for

fresh/chilled/frozen meat compared to processed products. Deployment of eye tracking showed that price, product quality and country-of-origin labelling received comparable attention – more than organic and quality assurance labels. Additional attributes on packaging did not diminish attention paid to the labels or reduce willingness to pay. Neither did the presence of a flag indicating country of origin, overall, although it did draw attention to country-of-origin labelling when packaging was more complex. These pre-tests of possible combinations of labels are informing further policy development on country-of-origin food labelling.

5.3 Consumer protection in Colombia's communications market

In a collaboration with the OECD, the Colombia Communications Regulator used behavioural insights to inform the redesign of its regulatory regime to protect consumers, who were often paying for services that failed to meet expectations. Based on the results of 25 consumer psychology experiments, the OECD made four recommendations to the regulator that covered principles governing how information should be communicated to customers with respect to consumption, customer service (including complaints and issues), and information on bundled services (OECD, 2017b). Following these studies the OECD recommended further pre-testing and analysis of the changes prior to implementation. One such test involved using eye tracking to trace the visual path that consumers took while reading a bill in order to assess how they attended to the information. Following implementation of the findings of all experiments, the new regime has overhauled the provision of information and steps for customer services to improve customer protection. One simplification was to change the contract provided to customers from a terms and conditions document that originally took 6 hours and 15 minutes to read to one that can be read in 12 minutes.

5.4 Calorie posting in Ireland

Eye tracking is becoming an increasingly popular tool in the arsenal of behavioural methodologies that can be used to pre-test consumer behaviour around policy interventions. In Ireland, the ESRI's BRU is currently using eye tracking to assess how consumers process calorie information on menus, whether the formatting of menus influences attention, and whether this in turn changes consumer behaviour. This is an experimental pre-test of a legislative proposal that is likely to affect thousands of businesses and almost all consumers at some point. In line with the argument of this paper, legislation to introduce calorie posting appears to be exactly the kind of substantial policy decision that experimental pre-testing has the potential to improve. Results of a first study are expected by autumn 2018.

There are of course some caveats to using eye tracking or other process-tracing methodologies. The first is whether it adds to the research question. In some situations eye tracking can provide valuable additional information about how consumers process information and this may be of key importance for the policy

question at hand. In other cases, the ‘whether’ is more important than the ‘how’ and thus adding eye tracking to a study adds cost in terms of time (consumers can only be tested individually rather than in groups) and equipment without a comparative benefit. The second caveat is that while tracking eye movements has been shown to be indicative of attention and predictive of choice, it is also clear that someone can attend to something while not looking at it. For this reason eye tracking should always be used in conjunction with other behavioural techniques. With these caveats in mind, if the policy research question would benefit from understanding how consumers process information and how this influences behaviour, then eye tracking and other process-tracing techniques can be used to record additional information, support and validation for the research question, improving the pre-test of policy impact.

6 CONCLUSIONS

The rapid expansion of behavioural science research as a policy tool over the past 10 years is testament to the value of experiments. It is also, due to the willingness of individual officers and executives to admit uncertainty about a policy outcome and to embrace experiments, a way to resolve the uncertainty during the process of policy development. While uncertainty about an outcome can be threatening, it is also the environment in which an experimental approach thrives, given that an experiment tests a specific effect with all else being held constant. The fact that an experimental outcome is another unknown may be a risk, but it is one that has to be weighed against the risk of a policy intervention that has no effect or, worse, a detrimental one. Government budgets are finite and contentious, so an approach that helps to promote effective interventions and to avoid costly mistakes is rightly gaining traction.

The research described in this paper illustrates the breadth of methods that can be used as tools to pre-test policy interventions. We have divided these into three categories that summarise much of the ongoing work, but there are subdivisions within these that could be further unpicked. Field trials, most often in the form of RCTs, have been the most common type of behavioural intervention used for policy. The Behavioural Insights Team has been instrumental in illustrating the value of field RCTs to test different forms of communication that can inform best practice for existing policies in guiding consumer behaviour and better decision making. The research carried out by the Energy Demand Research Project successfully used field trials to pre-test the effectiveness of smart meters on reducing energy consumption across the UK. Historically, laboratory experiments have been a less commonly used tool in behavioural research for policymaking, but they are increasing in both number and impact.

Controlled laboratory experiments such as those carried out by Ofcom have been able to show in fine-grained detail where a broader policy may be effective in

guiding consumer decision-making and where it may fail, giving an important insight into the other factors that a policymaker might need to consider when, for example, mandating changes to information or labelling. Triangulation of methods is an area that is rapidly expanding as the toolbox available to behavioural scientists and policymakers grows. Combinations of traditional data analyses and online, laboratory and field experiments, including RCTs, can be used to delve further into specific research questions, to hone policy questions and to validate findings that allow policymakers to be more certain about likely outcomes. In addition, the technological innovations that underpin process tracing now permit unobtrusive tracking of consumer decision-making in real time, such as through eye tracking, which provides further insight into how people process information and how this then guides behaviour.

These methodologies have been applied at all stages of the policy cycle, from research to design to implementation and evaluation. Yet there is an imbalance in this picture, with the vast majority being applied at the later stages of policy development. Within the subset applied to early policy development there is still only a minority of studies that seek to pre-test specific policy questions before implementation. This is perhaps inevitable given the success of early behavioural interventions to improve the administration of existing policies. There is also a lower risk involved in testing a small improvement to the implementation of a policy rather than a pre-test of the policy itself, which may have existing supporters and detractors. Such considerations must not mask the potential for the use of behavioural research to pre-test policies that are still in development. At present, high levels of expertise are sometimes being deployed to test relatively peripheral areas of policy which, while not unimportant, are not getting to the heart of what behavioural science can offer.

Much behavioural research for policy focuses on decision-making when consumers choose between products. This is natural given its foundations in economic decision-making and the progress that has been made in applying behavioural science to areas of financial decision-making that consumers typically find confusing and misleading. However, consumer decision-making is only one area that behavioural science can feed into. Given that many of the serious challenges faced by our communities, our countries and our planet are linked to specific forms of human behaviour, we have the potential to use behavioural insights to help find solutions in areas of pressing concern. These include over- and under-nutrition, physical activity, housing, education, inequality, parenting, medical services and the environment.

Beyond the understandable focus on decision-making, we can look at the context of people's behaviour, for example how specific environments may lead to feelings of inertia or to increased risk seeking, what behavioural barriers people face to accessing medical services, and how changes to early and late education can ease the way for better decision-making and healthier life choices. These are not simple

problems and they will not have simple solutions, but this is where experimental research offers an advantage. More often than not, experiments can provide a clear answer as to which of a small number of options are most likely to lead to a desired effect, all else being equal. One experiment will not solve the most complex problems, but a series of experiments that test, reassess, test and reassess can start to clear a path through what was initially a forest of uncertainty.

These benefits can only be obtained if behavioural science is applied more broadly throughout the policy development process, rather than to test minor amendments to existing policies or their implementation. In particular, if the techniques of behavioural science are deployed at the earliest stages of policy development, to provide guidance in understanding of behaviour, to pre-test where there is uncertainty, it can provide policymakers with a stronger tool that gives a scientific foundation to the policy development process. The approach of course requires us to embrace uncertainty openly and to test our assumptions. Admitting uncertainty requires some courage, but the experimental method offers the promise of greater certainty and, ultimately, better policy.

REFERENCES

- AECOM (2011), *Energy Demand Research Project: final analysis*, St Albans: AECOM.
<https://www.ofgem.gov.uk/ofgem-publications/59105/energy-demand-research-project-final-analysis.pdf>
- Amir, O., D. Ariely, A. Cooke, D. Dunning, N. Epley, U. Gneezy, B. Köszegi, D. Lichtenstein, N. Mazar, S. Mullainathan, D. Prelec, E. Shafir, and J. Silva (2005), 'Psychology, behavioural economics, and public policy', *Marketing Letters*, Vol. 16, pp. 443–454.
- Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shotland, and M. Walton (2017), 'From proof of concept to scalable policies: challenges and solutions, with an application', *Journal of Economic Perspectives*, Vol. 31, No. 4, pp. 73–102.
- Banerjee, A., R. Hanna, J. Kyle, B.A. Olken, and S. Sumarto (2018), 'Tangible information and citizen empowerment: identification cards and food subsidy programs in Indonesia', *Journal of Political Economy*, Vol. 126, No. 2, April, pp. 451–491.
- Behavioural Insights Team (2012), *Applying behavioural insights to reduce fraud, error and debt*, London: Cabinet Office.
- Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a, and J. Sandefur (2015), Interventions and institutions: experimental evidence on scaling up education reforms in Kenya, Working Paper.
- Carroll, J., S. Lyons and E. Denny (2014), 'Reducing household electricity demand through smart metering: the role of improved information about energy saving', *Energy Economics*, Vol. 45, pp. 234–243.
- Cartwright, N. and J. Hardie (2012), *Evidence based policy: a practical guide to doing it better*, Oxford: Oxford University Press.
- Codagnone, C., F. Bogliacino, and G. Veltri (2013), *Testing CO₂/car labelling options and consumer information*, London: London School of Economics and Partner Consortium.
https://ec.europa.eu/clima/sites/clima/files/transport/vehicles/labelling/docs/report_car_labelling_en.pdf
- Coleman, S. (2007), The Minnesota income tax compliance experiment: replication of the social norms experiment, MPRA Paper No. 5820. <https://mpra.ub.uni-muenchen.de/5820/>
- Crepon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora (2013), 'Do labor market policies have displacement effects? Evidence from a clustered randomized experiment', *Quarterly Journal of Economics*, Vol. 128, pp. 531–580.

- Deaton, A. (2010), 'Instruments, randomization, and learning about development', *Journal of Economic Literature*, Vol. 48, pp. 424–455.
- Duflo, E. and E. Saez (2003), 'The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment', *Quarterly Journal of Economics*, Vol. 118, pp. 815–842.
- Fraser, I.M., K. Balcombe, M. Hussein, and D. Bradley (2015), *Consumer preferences regarding country of origin labelling of meat – FA0156*. Project report. London: DEFRA.
- Glaholt, M.G. and E.M. Reingold (2011), 'Eye movement monitoring as a process tracing methodology in decision making research', *Journal of Neuroscience, Psychology, and Economics*, Vol. 4, No. 2, pp. 125–146.
- Haynes, L., O. Service, B. Goldacre, and T. Torgerson (2012), *Test, learn, adapt: developing public policy with randomised controlled trials*, London: Cabinet Office.
- Huck, S. and B. Wallace (2010), *Consumer switching: Experimental economics research*, London: London Economics. https://www.ofcom.org.uk/__data/assets/pdf_file/0021/68043/economics-research.pdf.
- Johnson, E.J. and D. Goldstein (2003), 'Do defaults save lives?', *Science*, Vol. 302, pp. 1338–1339.
- Kennedy, S., R. O'Carroll, M. Shirran, and K. Walsh (2017), *Applying behavioural science in tax administration – a summary of lessons learned*, Dublin: Revenue.
- Lochbuehler, K., M. Mercincavage, K.Z. Tang, C.D. Tomlin, J.N. Cappella, and A.A. Strasser (2017), 'Effect of message congruency on attention and recall in pictorial health warning labels', *Tobacco Control*, Vol. 27, pp. 266–271.
- Ludwig, J., J.R. Kling, and S. Mullainathan (2011), 'Mechanism experiments and policy evaluations', *Journal of Economic Perspectives*, Vol. 25, pp. 17–38.
- Lunn, P.D. (2014), *Regulatory policy and behavioural economics*, Paris: OECD Publishing.
- Lunn, P.D. and M. Bohacek (2017), 'Price transparency in residential electricity: experiments for regulatory policy', *Journal of Behavioral Economics for Policy*, Vol. 1, No. 2, pp. 31–37.
- Lunn, P.D. and Á. Ní Choisdealbha (2018), 'The case for laboratory experiments in behavioural public policy', *Behavioural Public Policy*, Vol. 2, pp. 22–40.
- Madrian, B.C. and D.F. Shea (2001), 'The power of suggestion: inertia in 401(k) participation and savings behaviour', *Quarterly Journal of Economics*, Vol. 116, pp. 1149–1187.
- McKenzie, C.R.M., M.J. Liersch, and S.R. Finkelstein (2006), 'Recommendations implicit in policy defaults', *Psychological Science*, Vol. 17, pp. 414–420.

OECD (2017a), *Behavioural insights and public policy: lessons from around the world*, Paris: OECD Publishing.

OECD (2017b), 'Impact update: what happened next?', *Protecting consumers through behavioural insights: regulating the communications market in Colombia*, Paris: OECD.

Purcell, K. (2016), *Applying behavioural economics in Irish policy*, Dublin: Department of Public Expenditure & Reform.

Sousa Lourenço, J., E. Ciriolo, S. Rafael Almeida, and X. Troussard (2016), *Behavioural insights applied to policy: European Report 2016*. EUR 27726 EN, Luxembourg: Publications Office of the European Union.

Sunstein, C.R. (2011), 'Empirically informed regulation', *University of Chicago Law Review*, Vol. 78, pp. 1348–1429.

Sunstein, C.R. (2013), 'Deciding by default', *University of Pennsylvania Law Review*, Vol. 162, pp. 1–57.

Thaler, R.H. and C.R. Sunstein (2008), *Nudge: improving decisions about health, wealth, and happiness*, New Haven, CT: Yale University Press.

van Bavel, R., B. Hermann, G. Esposito, and A. Proestakis (2013), *Applying behavioural sciences to EU policy-making*, Brussels: European Commission.

Whitaker Square,
Sir John Rogerson's Quay,
Dublin 2
Telephone **+353 1 863 2000**
Email **admin@esri.ie**
Web **www.esri.ie**
Twitter **@ESRIDublin**