Evidence and Evaluation in Social Policy

Broadening Perspectives on Social Policy
Series Editor: Bent Greve

The object of this series, in this age of re-thinking on social welfare, is to bring fresh points of view and to attract fresh audiences to the mainstream of social policy debate.

The choice of themes is designed to feature issues of major interest and concern, such as are already stretching the boundaries of social policy.

This is the seventeenth collection of papers in the series. Previous volumes include:

- Crime and Social Policy *H. Kemshall*
- The Times They Are Changing? Crisis and the Welfare State *B. Greve*
- Reforming Long-term Care in Europe *J. Costa-Font*
- Choice: Challenges and Perspectives for the European Welfare States *B. Greve*
- Living in Dangerous Times: Fear, Insecurity, Risk and Social Policy *D. Denney*
- Reforming the Bismarckian Welfare Systems *B. Palier and C. Martin*
- Challenging Welfare Issues in the Global Countryside *G. Giarchi*
- Migration, Immigration and Social Policy *C. Jones Finer*
- Overstretched: European Families Up Against The Demands of Work and Care *T. Kröger and J. Sipilä*
- Making a European Welfare State?: Convergences and Conflicts over European Social Policy *P. Taylor-Gooby*
- The Welfare of Food: Rights and Responsibilities in a Changing World *E. Dowler and C. Jones Finer*
- Environmental Issues and Social Welfare *M. Cahill and T. Fitzpatrick*
- The Business of Research: Issues of Policy and Practice *C. Jones Finer and G. LewandoHundt*
- New Risks, New Welfare: Signposts for Social Policy *N. Manning and I. Shaw*
- Transnational Social Policy *C. Jones Finer*
- Crime & Social Exclusion *C. Jones Finer and M. Nellis*

Evidence and Evaluation in Social Policy

Edited by

Ian Greener and Bent Greve

WILEY Blackwell

This edition first published 2014
Originally published as Volume 47, Issue 4 of *Social Policy & Administration*

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Library of Congress Cataloging-in-Publication Data applied for

Paperback ISBN: 978-1-118-81654-7

A catalogue record for this book is available from the British Library.

Cover image: © Dumitru Gornet/Alamy
Cover design by Design Deluxe

Set in 10.5 on 11 pt Baskerville MT by Toppan Best-set Premedia Limited

1 2014

CONTENTS

	List of Contributors	vii
	Introduction: Evidence and Evaluation in Social Policy *Ian Greener and Bent Greve*	1
1	Trials and Tribulations: The 'Use' (and 'Misuse') of Evidence in Public Policy *Christopher Deeming*	5
2	Understanding the Influence of Evidence in Public Health Policy: What Can We Learn from the 'Tobacco Wars'? *K. E. Smith*	29
3	Caught in the Same Frame? The Language of Evidence-based Policy in Debates about the Australian Government 'Intervention' into Northern Territory Aboriginal Communities *Emma Partridge*	47
4	A Systematic Review of Comparative Studies of Attitudes to Social Policy *Trude Sundberg and Peter Taylor-Gooby*	63
5	Public Opinion and Policy-making *Ray Pawson and Geoff Wong*	81
6	Obstacles to Evidence-based Policy-making in the EU Enlargement Countries: The Case of Skills Policies *Will Bartlett*	97
7	Understanding Employment Barriers for Lone Parents in Great Britain: Research Gaps and Missed Opportunities *Tina Haux*	115
8	Putting the Research Boot on the Policymakers' Foot: Can Participatory Approaches Change the Relationship between Policymakers and Evaluation? *Liz Richardson*	129
	Index	147

List of Contributors

Will Bartlett, European Institute, London School of Economics and Political Science, London, UK.

Christopher Deeming, Geographical Sciences, University of Bristol, Bristol, UK.

Ian Greener, Durham University, Durham, UK.

Bent Greve, Roskilde University, Roskilde, Denmark.

Tina Haux, University of Lincoln, Lincoln, UK.

Emma Partridge, Social Policy Research Centre, University of New South Wales, Sydney, Australia.

Ray Pawson, School of Sociology and Social Policy, University of Leeds, Leeds, UK.

Liz Richardson, School of Social Sciences, University of Manchester, Manchester, UK.

K. E. Smith, Global Public Health Unit, School of Social and Political Science, University of Edinburgh, Edinburgh, UK.

Trude Sundberg, University of Kent, Kent, UK.

Peter Taylor-Gooby, University of Kent, Kent, UK.

Geoff Wong, Centre for Primary Care and Public Health, Queen Mary, University of London, London, UK.

Introduction: Evidence and Evaluation in Social Policy

Ian Greener and Bent Greve

What counts as evidence in social policy, and how evidence does, or perhaps more often, does not, influence policy-making, have become central questions in the last 20 years. The evidence-based medicine movement gathered momentum in the 1980s and 1990s (Sackett *et al.* 1997), creating a framework for the assessment of research in that field and how it might lead to a more robust basis for clinical decision-making, and even health policy. This led to policymakers, especially the New Labour government in the UK, suggesting that the ideological and interest-based politics of the past were now to be jettisoned in favour of an approach that was based instead on a pragmatic, 'what works' basis instead (Davies *et al.* 1999).

Still, in times of austerity, policy suggestions that are based on solid evidence, especially where they offer opportunities for efficiency savings, will, all other things being equal, have a higher chance of being accepted. The counter-argument naturally being that this also implies a risk that policymakers will only use evidence pointing in the same direction as their existing biases, especially given that evidence is seldom straightforward in its interpretation to policy ends, and this is likely to result in policymakers wanting to be cautious and so tending towards policies that are based on conclusions they feel most comfortable with.

This last issue is especially prevalent in many areas of social policy services where the relation between user and producer, at least to a certain extent, is still a kind of black-box where it can be difficult from the 'outside' to see exactly what is going on, and so to track how policy is being influenced. However, this does not reduce the need for knowing more about what does and does not work in social policy as a way of not only using scarce resources in the best way possible, but also to be able to give users the best possible type of care.

From the perspective of 2013, looking back at a decade of evidence-based policy-making, we should be able to see clear examples of improved policy-making through an increased use of evidence, and a transformation in which research and evaluation drive the policy process, alongside methodological innovation and closer working between researchers and policymakers. Is this actually the case?

The chapters in this book, which were responses to a call asking researchers to specifically address the role of evidence and evaluation in policy-making, suggest a less rosy picture than we might hope.

Evidence and Evaluation in Social Policy, First Edition. Edited by Ian Greener and Bent Greve. Book compilation

Chris Deeming's chapter considers the role of randomized controlled trials (RCTs) in shaping public policy, exploring why RCTs have become regarded as the gold standard in many low- and middle-income countries, but not necessarily in the richer, liberal democracies. Deeming also considers a range of issues associated with the use of RCTs in social policy, including ethical and methodological issues, and the use and misuse of evidence and evaluation in social policy. It is a good introduction to the full range of issues covered in the book, as well as making a contribution in its own right through its comparison of the use of RCTs in the global north and developing south.

Katherine Smith's chapter on the 'tobacco wars' suggests that, rather than being based on research, policy-making with respect to public health and tobacco is perhaps better explained using the methods of political science, in which coalitions and interest groups work to secure their demands, and that evidence is enrolled to those ends rather than driving the process. Nevertheless, one could argue that without many years of substantive evidence on the link between smoking and health, changes would not have been possible.

From an entirely different context, Emma Partridge presents the case of policy-making from the Northern Territory of Australia where the language of evidence-based policy has been used by both the government and its critics. Similarly to Smith's work, in some respects, Partridge emphasizes the political basis of the relationship between knowledge, evidence and policy-making. More than this, in Partridge's view, using evidence-based policy as a prescriptive ideal can make policymakers lose sight of other possible strategies for improving policy towards indigenous people.

Trude Sunberg and Peter Taylor-Gooby's chapter calls into question the idea that systematic reviews can straightforwardly inform policy-making, suggesting that they require considerable care in their use because of potential biases toward US-based research, and the weaker reporting of book-based research compared to articles. Where secondary research is based on research summary and review, this can be a quick and effective way of generating evidence – but, as the authors demonstrate, it is not without significant problems.

Ray Pawson has been at the centre of methodological innovation with respect to the review and evaluation of social policy since the 1990s. Here, his chapter with Geoff Wong considers many of the concerns of Smith, Sunberg and Taylor-Gooby's work in considering the problems of measuring public support for policy interventions. Pawson and Wong suggest that systematic reviews can be useful, but that we need to take a more considered approach to measure not only snapshots of public opinion, but also how opinion is formed. To illustrate their approach, the authors explore public support for legislation banning smoking in cars carrying children. They show how public opinion data can supplement or even supplant a limited scientific research base through a diachronic assessment of public opinion formation that suggests that public opinion can play a wider role in policy-making than is often assumed to be the case.

Turning to a different set of challenges, Will Bartlett's chapter considers skills policies in EU enlargement areas, but even more saliently, the consequences of the economic crisis on countries in periods of rapid change for

policy-making more generally. He suggests that, in these circumstances, state capture and a coercive form of policy transfer occur that are a considerable distance from models of evidence-based policy-making. This apparently depressing story, however, can represent an opportunity for research to play a greater role in policy-making, especially for the evaluation of government programmes, but in a context where advocacy coalitions are a central feature of the political process.

Tina Haux's chapter is an illuminating case study of the relationship between evidence, research and social policy, considering work commissioned by the UK Labour government between 1997 and 2010 to explore the employment barriers for lone parents not in work. Haux finds that research often appears to have been commissioned in areas where it was not taken up by policy, but also that significant gaps occurred where policymakers needed research which was not commissioned. Even though, under the post-2010 Coalition government, policy has 'intensified', continuing and expanding many of Labour's ideas, there is still a lack of robust data and research supporting the government's approach, while at the same time the main research vehicle for monitoring changes to the composition and outlook of lone parents has come to an end, giving us incomplete data series. All in all, this is not an encouraging picture.

Lastly, Liz Richardson's chapter presents an innovative perspective in exploring how local government decision-makers can engage more fully with evaluation research – through the use of participative and co-productive methods. The research finds that these methods helped increase local politicians' awareness of their own ideological biases, and even an enhanced capacity for learning. It is important to note, however, the policymakers in the case study struggled to achieve robust research designs and methods, which may in turn have actually increased their appreciation of the value of academic expertise (something which we can enthusiastically endorse). Richardson suggests that politicians' belief in positivism does not have to be a barrier to more participative evaluation, and offers helpful suggestions for how this can be more successfully achieved. Her work offers the special edition an upbeat ending, showing the potential for bridging the gap between research and policy-making.

The work covered in the special edition, then, presents a very mixed picture of the role of evidence and evaluation in social policy in 2013. Politics remains central to decision-making, and in an environment of continued fiscal austerity, the opportunities for conducting funded research for central government are diminished. There is potential for the synthesis of existing research, but such syntheses need to be treated with care and caution rather than simply trying to import methods from systematic reviews and RCTs in scientific and clinical work. However, these kinds of reviews are often also useful in attempting to systematize what we know at least something about and what we do not know anything about, as well as highlighting areas where there is a stronger need for more research.

Naturally, we also need to take greater account of the specific policy and economic context when exploring the role of evidence in policy-making, and trying to assume such variables away does little to increase the persuasiveness

or potential to work of our research. Although this is not a counter-argument of using evidence, it is more a word of caution in the way one can use evidence. This also implies a stronger need for students of social policy and administration to know how and to what degree certain types of research are conducted, and how one can substantiate an argument in order to avoid stating that 'anything goes'. Solid knowledge and information are thus clearly still extremely important in order to ensure the best possible service for the users of the welfare state.

Lastly, Richardson's chapter especially stresses that it is possible for researchers and politicians to work together, but that it will involve compromises on both sides, and that smaller-scale, more localized work might be the way forward. At a time when funded opportunities in Europe appear to be moving towards large-scale, mega-projects, this is a salutary lesson indeed.

References

Davies, H., Nutley, S. and Smith, P. (1999), What Works? The Role of Evidence in Public Sector Policy and Practice, *Public Money and Management*, 19: 3–5.

Sackett, D., Richardson, W., Rosenberg, W. and Haynes, R. (1997), *Evidence-based medicine: how to practice and teach EBM*, London: Churchill Livingstone.

I

Trials and Tribulations: The 'Use' (and 'Misuse') of Evidence in Public Policy

Christopher Deeming

Introduction

> We can achieve a sort of control under which the controlled, though they are following a code much more scrupulously than was ever the case under the old system, nevertheless feel free. They are doing what they want to do, not what they are forced to do. That's the source of the tremendous power of positive reinforcement – there's no restraint and no revolt. By careful cultural design, we control not the final behaviour, but the inclination to behave – the motives, desires, the wishes. (Skinner 1948: 246–7)

Many observers have commented on the spread of behavioural experiments in global policy. This new and emerging form of public policy often involves conditioning the receipt of welfare – in the form of cash transfers, goods or services – according to specific individual behaviour. Few however, have touched on the contextual differences in terms of the application of research and evidence governing policy, especially social assistance in the developing world and social security within the developed world. Subsequently, the debates on the progress of public policy in the global north and south are not as well connected as they might be.

In LMICs, social experiments with conditionality are part of the drive for evidence-based policy (Fiszbein *et al.* 2009). Conditional cash transfer (CCT) programmes usually have an *a priori* evaluation design built into their operation, that embraces experimental or quasi-experimental features and RCT designs, for example. By contrast, in the advanced liberal democracies, there has been less direct appeal to research evidence gained using from robust evaluation in order to secure major welfare reform: experimentation and evaluation with RCTs has been less of a priority. Consequently, welfare states were reformed on ideological grounds, with an appeal to political theory as Mead and Beem (2005) observe. Putting this in stark terms, welfare conditionality in the south is, arguably, being driven by an evidence-based policy-making agenda, whereas in the north, political philosophy is clearly driving

Evidence and Evaluation in Social Policy, First Edition. Edited by Ian Greener and Bent Greve. Book compilation

welfare reform. This chapter seeks to shed new light on the relationship between evidence and evaluation within the different worldly contexts, by drawing out emerging arguments and counter-arguments about the 'use' and 'misuse' of evidence within public policy. The chapter is organized as follows: the first section examines the increase and nature of evidence-based policy for development; the second section considers how welfare policy has been transformed in the developed world; followed by a more detailed examination of some of the controversies in section three. Lastly, reflections and conclusions are drawn together in the fourth section.

Public Policy for Development: The Rise of Behavioural Economics

In the brave new world of the behavioural economist, achieving the Millennium Development Goals (MDGs) for health and well-being is ultimately about demonstrably changing people's behaviour for the better. Being able to demonstrate the effectiveness or impact of an intervention is thus the keystone for policy development. As a result of the global research effort, it is now well established that there are certain (desirable) human behaviour and conditions of living that are beneficial for our health and well-being that, arguably, should form the basis for public policy (Dean 2010). This point, implicit in the literature, certainly needs to be made more explicit, as biomedical and social research continues to establish food and dietary requirements for good health, education, housing and living standards, along with certain behaviour and practices that reduce the risk of ill-health and disease (e.g. WHO 2002). For far too long, the criticism has been that acceptance of incontrovertible evidence into functioning policy has been slow, partial and unsystematic; resulting in health deficits, waste of human potential and other associated costs to society (WHO 2011). The second point, about clearly demonstrating success and impact in social policy addresses the need for robust evaluation. In order to consider the effectiveness of a particular policy or programme, one really needs to know what outcomes would have been achieved had the programme not been in place. This is often referred to as the 'counterfactual outcome' (OECD n.d.). One way of overcoming this evaluation problem is through the use of RCTs, which are considered to be the 'gold standard' of all the methods available to researchers (Young *et al.* 2002). In the trials, participants are usually randomly assigned to intervention or control groups (Sibbald and Roland 1998), researchers then compare the outcomes between groups (see figure 1.1). A systematic review of RCT results can usually be found at the very pinnacle of the research evidence hierarchy (figure 1.2), as it provides a way of pooling evidence from different studies to provide an overview of outcomes (White and Waddington 2012).[1]

It may not be surprising, therefore, to find an influential group of economists advocating social experiments and in particular RCTs, as the main tool for studying the effectiveness of policy in development settings. In a recent trial in Malawi, for example, poor families were given cash on the condition that they send their children to school (Baird *et al.* 2000a, 2000b). Some villages were randomly assigned to the social assistance programme and

Figure 1.1

Randomized controlled trials (RCTs)

Randomized controlled trials (RCTs) are the most rigorous way of determining whether a cause-effect relation exists between treatment and outcome and for assessing the cost effectiveness of a treatment. They have several important features:

- Random allocation to intervention and control groups.
- Participants and trialists should remain unaware of which treatment was given until the study is completed (although such double blind studies are not always feasible or appropriate).
- All groups are treated identically except for the experimental treatment or intervention.
- The analysis is focused on estimating the size of the difference in predefined outcomes between intervention and control groups.

Other study designs, including non-randomized controlled trials, can detect associations between an intervention and an outcome. But they cannot rule out the possibility that the association was caused by a third factor linked to both intervention and outcome. Random allocation ensures no systematic differences between intervention groups in factors, known and unknown, that may affect outcome.

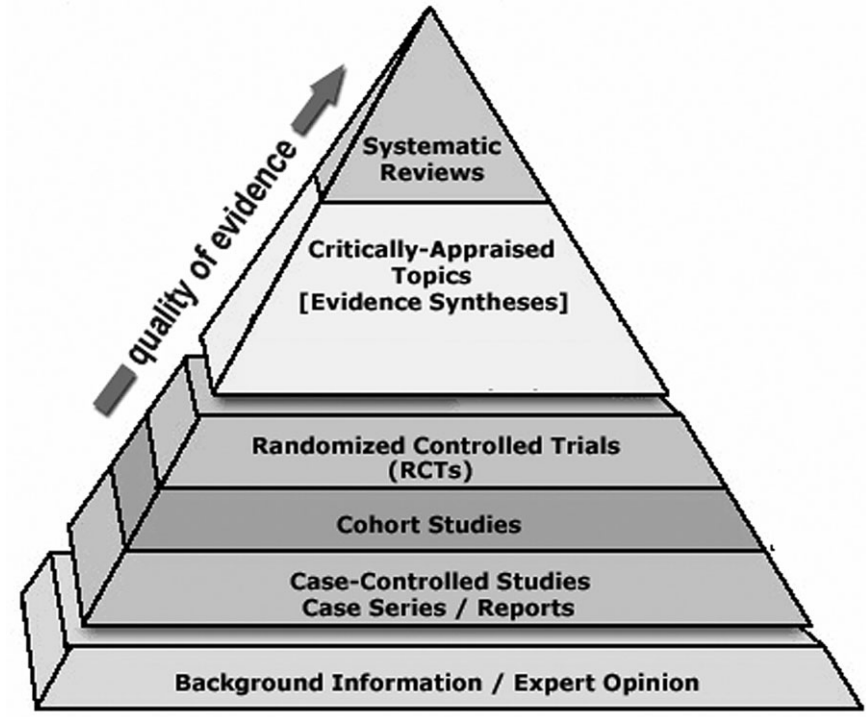
Source: adapted from Sibbald and Roland 1998: 201.

others were not (figure 1.3). Importantly, at the outset, each village had the same equal chance of being picked to receive the intervention; families were not told that they were part of an experiment, they were 'blind' or 'blinded': knowing this may affect their behaviour and the outcome of the study. The development economists – sometimes referred to as the 'randomistas' – argue that this type of randomized experiment is the only sure way of identifying impact, because it helps to eliminate bias and other confounding (hidden) factors. Other non-experimental methods found at the bottom of the evidence hierarchy are largely dismissed as they are considered unscientific and are best avoided (figure 1.2). The influence of the randomistas appears to be growing; non-governmental organizations (NGOs) (the World Bank, the International Monetary Fund, the World Trade Organization), philanthropic agencies and donors (e.g. DFID 2011b) are increasingly giving (explicit) preference to randomized designs and systematic reviews in evaluating public policy programmes and their impacts (Hickey *et al.* 2009; Hagen-Zanker *et al.* 2012).

According to behavioural economists like Banerjee and Duflo (2011), the major debates in international development can be boiled down to disagreements about the shape of a function in development theory (figure 1.4). The S-shaped curve on the left suggests poor people are 'trapped' in poverty and require a 'conditional push' to get out of their own poverty. The L-shaped curve on the other hand, suggests that poor people are gradually able to pull themselves out of poverty because they are not really 'trapped' at all. If we

Figure 1.2

Hierarchies of research evidence



Source: adapted from Davies *et al.* 2000: 48.

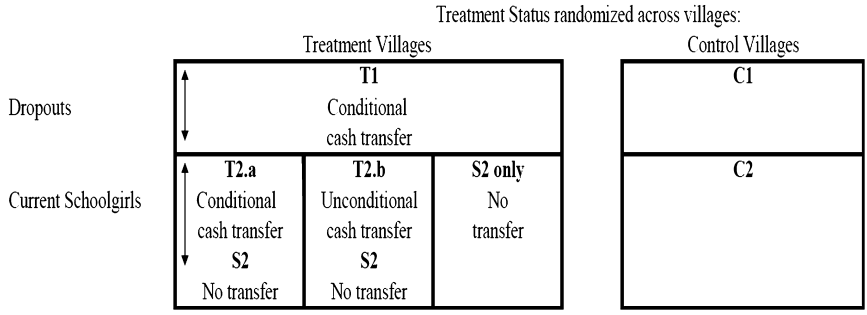
accept the premise of the development economist, the theoretical proposition and corresponding debate about which of the two graphs best represents the real world can only be settled experimentally, through the use of RCTs. Of course, not all social researchers agree with the world view depicted in these two charts or see the need for social experiments to solve the problem of global poverty. The randomistas, however, firmly believe that it is perfectly possible to make significant progress in tackling global social problems using experimentation; through the accumulation of small experimental steps, each well thought out, carefully tested and judiciously implemented. In effect, they claim to be saving lives with a well-placed behavioural 'nudge' and they offer plenty of evidence to support their position.

Saving lives with a well-placed nudge

Conditional social policy programmes are now firmly established across the developing world and shape the lives of millions of people. Many CCT

Figure 1.3

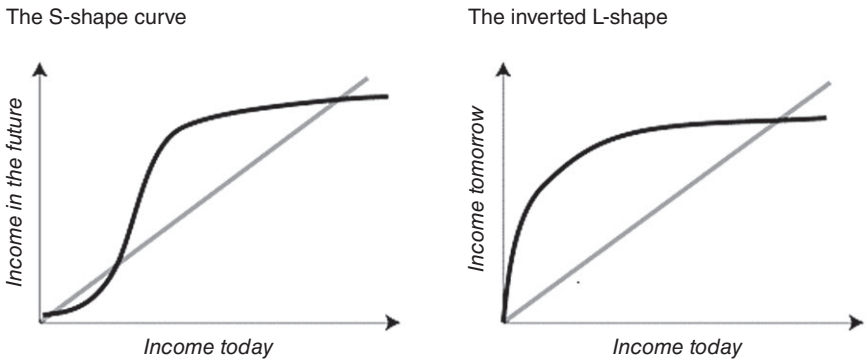
Malawi RCT design



Source: Baird *et al.* 2009a: 7.

Figure 1.4

Different prospects for the world's poorest



Source: adapted from Banerjee and Duflo 2011: 12–13.

programmes are large-scale and usually have an evaluation design built into their operation. *Oportunidades* in Mexico for example, covers about a quarter of Mexico's national population (1.5 million households). In Brazil, some 11 million families – 46 million people – receive regular transfers under the *Bolsa Família* programme. Mexico's *Progresa* (renamed *Oportunidades* in 2002), often seen as the seminal and model programme, began in 1997. This programme has demonstrated a range of benefits over the years, particularly in the areas of health and education (DFID 2006). For instance, some 70 per cent of

families participating in the scheme have shown improved nutritional status and stunting has been reduced; ante-natal care increased by 8 per cent, contributing to a 25 per cent drop in the incidence of illness in newborns; immunization rates improved as a result of preventive healthcare appointments; and school enrolment rates increased by over 20 per cent for girls and by 10 per cent for boys. These social assistance programmes also have fiscal appeal, at least according to national governments and NGOs such as the World Bank. In terms of national budget, Mexico and Brazil commit only 0.5 per cent of gross domestic product to their programmes. With the evidence-base for CCT programmes growing, Guatemala is one of the latest Latin American countries to embark on reform with *Mi Familia Progresa* (my family is moving forward), introduced in 2008 (Gaia 2010). In lower-income contexts, especially in Africa, CCT programmes and experiments are often on a smaller scale – as well as familiar conditions relating to school attendance and the use of health services, other conditions continue to be tested, for example adult education, micro-credit, housing and accommodation schemes, also bed net schemes to help protect people against malaria-carrying mosquitoes (DFID 2011a).

The evidence suggests CCT programmes can be particularly effective in promoting health and education services in LMICs (Lagarde *et al.* 2007). Improvements in the use of health services (table 1.1) is particularly striking as the effects are concentrated among families who are least likely to use the services in the absence of conditional reward. School enrolment rates also show significant gains with conditionality in place (table 1.2). In addition, there are demonstrable impacts on family food consumption. Table 1.3 shows significant increases in daily per capita food consumption: up by 12 per cent in Brazil, 6 per cent in Colombia and over 30 per cent in Nicaragua. Families in the intervention group also tend to spend significantly more on food, indicated by the findings on food budgets. For example, household food expenditures as a proportion of total household budgets increased by 4 percentage points among programme beneficiaries in Columbia, Ecuador and Nicaragua (table 1.3). Lastly, CCT programmes have secured significant reductions in levels of extreme poverty (table 1.4). In Mexico, for instance, severe poverty as measured by the 'squared poverty gap', which attempts to take account of the depth of poverty and inequality among poor people, has fallen by about 29 per cent. Systematic reviews continually look to refine and improve the effectiveness of CCT programmes (Gaarder *et al.* 2010; Yoong *et al.* 2012).[2] Despite the huge investments in large-scale trials, important questions remain unanswered, particularly concerning cost-effectiveness and cost-benefit, an issue to which we return below. Before this, however, we turn to consider the transformation of social policy in the advanced economies, where concerns about the paucity of evidence and robust evaluation continue to be raised.

Welfare Transformed: 'Workfare' States in the Developed World

The neo-liberal foundations of the welfare state can be traced back to the postwar period, particularly in the USA. The 1990s, however, witnessed the

Table 1.1

Impact of CTT programmes on children's attendance at health centres

Country	Country income categories	Programme	Baseline (%)	Impact[α]	Significance[β]	Evaluation design[γ]
Latin American and Caribbean countries						
Chile	Middle	*Chile Solidario*	17.6	2.4(2.7)		RDD
Colombia	Middle	*Familias en Acción*	N/A	33.2(11.5)	***	DID
Ecuador	Middle	*Bono de Desarrollo Humano*	N/A	2.7(3.8)		RCT
Honduras	Middle	*Programa de Asignación Familiar*	44.0	20.2(4.7)	***	RCT
Jamaica	Middle	Program of Advancement through Health and Education	0.2	0.3(0.08)	***	RDD
Mexico	Middle	*Oportunidades*	0.2	0.03(0.02)		RCT
Nicaragua	Low	*Atención a Crises*	70.5	6.3(2.0)	***	RCT
Nicaragua	Low	*Red de Protección Social*	55.4	13.1(7.5)	*	RCT

Source: Fiszbein *et al.* (2009: 19–20).
Notes: N/A = not applicable or comparable with control baseline; α = the column for 'impact' reports the coefficient and standard error (in parentheses); the unit is percentage points except Jamaica where the unit is the number of visits to the health centre in the past six months, and Mexico, where the unit is the number of visits to the health centre in the last six months; β = significance levels: * < 0.05, ** < 0.01, *** < 0.001; γ = RCT defined in figure 1.2; DID = difference-in-differences examines treatment effect by comparing the treatment group both before and after treatment, and to a control group; RDD = regression discontinuity design elicits intervention causal effects by exploiting a given exogenous threshold determining assignment to treatment. By comparing observations lying closely on either side of the threshold, it is possible to estimate treatment effect in situations where randomization was not feasible.

Table 1.2

Impact of CTT programmes on children's attendance at school

Country	Country income categories	Programme	Baseline (%)	Impact[α]	Significance[β]	Evaluation design[γ]
Latin American and Caribbean countries						
Chile	Middle	*Chile Solidario*	60.7	7.5(3.0)	***	RDD
Colombia	Middle	*Familias en Acción*	63.2	5.6(1.8)	***	DID
Ecuador	Middle	*Bono de Desarrollo Humano*	75.2	10.3(4.8)	**	RCT
Honduras	Middle	*Programa de Asignación Familiar*	66.4	3.3(0.3)	***	RCT
Jamaica	Middle	Program of Advancement through Health and Education	18.0	0.5(0.2)	**	RDD
Mexico	Middle	*Oportunidades*	45.0	8.7(0.4)	***	RCT
Nicaragua	Low	*Atención a Crises*	90.5	6.6(0.9)	***	RCT
Nicaragua	Low	*Red de Protección Social*	72.0	12.8(4.3)	***	RCT
Other countries						
Bangladesh	Low	Female Secondary School Assistance Programme	44.1	12.0(5.1)	**	FE
Cambodia	Low	Japan Fund For Poverty Reduction	65.0	31.3(2.3)	***	DID
Cambodia	Low	Cambodia Education Sector Support Project	65.0	21.4(4.0)	***	DID
Pakistan	Middle	Punjab Education Sector Reform	29.0	11.1(3.8)	***	DID
Turkey	Middle	Social Risk Mitigation Project	87.9	−3.0	*	RDD

Source: Fiszbein *et al.* 2009: 17–18.
Notes: α = the column for 'impact' reports the coefficient and standard error (in parentheses); the unit is percentage points except Jamaica where the unit is days; β = significance levels: * < 0.05, ** < 0.01, *** < 0.001; γ = RCT defined in figure 1.2; DID = difference-in-differences examines treatment effect by comparing the treatment group both before and after treatment, and to a control group; FE = a fixed effects statistical model represents the observed quantities in terms of explanatory variables that are treated as if the quantities were non-random, often used in the analysis of panel data; RDD = regression discontinuity design elicits intervention causal effects by exploiting a given exogenous threshold determining assignment to treatment. By comparing observations lying closely on either side of the threshold, it is possible to estimate treatment effect in situations where randomization was not feasible.

Table 1.3

Impact of CTT programmes on food consumption

		Brazil	Colombia		Ecuador	Honduras		Nicaragua		
		2002	2002	2006	2005	2000	2002	2000	2001	2002
Daily per capita food consumption	Control	0.45	0.6	0.65	0.73	0.53	0.47	0.44	0.35	0.35
	Impact (%)	12**	N/A	6**	N/S	N/A	N/S		38**	31**
Food budget shares	Control (%)	60	74	56	54	71	72	73	69	68
	Impact (% points)	0.02**	N/A	0.04**	0.04**	N/A	N/S		0.04**	0.04**

Source: Fiszbein *et al.* 2009: 113.
Notes: significance levels: * < 0.05, ** < 0.01, *** < 0.001; N/A = not applicable or comparable with control baseline; N/S = no significant or discernable impact.

Table 1.4

Poverty measures for number of people (in millions) below $1.25 a day in 2005 purchasing power parities

Country	Headcount		Poverty gap$^{\alpha}$		Squared poverty gap$^{\beta}$	
	Pre-transfer	Post-transfer	Pre-transfer	Post-transfer	Pre-transfer	Post-transfer
Brazil	0.2421	0.2369	0.0980	0.0901	0.0553	0.0471
Ecuador	0.2439	0.2242	0.0703	0.0607	0.0289	0.0235
Jamaica	0.2439	0.2329	0.0659	0.0602	0.0258	0.0224
Mexico	0.2406	0.2222	0.0847	0.0683	0.0422	0.0298

Source: Fiszbein *et al.* 2009: 110.
Notes: α = depth of poverty: this provides information regarding how far off households are from the poverty line; β = poverty severity: this takes into account not only the distance separating the poor from the poverty line (the poverty gap), but also the inequality among the poor. Thus, a higher weight is placed on those households who are further away from the poverty line.

full expression of this collective form of thinking. Mead and Beem (2005), for example, observe that major welfare reforms were undertaken in the advanced liberal democracies on ideological grounds, with an appeal to free-market political philosophy. Active labour market policies (ALMP), which have a tradition in the Nordic countries, were introduced into the liberal market economies with new duties (roles) and obligations (responsibilities) impressed on welfare benefit claimants. 'Workfare' (work-for-your-welfare) regulatory policies first appeared in the USA, thereafter work-based welfare reforms spread rapidly across the Organisation for Economic Co-operation and Development (OECD) countries (Nativel 2006). Notions of reciprocal obligation became firmly embedded in the various 'welfare-to-work' and 'activation' programmes. 'Activation' policy attempts to mobilize and enforce paid labour; work not only pays better than welfare, it is claimed, but also promotes well-being for individuals and nation states. Consequently, labour market conditionality now applies to most sections of the adult working age population across the developed world (Jacobsson and Noaksson 2010).

Currently, the primary obligation on social security recipients in most 'workfare' states is to actively look for paid employment, and in order to improve their chances, claimants must undergo training to develop new skills; they must engage with the employment services on offer for their own benefit. The original research that underpinned 'welfare-to-work' in the UK was encouraging, but certainly not convincing (Vincent *et al.* 1998).[3] The new 'workfare' policy, however, was designed to solve a problem about the effects of passive, rights-and-eligibilities-based welfare that arose, it was argued, because too many citizens had been reduced to 'welfare dependency'. Supporters of 'activation' argue the new duties impressed on welfare benefit claimants are perfectly justifiable as the social contribution now owed by

citizens to society (Mead 1986). There is thus a strong ideological basis to 'activation' that sits squarely within the free-market neo-liberal paradigm. However, the research evidence supporting this great transformation of social policy has been the subject of much debate.

Fragmentary evidence

Observers complain that research has not been driving welfare reform in the developed world in the way that it should (Haynes *et al.* 2012; Roberts *et al.* 2012). Aside from health, and possibly education (Oakley 2000), there appears a real lack of substantive investment in robust policy trials and systematic reviews. Research is only now beginning to piece together the fragmentary evidence relating to 'activation' in a systematic way (e.g. Kluve 2010; Marcia *et al.* 2012). The evidence from trials to date suggests only modest returns on certain activation programmes from unknown costs; notably, those that include job search activities accompanied by sanctions for non-compliance. However, the picture is as yet far from conclusive and findings may not be generalizable from one context to another (Dorsett *et al.* 2011; Hendra *et al.* 2011). Clearly 'welfare-to-work' programmes may have some obvious benefits as Lane *et al.* (2011) observe, providing tailored help and support to people who are looking for work, for instance. However, there may be structural concerns and policy limits. The present lack of employment opportunities is one (Dean 2012), child care costs are another (Davey and Hirsch 2011).

Until recently, 'welfare-to-work' evaluations have been rather limited in scope and design, particularly when considered against the large-scale trials in LMICs. Behavioural effects in the field of 'activation' are notoriously complex, difficult to observe and a challenge to monitor, particularly over time (Paz-Fuchs 2008). Often, there are indirect social and economic effects, unobserved consequences and population effects. Much of the evidence that has been generated seems patchy at best, meaning it is often inconclusive on key questions. For example, cost-benefit calculations are largely absent. 'Welfare-to-work' schemes in the UK have not delivered results, thus offering poor value to the British taxpayer (NAO 2012a). In some areas, 'activation' jobs are estimated to cost the public purse over £200,000 per new post (NAO 2012b). Subsequently, scandals make headlines in the British press and there have been high profile resignations amid allegations of fraud, excessive pay and bonuses for those charged with helping unemployed people back into paid work (Committee of Public Accounts 2012). Public confidence in the institutions may be shaken, but there may be popular support for many of the principles at stake here (discussed in the final section). Of course, politicians and policymakers are often keen to promote their policies and reforms. Thus, apparent and favourable impacts – improvements in employment rates or reductions in state expenditures, for instance – are seized upon to build public confidence and support. However, such information is usually uncontrolled knowledge. In other words, more often than not we know very little about the counterfactual outcome that would have been achieved had the programme not been in place.

Arguments Concerning the 'Use' and 'Misuse' of Evidence and Evaluation

The review of global social policy, albeit brief, highlights a number of competing tensions and complex interrelated issues for thinking about the 'use' and 'misuse' of evidence and evaluation in policy-making that require further elaboration: these concern the nature of 'evidence'; methodological and ethical issues; public attitudes and the nature of the policy process.

Evidence

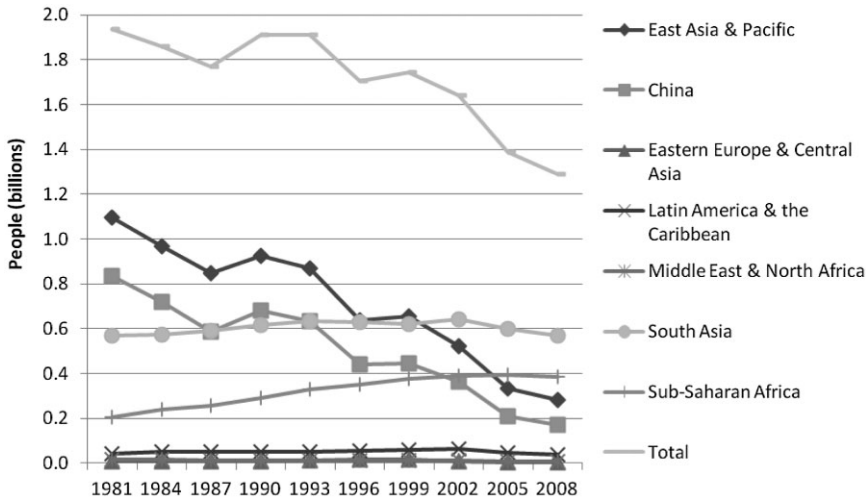
There is currently much debate over the nature and strength of the evidence emerging from CCTs. The evidence from LMICs, even for improvements on core programme objectives such as school attendance, is often unclear and context dependant, making it hard to generalize from one experiment to the next (Olken *et al.* 2012). Moreover, social interventions are often characterized by heterogeneity, delivered by different individuals operating in different social and geographical contexts. One implication of this heterogeneity is that average treatment and intervention effects, which are often the focus of attention, can be less useful than estimates of differential impacts across contexts (Plewis 2002). Increasingly, statistical models, and hierarchical models, are being used for estimating complex heterogeneity and different variances in the intervention and control groups; and this trend in modeling experimental and non-experimental social survey data is likely to continue (Byrne 2011).[4]

In the review we also find that evidence is weak or unproven in key areas of development, particularly for policies addressing child poverty (Yeates 2009) and child labour (Tabatabai 2009). Doubts are also expressed about the longer-term sustainability of programme benefits. Impacts tend to be evaluated in the short term and, as a consequence, little is known about longer-term effects. Observed benefits may be diluted over time. For example, evaluations tend to show the frailty of poverty reduction programmes (Lomelí 2009); jobs created by 'activation' programmes may not last (Kluve 2010). Structural issues extend far beyond individual agency and personal responsibility, as Wright (2012) observes. In times of high unemployment, for instance, there simply may not be enough jobs to go around (OECD 2012).

In addition, there are more fundamental, outright rejections of current neo-liberal policy prescriptions, which also appeal directly to the research evidence. 'Washington Consensus' policies, it is argued, are failing to make a real impact on chronic global poverty and deprivation (Townsend 2009; Anand *et al.* 2010). Extreme world poverty has fallen (Chen and Ravallion 2012), down from 1.94 billion people in 1981 to 1.29 billion people in 2008 (figure 1.5). However, nearly all of the poverty reduction during the period took place in China. Progress elsewhere, over nearly three decades, has been stubbornly slow, casting a long shadow of doubt over the 'evidence-base' supporting the present global policy response to poverty and deprivation (CSDH 2008). By contrast, a broad coalition of bodies under the

Figure 1.5

Estimates of chronic global poverty[α]



Source: Chen and Ravallion 2012: 5.
Note: α = number of people living below $1.25 a day in 2005 purchasing power parities.

United Nations recommends a global social protection floor to tackle the problem of chronic global poverty (ILO 2011), which would provide support to those not covered by adequate social security (four out of five people worldwide), for *unconditional* cash transfers are also effective and deliver many benefits for poor families. In Somalia, for example, the provision of cash grants to women not only improved food consumption levels and the use of health services, but also helped poor families to reduce their debts (Ali *et al.* 2005). Other countries, including India, are now beginning to trial and evaluate unconditional transfers; soon we may be able to assemble reliable comparative data to uncover the benefits of these schemes (Standing 2012).

Evaluations

Social experiments in this field are often complex in design (Samson *et al.* 2010). Problems over administration and coordination occur, including poor targeting, and the costs involved in conducting an evaluation can be substantial (Lomelí 2008). Even in well designed trials there are a range of methodological issues to consider, such as bias in the results. For example, researchers have recently detected a possible 'Hawthorne effect' in some of the trial data from Latin America (de Brauw and Hoddinott 2011). This implies that some

of the observed changes in behaviour during the course of a study may be related, at least in part, to the special conditions and monitoring systems in place during the period of investigation.

Despite the vast range of evaluations and RCTs, calculations of cost-benefit in this field are rare (St. Clair 2009). This makes it extremely difficult, if not impossible, to make clearly informed judgements about value for money, as Barrientos (2009) argues. The impact of conditionality is usually measured at the margins, by incremental gains. For example, secondary school enrolment rates for girls in Mexico's *Oportunidades* programme increased by 9 percentage points in two years, from a base of 67 per cent. Whether such modest gains are worth the costs is another matter, estimated at 2 per cent of the total social security programme budget in this case. Ideally, the overall (system) costs and benefits to individuals, families and society should be reflected in the evaluation. However, capturing the full range of information required is methodologically and technologically challenging, which itself comes at a cost. Conditionalities may, for example, impose non-trivial compliance costs on beneficiaries, which are not accounted for in their benefit levels. Conditionality imposes system costs as well as cost to individuals and families, although it is hard to know whether these costs are outweighed by the benefits, and if so, by how much. Including cost-benefit analysis in future evaluations would certainly help to inform the current debate. There is a compelling argument that *unconditional* transfers are cheaper, more efficient to administer and produce similar population benefits as previously discussed.

Rights

Critics argue that social policy is being eroded by this newer neo-liberal brand advocated by the behaviour economists (e.g. Standing 2011). The issue at stake here is less about the 'evidence-base' – as we saw above, conditionality can be effective in changing behaviour in certain circumstances. The problem, according to some observers, is that conditionality undermines people's rights, citizenship and important principles governing social solidarity. CCTs are thus divisive, in that they create and sustain social divisions within society. Poor people are treated differently from other fellow citizens and this has major implications, for their basic human rights may be violated. Several international instruments affirm that every human being has the right to social security. For example, Article 22 of the 1948 UN Universal Declaration on Human Rights (reconfirmed in 1993 and 1996) states that 'everyone, as a member of society, has the right to social security' and Article 25 states that 'everyone has the right to a standard of living adequate for their health and wellbeing'. If social security is only transferred on condition, then basic human rights are surely undermined. There are counter-arguments to this position, however, whereby exponents claim that conditionality represents an instrument of policy that, in practice, confronts the denial of basic human rights by ensuring that welfare reaches families most in need (see Lomelí 2008).

Power

Critics claim that conditionality in public policy reinforces entrenched inequalities within and between societies, and thus preserves existing social structures and power relations (e.g. Mcdonald and Marston 2005; Veit-Wilson 2009). This discourse attempts to move us on from fairly constrained debates about the 'evidence-base' of policy to much more basic concerns regarding the legitimacy of current policy programmes, and the ongoing power struggles between the dominant elites and the dominated poor. Thus, many policymakers seem to believe that poor people must be threatened with less money to incentivize them to follow predefined behaviour patterns, but this form of practice constitutes poor people as subjects engaged in a power struggle for resources, recognition and respect. Subsequently, we need to be much more critical about how 'social problems' are defined, as Bacchi (2009) argues, and the ways in which research 'evidence' is represented for policy-making.

Ethics

RCTs (as a method of evaluation) can raise profound ethical issues. RCTs involving conditionality (i.e. policy intervention) raise even more concerns because they antagonize people in two distinct camps. Critics of RCTs, for example, argue that it is unfair to experiment with the lives of poor vulnerable and disadvantaged people (Elizabeth and Larner 2009). Experiments with social security and assistance – by their very nature – often target the most vulnerable sections of society. Therefore, critics claim that conditionality undermines social inclusion or, worse still, reinforces existing prejudices in the community. Unemployed people are often stigmatized in the British media, for instance (e.g. Walters 2012). For this reason, the Roman Catholic Church in Australia principally opposes the inclusion of Indigenous families in Australia's new income management scheme (Quinlan 2010). In this programme, a range of benefits including unemployment, disability and single parenting payments, are granted on condition that recipients do not spend their money on goods such as alcohol, tobacco, pornography or gambling products (FaHCSIA 2010). The programme involves elaborate systems for monitoring compliance and controlling behaviour which, in some senses, seeks to attract a degree of social acceptability. Ethically, however, the scheme has been the subject of much controversy. In many ways, the key question turns on what counts as 'evidence', and to whom – and the extent to which – we can include notions of 'social justice' and 'human rights' in the framework of RCTs and evaluation. Ultimately, the design of welfare policy invokes our values and the values and attitudes of our fellow citizens, which may conflict and contrast in sophisticated ways.

Public attitudes

Whilst implementing welfare reforms with little or no clear evidence of their effectiveness may offend those interested in evidence-based policy-making,

and is itself unethical (Davies *et al.* 2000), we can recognize that 'evidence' comes in a variety of forms, and make room for broad conceptions of 'evidence' and ever-shifting social attitudes. Public attitudes and perceptions have an important place in the policy-making process. Standing (2008), for example, sees the CCT initiative in developing countries as a political device to legitimize cash transfers with middle-class voters and international agencies. In the developed world, Taylor-Gooby (2011) argues that welfare reform has become politically easier and more acceptable, thanks to recent changes in public attitudes. Analyzing the UK context, he observes hardening attitudes towards welfare, thus allowing the Coalition government to drive through radical reforms in an effort to cut public spending. Over half of the British public now think that welfare benefits are too high, with disincentive effects for work (Park *et al.* 2011). Reciprocity may help to legitimize the welfare state function according to some, but equally the shift can be seen to undermine the principles of social policy as social solidarity. A more mixed picture emerges across the developed world, where in the Nordic and Southern European countries, for instance, the majority of people appear to support the unconditional model of welfare, whereas in the Anglo-Saxon nations it is the conditional model that is currently in favour (Jæger 2011).

The issues involved here are notoriously complex, however, and can invoke a 'policy sciences of tyranny' (Dryzek 1989). For example, the British general public may support elements of conditionality in welfare policy, but large sections of the population who find themselves 'on condition' may object to the policy. Nearly two-thirds of lone parents in the UK, for instance, dislike or reject the conditions imposed upon them by the new 'welfare-to-work' regulatory regime: they would prefer to look after their own children themselves rather than be forced into work and have to pay for child care services (Rafferty and Wiggan 2011). Hence, the evidence seems to suggest their needs and interests, and possibly those of their children, are not respected under the new policy. Income management trials in Australia revealed strong support from welfare beneficiaries (FaHCSIA 2010): some 70–80 per cent of participants reported positive benefits for themselves and their families; six in ten participants said the scheme had helped in improving the lives of their children and just under half claimed it had done so significantly. These findings were used by policymakers to help justify the national policy, which was rolled out across Australia in 2012. Social surveys are not always value-free instruments, of course (Goerres and Prinzen 2012), but even if conditionality in policy improves levels of living for some families in ways that they approve of, it is highly questionable whether these same families would accept or consider appropriate the sanctions imposed when conditions are not met – particularly if child well-being was at stake.

The policy process

Conditionality in behavioural economics is often presented as a 'model' to be tested within the positivist tradition of social science. However, the implicit assumption of a linear relationship between research evidence and policy is

not easily supported (see Parsons 2002), especially where political ideology is inextricably tied up in the policy process, as Wiggan (2012) discovered. Work by Stubbs (2009) reveals how programme goals in this field are socially constructed. Drawing on CCT experiences in Eastern Europe, Stubbs observes how policymakers negotiate the aims and objectives of their programme, adapting trial evidence from one context (Latin America) to another (Europe). The historical, cultural and institutional connections here are tenuous at best, he argues, and families may suffer as a result of the policy translation process, as welfare is reformed. From this perspective, there is no real 'gold standard' approach; judgements about evidence and policy are required on a case-by-case basis (Cartwright 2007). Often, policy conclusions cannot simply be exported from one population or context to another.

Discussion

The world of public policy has been radically transformed in recent times, but here we have a tale of two worlds. The great transformation of welfare within the developed world is arguably a result of political thinking and ideology; today the research effort is largely focused on trying to embed complex 'workfare' systems, in an effort to make them work more effectively. In the global south, policy appears to be more driven by the evidence-based policy agenda for human development. Ultimately, this seems to be about ensuring scare funds are spent wisely to meet the MDGs, which on many levels is unobjectionable. Strengthening the evidence-base for policy decisions is therefore seen as a positive move. However, the whole 'evidence-based' approach to policy itself can be seen to be ideologically driven, as Packwood (2002) argues. In other words, CCT policies arguably support particular beliefs and values compatible with the dominant neo-liberal paradigm – promoted by the 'Washington Consensus' – that increasingly defines how people and society should function in the 21st century (Draibe and Riesco 2009).

So far, governments in the developed world have largely resisted the temptation to impose the sorts of behavioural conditions on welfare recipients seen in LMICs, although this may be changing. In Australia, for example, child benefit started to be conditioned on compliance with national healthcare requirements, while the new income management programme goes much further with behavioural conditioning. In the UK, the government is considering stronger 'workfare' measures for the long-term unemployed, who otherwise risk losing all their benefits. Social policy, it would seem, is becoming increasingly interventionist, and conditioned (e.g. Thaler and Sunsyein 2009; Behavioural Insights Team 2011). Governments are turning to behavioural economics and 'libertarian paternalism' to tackle complex 'social problems', often resulting from current lifestyles of apparent 'excess' (Le Grand 2008; HM Government 2010). In so doing, new systems of governance for monitoring and controlling our behaviour, designed to enhance social justice and human well-being, are firmly established, as the behavioural psychologist B. F. Skinner famously anticipated (Skinner 1948).

So what lessons might we draw from this review? A basic question within social policy, which is rarely addressed directly or well (the work by Saunders 2002, is exceptional in this regard), is to ask whether the ends of welfare justify their means. This review offers three different, and partly conflicting, answers to this question:

- The ends justify the means. There is no shortage of good solid evidence from robust evaluations and systematic-reviews in LMICs that show significant benefits and great strides in social progress and human flourishing (e.g. tables 1.1–1.4). Developed countries could learn much from social experimentation with social policies that include robust evaluations.
- The ends do not justify the means. The behavioural approach to tackling global social problems such as chronic poverty is failing (e.g. figure 1.5). Evaluations are often flimsy and evidence is overstated. At best, we find marginal gains for untold costs (e.g. tables 1.1–1.4).
- Maybe the ends justify the means. However, there is much uncertainty, results are often inclusive or conflicting and there are plenty of 'information gaps'. We simply do not know at present whether the ends justify the means. The lack of cost-benefit data and trials of alternative models (e.g. unconditional transfers) are major impediments in this field.

All three positions carry great weight in the social and policy sciences at present. Ultimately, then, the 'use' and 'misuse' of evidence and evaluation appear to lie in the eye of the beholder. Those advocating neo-liberal forms of policy, particularly with conditionality applied to social security or assistance, argue the 'use' of evidence and evaluation can be justified both in principle and practice. Those who argue for 'misuse' flatly reject this. The issues discussed here cannot easily be solved by appeals to the 'evidence' alone, at least how things stand, and while better evaluations may help, they are sure to stir and provoke our values and principles still further.

Acknowledgements

This research was supported by ESRC grant ES/K001353/1 'New Cultural Contradictions in Society'. I am also grateful to the Editors, reviewers, and to Jo Mulligan at Imperial College London and Ana Beaumont at Macquarie University for excellent suggestions and critical comments on an earlier draft of this chapter; the usual caveats apply.

Notes

1. To be included in a systematic review or meta-analysis, a study must usually meet strict research quality criteria relating to research design, often this usually means a randomization or a well designed before-and-after control study. A clear measure of at least one outcome is also required.
2. The International Development Coordinating Group website of the Campbell Collaboration has further information on systematic reviews completed and in-progress, http://www.campbellcollaboration.org/international_development/index.php (accessed 14 March 2013).

3. This study published in 1998 appears to be the first RCT in the field of social security in Britain.
4. A statistical model, underpinned by mathematical equations, is a formalization of relationships between variables. The model is statistical as the variables are not deterministically related but stochastically (best fit) related.

References

Ali, D., Fanta, T. and Tilleke, K. (2005), *Cash relief in a contested area: lessons from Somalia*, London: Overseas Development Institute.

Anand, S., Segal, P. and Stiglitz, J. (eds) (2010), *Debates in the Measurement of Global Poverty*, Oxford: Oxford University Press.

Bacchi, C. (2009), *Analysing Policy: What's the problem represented to be?* Frenchs Forest: Pearson Australia.

Baird, S., Chirwa, E., McIntosh, C. and Özler, B. (2009a), *Schooling, Income, and Sexual Behavior: The Design, Implementation, and Short-Term Impacts of a CCT Program for Schooling in Malawi*, Washington, DC: World Bank.

Baird, S., Chirwa, E., McIntosh, C. and Özler, B. (2009b), *The Short-Term Impacts of a Schooling Conditional Cash Transfer Program on the Sexual Behavior of Young Women*, Washington, DC: World Bank.

Banerjee, A. V. and Duflo, E. (2011), *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, New York, NY: Public Affairs.

Barrientos, A. (2009), Understanding Conditions in Income Transfer Programmes, *Global Social Policy*, 9, 2: 165–7.

Behavioural Insights Team (2011), *Annual update 2010–11*, London: Cabinet Office.

Byrne, D. (2011), *Applying Social Science: The Role of Social Research in Politics, Policy and Practice*, Bristol: Policy Press.

Cartwright, N. (2007), Are RCTs the Gold Standard? *BioSocieties*, 2, 1: 11–20.

Chen, S. and Ravallion, M. (2012), *An update to the World Bank's estimates of consumption poverty in the developing world*, http://siteresources.worldbank.org/INTPOVCALNET/Resources/Global_Poverty_Update_2012_02-29-12.pdf (accessed 14 March 2013).

Committee of Public Accounts (2012), *Department for Work and Pensions: The Introduction of the Work Programme*, London: The Stationery Office.

Commission on Social Determinants of Health (CSDH) (2008), *Closing the gap in a generation: health equity through action on the social determinants of health*, Geneva: World Health Organization.

Davey, C. and Hirsch, D. (2011), Childcare in Universal Credit: will work pay for single parents? *Journal of Poverty and Social Justice*, 19, 3: 289–94.

Davies, B., Nutley, S. M. and Smith, P. C. (eds) (2000), *What works? Evidence-based policy and practice in public services*, Bristol: Policy Press.

de Brauw, A. and Hoddinott, J. (2011), Must conditional cash transfer programs be conditioned to be effective? The impact of conditioning transfers on school enrollment in Mexico, *Journal of Development Economics*, 96, 2: 359–70.

Dean, H. (2010), *Understanding Human Need*, Bristol: Policy Press.

Dean, H. (2012), The Ethical Deficit of the United Kingdom's Proposed Universal Credit: Pimping the Precariat? *The Political Quarterly*, 83, 2: 353–9.

Department for International Development (DFID) (2006), *Using social transfers to improve human development*, London: DFID.

Department for International Development (DFID) (2011a), *Cash Transfers: Evidence Paper*, London: DFID.

Department for International Development (DFID) (2011b), *Systematic Reviews in International Development: An Initiative to Strengthen Evidence-Informed Policy Making*, http://

www.dfid.gov.uk/What-we-do/Research-and-evidence/case-studies/research-case-studies/2011/Systematic-Reviews-Background/ (accessed 14 March 2013).
Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) (2010), *Evaluation of the Child Protection Scheme of Income Management and Voluntary Income Management Measures in Western Australia*, Barton: FaHCSIA.
Dorsett, R., Rolfe, H. and George, A. (2011), *The Jobseeker's Allowance Skills Conditionality Pilot*, London: Department for Work and Pensions.
Draibe, S. and Riesco, M. (2009), Social Policy and Development in Latin America: The Long View, *Social Policy & Administration*, 43, 4: 328–46.
Dryzek, J. S. (1989), Policy Sciences of Democracy, *Polity*, 22, 1: 97–118.
Elizabeth, V. and Larner, W. (2009), Racializing the 'Social Development' State: Investing in Children in Aotearoa/New Zealand, *Social Politics*, 16, 1: 132–58.
Fiszbein, A., Schady, N., Ferreira, F. H. G., Grosh, M., Kelleher, N., Olinto, P. and Skoufias, E. (2009), *Conditional Cash Transfers: Reducing Present and Future Poverty*, Washington, DC: World Bank.
Gaarder, M. M., Glassman, A. and Todd, J. E. (2010), Conditional cash transfers and health: unpacking the causal chain, *Journal of Development Effectiveness*, 2, 1: 6–50.
Gaia, E. (2010), Mi Familia Progresa: Change and Continuity in Guatemala's Social Policy. In I. Greener, C. Holden and M. Kilkey (eds), *Social Policy Review 22: Analysis and Debate in Social Policy*, Bristol: Policy Press, pp. 199–223.
Goerres, A. and Prinzen, K. (2012), Can We Improve the Measurement of Attitudes Towards the Welfare State? A Constructive Critique of Survey Instruments with Evidence from Focus Groups, *Social Indicators Research*, 109, 3: 515–34.
Hagen-Zanker, J., Duvendack, M., Mallett, R., Slater, R., Carpenter, S. and Tromme, M. (2012), *Making systematic reviews work for international development research*, London: Overseas Development Institute.
Haynes, L., Service, O., Goldacre, B. and Torgerson, D. (2012), *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*, London: Behavioural Insights Team, Cabinet Office.
Hendra, R., Ray, K., Vegeris, S., Hevenstone, D. and Hudson, M. (2011), *Employment Retention and Advancement (ERA) demonstration: Delivery, take-up and outcomes of in-work training support for lone parents*, London: Department for Work and Pensions.
Hickey, S., R. Sabates-Wheeler, G. Guenther and I. Macauslan (2009), *Promoting Social Transfers: DFID the Politics of Influencing*, London: Department for International Development.
HM Government (2010), *Healthy lives, healthy people: our strategy for public health in England*, Norwich: The Stationery Office.
International Labour Office (ILO) (2011), *Social Protection Floor for a Fair and Inclusive Globalization*, Geneva: ILO.
Jacobsson, K. and Noaksson, N. (2010), From deregulation to flexicurity: The make-over of the OECD Jobs Strategy. In K. Martens and A. P. Jakob (eds), *Mechanisms of OECD Governance. International Incentives for National Policy Making?* Oxford: Oxford University Press, pp. 119–38.
Jæger, M. M. (2011), Do we all (dis)like the same welfare state? Configurations of public support for the welfare state in comparative perspective. In J. Kvist, J. Fritzell, B. Hvinden and O. Kangas (eds), *Changing Social Equality: The Nordic Welfare Model in the 21st Century*, Bristol: Policy Press, pp. 45–68.
Kluve, J. (2010), The effectiveness of European active labor market programs, *Labour Economics*, 17, 6: 904–18.

Lagarde, M., Haines, A. and Palmer, N. (2007), Conditional Cash Transfers for Improving Uptake of Health Interventions in Low- and Middle-Income Countries: A Systematic Review, *JAMA*, 298, 16: 1900–10.
Lane, P., Casebourne, J., Lanceley, L. and Davies, M. (2011), *Lone Parent Obligations: work, childcare and the Jobseeker's Allowance regime*, London: Department for Work and Pensions.
Le Grand, J. (2008), The giants of excess: a challenge to the nation's health, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171, 4: 843–56.
Lomelí, E. V. (2008), Conditional Cash Transfers as Social Policy in Latin America: An Assessment of their Contributions and Limitations, *Annual Review of Sociology*, 34, 1: 475–99.
Lomelí, E. V. (2009), Conditional Cash Transfer Programs: Achievements and Illusions, *Global Social Policy*, 9, 2: 167–71.
Marcia, G., Hilary, T., Kasia, B., Clare, B., Candida, F. and Lyndal, B. (2012), Welfare to work interventions and their effects on health and well-being of lone parents and their children, *Cochrane Database of Systematic Reviews*, 5.
Mcdonald, C. and Marston, G. (2005), Workfare as welfare: governing unemployment in the advanced liberal state, *Critical Social Policy*, 25, 3: 374–401.
Mead, L. M. (1986), *Beyond entitlement: The social obligations of citizenship*, New York, NY: Free Press.
Mead, L. M. and Beem, C. (eds) (2005), *Welfare Reform and Political Theory*, New York, NY: Russell Sage Foundation.
National Audit Office (NAO) (2012a), *The introduction of the Work Programme*, London: The Stationery Office.
National Audit Office (NAO) (2012b), *The Regional Growth Fund*, London: The Stationery Office.
Natível, C. (2006), From welfare-to-work to welfare-in-work: Concepts and policies. In Organisation for Economic Co-operation and Development (OECD) (ed.), *Skills upgrading: new policy perspectives*, Paris: OECD, pp. 33–83.
Oakley, A. (2000), A Historical Perspective on the Use of Randomized Trials in Social Science Settings, *Crime & Delinquency*, 46, 3: 315–29.
Organisation for Economic Co-operation and Development (OECD) (2012), *Economic Outlook 2012*, Paris: OECD.
Organisation for Economic Co-operation and Development (OECD) (n.d.), *Outline of principles of impact evaluation*, http://www.oecd.org/dataoecd/46/16/37671602.pdf (accessed 14 March 2013).
Olken, B. A., Onishi, J. and Wong, S. (2012), *Should Aid Reward Performance? Evidence from a field experiment on health and education in Indonesia*, Cambridge, MA: Poverty Action Lab, Massachusetts Institute of Technology. http://economics.mit.edu/files/6923 (accessed 14 March 2013).
Packwood, A. (2002), Evidence-based Policy: Rhetoric and Reality, *Social Policy and Society*, 1, 3: 267–72.
Park, A., Clery, E., Curtice, J., Phillips, M. and Utting, D. (eds) (2011), *British Social Attitudes 28* (2011–2012 edn), London: Sage.
Parsons, W. (2002), From Muddling Through to Muddling Up – Evidence Based Policy Making and the Modernisation of British Government, *Public Policy and Administration*, 17, 3: 43–60.
Paz-Fuchs, A. (2008), *Welfare to Work: Conditional Rights in Social Policy*, Oxford: Oxford University Press.
Plewis, I. (2002), Modelling impact heterogeneity, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165, 1: 31–8.

Quinlan, F. (2010), *A stronger, fairer safety net? Income management and the future of Australia's income support system*, Curtin: Catholic Social Services Australia.
Rafferty, A. and Wiggan, J. (2011), Choice and Welfare Reform: Lone Parents' Decision Making around Paid Work and Family Life, *Journal of Social Policy*, 40, 2: 275–93.
Roberts, H., Petticrew, M., Liabo, K. and Macintyre, S. (2012), 'The Anglo-Saxon disease': a pilot study of the barriers to and facilitators of the use of randomised controlled trials of social programmes in an international context, *Journal of Epidemiology and Community Health*, 66, 11: 1025–9.
Samson, M., van Niekerk, I. and Quene, K. M. (2010), *Designing and Implementing Social Transfer Programmes* (2nd edn), Cape Town: EPRI Press.
Saunders, P. (2002), *The Ends and Means of Welfare: Coping with Economic and Social Change in Australia*, Cambridge: Cambridge University Press.
Sibbald, B. and Roland, M. (1998), Understanding controlled trials: Why are randomised controlled trials important? *BMJ*, 316, 7126: 201.
Skinner, B. F. (1948), *Waldon Two*, New York, NY: Macmillan.
St. Clair, A. L. (2009), Conditional Cash Transfers: The Need for an Integrated and Historical Perspective, *Global Social Policy*, 9, 2: 177–9.
Standing, G. (2008), How cash transfers promote the case for basic income, *Basic Income Studies* 3, 1: 1–30.
Standing, G. (2011), Behavioural conditionality: why the nudges must be stopped an opinion piece, *Journal of Poverty and Social Justice*, 19, 1: 27–38.
Standing, G. (2012), *Cash Transfers: A Review of the Issues in India*, New Delhi: UNICEF India.
Stubbs, P. (2009), CCT as Policy Translation, *Global Social Policy*, 9, 2: 174–7.
Tabatabai, H. (2009), Conditional Cash Transfers and Child Labour, *Global Social Policy*, 9, 2: 179–82.
Taylor-Gooby, P. (2011), Security, equality and opportunity: attitudes and the sustainability of social protection, *Journal of European Social Policy*, 21, 2: 150–63.
Thaler, R. and Sunsyein, C. (2009), *Nudge: Improving decisions about health, wealth and happiness*, New Haven, CT: Yale University Press.
Townsend, P. (ed.) (2009), *Building Decent Societies: Rethinking the Role of Social Security in Development*, Houndmills: Palgrave Macmillan.
Veit-Wilson, J. (2009), Who Sets the Conditions? Conditionality, Reciprocity, Human Rights and Inclusion in Society, *Global Social Policy*, 9, 2: 171–4.
Vincent, J., Walker, R., Dobson, B., Stafford, B., Barnes, M. and Bottomley, D. (1998), *Lone Parent Caseworker Pilots evaluation: Final Report*, Loughborough: Centre for Research in Social Policy.
Walters, S. (2012), Cameron: My war on welfare culture, *The Mail on Sunday*, 24 June, pp. 1, 8–9.
White, H. and Waddington, H. (2012), Why do we care about evidence synthesis? An introduction to the special issue on systematic reviews, *Journal of Development Effectiveness*, 4, 3: 351–8.
World Health Organization (WHO) (2002), *The World Health Report 2002: Reducing Risks, Promoting Healthy Life*, Geneva: WHO.
World Health Organization (WHO) (2011), *The World Health Report 2012: No Health Without Research*, Geneva: WHO.
Wiggan, J. (2012), Telling stories of 21st century welfare: The UK Coalition government and the neo-liberal discourse of worklessness and dependency, *Critical Social Policy*, 32, 3: 383–405.
Wright, S. (2012), Welfare-to-work, Agency and Personal Responsibility, *Journal of Social Policy*, 41, 2: 309–28.

Yeates, N. (2009), Global Social Policy Forum, *Global Social Policy*, 9, 2: 163–5.
Yoong, J., Rabinovich, L. and Diepeveen, S. (2012), *The impact of economic resource transfers to women versus men: A systematic review*, London: EPPI-Centre, Institute of Education, University of London.
Young, K., Ashby, D., Boaz, A. and Grayson, L. (2002), Social Science and the Evidence-based Policy Movement, *Social Policy and Society*, 1, 3: 215–24.

2

Understanding the Influence of Evidence in Public Health Policy: What Can We Learn from the 'Tobacco Wars'?

K. E. Smith

Introduction

From New Labour's commitments to evidence-based policy (Cabinet Office 1999) to current efforts to promote research 'impact', the past 15 years have witnessed a growing belief that research should inform UK policy decisions. Within public health, this emphasis has been reinforced by links to evidence-based medicine and support from the World Health Organization (WHO 1998). Numerous studies now explore the relationship between research and policy in this field (e.g. Exworthy *et al.* 2003; Whitehead *et al.* 2004; Smith 2007) but, despite all this activity, most studies conclude that research evidence has played a limited role in health policy, or at least *more limited* than official commitments to 'evidence-based policy' implied (e.g. Capability Reviews Team 2007; Exworthy *et al.* 2003; Smith 2007).

One area of public health policy which is sometimes depicted as confounding this trend, at least in recent times, is tobacco control (Whitehead *et al.* 2004; Warner and Mendez 2010). Evidence concerning the health damaging consequences of tobacco began to be taken seriously in 1947 (see Berridge 2006, 2007) and, as evidence began to accumulate, calls for policy interventions grew louder (Berridge 2006, 2007). In the UK, Action on Smoking and Health (ASH) was established in 1971 (Berridge and Loughlin 2005) and subsequently played a high-profile part in advocacy efforts to encourage policy interventions to address what came to be known as the 'tobacco epidemic' (Lopez *et al.* 1994). A series of UK policy efforts to reduce tobacco use and/or the harms associated with smoking ensued, including health warnings on packs, restrictions, then bans, on tobacco advertising and, more recently, bans on smoking in indoor public places (Cairney 2007). By 2012, the UK was identified as having the most advanced tobacco control policies in Europe (Joossens and Raw 2011) with further interventions being implemented (e.g. product display restrictions) and considered (e.g. standardized, 'plain' packaging).

All this has been achieved despite the well-resourced efforts of the tobacco industry (Cairney 2007). Understandably, therefore, for many in

Evidence and Evaluation in Social Policy, First Edition. Edited by Ian Greener and Bent Greve. Book compilation

public health, tobacco control represents an example of the potentially positive influence of public health evidence (Whitehead *et al.* 2004; Warner and Mendez 2010; Proctor 2012). As such, tobacco is increasingly being positioned as a case study from which other areas of public policy, including food and alcohol, might learn (e.g. Freudenberg 2005; Brownell and Warner 2009).

This chapter builds on a small but growing number of studies that have sought to employ theoretical frameworks concerning policy change to illuminate deeper and more complex aspects of the 'tobacco wars' (e.g. Farquharson 2003; Berridge 2006; Breton *et al.* 2006; Givel 2006; Cairney 2007; Princen 2007; Larsen 2008; Young *et al.* 2012). It differs from most others by focusing specifically on the role of evidence and ideas in the decade-long debates about tobacco. It begins by assessing claims that tobacco control represents a (rare) example of evidence-based public health policy, before moving on to consider three further potential frameworks, each of which provides additional insights. It draws largely on the following sources (although additional texts are referred to in relation to specific points):

1. historical overviews of the 'tobacco wars', notably a series of UK-focused publications by the public health historian, Virginia Berridge (1999, 2003, 2006, 2007; Berridge and Loughlin 2005);
2. a 2012 20th-anniversary edition of the journal *Tobacco Control*; and
3. a systematic review of evidence relating to tobacco industry strategies to influence tobacco tax policies (Smith *et al.* 2012), the most effective policy lever for reducing tobacco consumption (IARC 2011).

The concluding discussion briefly summarizes the analysis and considers what insights this case study offers those interested in studying or improving the role of evidence in policy more broadly.

Four Ways of Understanding the UK's 'Tobacco Wars'

Rational evidence-focused approaches

Starting with the simplest approach, tobacco-related research might be understood as a rare example of a 'knowledge-driven model' of the relationship between research and policy, in which knowledge (derived from scientific research) helped identify a significant problem (the health harms caused by tobacco, first to smokers themselves and, more recently, to passive smokers) and then helped policymakers decide how to respond (see Weiss 1979 for a more detailed account of this way of conceiving the relationship between research and policy). One immediately obvious flaw with this approach is, as Larsen (2008) highlights, the failure to account for the significant delay between official recognition of the health harms associated with tobacco use and passive smoking (see Berridge 1999, 2006), and the policy interventions intended to reduce these harms. Some tobacco control interventions were, of course, put in place in the 1960s–1980s but, aside from tax increases, they were often voluntary and as Cairney (2007) argues, in legislative terms, UK tobacco

policy has been characterized by periods of significant stability, followed by rapid change, notably from the mid-1990s onwards. Comprehensive bans on tobacco advertising, for example, were only implemented in the UK in 2002 (Neuman *et al.* 2002) and bans on smoking in indoor public places were only introduced between 2004 and 2006 (Cairney 2007). To understand why there was such a long delay between the emergence of evidence about the health harms of tobacco and significant policy intervention to reduce those harms, other accounts of policy-making and the potential role of evidence within this, are therefore required.

One potential explanation comes from Caplan's (1979) thesis that 'cultural' and institutional divides separate researchers and policymakers, limiting the accessibility of evidence within policy. Many contemporary assessments certainly support the idea that communicative and institutional 'gaps' between researchers and policymakers plague public health (e.g. Hunter 2009). Yet, in examining British tobacco debates in the 1950s and 1960s, Berridge (Berridge and Loughlin 2005; Berridge 2006) claims policymakers largely understood the main messages of research on the health harms of tobacco and that resistance to taking policy action was only partly related to their assessment of this evidence. Indeed, Berridge points out that fundamental disagreements about tobacco existed *within* both the research and the policy communities at this time. For example, Ronald Fisher, a statistician who played a key role in developing randomized-controlled trial methodologies, was critical of researchers' call for action to discourage smoking, in part because he felt that correlation should not be taken as causation but also, Berridge (2006) argues, because policies to restrict smoking conflicted with his own libertarian views. Meanwhile, Berridge (2006) notes that divergent views about the need for policy action on tobacco were evident within both of the UK's main political parties, which she argues reflected differences in personal values, family backgrounds and election strategies as much as interpretations of the available evidence. Overall, focusing on differences between 'policymakers' and 'researchers' does not seem particularly helpful in explaining the delayed policy response to research concerning the harmful effects of smoking as it sidelines the role of actors' values and judgements (Cairney 2007; McQueen 2010). It also obscures the potentially important role of actors who do not fit neatly into the two 'communities' of 'researchers' and 'policymakers' (e.g. journalists, corporate lobbyists, health advocacy groups and government researchers).

Another possible explanation could be policy resistance to change. Historical institutionalism, for example, posits that policy tends to be resistance to change and that policy outcomes can only be understood by considering the historical and institutional context in which decisions are made (Béland 2005; Schmidt 2010). For example, to understand UK decisions about tobacco taxation, it is necessary to acknowledge that the Treasury (not the Department of Health) has primary responsibility, meaning the issue tends to be viewed through a revenue-, rather than health-, focused lens. A recent international review of evidence suggests tobacco tax increases usually lead to increased government revenue (IARC 2011) but such evidence is relatively recent and the tobacco industry has regularly argued that tax increases reduce revenue

(Smith *et al.* 2012). Back in the 1950s, Berridge (2006) notes that the Treasury's focus on tobacco tax revenue prompted significant anxiety about measures to reduce tobacco use. Attending to the institutional context in which evidence is interpreted and policy decisions reached certainly seems important. However, because historical institutionalism struggles to explain policy change (Schmidt 2010), it does little to help explain why tobacco tax increases were subsequently employed in the 1970s (Cairney 2007).

In sum, it does not seem easy to explain the significant time lag between the emergence of evidence about the harms caused by tobacco and subsequent policy interventions. A further fundamental problem with rational, linear conceptions of the relationship between evidence and policy in the 'tobacco wars' is that they do not distinguish between the multiple different kinds of evidence that might play a role in policy change. Yet much of the early (epidemiological and medical) evidence about the health harms of tobacco provided policymakers with little, if any, guidance as to what they might do to tackle this problem, let alone what might be most effective (or cost-effective) in a UK context. In other words, much of the evidence required to achieve 'evidence-based' policies (see Killoran and Kelly 2010) has emerged only recently (and gaps in the evidence-base remain). Furthermore, as Chapman (2007) acknowledges, some of the most recent 'advances' in tobacco control, such as the ban on smoking in public places, initially went beyond the available evidence (although evidence supporting these policies did subsequently emerge, e.g. Sims *et al.* 2012). Research alone, Weiss notes, is almost never 'comprehensive enough to be the sole source of policy advice' (Weiss 1990: 98).

This is an important point both descriptively and when considering the potential lessons that the 'tobacco wars' offers those interested in improving the use of evidence in policy. Public health researchers have often been strong advocates of the need for policy to be evidence-based (e.g. Macintyre 2011). Yet tobacco industry arguments against new tobacco control proposals often focus on the limits of the evidence. This is evident, for example, in recent tobacco company challenges to proposals for plain (unbranded) cigarette packs in Australia and the UK (see Jones 2012). Indeed, the world's largest two transnational tobacco companies have both been involved in campaigns for promoting the need for policies to be based on 'sound science', risk assessments and cost-benefit analyses, with the intention of using limitations in the available evidence to prevent (or at least delay) public health policies (e.g. Ong and Glantz 2001; Smith *et al.* 2010). In other words, an overly strong emphasis on the evidence in policy may unintentionally constrain public health policy innovation.

Overall, then, whilst evidence has clearly played an important role in promoting the need for tobacco control policies, it seems difficult to conclude that the 'tobacco wars' represent an example of a rational, linear relationship between evidence and policy. At best, the UK's recent strong performance in tobacco control 'league tables' (Joossens and Raw 2011) reflects decisions by the UK and devolved governments to pursue policy interventions which, in the context of varying levels and types of evidence, have attained sufficient policy support.

Value-orientated approaches (including the 'advocacy coalition framework')

Other authors have employed value-orientated approaches to studying the 'tobacco wars' (Farquharson 2003; Princen 2007). Most simply, some authors employ what Berridge (2006) has termed a 'heroes and villains' framework, depicting tobacco control advocates as David-like heroes who have achieved successes in some countries despite the Goliath-like tobacco industry's vast resources and willingness to employ devious and deceitful tactics. This kind of narrative is, for example, evident in some of the overviews presented in the anniversary issue of *Tobacco Control* (e.g. Proctor 2012). It is a potentially seductive framing which has a great deal in common with the notion that the 'tobacco wars' represent the ultimate victory of public health evidence over corporate interests, though it emphasizes public health *values* as much as evidence.

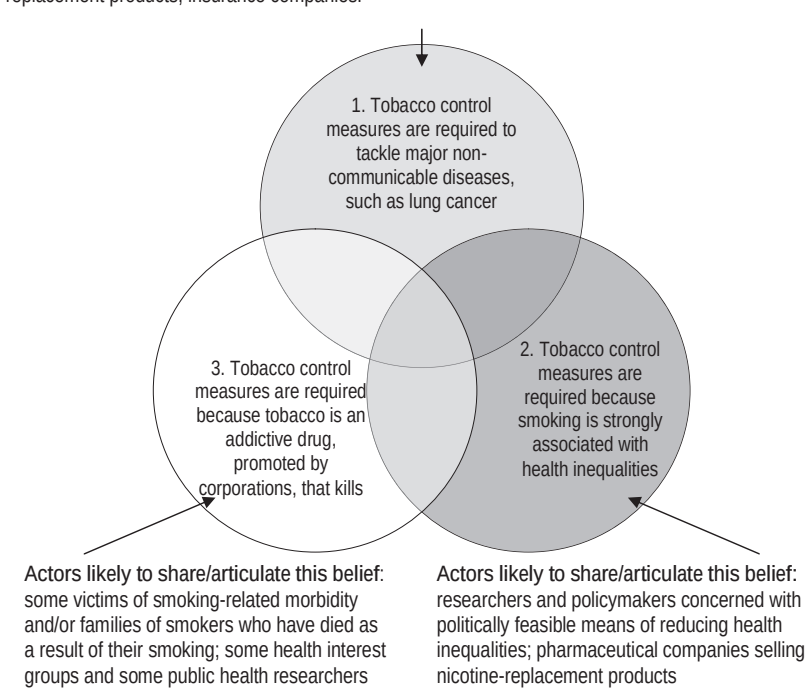
A somewhat more sophisticated, value-focused approach to understanding the 'tobacco wars' is provided in accounts employing Sabatier and Jenkins-Smith's (1993, 1999) 'advocacy coalition framework' (ACF) (e.g. Farquharson 2003; Givel 2006; Princen 2007). The ACF posits that networks of diverse actors (potentially including policymakers, researchers, think tanks, journalists, interest groups and others) compete to influence policy for particular issues (or 'policy subsystems', to use the language of ACF). It suggests that these competing networks tend to remain relatively stable because they form around 'policy core' beliefs which, in turn, reflect 'deep core' beliefs (deeply held ontological and normative beliefs). For example, beliefs concerning the balance between individuals' right to freedom versus social equality represent 'deep core' beliefs. These inform 'policy core' beliefs which include, for example, beliefs about the appropriate division of authority between governments and markets. Lastly, secondary policy beliefs concern beliefs about the optimum policy instruments to achieve agreed policy goals (e.g. whether to employ targets for smoking cessation). Changes in 'deep core' beliefs are extremely rare, being 'akin to religious conversion', whilst changes in 'policy core beliefs' are deemed only marginally more likely. In other words, the ACF foregrounds the importance of shared ways of viewing the world (over political and economic interests or evidence). When particular advocacy coalitions dominate, it is anticipated that policies will reflect their core beliefs and will remain stable. However, significant policy change may occur when a competing advocacy coalition's ideas are perceived to be so successful that some actors switch coalitions, shifting the balance of power in relation to the 'core ideas' driving policy. From this perspective, evidence is only likely to influence policy if it fits with the core beliefs of the dominant coalition.

The tobacco policy subsystem could be interpreted as being dominated by two clear advocacy coalitions:

1. proponents of stricter tobacco control measures (e.g. public health researchers, health advocacy groups and health policymakers); and
2. opponents of stricter tobacco control measures (e.g. tobacco industry and related interests, smokers rights' groups and policymakers responsible for business and trade interests).

Figure 2.1

The potentially divergent beliefs within coalitions favouring stronger tobacco control



The ACF has already been successfully applied to international tobacco control policy development (Farquharson 2003; Princen 2007), as well as to tobacco tax debates in Canada and the USA (Breton *et al.* 2006; Givel 2006). However, as Cairney (2007) points out, the ACF tends to attribute policy change of any magnitude to external shocks and has little to say about how or why coalitions lose or gain dominance over time or to the potential role of evidence within this.

This might be partly because in emphasizing the importance of core and policy beliefs, the ACF assumes a relatively high degree of coherence exists within opposing coalitions. In reality, tobacco debates in the UK have often been informed by temporary alliances between actors whose interests and/or beliefs overlapped for specific policy proposals (Cairney 2007). Indeed, as figure 2.1 outlines, the coalition supporting tobacco control interventions might be better understood as a convergence of overlapping principles and interests.

For many tobacco control interventions, including restrictions on marketing and bans on smoking in public places, the interests and beliefs outlined in figure 2.1 overlap sufficiently for these groups to be conceptualized as a unified coalition. At times, the existence of different interests within the coalition may benefit tobacco control (e.g. by enabling tobacco control measures to be promoted as means of reducing health inequalities, as discussed further in *Ideational and 'enlightenment' approaches*, below). However, for other tobacco control policies, this apparently cohesive coalition begins to unravel. For example, whilst tobacco control advocates widely support efforts to denormalize smoking, a leading health inequalities researcher has raised concerns about the negative impacts that stigmatization can have on poorer communities (Graham 2012).

Proposals to further increase tobacco taxes (which are already high in the UK) are another example of the differences depicted within figure 2.1. Poorer groups are more price sensitive and therefore more likely to quit, or reduce their consumption, as a result of tax-induced price increases, which is why IARC (2011) argues that tobacco tax increases help reduce inequalities in smoking. However, tobacco is an addictive habit usually taken up in childhood (Advisory Group of the Royal College of Physicians 2000) which means, even though most smokers want to quit (Robinson and Bugler 2008), many will be unable to and poorer smokers can find it particularly difficult to quit (Stead *et al.* 2001). Hence, despite price increases, some poor smokers will continue to smoke and will spend more of their (relatively lower) incomes on tobacco, leaving less available for other important living costs. This is why, in direct contrast to the IARC review, a UK government-commissioned review of approaches to health inequalities (led by health inequalities academics) judged further tobacco tax increases should *not* be supported (Marmot 2010). Such a contrast highlights the importance of interpretation and judgement in assessing the policy implications of evidence (Cairney 2007; McQueen 2010).

Proposals for harm reduction are yet another example of some of the divisions depicted in figure 2.1, although this time the differences are primarily between groups 1 and 3. For actors primarily concerned with reducing the health impacts of non-communicable diseases, harm-reduction measures such as legalizing 'snus' (a form of oral tobacco generally deemed less harmful than cigarettes) seems 'a promising public health policy' (Gartner *et al.* 2007). The same is true of long-term nicotine replacement therapy, which is strongly supported by pharmaceutical interests (e.g. Pfizer Ltd 2008). Yet, from the perspective of actors concerned with the health threats posed by business interests marketing addictive products, such harm reduction approaches can seem 'heretical' (Gartner *et al.* 2007). In addition, conflicting beliefs exist about whether harm reduction approaches complement or compromise risk elimination strategies (Gartner *et al.* 2007).

Differences are also evident in the coalition representing tobacco interests. This is perhaps most obvious with regards tobacco taxation and pricing, which are becoming increasingly important issues for tobacco companies in the UK (Gilmore 2012). For example, whilst virtually all tobacco interests tend prefer low tobacco taxes, and often lobby collectively on this issue (Smith *et al.* 2012), evidence suggests transnational tobacco companies may support tax

increases if, for example, they believe this will help them achieve changes to tax structures which competitively favour their brands (Shirane *et al.* 2012). Indeed, when it comes to tax structures the interests of the world's largest transnational tobacco company, Philip Morris, may have more in common with tobacco control advocates than with other tobacco companies, as both favour 'specific' taxation (IARC 2011; Smith *et al.* 2012). 'Specific' taxes involve a set monetary value being uniformly applied to packs of cigarettes (regardless of pack price), whereas *ad valorem* taxes are calculated as a percentage of product price (meaning a higher tax applies to more expensive cigarettes). The former, therefore, function to reduce price differences between cheaper and more expensive brands, to the benefit of Philip Morris' more expensive brand portfolio *and* to tobacco control advocates concerned about the potential for smokers to 'down-trade'.

These are merely illustrative examples of some of the complexities involved in debates about tobacco-related policies and interventions but they serve to highlight the limitations of approaches which frame tobacco control as a battle between two, clearly opposed, coalitions formed around coherent values. Even figure 2.1 is a simplified depiction of the variety of beliefs and interests that have been involved in promoting and resisting various tobacco-related proposals. Hence, whilst value-based, network approaches provide a better means of understanding the 'tobacco wars' than rational, evidence-based conceptualisations, they do not necessarily do enough to explain the varying interests and beliefs of actors within coalitions. This is important both because it is impacts on actors' interpretations of the policy consequences of the available evidence and because it is likely to inform the varying fortunes of coalitions over time.

Broader, network-based approaches (including Actor-Network Theory)

One of the particularly useful features of the ACF is its attention to diverse groups of actors and this is also a feature of the broader literature on 'policy networks', which ranges from 'iron triangles' involving stable relationships between politicians, interest groups and career civil servants (Overman and Don 1986) to larger, more fluid 'issue networks' (Heclo 1978). Marsh and Rhodes (1992) suggest that particular policy issues tend to be characterized by the existence of *either* an 'issue network' or a 'policy community'. However, Read (1992) argues that, during the 1980s, tobacco in the UK was characterized by the simultaneous existence of an 'issue network' (consisting of tobacco industry representatives and tobacco control lobbyists) and a 'policy community' (consisting of policymakers and tobacco interests but excluding tobacco control advocates). This highlights the potential limitations of focusing on particular kinds of policy networks and, given the breadth of the literature, this perhaps limits the utility of 'policy networks' as a concept. Moreover, like the ACF, the 'policy networks' literature generally offers few insights into the potential role of evidence and ideas in achieving policy change.

Alternative network-based approaches, developed in science studies, have also recently been applied to analyzing tobacco debates (Young *et al.* 2012).

Such theories, notably Actor-Network Theory (ANT), emphasize the importance of studying the processes involved in undertaking research and constructing, as well as disseminating, knowledge claims (Knorr-Cetina 1981; Latour and Woolgar 1986). ANT also places significant emphasis on the translation (as opposed to the transfer) of knowledge-claims (Latour 2005), encouraging analysts to trace how ideas change as they move between actors. It posits that the appearance of some actors as singular, discrete bodies (e.g. 'the government', 'the public health community' or 'the tobacco industry') is actually the effect of diverse underlying networks of actors which only become visible when they fail or when they are carefully uncovered through detailed anthropological observations (Latour 2005). Perhaps the most radical aspect of ANT is that the term 'actor' is extended to include non-humans, such as documents and technologies. As such, ANT usefully calls attention to the construction and enrolment of evidence within tobacco debates, and the potential for varying interpretations of knowledge claims, although unlike the ACF, it has little to say about the role of ethics or values.

Moreover, an assumption within ANT that the 'macro' is actually no different from the 'micro' (Law 1992) has led to criticisms that ANT may be politically disabling and uncritical, focusing analysts' attention on 'how' networks form and are performed, at the expense of considering why these networks are being produced and maintained (Bakker and Bridge 2006). This criticism is compounded by the fact that ANT is often promoted as a methodological approach requiring detailed anthropological research (Latour 2005) which tends, implicitly, to restrict the focus of research to small networks (or small parts of larger networks). This limits the possibility of applying ANT to an assessment of the UK's 'tobacco wars'.

Ideational and 'enlightenment' approaches

A growing acknowledgement of the complexity involved in evidence-translation (e.g. Sanderson 2006; Smith and Joyce 2012) has led some analysts to focus on *ideas* (rather than evidence) as the entity that moves between research and policy (Smith 2007), or across geographical locations (Stone 2004). Mirroring this shift, there has been a burgeoning interest in ideas within theories of policy change (Béland 2005; Schmidt 2010). Focusing on 'ideas' not only acknowledges the potential for translation, rather than transfer, thereby incorporating a key aspect of ANT, it can also be used to capture some of the interactions between politics, ethics, values and evidence (Sanderson 2006). However, the concept of 'ideas' is poorly defined (Blyth 1997), having been used to refer to ideologies, frames, norms, 'paradigms', explanatory theories *and* specific policy proposals. Further, because ideas 'do not leave much of a trail when they shift' (Hall 1993), it can be difficult to assess whether what appears to be the translation of a particular idea is merely another idea with similar characteristics (Smith 2007). Nevertheless, focusing on ideas helps emphasize that even evidence-informed messages can be continually translated as they become intertwined with politics, ethics and values (Sanderson 2006; Smith and Joyce 2012).

An analysis of the 'tobacco wars' focusing on *ideologies* might look very similar to the value-based approaches considered in the previous section. Alternatively, employing Weiss' (1977) 'enlightenment model' of the function of research draws attention to the potential influence of evidence-informed ideas. Such ideas may involve specific policy proposals but also, as Weiss argues, could involve policy influence via a process of gradual diffusion that leads to changes in the way the public and policy actors think about particular issues. Berridge (2006) has already applied this 'enlightenment' approach to historical tobacco control debates. It potentially helps to explain why evidence about the health harms of tobacco took so long to trigger any significant policy action and why, once public and policy perceptions had shifted, multiple policy developments were enabled, including those for which evidence was limited. For, once ministers were assured of sufficient public support for tobacco control, interventions to restrict tobacco use would have appeared far more viable (see Cairney 2007 for the importance of policy perceptions of public opinion about tobacco policy interventions).

However, it does not seem sufficient to attribute shifts in public and policy opinion to the gradual diffusion of evidence about the health harms of tobacco without also considering the evidence and ideas that tobacco interests constructed and employed in this period. Here, the concept of 'framing', which represents another way of thinking about the role of ideas in policy, seems more useful. This involves assessing the frames (or narratives) being used to portray particular issues (Scheufele and Tewksbury 2007). Policy frames can inform beliefs and ideas about particular issues, limiting how actors perceive potential policy options and, relatedly, informing the positions adopted by networks/coalitions. As such, they have been described as a 'weapon of advocacy' (Weiss 1989). Various authors have examined the 'frames' employed around specific tobacco-related policy developments (e.g. Larsen 2010; Weishaar *et al.* 2012) but they have not (to my knowledge) been used to assess tobacco debates more broadly (i.e. to understand what role particular framing devices might have played in the conceptual shift in public and policy opinion described above). Doing so highlights some important changes in the way tobacco control advocates have framed their arguments over time.

Initially, tobacco control advocates tended to focus exclusively on evidence relating to the health harms of tobacco (i.e. employing a health-orientated frame). In contrast, tobacco interests often employed economic and 'free personal choice' frames, helping to emphasize their economic contributions (e.g. through employment and revenue), whilst minimizing suggestions that policy interventions were needed to protect health (Cairney 2007; Warner 2000). Reflecting this, the policy audience for tobacco control advocates tended to be policymakers in the Department of Health, whilst tobacco interests focused on the Treasury and the Department of Trade and Industry, both of which tended to wield more power than the Department of Health (Cairney 2007; Read 1992). However, from the late 1970s onwards, tobacco control advocates began to develop their own economic frame, which challenged the industry one. This was an important development as an internal Philip Morris memo from 1978, outlining the basic premise of this shift, reflects:

> More industry antagonists are using an economic argument against cigarettes; i.e. cigarettes cause disease; disease requires treatment; major health coats are borne by the government; the taxpayers pay in the end. Thus, as health costs rise astronomically, the opposition becomes armed with more potent weapons. We must be prepared to counter this line of argument. (Saligman 1978)

Despite industry efforts, tobacco control advocates went on to develop persuasive claims about the economic (healthcare-related) costs of smoking (Warner 2000) and this played an important role in the US litigation cases brought against tobacco companies by various US states and health insurance companies (Warner *et al.* 1999). A 1999 World Bank report, which concluded that tobacco was economically damaging to all but a handful of tobacco-dependent agricultural economies, reinforced the credibility of tobacco control's economic frame (World Bank 1999). This kind of framing is now frequently evident in UK tobacco control advocacy material (e.g. ASH 2011).

At least two additional changes to the 'frames' employed in tobacco debates deserve consideration. First, as Cairney (2007) points out, growing evidence about the health harms caused by passive smoking enabled tobacco control advocates to challenge the industry's framing of smoking as a matter of personal choice. Second, more recently, tobacco control measures have been positioned as a means of reducing health inequalities (e.g. Gruer *et al.* 2009) as well as reducing smoking (despite the fact that some health inequalities researchers dispute this, e.g. Scott-Samuel 2009). This is important in the post-1997 UK era because reducing health inequalities was a policy priority that both the Department of Health and the Treasury were signed up to (see Cairney 2007; Smith and Hellowell 2012).

Overall, a focus on 'policy frames' supplements value-orientated, network-based and conceptual accounts of the 'tobacco wars' in the UK. It helps to demonstrate how tobacco control advocates' evidence and arguments shifted over time, in ways which appear to have helped attract greater policy and public support (or, in ACF terms, greater coalition support). The ability of tobacco control advocates to develop an economic frame seems likely to have been particularly important as this effectively re-presented tobacco control as a 'win-win' scenario for a variety of policy interests.

Concluding Discussion

Like Larsen (2008) and Berridge (2006), this chapter quickly concluded that rational, evidence-focused approaches are inadequate for understanding the fraught relationship between public health evidence about, and policy responses to, the UK's 'tobacco epidemic'. Such approaches leave key questions unanswered, notably why there was such a long delay between the emergence of evidence about tobacco related harms and significant policy action to reduce those harms. Once the multitude of different types of evidence is delineated (e.g. evidence concerning the health harms of smoking, from evidence relating to the potential costs and benefits of particular interventions, to evidence concerning public opinion), and the importance of

interpretation more widely acknowledged, focusing on the role of evidence in policy debates seems more useful. However, policy commitments to evidence-based policy in the UK (and, indeed, accounts within public health) have tended to ignore the diversity of evidence-types and the potential for differing policy interpretations, in favour of a simple, linear account in which evidence is positioned as the basis of policy change. The chapter argues that such linear, rational depictions are not only descriptively inaccurate but that commitments informed by such approaches can be used to restrict policy innovation.

Building on Farquharson (2003) and others (e.g. Breton *et al.* 2006; Princen 2007), the chapter went on to examine more value-orientated approaches to understanding the role of evidence in tobacco policy debates, notably Sabatier and Jenkins-Smith's (1993, 1999) ACF. Like Farquharson (2003), this chapter argues that the ACF provides a more convincing means of understanding the 'tobacco wars' than evidence-orientated accounts and, like Cairney (2007), it found the ACF's emphasis on diverse policy actors was useful. Indeed, the decision to employ the term 'tobacco wars' throughout the chapter reflects the extent to which the ACF seems to descriptively reflect UK tobacco debates from the 1950s onwards. However, the chapter also argues that, on closer examination, neither of the opposing 'coalitions' is necessarily as unified as the ACF literature implies. Indeed, recent debates about harm reduction within tobacco control seem so divisive (Gartner *et al.* 2007) that it could be argued the 'tobacco control' coalition is at risk of experiencing a 'civil war'. Meanwhile, positions on tobacco taxation vary within both health and industry coalitions. This highlights the extent to which apparently cohesive coalitions can unravel as policy proposals evolve, potentially limiting the utility of the ACF. Moreover, the ACF offers few insights into the role of evidence within policy debates and does not sufficiently explain how particular coalitions gain and lose support over time.

Conscious of the need to maintain a focus on multiple kinds of actors, the chapter briefly considered how other network-based approaches might aid analyses of the 'tobacco wars'. It concluded that the literature on 'policy networks' was too diverse to be useful but that ANT helped draw attention to the potential fragility of networks and to the social construction of knowledge-claims (ideas, evidence and arguments). However, ANT's concern with the 'micro' makes it difficult, if not impossible, to apply to the decades-long, multi-sectoral 'tobacco wars'. Moreover, ANT says little about the values and beliefs that seemed so useful when employing the ACF as a theoretical framework.

In the last section, the chapter focused on ideational approaches (Béland 2005) to understanding the 'tobacco wars'. Specifically, it considered how conceptual changes may occur over time and how coalitions can develop and employ particular 'frames' to help attract support. It is argued that this approach retains some of the sociological interest in the construction and translation of knowledge-claims evident in ANT, but better captures the role of values, interests, interpretation and advocacy in the translation of evidence. All this suggests that ideas (rather than evidence) may be the more appropriate unit of analysis when studying the relationship between evidence and policy, which also reflects the broader turn to 'ideas' within policy studies (Blyth 1997; Béland

2005; Schmidt 2010). This seems to offer a particularly fruitful approach to thinking about the potential role of evidence in policy when combined with an analysis of relevant networks, such as that offered by the ACF.

Although this chapter has focused solely on tobacco, it seems plausible that an ideational, network-orientated approach to examining the relationship between evidence and policy may have broader application. Indeed, similar approaches have already been applied in studies exploring health inequalities and climate change debates (Smith 2007; Blok 2010). Other contemporary public health concerns, arguably, have even more in common with tobacco (e.g. Freudenberg 2005; Brownell and Warner 2009), although the ACF is likely to be less useful for policy issues which are not marked by strong oppositions. If the approach outlined in this chapter does, indeed, have broader applicability, further research might explore how evidence and ideas shape, and are shaped by, coalition strategies and tactics (e.g. Cairney 2007 notes the potential importance of multi-level governance and 'venue-shopping' in explaining the evolution of UK tobacco policy). More ambitiously, it might be possible to assess whether the characteristics of different ideas (including their relationship with evidence), and/or the strategies used to promote those ideas, can be categorized in ways which help to explain why some (but not all) ideas enable policy change.

Acknowledgements

This chapter has benefited from discussions in the University of Edinburgh's Tobacco Control Research Group, extremely useful comments on an earlier version by Professor Jeff Collin, plus feedback from the editors of this special issue. The author was supported by a joint Medical Research Council and Economic and Social Research Council Postdoctoral Fellowship grant (PTA-037-27-0181).

References

Advisory Group of the Royal College of Physicians (2000), *Nicotine addiction in Bristain? A report of The Tobacco Advisory Group of the Royal College of Physicians*, London: Royal College of Physicians.

Action on Smoking and Health (ASH) (2011), *Tobacconomics: How big tobacco uses dodgy data to 'throw sand in the gears' of global health policy*, London: ASH.

Bakker, K. and Bridge, G. (2006), Material worlds? Resource geographies and the 'matter of nature', *Progress in Human Geography*, 30, 1: 5–27.

Béland, D. (2005), Ideas and Social Policy: An Institutionalist Perspective, *Social Policy & Administration*, 39, 1: 1–18.

Berridge, V. (1999), Passive smoking and its pre-history in Britain: policy speaks to science? *Social Science & Medicine*, 49, 9: 1183–95.

Berridge, V. (2003), Post-war smoking policy in the UK and the redefinition of public health, *Twentieth Century British History*, 14, 1: 61–82.

Berridge, V. (2006), The policy response to the smoking and lung cancer connection in the 1950s and 1960s, *The Historical Journal*, 49, 4: 1185–209.

Berridge, V. (2007) *Marketing Health: Smoking and the Discourse of Public Health in Britain, 1945–2000*, Oxford: Oxford University Press.

Berridge, V. and Loughlin, K. (2005), Smoking and the new health education in Britain 1950s–1970s, *American Journal of Public Health*, 95, 6: 956–64.
Blok, A. (2010), Topologies of climate change: actor-network theory, relational-scalar analytics, and carbon-market overflows, *Environment and Planning D*, 28, 5: 896–912.
Blyth, M. (1997), 'Any more bright ideas?' The ideational turn of comparative political economy, *Comparative Politics*, 29, 2: 229–50.
Breton, E., Richard, L., *et al.* (2006), Fighting a Tobacco-Tax Rollback: A Political Analysis of the 1994 Cigarette Contraband Crisis in Canada, *J Public Health Pol*, 27, 1: 77–99.
Brownell, K. D. and Warner, K. E. (2009), The Perils of Ignoring History: Big Tobacco Played Dirty and Millions Died. How Similar Is Big Food? *Milbank Quarterly*, 87, 1: 259–94.
Cabinet Office (1999), *Modernising Government*, London: The Stationery Office.
Cairney, P. (2007), A 'multiple lenses' approach to policy change: the case of tobacco policy in the UK, *British Politics*, 2: 45–68.
Capability Reviews Team (2007), *Capability Review of the Department of Health*, London: Cabinet Office.
Caplan, N. (1979), The two-communities theory and knowledge utilization, *American Behavioral Scientist*, 22: 459–70.
Chapman, S. (2007), *Public Health Advocacy and Tobacco Control – Making Smoking History*, Oxford: Blackwell.
Exworthy, M., Stuart, M., Blane, D. and Marmot, M. (2003), *Tackling Health Inequalities since the Acheson Inquiry*, Bristol: Policy Press (for the Joseph Rowntree Foundation).
Farquharson, K. (2003), Influencing policy transnationally: pro- and anti-tobacco global advocacy networks, *Australian Journal of Public Administration*, 62, 4: 80–92.
Freudenberg, N. (2005), Public health advocacy to change corporate practices: implications for health education practice and research, *Health Education & Behavior*, 32, 3: 298–319.
Gartner, C. E., Hall, W. D., Chapman, S. and Freeman, B. (2007), Should the health community promote smokeless tobacco (snus) as a harm reduction measure? *PLoS Medicine*, 4, 7: e185.
Gilmore, A. B. (2012), Understanding the vector in order to plan effective tobacco control policies: an analysis of contemporary tobacco industry materials, *Tobacco Control*, 21, 2: 119–26.
Givel, M. (2006), Punctuated equilibrium in limbo: the tobacco lobby and US State policymaking from 1990 to 2003, *Policy Studies Journal*, 34, 3: 405–18.
Graham, H. (2012), Smoking, Stigma and Social Class, *Journal of Social Policy*, 41, 1: 83–99.
Gruer, L., Hart, C. L., Gordon, D. and Watt, G. (2009), Effect of tobacco smoking on survival of men and women by social position: a 28 year cohort study, *BMJ*, 338: b480.
Hall, P. A. (1993), Policy paradigms, social learning, and the state: the case of economic policymaking in Britain, *Comparative Politics*, 25, 3: 275–96.
Heclo, H. (1978), Issue networks and the executive establishment. In A. King (ed.), *The New American Political System*. Washington, DC: American Enterprise Institute for Public Policy Research, pp. 87–124.
Hunter, D. J. (2009), Relationship between evidence and policy: A case of evidence-based policy or policy-based evidence? *Public Health*, 123, 9: 583–6.
International Agency for Research on Cancer (IARC) (2011), *Handbook Volume 14 – Effectiveness of Tax and Price Policies for Tobacco Control*, Lyon: IARC.
Jones, D. (2012), Japan Tobacco feels UK has decided on plain packets, *Reuters*, http://uk.reuters.com/article/2012/07/06/uk-tobacco-britain-plainpackaging-idUKBRE8650P220120706 (accessed 11 July 2012).

Joossens, L. and Raw, M. (2011), *The Tobacco Control Scale 2010 in Europe. Association of European Cancer Leagues*, http://www.ensp.org/sites/default/files/TCS_2010_in_Europe_FINAL.pdf (acessed 25 March 2011).
Killoran, A. and Kelly, M. P. (2010), Introduction: effectiveness and efficiency in public health. In A. Killoran and M. P. Kelly (eds), *Evidence-based Public Health: Effectiveness and Efficiency*, Oxford: Oxford University Press.
Knorr-Cetina, K. (1981), *The Manufacture of Knowledge: An essay in the constructivist and contextual nature of science*, Oxford: Pergamon Press.
Larsen, L. T. (2008), The political impact of science: is tobacco control science- or policy-driven? *Science and Public Policy*, 35, 10: 757–69.
Larsen, L. T. (2010), Framing knowledge and innocent victims. Europe bans smoking in public places, *Critical Discourse Studies*, 7, 1: 1–17.
Latour, B. (2005), *Reassembling the Social: An Introduction to Actor-Network Theory*, Oxford: Oxford University Press.
Latour, B. and Woolgar, S. (1986), *Laboratory Life – The Construction of Scientific Facts*, Princeton, NJ: Princeton University Press.
Law, J. N. (1992), Notes on the Theory of the Actor-Network: Ordering, Strategy, and Heterogeneity, *Systemic Practice and Action Research*, 5, 4: 379–93.
Lopez, A. D., Collinshaw, N. E. and Tiha, P. (1994), A descriptive model of the cigarette epidemic in developed countries, *Tobacco Control*, 3: 242–7.
Macintyre, S. (2011), Good intentions and received wisdom are not good enough: the need for controlled trials in public health, *Journal of Epidemiology and Community Health*, 65: 564–7.
Marmot, M. (2010), *Strategic review of health inequalities in England post-2010. Marmot review final report*, London: University College London.
Marsh, D. and Rhodes, R. A. W. (1992), *Policy Networks in British Government*, Oxford: Oxford University Press.
McQueen, D. V. (2010), Ethics and evidence in health promotion. In A. Killoran and M. P. Kelly (eds), *Evidence-Based Public Health – Effectiveness and Efficiency*, Oxford, Oxford University Press.
Neuman, M., Bitton, A. and Glantz, S. (2002), Tobacco industry strategies for influencing European Community tobacco advertising legislation, *The Lancet*, 359, 9314: 1323–30.
Ong, E. K. and Glantz, S. A. (2001), Constructing 'sound science' and 'good epidemiology': Tobacco, Lawyers, and Public Relations Firms, *American Journal of Public Health*, 91, 11: 1749–57.
Overman, E. S. and Don, F. S. (1986), Iron triangles and issue networks of information Policy, *Public Administration Review*, 46: 584–9.
Pfizer Ltd (2008), *Memorandum by Pfizer Limited (HI 92) The Contribution of the NHS to Reducing Health Inequalities (Select Committee on Health Written Evidence)*, London: Select Committee on Health.
Princen, S. (2007), Advocacy coalitions and the internationalization of public health policies, *Journal of Public Policy*, 27, 1: 13–33.
Proctor, R. N. (2012), The history of the discovery of the cigarette-lung cancer link: evidentiary traditions, corporate denial, global toll, *Tobacco Control*, 21, 2: 87–91.
Read, M. D. (1992), Policy networks and issue networks: the politics of smoking. In D. Marsh and R. A. W. Rhodes (eds), *Policy Networks in British Government*, Oxford: Oxford University Press, Chapter 6.
Robinson, S. and Bugler, C. (2008), *Smoking and drinking among adults* (General Lifestyle Survey 2008), Newport: Office for National Statistics.
Sabatier, P. A. and Jenkins-Smith, H. C. (1993), *Policy Learning and Change: an advocacy coalition approach*, Boulder, CO: Westview Press.

Sabatier, P. A. and Jenkins-Smith, H. C. (1999), The Advocacy Coalition Framework: an assessment. In P. A. Sabatier (ed.), *Theories of the Policy Process*, Oxford: Westview Press: 117–66.
Saligman, R. B. (1978), Meeting in New York, source: Philip Morris, 20 November, Bates number(s): 2074827747, http://legacy.library.ucsf.edu/tid/uvi05a00 (accessed 5 January 2013).
Sanderson, I. (2006), Complexity, practical rationality and evidence-based policy making, *Policy & Politics*, 34, 1: 115–32.
Scheufele, D. and Tewksbury, D. (2007), Framing, agenda setting, and priming: the evolution of three media effects models, *Journal of Communication*, 57, 1: 9–20.
Schmidt, V. (2010), Taking ideas and discourse seriously: explaining change through discursive institutionalism as the fourth 'new institutionalism', *European Political Science Review*, 21, 1: 1–25.
Scott-Samuel, A. (2009), What the Renfrew/Paisley data really tell us about tackling health inequalities: the need to refocus upstream, *BMJ*, http://www.bmj.com/rapid-response/2011/11/02/what-renfrew-paisley-data-really-tell-us-about-tackling-health-inequalitie (accessed 5 January 2013).
Shirane, R., Smith, K., Ross, H., Silver, K. E., Williams, S. and Gilmore, A. (2012), Tobacco industry manipulation of tobacco excise and tobacco advertising policies in the Czech Republic: an analysis of tobacco industry documents, *PLoS Medicine*, 9, 6: e1001248.
Sims, M., Bauld, L. and Gilmore, A. (2012), England's legislation on smoking in indoor public places and work-places: impact on the most exposed children, *Addiction*, 107, 11: 2009–16.
Smith, K. E. (2007), Health inequalities in Scotland and England: the contrasting journeys of ideas from research into policy, *Social Science & Medicine*, 64, 7: 1438–49.
Smith, K. E. and Hellowell, M. (2012), Beyond rhetorical differences: a cohesive account of post-devolution developments in UK health policy, *Social Policy & Administration*, 46, 2: 178–98.
Smith, K. E. and Joyce, K. E. (2012), Capturing complex realities – understanding efforts to achieve evidence-based policy and practice in public health, *Evidence & Policy*, 8, 1: 59–80.
Smith, K. E., Fooks, G. Collin, J., Weishaar, H. and Gilmore, A. (2010), Is the increasing policy use of Impact Assessment in Europe likely to undermine efforts to achieve healthy public policy? *Journal of Epidemiology and Community Health*, 64, 6: 478–87.
Smith, K. E., Savell, E. and Gilmore, A. (2012), What is known about tobacco industry efforts to influence tobacco tax? A systematic review of empirical studies, *Tobacco Control*: doi:10.1136/tobaccocontrol-2011-050098.
Stead, M., MacAskill, S., MacKintosh, A.-M., Reece, J. and Eadie, D. (2001), 'It's as if you're locked in': qualitative explanations for area effects on smoking in disadvantaged communities, *Health & Place*, 7, 4: 333–43.
Stone, D. (2004), Transfer agents and global networks in the 'transnationalization' of policy, *Journal of European Public Policy*, 11, 3: 545–66.
Warner, K. E. (2000), The economics of tobacco: myths and realities, *Tobacco Control*, 9: 78–89.
Warner, K. E. and Mendez, D. (2010), Tobacco control policy in developed countries: Yesterday, today, and tomorrow, *Nicotine & Tobacco Research*, 12, 9: 876–87.
Warner, K. E., Hodgson, T. A. and Carroll, C. E. (1999), Medical costs of smoking in the United States: estimates, their validity, and their implications, *Tobacco Control*, 8, 3: 290–300.

Weishaar, H., Collin, J., Smith, K. E., Grüning, T., Mandal, S. and Gilmore, A. (2012), Global health governance and the commercial sector: a documentary analysis of tobacco company strategies to influence the WHO Framework Convention on Tobacco Control, *PLoS Medicine*, 9, 6: e1001249.

Weiss, C. (1977), Research for policy's sake: the enlightenment function of social research, *Policy Analysis*, 3: 531–47.

Weiss, C. (1979), The many meanings of research utilization, *Public Administration Review*, 39, 5: 426–31.

Weiss, C. (1990), The uneasy partnership endures: social science and government. In S. Brooks and A.-G. Gagnon (eds), *Social Scientists, Policy, and the State*, New York, NY: Praeger, pp. 97–111.

Weiss, J. A. (1989), The powers of problem definition: the case of government paperwork, *Policy Sciences*, 22, 2: 97–121.

Whitehead, M., Petticrew, M., Graham, H., Macintyre, S., Bambra, C. and Egan, M. (2004), Evidence for public health policy on inequalities: 2: Assembling the evidence jigsaw, *Journal of Epidemiology and Community Health*, 58, 10: 817–21.

World Bank (1999), *Curbing the Epidemic: Governments and the economics of tobacco control*, Washington, DC: World Bank.

World Health Organization (WHO) (1998), *World Health Assembly resolution WHA51.12*, Geneva: WHO.

Young, D., Borland, R. and Coghill, K. (2012), Changing the tobacco use management system: blending systems thinking with Actor-Network Theory, *Review of Policy Research*, 29, 2: 251–79.

3

Caught in the Same Frame? The Language of Evidence-based Policy in Debates about the Australian Government 'Intervention' into Northern Territory Aboriginal Communities

Emma Partridge

Introduction

The Northern Territory Emergency Response (NTER), or 'Intervention' was launched by the Howard government in 2007, and has continued under the Rudd and Gillard Labor governments with some modifications. The Intervention is a complex package of measures in Aboriginal communities in the Northern Territory (NT), concerning issues as diverse as land tenure, policing, law and order, health, housing, education, employment, welfare payments and governance. It has been labeled 'a governmental intervention unmatched by any other policy declaration in Aboriginal affairs in the last forty years' (Hinkson 2007: 1). Given this complexity, what follows is a simplified overview.

When launched, the Intervention was described as a 'national emergency response to protect Aboriginal children in the NT', following an NT government inquiry into child sexual abuse (Wild and Anderson 2007). The NTER legislation defined over 500 Aboriginal communities as 'prescribed areas'. To enable its application *exclusively* in these communities, the government suspended the Racial Discrimination Act 1975 (RDA). Measures introduced included: alcohol and pornography restrictions, quarantining half of welfare recipients' payments for use only on 'approved' purchases, compulsory child health checks, government acquisition of land on five-year leases, exemptions to the Aboriginal land permit system, increased policing, 'work for the dole' arrangements, linking welfare payments to school attendance, excluding consideration of customary law and cultural practice in bail and sentencing, and appointing 'government business managers' in communities. Some were implemented immediately and are still in force, some were modified or discontinued after the change of government in 2007 and some have only recently commenced. In 2010, the Labor government reinstated the RDA, although many of the measures previously considered in breach of it were retained and, the government claimed, 're-designed to more clearly be special

Evidence and Evaluation in Social Policy, First Edition. Edited by Ian Greener and Bent Greve. Book compilation

measures' under that Act. Welfare quarantining (or 'income management') was extended across the NT, allowing the government to claim that the previously race-based measure was now 'non-discriminatory' (Macklin 2009b).[1] New legislation passed in 2012 extended various measures for another ten years and expanded the pilot SEAM[2] programme enabling a parent's welfare payment to be suspended if he or she does not comply with compulsory school attendance plans for his or her children.

The Rudd and Gillard Labor governments[3] have repeatedly stated that their approach to the Intervention is one of evidence-based policy. In 2008, the Minister for Indigenous Affairs, Jenny Macklin, said 'my whole approach in Indigenous affairs will be that based on evidence. I'm not interested in ideology. What I'm interested in is what works' (Schubert 2008). Macklin's assessments of the policy would be 'based on data and hard facts' and future decisions 'grounded in a clear evidence base', in line with the government's broader 'commitment to an evidence-based approach to Indigenous policy' (Macklin 2008d). More recently, Prime Minister Gillard claimed that this approach had 'allowed Australia to move beyond anecdote and intuition and instead to act on the best evidence we can get' (Gillard 2011). Numerous government reports, policy statements and Ministerial comments have repeated the claim that the Intervention is and will continue to be 'evidence-based', rigorously evaluated and 'driven by hard facts and evidence' (Macklin 2008a).

While the government has framed the Intervention as evidence-based policy, many critics have also mobilized this same discourse to criticize the policy. These include academics (Cox 2012, 2011; Behrendt 2010, 2008; Altman 2011a, 2011b, 2008; Altman and Johns 2008), numerous Indigenous and non-Indigenous community organizations and service providers (see public submissions to Senate Community Affairs Committee 2010; NTER Review Board 2008), the Australian Greens (Siewert 2010), one of the NT Board of Inquiry chairs (Anderson 2011, 2007) and activists (Stop the Intervention Collective Sydney 2011). These critics point to a 'dearth of evidence' for the policy (Cox 2011: 2), a 'huge gap' between the government's assertions of 'progress' and the available evidence (Altman 2011a) and an approach that runs 'contrary to the evidence of what works' in Indigenous policy (Behrendt 2010: 128). They cite instances when the government has ignored, dismissed, discredited or contradicted evidence that did not support its position, or sought supporting evidence retrospectively, 'cherry-picking' it from a sea of contradictory findings, using it selectively, overstating it or taking it from one context and applying it inappropriately in another.

Critics also argue that the policy *process* has not conformed to the principles of evidence-based policy. They criticize the way the government has collected, assessed and responded to evidence, including both qualitative and quantitative data and consultations and inquiries. They are highly critical of the monitoring and evaluation process, citing poorly designed evaluation methods that lack baseline data and fail to collect meaningful empirical data to assess outcomes (Siewert 2010; Altman 2011b). An analysis by the Australian Institute of Health and Welfare (AIHW) of the government's income management data (AIHW 2010) is cited by many as illustrating the Intervention's poor evidentiary processes (Siewert 2010; Cox 2011; Behrendt 2010).

For many critics, despite the government's evidence-based policy 'mantra', the Intervention is a case of 'ideology replacing evidence' (Fisher 2009: 2, 10). Many urge the government to adhere to evidence-based policy principles, arguing that 'evidence, rather than prejudice or uneducated guesswork, needs to be the basis of policy' (Amnesty International Australia 2010: 8), and that Indigenous policy 'needs hard evidence, transparent and widespread community consultation and the advice of experts' (Altman 2008), 'less emotion, more evidence', and a move 'away from failed ideological policies to an evidence-based approach' (Behrendt 2008).

This chapter suggests that the use of the same discursive framework by the government and its critics is worthy of reflection. This commonality of language reflects the current dominance of the evidence-based policy paradigm in Australia. Following international trends, Australian public policy discourse has in recent years been dominated by commitments to, and pleas for, 'a robust, evidence-based policy making process' (Rudd 2008) in order to address complex social challenges and 'ameliorate or neutralize political obstacles' to policy (Banks 2009: 2, 6). These ideas, particularly the need to move 'beyond ideology' and 'assess the evidence' to determine 'what works', are now so common that this has been described as an era in which 'all sides of politics genuflect before the altar of "evidence-based" policy' (Shergold 2011) and in which it is 'increasingly difficult to bypass the imperative to fit one's research to an evidence-based framework' (Bacchi 2009: 107). It is, therefore, not surprising that both the government and its critics use evidence-based policy language. However, this framework is heavily criticized in the policy literature, suggesting a need for critical reflection on its use.

Critiques of Evidence-based Policy Discourse

'Evidence-based policy' is often presented as sensible, rational and essential to ensure effectiveness, value for money and optimum resource distribution (Sanderson 2006). This implies an objective process in which decision-makers assess 'the best evidence' and determine appropriate policy 'based on' that evidence. This 'intuitive, common sense logic' helps explain how this framework has become 'naturalized' across so many policy settings (Marston and Watts 2003: 144). However, there is now a significant body of literature that problematizes the evidence-based policy model. Critiques come from both post-positivist and pragmatist perspectives (Geyer 2011), pointing to two sets of related and problematic assumptions underpinning the idea of evidence-based policy. Together these two 'follies' are seen by critics to represent the 'inappropriate idealism' of the very idea that policy is evidence-based (Nutley *et al.* 2009: 7–9). The first concerns the nature of evidence as a form of knowledge, and the second concerns the politics of evidence in the policy process.

First, this critical literature argues that the question of what constitutes 'evidence' is not as straightforward as the evidence-based policy framework implies, because rather than being an objective set of relevant facts, policy evidence, like all 'knowledge' is socially constructed, context dependent and highly contested. Multiple, diverse and contradictory kinds of 'evidence' and 'knowledge' are considered relevant or valid by different stakeholders.

Evidence is also understood and interpreted differently by different stakeholders as a result of social, cultural and political dynamics, and in relation to existing values and ideological commitments.

The second criticism is that the framework of evidence-based policy assumes a linear and rational policy process. Yet this instrumental, or technocratic model has long been criticized, for oversimplifying the policy process, which is actually complex, messy, context dependent, interactive and dynamic. Further, from a post-positivist perspective, it depoliticizes what is inevitably a political and interpretive process in which values, power relations, vested interests, political agendas and ideologies influence the ways in which 'evidence' is interpreted and used, policy narratives constructed, arguments made and policy developed. By obscuring this, the evidence-based policy paradigm allows governments to legitimize their decisions as technical and rational rather than ideological.

Evidence as Knowledge in the Intervention

Rather than asking whether the Intervention is evidence-based policy then, we might view it as an illustration of these two major critiques of that framework. First, debates about the Intervention have clearly involved a contest over what kinds of 'knowledge' constitute relevant and valid evidence.

A central argument in the critical literature is that policymaking is informed by multiple and diverse kinds of 'evidence' and different types and sources of knowledge (Head 2008; Mulgan 2005; Davies *et al.* 2008), not limited to the systematic research evidence prioritized by the evidence-based policy notion of 'knowledge transfer' (Davies *et al.* 2008: 189). Policy formulation involves many kinds of knowledge, including consultation findings, ideas, opinions, political and economic considerations (Bowen and Zwi 2005) and practical knowledge (Head 2008).

This argument is clearly illustrated both by the government's Intervention policy-making process, and by the associated public debate. The government has been influenced by many different kinds of 'evidence' besides formal research and evaluation. Indeed, the latter has arguably had the *least* influence. For example, in November 2011, the government announced that future Intervention legislation would include a new measure to suspend parents' welfare payments if their children's school attendance was not satisfactory. The Minister had claimed the evaluation of a trial of this measure would 'help to inform' the new legislation (quoted in Wilson and Rout 2011). However, the government announced the legislation before the evaluation was released. When it was, it not only failed to provide sufficient data to evaluate the effect of payment suspension (the sanction was applied to only six NT parents in the trial), but found that in most cases, children's absences actually *increased* following suspension of their parents' payment. The report concluded that 'overall sanctions had no impact on reducing unauthorized absence rates for the small number of families affected' (Department of Education, Employment and Workplace Relations 2012: 52–3). The Minister's media release did not acknowledge this finding, instead stating that the results were 'encouraging' and the scheme was 'having a positive impact' (Macklin 2012).

This example suggests not only that research evidence had little influence on the policy, but also that other factors determined the government's decision. The critical literature suggests these are likely to have been value-based political judgements, as discussed below.

The critical literature suggests not only that policy is influenced by various kinds of 'knowledge', but that all knowledge is inherently contestable, historically contingent, indeterminate, contradictory, partial and provisional (Mulgan 2005; Head 2008; Davies *et al.* 2008). Rather than being 'objective, stable and acontextual' (Davies *et al.* 2008: 189), what counts as evidence is often not clear or easy to determine but is context dependent (Nutley *et al.* 2009) and varies for different stakeholders (Head 2008). Further, the many kinds of knowledge or evidence available do not 'inform' the policy process in any straightforward or technical way, nor is evidence simply 'translated' into policy. Rather, evidence must be interpreted, and judgments made about its relevance, status and use.

Again, the Intervention illustrates this inevitable contestation over evidence. Various aspects of the policy have been subject to claim and counterclaim by government representatives, academics, commentators, community organizations and activists. Different participants have provided competing and contradictory interpretations of what or whose evidence is relevant and what policy conclusions should be drawn from particular evidence. This shows how evidence is seen differently through the various 'lenses' of different actors located in different institutional settings (Head 2008: 9) and that its interpretation is influenced by histories, experiences, beliefs and values at the individual, organizational and system levels (Bowen and Zwi 2005). Consequently, decision-making is far more complex than determining 'what works'. This contestation also illustrates the argument that knowledge is not simply 'information' (Bacchi 2009: 242), but is socially constructed, meaning 'evidence' can never be objective because its interpretation is always value-based and occurs within a context of uneven power relations. This constitutes a significant objection to technocratic evidence-based policy models that largely fail to acknowledge the influence of values on the policy process (Bowen and Zwi 2005). In practice, what constitutes 'valid social knowledge' is always contested (Sanderson 2002: 5) because diverse forms of lived experience generate a diversity of understandings 'not only about what works, but also about what is worthwhile and meaningful' (Head 2008: 9).

An example of these dynamics in operation in this case is the contestation over two consultation processes the government undertook on the Intervention, in 2009 and 2011. While the consultations are one source of 'evidence' about policy impact, there are widely varying views about the meaning and validity of this evidence. The government interpreted the findings as largely supportive of the policy. Referring to the 2009 consultations, Macklin stated that while views were mixed, 'many participants reported that income management had delivered discernible benefits' and 'the majority' of those consulted supported its continuation (Macklin 2009c: 12784, 12786). However, others disputed this, arguing that dissenting voices had been ignored and participants' comments misinterpreted (Nicholson *et al.* 2009; Harris 2010)[4] or that the consultations were invalid because the process was flawed. The 2011

consultations faced particular criticism, with activists labeling them 'a sham' (Intervention Rollback Action Group *et al.* 2011) because they did not raise the specific proposed policy changes for discussion. The Australian Human Rights Commission agreed that this omission, as well as a rushed timeframe, and a failure to translate the government's discussion paper into community languages, rendered the consultations 'inadequate' (AHRC 2012: 13). This example illustrates how significant the contestation can be over what constitutes 'valid' evidence and what that evidence 'means'.

Evidence as Politics in the Intervention

The second major critique of the evidence-based policy framework is that the use of particular 'knowledge' as evidence for policy is inherently political. Arguing that rationalist policy models obscure this, the critical literature instead conceives of policy-making as an explicitly political process in which 'research interacts with values and vested interests to determine policy outcomes' (Marston and Watts 2003: 146) and 'political factors loom large in policy discussions, including (sometimes especially) those claimed to be based on evidence' (Bryson and Mowbray 2005: 101). The Intervention illustrates this – while the government uses the language of evidence-based policy to frame its approach as apolitical and based on 'hard evidence', in practice its definition of what evidence 'counts' has been highly political.

The government's evaluation of income management shows that the data that counts as 'evidence' in policy evaluation is often far removed from any rationalist definition of valid policy evidence. Assessing the government's evaluation and data collection methods, the AIHW concluded they sat 'towards the bottom of an evidence hierarchy' (AIHW 2010: 16). It pointed to the absence of baseline data, comparison groups and empirical indicators, limited and poor quality quantitative data, poor interview sampling methods and poor focus group reporting, which made it 'difficult to identify whose views were reported, or whether they applied to the majority of stakeholders in the focus groups' (AIHW 2010: 16–17). It concluded that the government's own evaluation showed that 'the overall evidence about the effectiveness of income management was not strong' (AIHW 2010: 16–17). This assessment contrasted significantly with the government's overwhelmingly positive media release on the report, which cited only those findings that supported the policy (Macklin 2009a). This example illustrates the contrast between idealistic assumptions about disinterested, evidence-based decision-making and a reality that is both far messier and inherently political (Nutley *et al.* 2003). It suggests that 'how and when evidence is used often depends upon the political agenda and ideology of the government of the day, not on the nature of the evidence, however compelling' (Bowen and Zwi 2005: 601–2).

The use of an evidence-based policy framework also appears to imply a preference for quantitative evidence, with its connotations of 'objectivity'. As noted earlier, Rudd, Macklin and Gillard explicitly eschew 'ideology', 'anecdote' and 'intuition', referring instead to 'hard data', 'rigorous evaluation' and a 'methodical approach', to establishing a 'clear evidence base' for policy decisions. Yet this framing contrasts with the kind of 'evidence' the govern-

ment later cites. Consider Macklin's response when pressed for evidence for the efficacy of suspending parents' welfare payments if their children are not attending school:

> Interviewer: Using welfare measures to improve school attendance is already being trialed in some communities in the Northern Territory. What's the evidence that it's working?
> Macklin: Well the evidence is really coming from Aboriginal people themselves. We've just spent the last few months speaking with many, many people in lots of different communities in the Northern Territory and I've heard myself from Aboriginal people, especially older people, saying to me we want to do everything possible to get our children to school.
> Interviewer: In those communities though where you've trialed it, are children attending class more often?
> Macklin: We've certainly seen some improvements but we know we've got a lot further to go (Macklin 2011a).

Here, instead of referring to data from the government's own trial, Macklin cites as evidence people's general views about the importance of school attendance. Her response both elevates (some people's) opinions above the quantitative data, and implies that the trial results were positive when, as noted above, the evaluation found the measures had 'no impact'.

The use of evidence in the Intervention also supports the proposition that policy-making is an argumentative process in which information or data is selected and presented as evidence as part of a political process of persuasion (Majone 1989). Rather than assessing evidence in a 'rational' way, policymakers use it selectively 'to create persuasive policy stories' or 'policy narratives' to help 'sell the policy' (Stevens 2011: 242). An example is the way the government has gradually built a narrative about the benefits of withholding parents' welfare payments to sanction unsatisfactory school attendance. Its *Stronger Futures* discussion paper established the narrative by stating that 'initial advice shows that [the trial] is having a positive impact on parents ensuring their children are enrolled and regularly attending school' (Australian Government 2011a: 10). Subsequent community consultations introduced the issue of 'school attendance and educational achievement' by asking: 'what can the Government do to encourage this, *for example through links to welfare payments*?' (Australian Government 2011a: 11, emphasis added). This clearly leading question pointed discussion towards the policy that the government's narrative had already constructed as the solution. The government then interpreted the consultation results in line with its pre-existing narrative, suggesting that those consulted 'commented *relatively frequently*' that parents should have welfare payments withheld or reduced if they did not send their children to school, and that '*fewer respondents*' disagreed (Australian Government 2011b: 23, emphasis added). This was despite an independent review of the consultations that found meeting records were not sufficiently detailed to determine levels of community participation, or whether comments reflected a common view (Cultural and Indigenous Research Centre Australia 2011: 24). Macklin also

stated that 'people believe education is vitally important for children and that parents should be responsible for their children's regular attendance at school' and (despite the Cultural and Indigenous Research Centre Australia's finding that records made it impossible to judge participants' priorities) that 'getting more children to school' was one of three 'top priorities for local people' (Macklin 2011b). Even if these claims are accurate, they do not necessarily constitute support for the specific policy of suspending parents' welfare payments. They do, however, reinforce the impression that this is an inevitable, commonsense policy that is not only supported by Aboriginal people but responds to *their* priorities. As such, the process looks less like a rational assessment than a 'selective, narrative use of evidence' intended to reduce uncertainty and strengthen the government's policy narrative (Stevens 2011: 237).

While ideology and evidence are often represented dichotomously, evidence needs to be understood in an ideological context, especially in Indigenous affairs (Sanders 2009). The Intervention appears to support the contention that strongly held ideological positions are not easily changed by the emergence of contrary evidence – as those who point to examples of 'policy-based evidence' argue (Bryson and Mowbray 2005; Marmot 2004). In 2008 a government-appointed Review Board consulted with affected communities, commissioned independent research and assessed over 200 public submissions. It found that blanket imposition of income management 'was not based on any assessment of a person's capacity to properly meet their family responsibilities' and that this race-based measure had caused 'widespread disillusionment, resentment and anger' among Indigenous people (NTER Review Board 2008: 20). It recommended that while a voluntary measure might have merit, the blanket application of compulsory income management across all NT Aboriginal communities should cease.

The government responded by announcing that the measure would continue 'because of its demonstrated benefits for women and children'. It did not provide evidence for this except to state that 'women say that income management means they can buy essentials for their children such as food and clothes' and 'shopping habits in licensed stores have changed – more is being spent on fresh food, sales of cigarettes have halved and the incidence of "humbugging"[5] has fallen' (Macklin 2008b). Asked in an interview for 'clear definable measurable evidence' for income management, Macklin pointed to a survey of store owners and 'anecdotal' comments made by 'all of the women that I have spoken to in many, many communities' (Macklin 2008c). No source was given for the claim about the 'incidence of humbugging', and the reliability of the store data was widely questioned (Altman 2008) particularly after it was revealed to be based on a survey of just ten stores.[6] Asked why she had ignored the Review Board's recommendation, Macklin stated 'we do want to see the development of strong social norms in these communities, and as far as I'm concerned the evidence is very strong that it's coming from compulsory income management' (Macklin 2008c).

This example suggests that the government had a pre-existing ideological commitment to income management that required its continuation in spite of the Review's findings. This is a process far removed from the linear, rational

one suggested by the framework of evidence-based policy, or indeed from the Review Board's terms of reference, which explicitly framed it as part of an evidence-based policy process (NTER Review Board 2008: 68). Examining the Intervention for its ideological basis, rather than its 'evidence base', suggests the government is committed to an ideological agenda based on the development of 'social norms'.[7] Macklin selectively cites evidence that supports her belief that income management helps to achieve this – even if it is anecdotal opinion or generalization, one supportive submission among many opposing ones (Cox 2011: 36–8), or based on a small, unreliable sample.

Arguably, the most serious problem with the evidence-based policy framework is that it allows governments to *claim* a rational process. It does this by obscuring ideological agendas and power relations (Naughton 2005; Lattas and Morris 2010), government influence over research agendas (Bacchi 2009: 146), vested interests seeking to influence policy irrespective of the evidence (Marston and Watts 2003: 146) and the exploitation of evidence by those seeking to retain power (Majone 1989) or minimize political sensitivities (Guenther *et al.* 2010: 2). Because the term 'evidence-based' acts as a synonym for 'scientific', 'scholarly' and 'rational' (Marston and Watts 2003: 144–5), governments can use it to present their approach as practical, 'informed social governance' (Lattas and Morris 2010: 79) and to legitimize policy decisions as rational rather than ideological or political (Naughton 2005; Nutley *et al.* 2009; Sanderson 2002).

Macklin has consistently presented the Intervention in this way, telling anti-Intervention activists, 'look at the evidence. This has nothing to do with ideology or politics, it is about what people need' (in Wilson and Karvelas 2011). Launching a review of the policy, Macklin stated 'our test will be, what's the evidence of what works' (Macklin 2011c). Such comments demonstrate how evidence-based policy discourse acts to 'depoliticize politics', allowing governments to present policy as 'practical, technical measures' (Naughton 2005: 51). Denying ideological commitments, Macklin presents the Intervention as based on 'the evidence' of 'what people need', and the 'test' of 'what works'. This obscures the political, value-laden, ideological and contested nature of Aboriginal policy, representing the Intervention as rational and needs-driven. Yet the examples discussed above suggest it is difficult if not absurd to characterize the Intervention as a technical exercise where rational policy decisions are based on evidence of 'what works'.

Evidence-based Policy as a Critical Framework? Limitations and Risks

The critiques of the evidence-based policy framework discussed above suggest it is worth considering the implications of framing criticism of the Intervention within this discourse. This is not to dismiss the importance of such critiques, many of which have been made by Aboriginal people and organizations, challenging the government and presenting alternative evidence and proposals, and striving to transform the Intervention into a positive opportunity. Constructive and engaged resistance is an important example of the potential 'power of the community, to argue, to persuade, to use the evidence of their

own experience, and to transform the world' (Anderson 2009: 29). Independent and critical assessments of whether government policies are 'evidence-based' are extremely valuable and can enable the language of evidence-based policy to be used as an 'asset', to challenge a particular 'regime of governance' (Bacchi 2009: 147). In this case, they have created a powerful body of counter-argument to the Intervention, that seriously undermines the government's claims about its evidentiary basis. Notwithstanding these strengths, however, this kind of critique also has limitations and risks.

First, this kind of criticism risks reinforcing the 'non-ideological' model of policy-making implied by evidence-based policy discourse. Many of the Intervention's critics clearly agree that evidence-based policy is not a useful *description* of how policy is made. Cox, for example, suggests the Intervention 'illustrates with unusual clarity how little attention governments pay to evidence when they are driven by prior prejudices and beliefs' (Cox 2011: 2). Yet basing a judgment of the Intervention on whether or not it *is* evidence-based appears to imply that policy *can and should be* evidence-based. This risks leaving intact the rational policy-making model and reproducing the 'evidence'/'ideology' dichotomy. Such a position can easily slide into an argument that assumes 'evidence-based policy' is the norm that the Intervention has departed from. This is illustrated by Cox's suggestion that the policy process produced an 'irrational outcome' (Cox 2011: 54) and that what is remarkable about the Intervention is the way it differs from the *usual* policy-making process:

> This policy process is quite different from the acceptable norm in policy making. In addition to an unusual lack of prior serious discussion and consultation on the merits and risks of such changes, there is a dearth of evidence that the process has net benefits to justify the financial and social costs (Cox 2011: 1)

This implies that 'rational' policy-making, whereby proposals are subjected to careful discussion, consultation and cost-benefit analysis, is the norm, despite this 'textbook' model having been criticized as 'little more than a myth' (Guenther *et al.* 2010). Myths, however, are powerful. The rationalist, instrumental, linear model of evidence-based policy persists despite significant criticism (Freiburg and Carson 2010; Sanderson 2002) and remains a 'common-sense' understanding of policy-making. Freiburg and Carson observe that while the linear/rational model of policy-making is rarely stated in a naive form:

> it is surprising that its ghost so frequently haunts the corridors of parliaments, bureaucracies and academe. In this sense it is indeed an 'imaginary'; unattained and unattainable, but still practically consequential in that it so powerfully guides the attitudes adopted by those who crucially influence the development of policy (Freiburg and Carson 2010: 155)

In reproducing this model, critics risk reinforcing the myth. The portrayal of 'evidence-based policy' as normative is unlikely to be the intent of Intervention critics, but by contrasting the Intervention to an implied 'norm', and

repeating calls for it to be evidence-based rather than ideological, they risk validating the model, at least as a prescriptive one.

A second reason for ambivalence about these kinds of arguments is that their *effectiveness* appears limited. Despite repeated criticism, the Intervention has changed little over five years, and it is difficult to identify instances where the government's intended course has been altered by counter-evidence. That it continues to frame its approach as 'evidence-based' suggests that the many criticisms – however warranted – have had little effect. Because evidence is contested and contestable, arguments about whether or not the policy is evidence-based become circular and endless. Government can always point to evidence, and opponents to counter-evidence. A final indication of the malleability of arguments calling for evidence-based policy is that they are also used by *supporters* of the Intervention, who for example, accuse critics of 'a mistaken denial of the evidence of the extent of child abuse and neglect and the conditions in which they flourish' (Langton 2010: 93–4).

A further risk of arguments advocating 'evidence-based' policy is that they may obscure other, potentially more effective strategies for improving the policy process in Indigenous affairs. The chapter concludes by pointing to some of these.

Conclusion: Alternative Strategies

Proponents of evidence-based policy make many suggestions about how to improve policy processes. These include making evidence collection, assessment and use more rigorous, transparent or objective, and applying 'best practice' monitoring and evaluation to measure policy impact against specific goals. Such measures are undoubtedly important. While there needs to be more acknowledgement that 'research and improved government data collections are not the saviour; policy development will never be depoliticized or devoid of conflict' (Anderson 2003: 235), there is nevertheless much that can and must be done to improve the Australian government's approach to 'evidence' in Indigenous policy. In the case of the Intervention, critics have shown it is a long way from meeting even the minimum standards that its own purported framework of evidence-based policy requires.

However, while such attempts to hold the government to account are necessary and important, this chapter has argued that over-reliance on this strategy entails risks and limitations. This suggests that other kinds of strategies may ultimately be more important. These might have two critical dimensions. The first is a focus on *values* in policy. In contrast to rationalist approaches that marginalize discussion of values, a 'value-critique of policy' foregrounds both the values on which a policy is based, and the wider social and political values and power relations that shape the policy context and give it political meaning (Taylor 2006: 253). This involves a consideration of what issues of power and knowledge are apparent in the policy, and what social and political values determine its evaluation (Valentine 2009: 451). Rejecting a distinction between technical and political questions, it does not propose values as something in opposition to 'evidence', but sees the two as mutually constitutive (Valentine 2009: 459).

The second dimension of an alternative strategy would address the fact that policy is (legitimately) developed in a social, political and cultural context in which there will always be a contest of views, values and interests. This approach draws on theoretical and practice-based literature that follows the 'deliberative turn' in political theory (e.g. Dryzek 2000; Kahane *et al.* 2010), arguing that democratic legitimacy requires genuine deliberation on the part of those affected by policy. These ideas might be usefully developed and adapted to respond to calls from many Indigenous sources for more genuine participation of Indigenous people in dialogue and *negotiation* with the state (Anderson 2011; Gooda 2011; NTER Review Board 2008: 8) and for community–government partnerships that enable not only more collaborative policy but new forms of community governance and self-determination (Calma 2007; AHRC 2012; Anderson 2009, 2007; Aboriginal Peak Organisations Northern Territory 2011; National Congress of Australia's First Peoples 2012; Secretariat of National Aboriginal and Islander Child Care 2012). Many Indigenous authors and organizations point to the need for a deep restructuring of the relationship between Indigenous Australians and the state. Understanding policy as necessarily values-based, and exploring whether and how contemporary experiments in deliberative, discursive and participatory governance might be relevant to such a restructuring are approaches worthy of further attention.

Notes

1. This claim was met with scepticism, following reports that as of February 2011, 94 per cent of the 16,000 NT residents on income management were Indigenous, while more than three-quarters of those granted exemptions were non-Indigenous (Drape 2011).
2. The 'Improving School Enrolment and Attendance through Welfare Reform Measure' (SEAM) was trialled in some NT and Queensland schools from 2009. The programme makes parents' welfare payments conditional on their children's satisfactory school attendance.
3. While the Howard government made some use of ideas of 'evidence' to justify the policy, it lost office just five months after launching it. It is under the subsequent Labor governments that I observe a shift to a more explicit use of the framework and language of 'evidence-based policy' to justify the Intervention, and Indigenous policy generally, and it is this period of the policy that I largely refer to in this chapter.
4. For similar contesting interpretations of the 2011 consultations, see Concerned Australians (2012, 2011).
5. 'Humbugging' refers to pestering friends and relatives for money.
6. The government admitted this when pressed in a Senate hearing for details of the relevant 'evidence' (Fisher 2009).
7. The government has expressed the Intervention's goals as to 'promote personal responsibility and rebuild community norms', to 'generate greater individual and community responsibility' and to build 'increased levels of personal and community responsibility', and describes income management as one means of achieving these goals (see Macklin 2008b).

References

Aboriginal Peak Organisations Northern Territory (2011), *Response to Stronger Futures*, submission to the Senate Standing Committee on Community Affairs regarding the Stronger Futures in the Northern Territory 2011 Bill, the Stronger Futures in the Northern Territory (Consequential and Transitional Provisions) Bill 2011, and the Social Security Legislation Amendment Bill 2011, http://www.clc.org.au/files/pdf/APONT_submission_to_Stronger_Futures_-_Final.pdf (accessed 15 March 2013).

Altman, J. (2011a), Evidently: Yes, no, maybe. . . Prime Minister, *Tracker*, 10 August.

Altman, J. (2011b), The intervention is dead, long live the intervention, *Crikey*, 1 November.

Altman, J. C. (2008), *Submission to The Northern Territory Emergency Response Review*, CAEPR Topical Issue, 10, Centre for Aboriginal Economic Policy Research, Canberra: Australian National University.

Altman, J. and Johns, M. (2008), Indigenous welfare reform in the Northern Territory and Cape York: a comparative analysis, CAEPR working paper, 44, Centre for Aboriginal Economic Policy Research, Canberra: Australian National University.

Amnesty International Australia (2010), *Submission to the Senate Standing Committee on Community Affairs regarding the Inquiry into Social Security and other Legislation Amendment (Welfare Reform and Reinstatement of Racial Discrimination Act) Bill 2009 and the Families, Housing, Community Services and Indigenous Affairs and other Legislation Amendment (2009 Measures) Bill 2009*, http://www.aph.gov.au/Parliamentary_Business/Committees/Senate_Committees?url=clac_ctte/completed_inquiries/2008-10/soc_sec_welfare_reform_racial_discrim_09/submissions/sublist.htm (accessed 15 March 2013).

Anderson, I. (2003), Evidence and Aboriginal health policy. In V. Lin and B. Gibson (eds), *Evidence-based health policy: problems and possibilities*, Oxford: Oxford University Press, pp. 224–36.

Anderson, P. (2007), Douglas Gordon oration, Public Health Association of Australia conference, Alice Springs, 25 September.

Anderson, P. (2009), 'The Intervention': some reflections two years on, Social Justice Lecture, Melbourne, University of Melbourne, 24 June.

Anderson, P. (2011), Research for a better future, keynote address to 3rd Aboriginal health research conference, Sydney, 5–6 May.

Australian Government (2011a), *Stronger Futures in the Northern Territory: discussion paper*, Canberra: Commonwealth of Australia.

Australian Government (2011b), *Stronger Futures in the Northern Territory: report on consultations*, Canberra: Commonwealth of Australia.

Australian Human Rights Commission (AHRC) (2012), *Stronger Futures in the Northern Territory Bill 2011 and two related bills*, submission to the Senate Community Affairs Legislation Committee regarding the Stronger Futures in the Northern Territory 2011 Bill, the Stronger Futures in the Northern Territory (Consequential and Transitional Provisions) Bill 2011, and the Social Security Legislation Amendment Bill 2011, http://humanrights.gov.au/legal/submissions/2012/20120206_stronger.html (accessed 15 March 2013).

Australian Institute of Health and Welfare (AIHW) (2010), *Evaluation of income management in the Northern Territory*, Occasional paper No. 34, Australian Government Department of Health and Ageing, Canberra: Commonwealth of Australia.

Bacchi, C. L. (2009), *Analysing policy: what's the problem represented to be?* Frenchs Forest: Pearson Education.

Banks, G. (2009), *Challenges of evidence-based policy making*, Productivity Commission and Australian Public Service Commission, Canberra: Australian Government.

Behrendt, L. (2008), A successful intervention needs less emotion, more evidence, *Sydney Morning Herald*, 16 October.
Behrendt, L. (2010), Closing the evidence gap. In M. Davis and M. Lyons (eds), *More Than Luck: Ideas Australia Needs Now*, Sydney: Centre for Policy Development, pp. 117–28.
Bowen, S. and Zwi, A. (2005), Pathways to 'evidence-informed' policy and practice: a framework for action, *PloS Medicine*, 2, 7: 600–5.
Bryson, L. and Mowbray, M. (2005), More spray on solution: community, social capital and evidence based policy, *Australian Journal of Social Issues*, 40, 1: 91–106.
Calma, T. (2007), Tackling child abuse and inequality. In J. Altman and M. Hinkson (eds), *Coercive reconciliation: stabilize, normalize, exit Aboriginal Australia*, North Carlton: Arena, pp. 273–86.
Concerned Australians (2011), *Cuts to Welfare Payments for School Non-Attendance Requested or Imposed?* Melbourne: Concerned Australians.
Concerned Australians (2012), *NT consultations report 2011: by quotations*, Melbourne: Concerned Australians.
Cox, E. (2011), Evidence-free policy making? The case of income management, *Journal of Indigenous Policy*, 12: 1–98.
Cox, E. (2012), What's data got to do with it? Reassessing the NT intervention, *The Conversation*, 25 January.
Cultural and Indigenous Research Centre Australia (2011), *Department of Families, Housing, Community Services and Indigenous Affairs report on Stronger Futures consultation*, Sydney: Cultural and Indigenous Research Centre Australia.
Davies, H., Nutley, S. and Walter, I. (2008), Why 'knowledge transfer' is misconceived for applied social research, *Journal of Health Services Research & Policy*, 13, 3: 188–90.
Department of Education, Employment and Workplace Relations (2012), *Improving School Enrolment and Attendance through Welfare Reform Measure (SEAM): Evaluation Report for 2010*, Canberra: Australian Government.
Drape, J. (2011), Welfare management 'discriminatory': Gooda, *Sydney Morning Herald*, 10 March.
Dryzek, J. S. (2000), *Deliberative democracy and beyond*: *Liberals, Critics, Contestations*, Oxford: Oxford University Press.
Fisher, I. (2009), Closing the gap: is the evidence really there?: racism and the NT Intervention, Speech to University of Technology, Sydney, 22 March, http://stoptheintervention.org/uploads/files_to_download/Irene-Fisher-speech-22-3-09.pdf (accessed 15 March 2013).
Freiburg, A. and Carson, W. G. (2010), The limits to evidence-based policy: evidence, emotion and criminal justice, *Australian Journal of Public Administration*, 69, 2: 152–64.
Geyer, R. (2011), Can complexity move UK policy beyond 'evidence-based policy making' and the 'audit culture'? Applying a 'complexity cascade' to education and health policy, *Political Studies*, 60, 1: 1–24.
Gillard, J. (2011), This work will go on, Speech to the House of Representatives, 9 February, Canberra.
Gooda, M. (2011), Presentation to FaHCSIA (NSW/ACT) staff, 11 April, Australian Human Rights Commission, Sydney.
Guenther, J., Williams, E. and Arnott, A. (2010), The politics of evaluation: evidence-based policy or policy-based evidence? paper presented to the North Australia Research Unit public seminar series, Darwin, 30 November, http://covaluator.net/docs/S2.1_the_politics_of_evaluation.pdf (accessed 15 March 2013).
Harris, M. (ed.) (2010), *This is what we said: Australian Aboriginal people give their views on the Northern Territory Intervention*, Melbourne, Concerned Australians.

Head, B. W. (2008), Three lenses of evidence-based policy, *Australian Journal of Public Administration*, 67, 1: 1–11.
Hinkson, M. (2007), Introduction: in the name of the child. In J. Altman and M. Hinkson (eds), *Coercive reconciliation: stabilize, normalize, exit Aboriginal Australia*, North Carlton: Arena, pp. 1–12.
Intervention Rollback Action Group, Darwin Aboriginal Rights Coalition and Stop the Intervention Collective Sydney (2011), *Stand against Macklin's decade of discrimination – No second Intervention!* media release, 25 November.
Kahane, D., Weinstock, D., Leydet, D. and Williams, M. (eds) (2010), *Deliberative democracy in practice*, Vancouver: UBC Press.
Langton, M. (2010), The shock of the new: a postcolonial dilemma for Australianist anthropology. In J. Altman and M. Hinkson (eds), *Culture crisis: anthropology and politics in Aboriginal Australia*, Sydney: UNSW Press, pp. 91–115.
Lattas, A. and Morris, B. (2010), The politics of suffering and the politics of anthropology. In J. Altman and M. Hinkson (eds), *Culture crisis: anthropology and politics in Aboriginal Australia*, Sydney: UNSW Press, pp. 61–87.
Macklin, J. (2008a), Closing the gap between Indigenous and non-Indigenous Australians, Budget statement, May, Canberra: Commonwealth of Australia.
Macklin, J. (2008b), Compulsory income management to continue as key NTER measure, media release, 23 October.
Macklin, J. (2008c), interview transcript, 7.30 Report, Australian Broadcasting Corporation, 23 October.
Macklin, J. (2008d), Foreword. In Australian Government (2008) *Northern Territory Emergency Response: One Year On*, Canberra: Australian Government.
Macklin, J. (2009a), Children eating healthier food in NTER communities, media release, 15 December.
Macklin, J. (2009b), Strengthening the Northern Territory Emergency Response, media release, 25 November.
Macklin, J. (2009c), *Social Security and other legislation amendment (Welfare reform and reinstatement of Racial Discrimination Act) Bill 2009, Second Reading Speech*, 25 November, Parliament House, Canberra.
Macklin, J. (2011a), Transcript of interview, ABC Radio National, 14 November.
Macklin, J. (2011b), Aboriginal people clear about their priorities for building stronger futures in the Northern Territory, media release, 18 October.
Macklin, J. (2011c), Launch of Northern Territory Emergency Response discussion paper, joint press conference with Gillard, J, 22 June.
Macklin, J. (2012), Good progress in improving school enrolment and attendance in Queensland and the Northern Territory, media release, 2 February.
Majone, G. (1989), *Evidence, argument and persuasion in the policy process*, New Haven, CT: Yale University Press.
Marmot, M. G. (2004), Evidence based policy or policy based evidence? *BMJ*, 328, 7445: 906–7.
Marston, G. and Watts, R. (2003), Tampering with the evidence: a critical appraisal of evidence-based policymaking, *The drawing board*, 3, 3: 143–63.
Mulgan, G. (2005), Government, knowledge and the business of policy making: the potential and limits of evidence-based policy, *Evidence & Policy*, 1, 2: 215–26.
National Congress of Australia's First Peoples (2012), *Statement to the Senate Standing Committee on Community Affairs on conditions affecting Aboriginal communities in the Northern Territory including the proposed Stronger Futures in the Northern Territory Bill (2011)*, Strawberry Hills: National Congress of Australia's First Peoples, http://nationalcongress.com.au/wp-content/uploads/2012/02/CongressStrongerFutures.pdf (accessed 15 March 2013).

Naughton, M. (2005), 'Evidence-based policy' and the government of the criminal justice system – only if the evidence fits! *Critical Social Policy*, 25, 1: 47–69.
Nicholson, A., Behrendt, L., Vivian, A., Watson, N. and Harris, M. (2009), *'Will They Be Heard?' – a response to the NTER consultations June to August 2009*, Sydney: University of Technology, November.
Northern Territory Emergency Response Review Board (NTER Review Board) (2008), *Report of the NTER Review Board*, Canberra: Commonwealth of Australia.
Nutley, S., Davies, H. and Walter, I. (2003), Evidence-based policy and practice: cross-sector lessons from the United Kingdom, *Social Policy Journal of New Zealand*, 20: 29–48.
Nutley, S., Walter, I. and Davies, H. (2009), Past, present, and possible futures for evidence-based policy. In G. Argyrous (ed.), *Evidence for policy and decision-making*, Sydney: UNSW Press, pp. 1–23.
Rudd, K. (2008), Address to heads of agencies and members of Senior Executive Service, 30 April, Canberra: Australian Public Service Commission.
Sanders, W. (2009), Ideology, evidence and competing principles in Australian Indigenous affairs: from Brough to Rudd via Pearson and the NTER, CAEPR discussion paper 289, Canberra: Centre for Aboriginal Economic Policy Research.
Sanderson, I. (2002), Evaluation, policy learning and evidence-based policy making, *Public Administration*, 80, 1: 1–22.
Sanderson, I. (2006), Complexity, 'practical rationality' and evidence-based policy making, *Policy and Politics*, 34, 1: 115–32.
Schubert, M. (2008), Media, contractors escape Aboriginal access bans, *The Age*, January 18.
Secretariat of National Aboriginal and Islander Child Care (2012), *Submission to the Senate Community Affairs Legislation Committee Inquiry into the Stronger Futures in the Northern Territory Bill*, North Fitzroy: Secretariat of National Aboriginal and Islander Child Care, http://www.snaicc.org.au/_uploads/rsfil/02785.pdf (viewed 15 March 2013).
Senate Community Affairs Committee (2010), *Inquiry into Social Security and Other Legislation Amendment (Welfare Reform and Reinstatement of Racial Discrimination Act) Bill 2009* and *Families, Housing, Community Services and Indigenous Affairs and Other Legislation Amendment (2009 Measures) Bill 2009* and *Families, Housing, Community Services and Indigenous Affairs and Other Legislation Amendment (Restoration of Racial Discrimination Act) Bill 2009*, Canberra: Australian Parliament.
Shergold, P. (2011), Seen but not heard, *The Australian*, 4 May.
Siewert, R. (2010), Dissenting report in Senate Community Affairs Legislation Committee, *Social Security and Other Legislation Amendment (Welfare Reform and Reinstatement of Racial Discrimination Act) Bill 2009 [. . .]*, Canberra: Australian Parliament, pp. 75–95.
Stevens, A. (2011), Telling policy stories: an ethnographic study of the use of evidence in policymaking in the UK, *Journal of Social Policy*, 40, 2: 237–55.
Stop the Intervention Collective Sydney (2011), Government discussion paper will keep NT Intervention architecture in place, media release, 22 June.
Taylor, D. (2006), Critical policy evaluation and the question of values: a psychosocial approach, *Critical Social Policy*, 26, 1: 243–67.
Valentine, K. (2009), *Critical Social Policy*, 29, 3: 443–64.
Wild, R. and Anderson, P. (2007), *Ampe Akelyernemane Meke Mekarle 'Little children are sacred': Report of the Northern Territory Board of Inquiry into the protection of Aboriginal children from sexual abuse*, Darwin: Northern Territory Government.
Wilson, L. and Karvelas, P. (2011), Evidence backs intervention, Macklin insists, *The Australian*, 11 November.
Wilson, L. and Rout, M. (2011), Greens jump on 'mixed' NT school truancy results, *The Australian*, 21 October.

4

A Systematic Review of Comparative Studies of Attitudes to Social Policy

Trude Sundberg and Peter Taylor-Gooby

Introduction

Systematic review (SR) has expanded rapidly during the past two decades as a means of evaluating research findings within topic areas. Advocates believe that SR has considerable value in summarizing key points across rapidly expanding fields of knowledge and is of particular relevance to policymakers. Critics see it as positivistic, limited and a creature of audit culture. The approach has been applied mainly to specific questions within relatively narrow fields of research in medicine, healthcare and education. However, the problems of dealing with large volumes of research activity and of understanding patterns, direction and the balance of evidence apply more generally in social science. The techniques and databases to enable rapid structured searches of large quantities of material have also developed rapidly. For these reasons, an examination of the potential for SR across fields where it has not yet been applied is timely. We add to current knowledge by critically assessing the use of SR as a tool for the evaluation of evidence, focusing on the first steps of a review.

This chapter examines the value of SR techniques in investigating the structure and directions of research across broader areas in social science to advance academic knowledge. Problems may arise when SR relies too readily on conveniently available sources of information. We take a middle position between advocates and critics of the approach, and argue that SR is of value in providing an overview of research and also in tracing issues in the way knowledge develops within an academic field. SR may help synthesize existing knowledge as a basis for further analysis. However, it must be applied cautiously so as not to induce a bias, taking into account the limitations of the available material, the shortcomings of databases and the importance of books versus articles in most social sciences. This chapter demonstrates the value and limitations of SR in a review of cross-national comparative studies of attitudes to social policy. It falls into three sections: the first is a discussion of recent debates about SR; the second is an analysis of the strengths and weaknesses of the approach, illustrating how these strengths and weaknesses

Evidence and Evaluation in Social Policy, First Edition. Edited by Ian Greener and Bent Greve. Book compilation

emerge in relation to social science through application of the method in a pioneering SR review of cross-national attitude surveys; and the third is a demonstration of the value of SR when used with caution in a presentation of substantive findings from the review. We draw conclusions about the field and about the potential for using SR more broadly in substantive areas of social science as a technique for the evaluation of evidence.

Systematic Review

SR could be described as 'one of the success stories of the 1990s. In the space of ten years the movement has had a significant impact on health care and policy' (Trinder and Reynolds 2000: 1). Some see its growth as a result of the development of meta-analysis as an alternative to traditional, narrative reviews of research literature (Hammersley 2001: 543). Others point to the increasing pressure to ensure that policy and practice rest securely on research evidence (Davies *et al.* 2000). Policymakers themselves increasingly seek research projects that can provide input and advice to help shape new policy interventions. SR has become a minor academic industry flourishing in the climate of evidence-based policy-making and driven forward by the demands for more rational resource allocation in healthcare, education and elsewhere, epitomized by the work of the National Institute for Clinical Excellence. It formed the core of the invitation of David Blunkett (Secretary of State for Education and Employment) to 'social scientists to work with us to find out what works and why and what types of policy initiatives are likely to be most effective. This is crucial to our agenda for modernising government' (Blunkett 2000). Academic engagement and research funding have established an institutional framework. The Economic and Social Research Council (ESRC) promotes work on SR through the National Centre for Research Methods (NCRM 2010). The UK Cochrane Centre and the EU- and ESRC-funded Evidence for Policy and Practice Information and Co-ordinating Centre were both set up in 1993, the National Health Service Centre for Reviewers and Dissemination (CRD) was set up in 1994 and the Campbell Collaboration in 1999.

Formal definitions of SR highlight structure, focus and transparency, 'reviews of this type: use a formal protocol to guide the process; focus on a specific question; identify as much of the relevant literature as possible; appraise the quality of the studies in the review and synthesise their findings' (Grayson and Gomersall 2003: 3). They rest on the assumption that it is possible to achieve unambiguous, comprehensive and unanimous agreement on the quality of a large number of research projects and apply this to achieve results:

> A systematic review has clearly identified questions, searches for relevant research following a procedure specified in advance, has criteria for which studies to include, has criteria for the information to be considered from these studies, appraises studies using clear criteria for what is good and what is less good research and synthesises the results in a transparent fashion, sometimes using statistical pooling (Bradshaw *et al.* 2000: 6)

From one perspective this is simply the application of reason to the evaluation of research evidence, 'systematic reviews are needed because they accumulate evidence from multiple studies which may prevent practitioners and policymakers from basing their decisions on single studies that might be flawed and biased' (CRD 2009: v). An alternative approach analyses SR from a sociological viewpoint as a shift in intellectual practice which contributes to the restructuring of the way research evidence is understood and thus to the way in which it can influence policy-making. In this view, one of the most striking developments during the past 30 years has been the development of a culture of audit and risk avoidance (Power 2004). As Pawson and Tilley (1997: xii) argue, 'we live in a knowledge-centred, value-adding, information-processing, management-fixated world which has an obsession with decision-making'. Analysts of risk society argue that individuals are now both better-informed and less deferential to experts (Beck 1992: Chapter 1; Wynne 1996). As a response to these societal changes, SR provides a technology to maximize the value generated from research and, at the same time, to demonstrate the authority of the evidence that informs policy. The point of SR is summed up in the claim that, in a less respectful world, 'organised distrust generates trustworthy reports' (Campbell 1984: 38).

Critics of SR often emphasize four issues: the principle of replication, the role of judgement, the use of a hierarchy of evidence, and the failure to understand the contextual nature of knowledge.

Replication is one of the defining principles of a systematic review, and is achieved through rigorous documentation and by making every step of the review transparent. One of the most influential critics of SR, Ray Pawson, argues that it is impossible to account for and document every point of a review. Furthermore, procedural uniformity may be counterproductive as flexibility and iteration may be crucial to appreciate the variety of evidence found. He advances a 'model of validity that rests on refutation rather than replication' (Pawson 2006: 26), as does this chapter.

Due to the wide range of methodological approaches and less standardized research processes in social science compared to natural sciences, the replication principle combined with hierarchical evidence approaches may exclude and simplify rather than systematize and enhance knowledge. We underline the importance of *judgement* in including and interpreting heterogeneity in studies, instead of eliminating it as an anomaly in the way that some meta-analysis software does (Petticrew and Roberts 2006: 197). This is because heterogeneity in itself may be valuable and expected in social policy. SR in this area needs to recognize this or risk reinforcing a skewed research synthesis which does not take real-world complexities into account.

Many studies organize findings according to a *hierarchy of evidence*, determined by claims about internal validity. From this perspective, SR and meta-analysis are given the greatest weight, followed by randomized controlled trials (RCT) with definitive results and then by RCTs which fail to generate definitive conclusions (Petticrew and Roberts 2006: 58). Knowledge obtained by other means (e.g. quantitative survey or qualitative analysis) is accorded correspondingly less weight.

This approach follows a natural science model and builds on techniques developed to test the value of interventions in medical settings where everything beyond a limited range of measured variables is assumed to be either controlled or irrelevant. Critics suggest that important research may be ignored (Hammersley 2001: 546; Evans and Benefield 2001) and the diversity of methods reduced as research is driven towards the medical model (Evans and Benefield 2001). They underline the importance of diversity in social sciences' methods that may bring about heterogeneity in a review, and argue that this should not be seen as a weakness but as contributing to understanding. More recently, manuals on SR have encouraged practitioners to consider explicitly 'which type of study is most appropriate for answering your review question', and have encouraged reviewers to take account of studies using a range of methods, including qualitative approaches, as appropriate (Petticrew and Roberts 2003: 59). However, positivistic models of replication and transparency still tend to predominate.

The fourth criticism underlines the need to consider *context* when assessing evidence of the effectiveness of an intervention (Kitson *et al.* 1998; Pawson 2006). SR traditionally focuses on 'what works' without taking into account the context in which interventions have taken place. Realist synthesis seeks to include these factors. It is concerned with explanation and is driven by the question 'What works for whom in what circumstances and in what respect?' (Pawson 2006: 18). It rejects the principle of replication and emphasizes that of refutation: studies are considered not so much in terms of whether they reinforce a consensus as whether their findings are disproved. We argue that issues of context also apply to the academic and scholarly context in which studies are shaped and conducted, for example in terms of dominant schools of thoughts and methodological approaches, which may shape their findings. Pawson argues that research studies are conducted in a complex social context and are never carried out in exactly the same way or in exactly the same setting (Pawson 2006: 6).

SR essentially sums the number of studies in the literature which point to specific conclusions and which meet particular quality criteria, often resting on assumptions about the hierarchy of evidence generated through particular methods. Realist synthesis argues, instead, that this process may discard evidence that is of value, despite limitations of method, and which can be used to support diverse interventions and programmes as appropriate in various contexts (Pawson 2006: 12). It sees SR as an iterative process that should be driven by scepticism and the principle of refutation, and where interpretation and theory is crucial. We agree with Pawson's point that the judgement of the researcher cannot be written out of the process, but must be made explicit and transparent:

> Science is more than excellence of execution, more than durability of data. Neither is it simply a product of following a pre-determined investigatory pathway or its course would be mapped out already. The stark, staring implication for this study is that there will never be a simple litmus-test for science; there are no instantaneous warrants for declaring certain procedures as valid science (Pawson 2003: 6)

SR has generated enormous interest at the interface between social science and policy-making. The above comments indicate that there are strong grounds for considering the appropriateness of particular methodologies and kinds of evidence in particular contexts, and that there is controversy about the limitations to positivism. SR needs to ensure that it does not exclude studies that point to relevant findings, while retaining a critical perspective that enables it to eliminate methodologically unsatisfactory work and appreciating the heterogeneity found in social science. It remains helpful for researchers to take an overview of work across a field, discover the key patterns within it, point to areas where work is concentrated and where there are fewer studies and report the chief findings supported by various approaches. As the Social Care Institute for Excellence argues, 'Systematic review methods can be applied to any type of question . . . transparent and replicable methodology should be applied to all forms of literature review in the interests of quality and reliability' (Coren and Fisher 2006: 2).

In this chapter we seek to apply techniques from SR in a broader field, the comparative analysis of attitude surveys. We use this work to illustrate problems with the use of SR in social policy. We focus on the early stages of SR – data collection rather than interpretation – since these processes have tended to receive less attention. The chapter ascribes to a position where explanatory schemes are not 'arrived at on the basis of the data, but rather on the basis of knowledge, theoretical formulations and assumptions, and logical analysis' (Pedhauzer 1983: 579). SR cannot substitute theory.

We argue in favour of the importance of an iterative process and of judgement, context and scepticism when conducting SR in social policy to acknowledge the diversity of the field. We manage large volumes of data through the use of random sampling. We reject assumptions about a hierarchy of methodology and seek to include the widest range of research, evaluated as appropriate to the various studies. However, the strengths and weakness of particular pieces of work must be taken into account. We also argue that the assumptions about replication and about formalized protocols often made in SR reviews of more technical literatures may be problematic in a social science context. Such issues may arise in more narrowly focused and natural science dominated fields due to differences in the ways databases are constructed and particularly in the rules governing the presentation of information in journal abstracts. They are much more relevant in social science due to differences in vocabularies and the definitions of terms between writers, the lack of any common structure to abstracts and the gaps in coverage of databases. These problems require a broad and pragmatic approach as our empirical work shows.

SR is not a royal road to knowledge, it is simply a convenient technique for the review of large numbers of studies, 'that aims to provide an account of the literature that is comprehensive, capable of replication and transparent in its approach' (Bryman 2008: 700). This approach does not follow assumptions about the hierarchy of methods. It requires considerable amounts of time to be spent in evaluating the contribution of particular studies. Its value is that it points to methods for systematically searching databases in order to reduce bias in the selection and management of findings from a number of studies.

This leads to an important pragmatic concern. Our research indicates that there are real shortcomings in the most conveniently available databases – some important material is omitted, thereby damaging literature reviews based on them. This suggests that SR may generate problems that are not immediately obvious to those using the approach less critically. These problems need to be addressed by using as wide a range of databases as possible and by supplementing them from other sources, rather than simply applying a protocol mechanistically to the material available from the most convenient sources. Transparency about the techniques used to generate the review is essential for the outcome of the approach to be fully understood by the academic community.

Illustration of Strengths and Weaknesses: The Attitude Survey Study

In this section we discuss attitude survey-based studies which serve as an illustration of potential problems with SR. This is a suitable area for examination because it is highly international and spans several databases. It illustrates the issues surrounding the adequate recognition of material from a range of sources. It has also expanded rapidly as an area of study in recent years. The studies for the most part centre on a relatively small number of major international surveys, so that there is a strong need for systematic comparison of different interpretations of the data. SR can contribute to this. We recognize that a pilot study may have some methodological shortcomings due to limitations of scope and sample (Teijingen and Hundley 2001), but show that the sample is an adequate foundation for a critical discussion of some central issues in applying SR to social policy.

Our aim in this research was to cover the full range of material generated in the field, ensuring that the studies included are comparative and cross-national and deal with material from attitude surveys. For practical reasons the research was limited to studies published in the English language covering Europe and the USA, Canada, New Zealand and Australia where the study also included European countries. Expert advice suggests that the most important omission is the German-language literature, but that the vast majority of published work is included.

To cover the full range of publications, the review included the literature listed in the International Bibliography of the Social Sciences (IBSS) and ISI Web of Science (WoS) databases (which are largely limited to articles in the journals included by the databases, with some conference papers), plus hand searches of the reference lists from the chapters of a recently published book which reviews the field and a research paper to control for differences in items identified through this method. We chose these two databases to include the leading international database with coverage across a wide range of disciplines, and a database more specifically focused on social science. We did not include Elsevier's Scopus database due to its natural science focus, 'two per cent of the material in Scopus is from social science, as against 14 per cent in WoS' (Jasco 2005: 1540). For this reason, and since we were under time pressure, we chose one database with wide coverage of subjects and journals

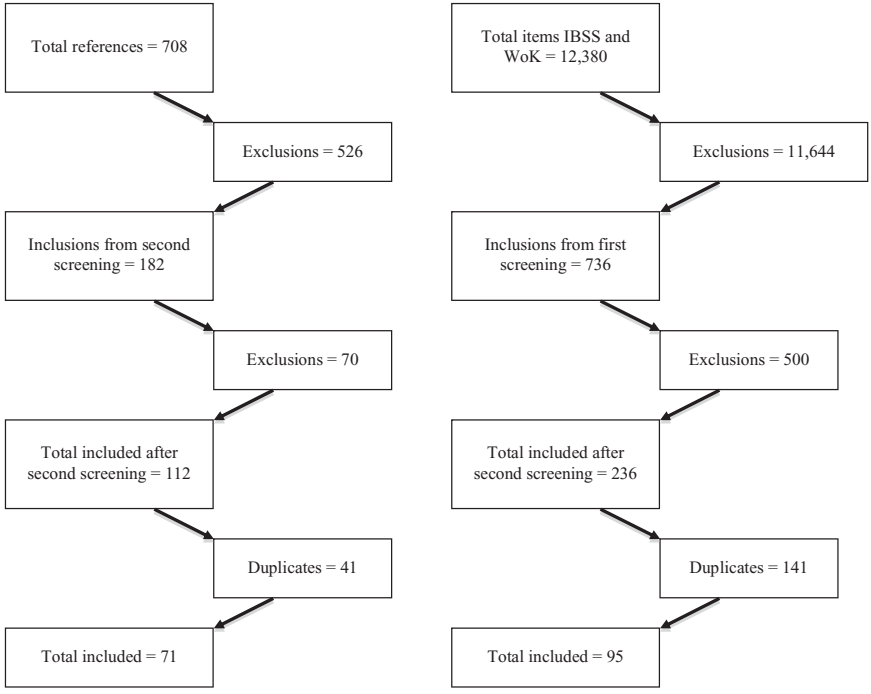
and one social science-focused database: WoS and IBSS. As Jasco argues, 'WoS comes the closest to be a genuinely pan-disciplinary database' (Jasco 2005: 1540). According to his calculations the distribution of records among the major component of WoS is: science 77 per cent, social sciences 14 per cent, and arts and humanities 9 per cent; whilst for scopus it is: health and life science 60 per cent, chemistry, physics, maths and engineering 25 per cent, biological, agricultural, earth and environmental sciences 13 per cent, and social sciences, psychology and economics 2 per cent (Jasco 2005: 1540).

Ninety-three search terms were identified in the field of welfare, state provision, welfare finance, citizenship, poverty, inequality, need, social justice and related normative issues, and combined using Boolean operators (see Appendix 2). The list of terms was refined from a longer list through an iterative process of consultation with 15 academics in social policy, sociology, criminology, economics, politics and psychology at the University of Kent.

The search was carried out in November and December 2009 covering the 15-year period between 1994 and 2009 and achieving more than 12,000 hits. Figure 4.1 summarizes the SR process. The material from each database and

Figure 4.1

Summary of inclusions and exclusions



that identified through the lists of references from books and chapters in the field was screened twice, once using the keywords and abstracts provided by the database and once against the inclusion criteria using the full study and the judgement of the researcher. Duplicate material was identified and removed. This follows the procedure recommended by the CRD (2009: 13).

As can be seen from figure 4.1, the initial search generated a high proportion of non-relevant hits, nearly 99 per cent being excluded by the screening process. One reason for this was the use of an inclusive search protocol that ranged over a considerable area in social science. Examination of the Cochrane and Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) databases indicates that proportions as high as this are common in more tightly-focused reviews in the medical and educational literature. In any case, a low proportion of relevant studies is to be expected in social science due to the fact that many journals do not impose rigour in titles, do not use keywords and do not insist on the inclusion of methodological information in abstracts (Petticrew and Roberts 2006).

Key findings

Five points emerge from the review and in studies of the databases:

- databases tend to over-represent US and under-represent other literatures;
- there are substantial differences in coverage between the different databases;
- WoS, with its assumptions and biases, tends to dominate database work;
- the vocabularies of key search terms tend to differ between disciplines so that a search from one perspective may miss relevant work from another perspective; and
- books, book chapters and research papers are systematically under-represented in all databases, despite being of considerable importance in many areas of social science.

We will present an overview of these issues and then illustrate and refine them, using the material from our own systematic review of the literature on attitudes to welfare. Our findings emphasize practical issues underlining the importance of using an iterative research process in which knowledge and judgement are crucial at every step when conducting SR in social sciences.

US over-representation

First and perhaps most importantly, much of the data available is biased towards the USA. This reflects the use made of the WoS database, which has much better coverage of US journals. This points the importance of gate-keeping processes regulating the inclusion or exclusion of journals from the databases. Braun and Dióspatony (2005), analyzing natural science publications, show that the USA, which has 54 per cent of the gate-keepers, contributes 32 per cent of the papers included. The UK is in second place with 10 per cent of the gate-keepers and correspondingly fewer papers, only 8 per cent

(Braun and Dióspatony 2005: 1549). On the other hand Almeida *et al.* (2009) argue that size of country and size of institution influence productivity, in terms of published papers, and thus numbers of citations, 'this is consistent with recent observations from several quality and efficiency criteria showing that the United States are outperforming Europe in terms of science' (Almeida *et al.* 2009: 135). About 36 per cent (4,142) of the titles found on WoS are US titles. Seventeen per cent (504) of IBSS titles are from the USA (numbers from JISC, Academic Database Assessment Tool) again showing the difference in coverage of the USA, in particular for the biggest database, WoS. Scholars need to take this bias into account in reading and conducting SRs.

Differences between databases

The databases differ substantially in their coverage of titles from different European countries (see table 4.1). This is of interest since comparative attitude studies are published in a number of countries and it is necessary to consider whether there are different national traditions. IBSS is rather weaker than WoS in German journals but particularly strong in those from France. Another interesting finding is the low overlap between WoS and IBSS in the titles covered: 16 German titles out of 368 are identified in both databases and 32 out of 839 French ones. Limiting inclusion to studies covering European countries ensured comparability between the social science surveys which the studies analyze. However, this creates problems when there are different national traditions in social science. Databases that vary in their coverage of a country's scientific contribution may bias the reviews based on them. For example, a comparison of the top 20 Polish sociologists cited in the Social Sciences Citation Index (SSCI) and a Polish index had only 12 names in common (Hicks 1999: 203). Hicks argues that this is due not only to the fact that some journals are not included in international databases, but also to the fact that studies with a national character, not addressed to an international audience, may get less attention. However, one might argue that any analysis of a phenomenon might be interesting and of use to others studying a similar theme or applying similar methodologies.

Table 4.1

English, German and French titles

Database	English	German	French
IBSS	65%	7%	28%
	1,570	161	680
WoS	96%	2%	2%
	9,398	207	159

Source: JISC, Academic Database Assessment Tool, data from 2010.

National differences in terms of coverage, and in the production of articles on certain topics, may result in a bias in the material included both in systematic reviews and other literature reviews. Citations and databases may not reflect the most important contributions and authors within a field, as Hicks shows. Citation rankings may influence the financing of institutions, judgements of the quality of a university, and evaluations of the strength of research in a country (Braun and Dióspatonyi 2005: 1548).

The databases searched will have an impact on what evidence is found in an SR. This underlines the importance of judgement and context when interpreting findings, and of an iterative process where changes may be made to a review protocol during the research process.

The dominance of Web of Science

The third point is the domination of WoS in the process of SR and in its coverage of a vast amount of journal titles. WoS provided the majority (8,332) of all initial hits in the attitude survey, roughly twice as many (4,048) as from IBSS, as can be seen in table 4.2. There was substantial duplication between the databases: 141 out of 236 items were excluded as duplicates when they were combined. The high quantity of hits made when searching WoS reflect the large number of journal titles covered by WoS so that it is essential to include this database in any review.

However, exclusive reliance on WoS may be misleading. To analyze the dominance of WoS, the JISC Academic Database Assessment Tool, which facilitates comparisons of databases, was used. The testing was done on WoS rather than Web of Knowledge (WoK), as WoS also includes conference proceedings, patents and open access material (http://wokinfo.com/about/faq/ [accessed 13 June 2010]) but the tool gives a good insight in the different coverage of the two databases. Table 4.3 shows that the two databases vary dramatically. WoS is roughly four times as large as IBSS, including 11,456

Table 4.2

Inclusions and exclusions in the review

	WoS	IBSS	Review of books and chapters
Initial references	8,332	4,048	708
1st screening exclusions	8,074	3,570	526
2nd screening exclusions	168	332	70
Total included	90	146	71
Duplicates between databases	141		–
Duplicates between databases and books and chapter items	14		
Total for analysis	152		

Table 4.3

Geographical coverage of IBSS and WoS

	IBSS	WoS
UK	25%	21%
Europe	39%	27%
World	36%	52%

Source: JISC, Academic Database Assessment Tool.

titles against IBSS's 2,712. The limited overlap is also striking. Less than 10 per cent of WoS titles are included in IBSS, making up 40.7 per cent of that database's titles. IBSS is clearly more centred on European (including UK) literature (table 4.3). This reinforces the earlier point that researchers should search more than one database to guard against bias. It is also important to include books for acceptable coverage of the areas in question. WoS does not include all the titles in IBSS, again underlining the importance of going beyond this database.

Table 4.3 also shows that IBSS has better coverage of Europe than WoS, but a more limited coverage of US material. Searches limited to WoS risk finding out what works only in a US setting.

Different traditions and search terms

It is advisable to include wide-ranging terms in order to ensure that the topics are fully covered. One issue that emerged was the relevance of terms that have emerged in other disciplines but are at the periphery of social policy debates (e.g. 'social capital', 'citizenship', 'legitimacy', 'identity') to the current interests of the subject (see Grayson and Gomersall 2003 for a discussion of changing emphases in social science literatures). The field is relatively broad, with changing emphases, and accommodating academic interests from different directions.

The under-representation of books

Substantial differences emerged between material identified from the screening of books, chapters and papers, and that from the databases which included almost exclusively articles (with some conference papers covered by WoS from 2009). This shows the importance of including such material in database searches in social science (see Hicks 1999: Table 4 'Overlap between IBSS and WoS'). 'The centrality of books in scholarly communication in the social sciences contrasts with their absence in literature databases, including the SSCI . . . journal based social science indicators will be problematic given the heterogeneous literature of the social sciences' (Hicks 1999: 201–2).

This point is rarely discussed in SR literature but matters if the technique is to be extended outside narrow medical and science-based literatures. Almost 10 per cent of the references culled from book and chapter reference lists were included in the final study after the two waves of screening, as against less than 1 per cent of those from the databases. The 112 book references remaining after the second screening included only 41 duplicates (37 per cent), compared with 141 out of 236 (nearly 60 per cent) for the databases. When the final material from the book references (71 items) and the databases (95 items) were combined, only 14 duplicates were found, less than 10 per cent.

The different databases cover substantially different literatures. WoS is much larger and US focused, IBSS has more on Europe and particularly on French literature. In addition, there is much material in reference lists, especially in books, that must be included because it is not covered in the databases. Thus an unbiased SR needs to use the full range of available databases, to be alert to overlaps and also to include as much material as possible from non-database sources, such as reference lists from book chapters.

We now develop these points further in the context of a small-scale study of attitudes to welfare, carried out to illustrate the potential of SR for bringing together material across a broader field to develop knowledge in social science.

A Demonstration of the Value of Systematic Review in Social Science and of the Issues of Quality and Bias

Our study was carried out to illustrate and develop methods. It emphasizes the importance of judgement and interpretation and, therefore, the laborious and time-consuming nature of the review exercise. For this reason, further analysis was based on the detailed review of a sub-sample of 24 studies, chosen at random from the studies remaining after the second stage of exclusion. A preliminary review of the content of the sample considered methodology in order to eliminate material where the method was unable to support the findings. However, some studies give few details of method, especially of justification for their choice of countries, and do not discuss issues of sample size and populations within countries. Second, technical advances, particularly the spread of multi-level modelling, time-series and fuzzy logic analyses, enable researchers to interrogate the data more searchingly. Since there is no commitment to a methodological hierarchy of evidence, studies were considered on their merits, in relation to the questions they sought to answer. Third, most researchers were limited to the analysis of the standard questions provided by major cross-national surveys such as the International Social Survey Programme, Euro-Barometer, the World Values Survey, the European Values Survey and the European Social Survey (see Appendix table 4.A1). These questions were often interpreted rather differently and applied to different themes and hypotheses. Again, commentary is based on the merits of the individual study.

The review of the content of the papers identified a number of points which reinforce and extend the general findings discussed in the previous section. These points concern the scope and orientation of studies and the importance of disciplinary perspective in this process.

Explanatory factors: individual characteristics and regime type

Studies often treat attitudes as related to individual characteristics (social class, gender, age, faith, political orientation, and so on) and features of welfare states (regime type, constitutional and institutional structure, variety of capitalism, security of employment, gender policies, and so on). In general, the studies treat individual characteristics as embedded within the macro-level features of welfare states. Multi-level modelling provides a methodology which enables researchers to take this approach much further and which is now being enthusiastically embraced (Van Oorschot and Uunk 2007; Esser 2005; Van Groezen *et al.* 2009; Anderson and Singer 2008). The most important macro-level characteristics used in the research concern regime type, following the model associated with Esping-Andersen (1990). Seven out of the 24 studies were based explicitly on tests of this approach and all found some confirmation of it.

The macro-level features of welfare states were used to combine characteristics in different ways and to examine different issues. Alesina and Glaeser (2003) found that ethnic heterogeneity has the strongest explanatory power in relation to level of social spending, while political institutions, geography, and presence of war also influence redistribution. Vala *et al.* (2004) found that whether a country had a tradition of emigration or immigration had an impact on solidarity with immigrants, while Van Oorschot and Uunk (2007) found that the proportion of immigrants in a country affects the level of public concern about immigration. Anderson and Singer (2008) show that level of inequality in a country impacts on satisfaction with the political system. Iversen and Soskice (2001) found that income equality, social spending and a nation's education system are related all help shape skills sets which, in turn, influence level of social policy preferences. Esser (2005) tested a range of macro-level indicators and found that regime type and unemployment rates influence attitudes to retirement.

Huseby (1995) showed that economic differences between countries and gross domestic product per capita both have explanatory power when analyzing attitudes about size and responsibility of governments. Roller (1995b) found that countries, standards of living, income inequality, wealth of a country and political agendas all have effects on attitudes towards socio-economic equality. Newton and Confalonieri (1995) and Pettersen (1995) found that political circumstances and national political mood impacts on support for paying taxes and attitudes to socio-economic security.

The influence of academic disciplines

Analysis of the studies by discipline of researcher shows a general pattern. Those from political science tend to focus on the political sphere and political

circumstances rather than the welfare state system and welfare outcomes (e.g. Borre and Scarborough 1995; Iversen and Soskice 2001; Anderson and Singer 2008; Gijsberts and Nieuwbeerta 2000). Economists tend to analyze attitudes to welfare in relation to self-interest and the provision and finance of benefits (Alesina and Glaeser 2003; Van Groezen *et al.* 2009; Iversen and Soskice 2001), while sociologists pay more attention to social contacts, social exclusion and deservingness (Mau 2003; Van Oorschot and Uunk 2007; Roller 1995a, 1995b).

The review of the content of the articles underlines the importance of a theory-led approach to synthesis to incorporate the heterogeneity of the studies and to use different findings from different approaches to advance knowledge. In a larger study, currently in train, this approach will be further developed in order to map the range and structure of attitudes, the circumstances in which they are found and who displays them.

Conclusions

Taken together, the findings from the general review of the databases and the study of the attitudinal literature reinforce the point that different researchers pursue cross-national comparative attitudinal studies from different perspectives and in order to answer different questions. Most SR is directed at identifying what social science research has to say about a specific problem or issue. It seeks the answer or answers that are best supported by the balance of evidence. The domain we have examined is broader and less conveniently structured. The research supports two points relevant to the extended use of SR techniques as methods for evaluating evidence in order to advance academic literature and theory-building rather than as simply a tool for assessing 'what works' in a defined context as a guide to policy-making:

- First, the most widely used databases are limited to particular literatures and differ in coverage. Searching within them suffers limitations in their capacity to identify material relevant to European contexts. Academic work must approach databases critically, ensure that a range of databases are included and also consider literature outside the databases as currently constructed.
- Second, books, book chapters and other material offer important sources of information that is, to some extent, independent from that contained in the articles that comprise by far the greatest proportion of the databases. SR review may be misleading if it focuses exclusively on the material conveniently available in electronic databases. There appears to be no current substitute for hand searches of bibliographies. It is possible to combine these with SR of the databases, provided these are approached critically and the hand searches are conducted systematically according to similar protocols.
- There is a risk that uncritical use of the electronic databases in SR academic work (and more broadly in literature reviews) may produce a perspective limited to material published in journals, which does not cover

all of social science knowledge and research, although it includes an important part of it, and fails to give adequate weight to any differences between US and other literatures.

Further points concern the substantive study of attitudes to social policy. Discipline pays a noteworthy role in directing the interests and approach of researchers. There is considerable variation in practice across the field. SR shows that comparative cross-national attitude studies, despite the number and sophistication of relevant publications, have yet to develop into a domain of study where it is possible to point to shared findings that provide a secure foundation for hypothesis-building and further development. The ideas tested come from other areas and reflect wide-ranging interests.

SR is likely to play an increasing role in evaluating social science evidence as the quantity and range of material to be synthesized grows, as it comes to be structured and cross-indexed within databases, and as databases enlarge their coverage. It has the potential to be extended from closely policy-related work to more theoretical and academic studies where the capacity to synthesize a very large range of material is attractive. We have shown a number of weaknesses in naïve approaches to SR particularly concerning databases: that different databases over-represent and under-represent different literatures; that all tend to ignore one of the more important channels of social science communication – books; and that disciplinary perspective shapes both the orientation of studies and of searches. We hope that these points will serve both as words of caution and also as encouragement to future researchers to use these methods alongside others, and to employ them critically, as they would other methods.

Appendix 1

Table 4.A1

Cross-national surveys used by the studies reviewed

Survey	Number of studies using the survey
ESS	2
ISSP	10
Eurobarometer	9
ISJP	1
ISEA	1
Political Action	3
EVS	1
WVS	2
Other cross-national	2
National surveys	6

Appendix 2: Search Strategy

(Attitudes AND (Welfare OR Religion OR Trust OR Political efficacy OR quality of service OR public officials OR political authorities OR political actors OR political system OR Political participation OR Social capital OR Individuals beliefs OR personal beliefs OR Postmaterial beliefs OR Citizenship OR Civic duty OR Civic virtue OR Civil liberties OR Deviance OR cheating on taxes OR bribery OR Abuse of welfare OR Social benefits make people lazy OR Pension OR Pensioners OR care for the elderly OR Ageing OR Class OR Poverty OR Social exclusion OR Entitlement to benefit OR State responsibility OR individual responsibility OR Government intervention OR ownership of social service institutions OR welfare support OR more government spend OR Less government regulation OR tax OR progressive taxation OR government spend OR welfare regime's generosity OR cuts and reductions in public services OR service delivery by private companies/civil society OR welfare provision by the state OR financing welfare OR Responsibilities of EU OR Employment OR Social protection OR Social justice OR Distributive justice OR Redistribution OR Inequality OR Opportunity OR Fairness OR Solidarity OR Social legitimacy OR legitimacy OR Social inclusion OR Social cohesion OR Risk OR Religious identity OR social identity OR Class identity OR National identity OR European identity OR Family OR new forms of family life OR traditional forms of family life OR Maternity leave OR Parental leave OR Duty of women OR Children OR Youth welfare OR Gender OR Women's rights OR Mental health OR Immigration OR migration OR Diversity OR Intolerance OR Cultural integration OR integration OR race OR disabled OR long-term sick benefits OR Housing benefits OR Future of welfare OR Future of public sector OR Financial concerns OR Confidence in the economy) AND Survey NOT (Japan OR Russia OR China).

Searches were limited to English language publications, published 1/1994–12/2009.

References

Alesina, A. and Glaeser, E. (2003), *Fighting poverty in the US and Europe*, Oxford: Oxford University Press.

Almeida, J. A. S., Pais, A. A. C. C. and Formosinho, S. J. (2009), Science indicators and science patterns in Europe, *Journal of Informetrics*, 3, 2: 134–42.

Anderson, C. and Singer, M. (2008), The sensitive left and the impervious right, *Comparative Political Studies*, 41, 4–5: 564–99.

Beck, U. (1992), *Risk Society*, London: Sage.

Blunkett, D. (2000), *Influence or irrelevance?* http://www.timeshighereducation.co.uk/story.asp?storyCode=150012&sectioncode=26 (accessed 13 June 2010).

Borre, O. and Scarborough, E. (eds) (1995), *The Scope of Government*, Oxford: Oxford University Press.

Bradshaw, J. R., Baldwin, S. *et al.* (2000), *Methodologies for socially useful systematic reviews in social policy*, York: University of York, http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=2&ved=0CDkQFjAB&url=http%3A%2F%2Fwww.esrc.ac.uk%2Fmy-esrc%2Fgrants%2FL141250001%2Foutputs%2F

Download%2F1566fec1-d035-48d1-844f-4a5fb65d68d8&ei=AQNKUfueGNCXo QWA8oHIDg&usg=AFQjCNFZG8-pX20_j-gByP1u6fGAfJ_dXQ&sig2=DofX-YerN9VRWnoFGiijcQ (accessed 20 March 2013).
Braun, T. and Dióspatony, I. (2005), The counting of core journal gatekeepers as science indicators really counts, *Scientometrics*, 52, 3: 297–319.
Bryman, A. (2008), *Social Research Methods*, Oxford: Oxford University Press.
Campbell, D. (1984), Can we be scientific in applied social sciences? In R. F. Connor, D. G. Altman and C. Jackson (eds), *Evaluation Studies Review Annual*, Newbury Park, CA: Sage, pp. 26–48.
Centre for Reviews and Dissemination (CRD) (2009), *Systematic Reviews: CRD's guidance for undertaking reviews in healthcare CRD*, York: University of York.
Coren, E. and Fisher, M. (2006), *The conduct of systematic research reviews for SCIE knowledge reviews*, London: Social Care Institute for Excellence.
Davies, H. T. O., Nutley, S. M. and Smith, P. (2000), *What works?* Bristol: Policy Press.
Esping-Andersen, G. (1990), *The Three Worlds of Welfare Capitalism*, Cambridge: Polity Press and Princeton, NJ: Princeton University Press.
Esser, I. (2005), *Why work? – Study 3*, PhD thesis, Stockholm: Swedish institute for social research.
Evans, J. and Benefield, P. (2001), Systematic Reviews of Educational Research, *British Educational Research Journal*, 27, 5: 527–41.
Gijsberts, M. and Nieuwbeerta, P. (2000), Class cleavages in party preferences in the new democracies in Eastern Europe, *European Societies*, 2, 4: 397–430.
Grayson, L. and Gomersall, A. (2003), *A difficult business*, ESRC UK Centre for Evidence Based Policy and Practice, WP 19, London: Queen Mary University of London.
Hammersley, M. (2001), On 'Systematic' Reviews of Research Literatures: a 'narrative' response to Evans & Benefield, *British Educational Research Journal*, 27, 5: 543–54.
Hicks, D. (1999), The difficulty of achieving full coverage of international social science literature and the bibliometric consequences, *Scientometrics*, 44, 2: 193–215.
Huseby, B. M. (1995), Attitudes towards the size of Government. In O. Borre and E. Scarborough (eds), *The Scope of Government*, Oxford: Oxford University Press.
Iversen, T. and Soskice, D. (2001), An asset theory of social policy preferences, *APSR*, 95, 5: 875–93.
Jasco, P. (2005), As we may search, *Current Science*, 89, 9:1537–47.
Kitson, A., Harvey, G. and McCormack, B. (1998), Enabling the implementation of evidence based practice, *Quality in Health Care*, 7, 149–58.
Mau, S. (2003), *The moral economy of welfare states: Britain and Germany compared*, London: Routledge.
National Centre for Research Methods (NCRM) (2010), *NCRM Methods for Research Synthesis Programme*, http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=188 (accessed 10 August 2010).
Newton, K. and Confalonieri, M. A. (1995), Politics, Economics, Class and Taxation. In O. Borre and E. Scarborough (eds), *The Scope of Government*, Oxford: Oxford University Press.
Pawson, R. (2003), *Assessing the quality of evidence in evidence-based policy*, WP 1, ESRC Research Methods Programme, Leeds: University of Leeds.
Pawson, R. (2006), *Flying 'The Tattered Flag of Enlightenment' in Evidence-based Policy*, London: Sage, http://www.leeds.ac.uk/realistsynthesis/resources.htm (accessed 13 June 2010).
Pawson, R. and Tilley, N. (1997), *Realistic Evaluation*, London: Sage.
Pedhauzer, E. J. (1983), *Multiple Regression in Behavioral Research: Explanation and Prediction* (2nd edn), Orlando, FL: Harcourt College Pub.

Pettersen, P. A. (1995), The Welfare State. In O. Borre and E. Scarborough (eds), *The Scope of Government*, Oxford: Oxford University Press.
Petticrew, M. and Roberts, H. (2006), *Systematic Reviews in the Social Science*, Oxford: Blackwell Publishing.
Power, M. (2004), *The Risk Management of Everything*, London: Demos.
Roller, E. (1995a), Political Agendas and Beliefs about the Scope of Government. In O. Borre and E. Scarborough (eds), *The Scope of Government*, Oxford: Oxford University Press.
Roller, E. (1995b), The Welfare State. In O. Borre and E. Scarborough (eds), *The Scope of Government*, Oxford: Oxford University Press.
Teijingen, E. and Hundley, V. (2001), *The Importance of Pilot Studies*, http://sru.soc.surrey.ac.uk/SRU35.html (accessed 18 May 2012).
Trinder, L. and Reynolds, S. (eds) (2000), *Evidence-Based Practice*, London: Wiley-Blackwell.
Vala, J., Lima, M. and Lopes, D. (2004), Social values, prejudice and solidarity in the European Union. In W. Arts and L. Halman (eds), *European Values at the Turn of the Millennium*, Leiden: Brill.
Van Groezen, B., Kiiver, H. and Unger, B. (2009), Explaining Europeans' preferences for pension provision, *European Journal of Political Economy*, 25, 2: 237–46.
Van Oorschot, W. and Uunk, W. (2007), Welfare spending and the public's concern for Immigrants, *Comparative Politics*, 40, 1: 63–82.
Wynne, B. (1996), May the sheep safely graze? In S. Lash, B. Szerszynski and B. Wynne, *Risk Environment and Modernity*, London: Sage.

Systematic review institutions and resources

Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-centre), http://eppi.ioe.ac.uk/cms/ (accessed 13 June 2010).
JISC, Academic Database Assessment Tool, http://www.jisc-adat.com/adat/home.pl (accessed 13 June 2010).
The Campbell Collaboration, http://www.campbellcollaboration.org/ (accessed 13 June 2010).
UK Cochrane Centre, http://www.cochrane.co.uk/en/index.html (accessed 13 June 2010).

5

Public Opinion and Policy-making

Ray Pawson and Geoff Wong

Introduction

> No written law has ever been more binding than unwritten custom supported by popular opinion (Carrie Chapman Catt)[1]

Studies of the relationship between public opinion and policy-making are legion. The literature on 'representative democracy' debates a normative question on the extent to which public policy ought to mirror the public mood (e.g. Burstein 1998). Another body of inquiry examines in great empirical detail the extent of the correspondence between policy commitments and the preferences of the public (e.g. Monroe 1998). Still further work has followed the nature of the linkage – in what circumstances and by which processes does public opinion impact on public policy (e.g. Page 1994)? A rather interesting methodological issue has yet to surface. Given that public opinion can be most fickle and given all the vexatious issues in gauging it, is the typical evidence on the collective conscience trustworthy enough to be relied upon as a vital consideration in policy-making?

This chapter seeks to answer this question and to suggest a strategy for adding authority to public opinion data. It does so via a case study in which we consider the empirical evidence available to policymakers contemplating the introduction of a law banning smoking in cars carrying children. To add a little verisimilitude and yet some anonymity to our account, we present the episode in a 'typical' policy context. Let us posit that a research review has been commissioned by the Department of Public Health & Safety in Her Majesty's Government in the ancient Kingdom of Erewhon.[2]

The research brief would undoubtedly cover a wider range of pertinent issues – whether in-car toxic risks justify legislation, whether the passage of law onto the statute books will survive lobbying by the tobacco companies, and whether the law once enacted would be enforceable. Evidence on these scores has been presented elsewhere (Wong *et al.* 2011) as has an account of the method, realist synthesis, which we used in conducting the review (Pawson 2006). Here, we concentrate entirely on what the evidence has to say about public support for this particular ban.

Evidence and Evaluation in Social Policy, First Edition. Edited by Ian Greener and Bent Greve. Book compilation

Erewhon has a significant per cent of adults who are long-term smokers and its own tobacco company (Erewhon International) to feed the habit. Whilst not in the forefront of tobacco control legislation, laws have already been introduced banning smoking on public transport, followed by indoor workplace restrictions, smoke-free restaurants, bars and pubs. There is a feeling in government circles, but no more than that, that these bans have met with public approval despite some initial pockets of resistance. There is a further hunch that child protection legislation governing, for instance, the use of car safety harnesses has produced high levels of compliance. Against this background, we arrive at the burgeoning policy predicament – is public opinion primed for the next legislative frontier, namely outlawing smoking in private cars carrying children?

Overview of the Evidence

There is a considerable battery of existing studies which can be called upon in gauging levels of support for banning smoking in cars carrying children. We begin by describing the broad contours. Some of the available evidence comes from studies tailor-made to answer this specific question, but most of it comes from fragments of data culled from large-scale population surveys, often telephone inquiries, seeking to gauge public opinion on a number of public health issues (Norman *et al.* 1999; McMillen *et al.* 2004; Walsh *et al.* 2002; Leatherdale *et al.* 2008). As well as monitoring the overall magnitude of support for the ban, there is a great deal of evidence charting the footprint of endorsement for in-vehicle smoking bans. There are studies relating support, or lack of it, to gender, education, social class, employment, poverty, region, cultural background, nationality, health status, and so on (Gonzales *et al.* 2006; Kegler *et al.* 2008; Kegler and Malcoe 2002; King *et al.* 2005; McMillen *et al.* 2004). Policymakers estimating support for their legislation need to pay heed to the voices of the relevant constituencies. Since a crucial task is to gauge headwinds to the passage of the law through this climate of public opinion, this body of data also relates to the attitudes of the most palpable foes, namely smokers. A further tranche of studies charts comparative levels of endorsement of bans in different micro-locations, a particular preoccupation being to juxtapose levels of support in the 'private' bastions of the car and the home (Norman *et al.* 1999; Kegler and Malcoe 2002; Kegler *et al.* 2008; Binns *et al.* 2009; King *et al.* 2005; McMillen *et al.* 2004; Gonzales *et al.* 2006; Leatherdale *et al.* 2008; Dunn *et al.* 2008). We have no space to set down and review all of the above-mentioned studies, therefore, four snapshots follow exemplifying some 'typical evidence' on the overall magnitude of support as well as some of its nuances. This is followed by a grand summary matrix of declared support for and/or the voluntary practice of bans by smoking status in table 5.1:

1. In the first major attempt to gauge opinion, Bauman *et al.* undertook a random household survey in New South Wales using 'standard methods for assessing the prevalence of smoking and attitudes. They reported their

Table 5.1

Levels of self-declared support for and/or practice of smoking bans in vehicles carrying children reported in included studies by smoking status (in date order)

Study	Year of survey data collection	Type of data reported (support and/or practice of ban)	Country	Cars		
				Non-smokers and smokers support/practice %	Non-smokers support/ practice %	Smokers support/ practice %
(Bauman *et al.* 1995)	1994	Support	Australia, NSW	72		63
(Norman *et al.* 1999)	1996/97	Practice	USA, California	66.0 (16.0*)		
(Walsh *et al.* 2002)	2000	Support and practice [in brackets]	Australia, NSW		58.8 [86.7]	44.7 [39.8]
(Kegler and Malcoe 2002)	2000	Practice	USA, Oklahoma State (North East)		67.4	12.8
(Binns *et al.* 2009)	2001	Practice	USA, Chicago		83.0	58.0
(King *et al.* 2005)	2001	Practice	USA, nationwide (African Americans only)		84.1	21.4
(McMillen *et al.* 2004)	2002	Practice	USA, nationwide	83.2 (urban) 68.7 (rural)**		
(Gonzales *et al.* 2006)	2003/04	Practice	USA, Albuquerque (Hispanics only)	81.0		
(Walsh *et al.* 2008)	2004	Support	Australia, NSW		55.6	50.5
(Leatherdale *et al.* 2008)	2004	Support	Canada, nationwide		91.8	72.9
(Kegler *et al.* 2008)	2004/05	Practice	USA, Georgia (South West)	36.8 (40.4*)		
(Leatherdale and Ahmed 2009)	2006	Support	Canada, nationwide		90.3	79.2
(Jalleh *et al.* 2006)	2006#	Support	Australia, Western Australia		87	80
(Thomson *et al.* 2008)	2007/08	Support	New Zealand, nationwide			95.9
(Dunn *et al.* 2008)	2008	Support	Australia, Queensland		82.9***	76.9***

Notes: * = % support if ban was only partial; ** = % support differentiated by rural and urban responders; *** = % support refers only to children aged 12 or under present in the car; # = year of publication.

results for the question, 'Do you think it should be illegal to smoke in cars when travelling with children?' as follows: 'of the 1461 adult responders, 72% agreed, 27% disagreed and 1% were undecided' (Bauman *et al.* 1995).

2. Walsh *et al.* (2002) report on a sub-sample of data from a cross-sectional (N = 656) omnibus survey from the Australian Cancer Education Research Program. This study provides an insight into the difference between the voluntary enforcement of bans and support for formal legislation mandating bans. In terms of voluntary bans, 86.7 per cent of non-smokers compared to 39.8 per cent of smokers enforced a complete ban on smoking in their cars. However, when asked if they would then support legislation to ban smoking in cars carrying children, support was reported at 58.8 per cent for non-smokers and 44.7 per cent for smokers.
3. A later Australian telephone survey assessed knowledge, attitudes and self-reported behaviour with regards to second-hand smoke in cars and homes (Dunn *et al.* 2008). Using random digit dialling they interviewed 1,026 adults in Queensland. Overall support for smoking ban in cars was high (although varying depending on the type of passenger) – 63.8 per cent (non-smoking adult passengers) to 80.5 per cent (children under 12 and pregnant women passengers).
4. The highest levels of support emerge from Thomson *et al.*'s public health survey of a national sample of 1,376 adults in the New Zealand *Smokers* survey (Thomson *et al.* 2008). This contained the item 'Do you think smoking should be allowed in cars with preschool children in them?'. Results which were weighted to reflect the national population of smokers showed that, '. . . 95.9% disagreed . . . and only 3.0% agreed with this question', leading the authors to conclude that, 'there appears to be almost universal support for not allowing smoking in cars carrying children, from smokers themselves'.

Uncertainties in the Data

Clearly, there are differences in reported levels of support for the proposed ban across these studies. The reviewer's task, however, is to seek pointers for policy implications and he or she might well report of the whole body of data that: there are impressive levels of general support for the ban (certainly against the mixed picture that emerges when other contentious issues are polled); smokers themselves are not overly hostile (and in some circumstances favour a ban); the broad 'direction of travel' seems to be moving in favour of a ban. Accordingly, it might be suggested, with the reviewer's customary caution, that there is evidence of prima facie public support for legislation.

Or is there? An important step in systematic review is the issue of 'quality appraisal' (Petticrew and Roberts 2006). Social research is prey to all kinds of error and bias. Indubitably, evidence-based policy cannot rely on any old evidence. Normally, a systematic review would undertake quality appraisal study by study. In lieu of this and for the sake of brevity, we consider the 'typical' methodological problems that inhere in all of the 'typical' studies

reviewed in the previous section. There will be few surprises here for denizens of public opinion polling. All studies suffer:

1. the potential pitfalls of question wording;
2. the lurking threats from the social desirability effect; and
3. the ever-present danger of sampling bias.

We rehearse the problems briefly here, not to rake over old methodological coals, but to assess the overall dependability of the evidence that might be fed back to the Ministry.

Problem 1, question wording, has proved a pitfall from the day questionnaires were invented, and the responses to questioning on health matters has been shown to be particularly sensitive to nuances of item wording, issue framing, scale labelling, and so on (McColl *et al.* 2001). Quick eyes will spot that (where reported) the above questions carry subtle differences of emphasis that might shape the willingness to support a ban. Public compassion might well differ for 'children', 'children under 12' and 'pre-school' children. Reponses on health preferences also differ significantly according to whether the choices put to respondents are one dimensional or offer comparators (Froberg and Kane 1989). Surveys carried out independently will rarely contain the same scenarios and response options. In this respect we reprise the variations noted above. Some but not all of the surveys raised the issue of smoking behaviour in cars across different risk groups. Some but not all asked about different locations in which smoking might be banned. Lastly, note that the key substantive question put to respondents elicits their ideas on the legitimacy of legalization. This issue may be phrased, as in the various examples above, as an item on whether smoking should continue to be 'allowed' or be 'banned'. Response on the former might be expected to be the mirror-image of the latter – but long and painful experience with the 'status quo' bias tells us that this is not necessarily so (Samuelson and Zeckhauser 1998). Put all of these difficulties together and, clearly, the decimal-point exactitude of the estimates claimed in some studies is severely blunted.

Problem 2 is another pernicious problem with such survey findings, the ancient bugbear of the 'social desirability effect' – the commonplace defensive tendency of respondents to portray their ideas and behaviour in keeping with perceived cultural norms (Phillips and Clancy 1972). In a nutshell, the problem here is that smoking addicts, who suffer routine stigma on top of slow poisoning, may well have chosen to dissemble. It is impossible, certainly through secondary analysis, to determine the extent of 'faking good' in the above research. What may be considered, however, are the circumstantial factors surrounding these surveys, all of which heighten the probability that many smokers may have preferred to pronounce on the side of the angels. Studies of the social desirability effect have shown that it is particularly prevalent when subjects belong to marginalized social groups (Ross and Mironsky 1984) and when the topics under investigation relate to personal and intimate activity in such areas as health behaviour, physical activity, drug and alcohol use (Aquilion 1994; Adams *et al.* 2005). Moreover, a landmark study on 'mode

effects' comparing face-to-face with telephone surveys has shown that 'telephone respondents were more suspicious about the interview process and more likely to present themselves in socially desirable ways' (Holbrook *et al.* 2003).

Problem 3, devising the appropriate sample, is always vital in any claim that research has captured public opinion and in this respect the secondary usage of existing survey material faces the sternest test. Trends that might apply in the open swards of New Zealand or in smoke-averse Australia might not apply to the policymaker's bailiwick. In general, it is true to say that the studies reviewed in the above section were all made in jurisdictions which were at least in the initial stages of contemplating such a ban. The poll that Erewhonian policymakers require is a sample of 'their patch, right now'. Even if we consider that these nations provided useful, indicative parallels, there is a further sampling issue – the ubiquitous problem of non-response bias. Especially when response rates are low (as in several of the studies noted above), there is the worrying possibility that findings are dominated by the opinions of those with an axe to grind. This relates closely to what health researchers refer to as 'systematic exposure misclassification bias' (Tripepi *et al.* 2010) – the notion that pre-existing experiences and commitments of participants heavily influence whether and how they engage in research (be it a survey or, indeed, a clinical trial). Different constituencies of smokers, ex-smokers and anti-smokers may, of course, be the axe-grinders in this instance, with a distinct possibility that pugnacious tobacco control supporters are over-represented.

Put all three issues together and the wise reviewer would probably conclude that, despite the strong prima facie evidence on endorsement for legislation, it is unwise to suppose that 'public support' is a singular concept and unwise to assume that public opinion research can provide us with a precise threshold measure of that support. We turn, accordingly, to potential solutions. The warrant for the proposed ban hardened under our initial review of the evidence, only to soften under closer methodological scrutiny. Is there a way of rethinking and retightening the policy inference? There are, of course, standard technical ameliorations to this trio of problems, which again we briefly acknowledge before considering their relevance in the systematic review process:

1. Solutions to wording problems are, of course, specific to the questionnaire items under scrutiny. Long established rules of thumb, however, tell us that one can handcraft better questionnaires by ensuring: that indicators and measure have an extensive programme of piloting and validation; that the multiple conjoined, components of attitudes are put under separate and systematic investigation; and that comparative and longitudinal studies are greatly improved by the use of common and standardized instruments (Converse and Presser 1986).
2. The social desirability effect may sometimes be detected by checking the internal consistency of a portfolio of questions on the behaviour under study and/or by seeding further desirable but improbable scenarios into the questions to gauge the respondent's general propensity to fake good (Nederhof 1985). It may be counteracted, up to a point and in some

respects, by the use of self-administered questionnaires, the selection of skilled and experienced interviewers, the use of proxy subjects and the 'bogus pipeline' technique (Podsakoff *et al.* 2003).

3. Non-response bias can be detected by comparing the demographic characteristics of respondents with population profiles from official statistics and insurance records (Korkeila *et al.* 2001). Findings then may be reweighted to fall in line with known risk factors. Another approach is the more aggressive follow-up of non-respondents, contacting them by more personal approaches, thus permitting a direct comparison of the opinions of willing and recalcitrant subjects (Hill and Roberts 1997).

Noting in passing that such strategies are seen as mitigations rather than solutions to stubborn problems, we return to the plight of the reviewer faced with a mass of survey research of uncertain quality. Do these strategies of methodological diagnosis and remedy help us with the quality problem? The answer, alas, is not one bit. Systematic review, by definition, is secondary research. In looking across dozens of primary studies it is possible to glean information on the broad designs in use and something about the population and context under study. In some cases, questionnaire items are published and one can often decipher unreported response categories from the statistical analysis. However, journal publication norms and word limits allow for precious little discussion of the nuances of such matters as question wording, desirability effects and response bias, and what is not there cannot be appraised (Tieman *et al.* 2010).

In the review in question, the research team gained the overall impression that very little attention was paid to the particular challenges we have highlighted here. Remember that much of the specific data on attitudes to in-car smoking come from one-off fragments of specially commissioned, broadly based public health surveys and it is unlikely that they carried any extensive item-by-item piloting, follow-up and standardization recommended above. In short and quite typically, the reviewer has the enormous advantage of the wide purview of evidence but this is counteracted by the lack of microscopic precision on the specifics of the research act.

This is a routine impasse in review work, and one (sadly common) solution is to conclude an exhaustive research review with the summons that 'more research is needed'. This might seem a particularly apt call in locations such as Erewhon in view of the fact that the majority of available research is from other jurisdictions; however, we consider it a poor and defensive response. The whole point of research synthesis is to bring timely evidence to decision-making. Commissioning further detailed research could provide the necessary local measure of local public opinion, but to do so with the requisite methodological sophistication would require pilot inquiries and an extensive multi-method strategy, which would not square with the high tempo of policy-making or the patience of the typical policymaker. We find ourselves back at square one – useful prima facie evidence on endorsement for legislation, based on data which, on closer inspection, has some inherent but intangible limitations. Some lateral thinking is needed to make more of the available evidence.

Dealing with Uncertainty

We arrive at the central thesis of the chapter – the most authoritative attitudinal evidence to support policy is not a matter of taking contemporary, error-free snapshots of public opinion but derives from a strategy of building and testing theories of how public attitudes are shaped. A more compelling source of certainty lies, we submit, in the process of *explanation building*. The import of public opinion on tobacco control or any other controversial measure is not just a matter of the percentage of affirmation here and the degree of support there. Attitudes are rooted in reasoning and those roots are open to investigation in systematic review. In the present case, this involves asking rather different questions – what accounts for these seemingly high levels of support, what in particular drives public concern, why is legislative compulsion deemed appropriate and, above all, why might smokers sanction restrictions on their own behaviour?

Responding to such questions prompts a different approach to systematic review – searching, realist-style, for potential theories that might explain the grounds for public support for legislation and then gathering evidence to test their relevance and explanatory power. This approach takes a review beyond inquiries that have focused on the rear seat of the car and thus examines original research on the components that might make up the supportive posture – beliefs about smoking, judgements about child welfare and experiences of tobacco control. Again, we omit details of how we sorted and sifted the many potential hypotheses (Pawson 2013). Here we concentrate on presenting the evidence on what we consider the three most telling explanations. Note once again that we present only 'typical' fragments of data rather than the full portfolio of evidence.

The 'invincible sub-text' of child vulnerability

Many studies, both quantitative and qualitative, indicate that smoking is already significantly modified in the presence of children. This pre-existing condition may help to account for the charitable view of a future in-car ban. One example is Gillespie *et al.*'s New Zealand survey, which reports on smokers' attempts to manage the risks of second hand smoke (SHS) (Gillespie *et al.* 2005). Questioned on the reasons for this selective abstinence, nearly half (46.5 per cent) said that this was because they recognized the 'vulnerability of children to SHS'. Setting a 'good example for children' was also reported as an important reason for not smoking in their presence (25.6 per cent).

To be sure, these data are also subject to distortions of question wording, desirability effects and response bias. They are buttressed, however, by research employing other methods. The 50 Scottish smokers in Phillips *et al.*'s qualitative study were questioned about the informal rules and tacit reasoning which guided their cigarette consumption. Many considered themselves 'considerate smokers' with the most important consideration being not to harm children, 'Many thought that children were particularly at risk because they were still developing'. Rather sophisticated 'hierarchies of avoidance' were also described. 'All respondents . . . described how they would temporarily

modify [smoking restrictions] in particular circumstances. For example, partial restrictions would become stricter in the presence of children and grandchildren, or relaxed if adult visitors were smokers' (Phillips *et al.* 2007).

This widespread perception of the yoke between the 'toxic risk' and 'vulnerability of children' is often seized upon in jurisdictions which have implemented the in-car ban. Several studies of policy process in smoking control have observed this motif. Moulton reports how a 'save the children' ethic was prominent during the preparation and enactment of the 2008 Bill banning smoking in cars carrying children in Nova Scotia, Canada. Local officials are reported as saying:

> The council . . . has once again reflected the conscience of the community while dealing with an issue on behalf of an element of the population, that is, children, that don't often have the right to change the environment in which they might find themselves . . . children don't often have a voice for themselves on health issues . . . cars are an environment where they can't walk away. (Moulton 2008)

The same thread emerges from a 'qualitative narrative review' by Freeman *et al.* summarizing the 12-year Australian history of advocacy for smoking bans in cars. The authors analyzed data from the print news media seeking out the ascendant themes. A key motif framed the issue as a concern for the protection of defenceless children with no choice about being exposed to SHS. This theme was explicit in 35 per cent of articles. Highly emotive terms such as 'gas chamber' and 'child abuse' (eight and ten instances respectively) were used to describe the conditions experienced by a child inside a smoky car, prompting the authors to conclude that 'Invoking the protection of vulnerable children in the debate about smoking in cars called up an almost invincibly powerful sub-text for advocates' (Freeman *et al.* 2008).

Even in these four short summaries, one can begin to see the foundation of a strong explanatory theme. Those smokers who reckon themselves 'considerate' and aware of their responsibilities for child protection are unlikely to obstruct a law framed and promoted explicitly as a child welfare issue.

The 'near universal expression of regret' about smoking

A different body of evidence teaches us that smokers, as a collective, are most unusual in the ranks of peer-groups and interest-groups in as much as they have an ambivalent attitude to membership of their group. Many studies, over the years, have demonstrated that significant proportion of them want to quit, and several different estimates exist in the literature. For example Lee and Kahende suggest that in the USA in any given year nearly two out of five cigarette smokers try to quit, but fewer than 10 per cent succeed (Lee and Kahende 2007).

Many studies have tried to understand the dissonance that follows as smokers come to terms with cessation failure in the face of addiction. We highlight the largest investigation attempting to gain a measure of the degree of that contrition. Fong *et al.* conducted a random-digit-telephone survey of

8,000 smokers across four countries (the USA, the UK, Canada and Australia) (Fong *et al.* 2004). Respondents indicated whether they strongly agree, agree, neither agree nor disagree, disagree or strongly disagree with the statement, 'If you had to do it over again, you would not have started smoking'. The proportion of respondents who agreed or strongly agreed was 90 per cent and nearly identical in each country:

> Regret was more likely to be experienced by older smokers, women, those who had tried to quit more often, those who perceive quitting as conferring health benefits, those with higher levels of perceived addiction, those worried about future damage to health, those who perceived smoking as lowering the quality of life, those who perceived the higher monetary costs of smoking, and those who perceived that smoking is not socially accepted. This predictive model was the same in all countries (Fong *et al.* 2004)

On the basis of this evidence, the authors declare that, 'Regret is a near-universal experience among smokers'.

Methodological caution about potentially leading question-wording, superficial telephone interviews and, in this instance, the possibility of 'faking sorrow' must again be expressed. But, as before, the validity of this particular thesis does not lie with this study alone but with the evidence from many other inquiries uncovering the dissonance in smoker's attitudes. For instance, Mullins and Borland's study began with a similar finding that only 12 per cent of smokers had no interest in quitting (Mullins and Borland 1996). Further investigation of those receptive to the idea of giving up demonstrated that most (75 per cent) are inhibited from taking action by the perceived enormity of the task. A later study of those wanting to quit reveals more detail on this ambivalence (Balmford and Borland 2008). Approximately one-third of respondents consider themselves simply too addicted to quit. The majority (70 per cent) believe that they lack what they perceive as the overriding virtue needed in smoking cessation, namely – willpower and the unswerving desire to quit.

We have a portrait in miniature here of two powerful apprehensions, the 'near universal expression of regret' set in the midst of 'inescapable personal characteristics'. Whilst this perceptual straightjacket presents a fearsome challenge for cessation programmes, it provides a sturdy platform for the legislation under contemplation here. A straightforward inference is that remorse about their own habit means that even hardened smokers will have little interest in conscripting a new generation of smokers. In short, smokers do not and will not tend to proselytize. 'Regret' may be a significant part of the explanation for the muted levels of opposition to the ban on smoking in cars carrying children and thus becomes a serviceable fragment of evidence explaining the lack of opposition to such a ban.

The steady march of 'denormalization'

It has already been noted that the vast majority of the available research emanates from counties in which tobacco control has already gained

purchase. An obvious corollary is that support for a particular ban is elevated the more extensive the previous legislation in that policy domain. If smoking bans have been enacted on public transport, followed by office and indoor workplace restrictions, smoke-free restaurants and, lastly, bars, pubs and gambling venues, then we may forward the hypotheses that public opinion is softened up for the next location, namely private cars. Known colloquially as 'domino theory' and in tobacco control circles as 'denormalization theory', this thesis has come under considerable research scrutiny.

Hammond *et al.* analyze further results from the four-nation study reported on earlier (Fong *et al.* 2004) that pertain directly to the smokers' perceptions of denormalization (Hammond *et al.* 2006). Most smokers (81 per cent) agreed with the statement that 'there are fewer and fewer places I feel comfortable smoking'. An identical percentage agrees that, 'society disapproves of smoking'. The cohort of respondents in the four-nation study was followed up in two further annual waves, and Hyland *et al.* report that in all countries there is increasing year-on-year support for comprehensive, smoke-free public places (Hyland *et al.* 2006).

Further research from the busy Fong, but this time in Ireland, probes the smoker's direct experience of legislation (Fong *et al.* 2006). Adult smokers were surveyed by telephone before and after the implementation of the Irish workplace smoke-free law in March 2004. The relevant and paradoxical finding was that even when smokers did not support such a ban at inception, they often did so after it was introduced:

> Increase in support was most dramatic in venues where pre-policy support was lowest (for example, bars/pubs and restaurants), suggesting that policy-makers that stay the course in implementing comprehensive smoke-free policies are likely to experience increased support among smokers after implementation. (Fong *et al.* 2006)

Whilst we have growing evidence here that smokers know that they face increasing societal disapproval and are inclined to limit tobacco use under that pressure, the domino theory is still in need of refinement. Insufficient studies exist to indicate whether legislative incrementalism has a precise cadence (an optimal order in which the dominos must fall, with car bans at point x of the sequence) or whether smokers will always concede under the brute force of tightening legislation. The exact grounds under which smokers accept bans remain unclear. A reprise of Phillips *et al.*'s qualitative study on 'smoking ethics' gives an indication that there may be a sticking point:

> Two factors emerged as important in how smoking restrictions were managed or moderated. These related to the meaning of the home and social identity. The home was seen as being a private space, a place of relaxation and comfort. This seemed to bring with it notions of choice about when and where respondents smoked in the home and about how others should respect their views (Phillips *et al.* 1972)

Such evidence suggests that smokers may *at some point* begin to resist the steady incrementalism of smoking bans. The vital question for our review is whether the ban on smoking in cars carrying children represents that point. Further data from Phillips *et al.*'s study suggest that it does *not*, 'Smoking restrictions in the car seemed to be more robust than in the home, suggesting that the car occupies an intermediary position between public and private space; its confined nature also seemed to encourage stronger rules' (Phillips *et al.* 2007).

The studies reviewed here begin to tackle an issue of great subtlety, trying to chart and interpret smokers' reactions to the tightening grip of smoking bans. Strong evidence exists demonstrating that smokers internalize and respond to the progressive weight of public opinion as embodied in legislation and, from that finding alone, there is support for the domino theory. We have further nuggets of evidence showing that this tendency to comply with bans may well extend to all places in the public purview.

Evidence for Policy

We come to our conclusions, particular and general. On the matter of smoking control in cars carrying children, an initial examination uncovers significant evidence from a great number of public opinion studies demonstrating support for a ban. However, all these studies betray the standard methodological weaknesses of survey research in trying to bridge the imponderable gap between what people say and what they do. Moreover, this evidence is narrow in its population base, emanating from localized studies in specific time periods. We conclude that *direct evidence* on public support for a ban must be considered a matter of tendencies and probabilities rather than uniformities and certainties.

We have suggested that a more compelling justification for such an intervention lies in the evidence explaining the grounds for public endorsement. In this case, the solidity of support is attested and reinforced in further research uncovering: beliefs about the vulnerability of children to second-hand smoke; sentiments of regret about taking up smoking; and acknowledgement of the burden of maintaining the smoking habit under incremental legislation.

More generally, this episode provides important lessons about the nature of the 'test of public opinion'. We have attempted to show that the warrant of public support for public policy is not a matter of demonstrating that public opinion has overhauled a particular popularity threshold. Rather, our case is made in terms of the *realpolitik* of policy-making. Time and again, political science has demonstrated that policies emerge from a process of 'muddling through' (Linblom 1959), of balancing many contingencies (Weiss 1980) and of trying to produce partial coherence out of contenting objectives (Mintzberg *et al.* 1976). Policy-making and legislation are always grounded in these broad 'decision agendas' rather than focused on singular 'decision points' (Kingdon 1995).

We should not suppose that the public are incapable of wresting with the same contingencies and thus posit that evidence on public opinion should examine 'popular decision agendas'. The reflections of a smoker deciding whether to comply with legislation forbidding smoking in cars carrying chil-

dren will be forged in his or her views on children's health, on his or her smoking identities and biographies, and on his or her coping strategies in the face of increasing restrictions. It is in the balance of such considerations that attitudes coalesce, and there exists a powerful research base on how the contending pressures are balanced.

Accordingly, our thesis is that the best evidence on offer to the policymaker does not take the form of a verdict, a thumbs-up (or down) from public opinion. Rather, it takes the form of an if-then proposition. If elements a, b, c, d . . . are in place, then the intervention x is more likely to be supported. We further suppose that advice on such explanatory configurations will be particularly useful for those policymakers, as in the case of Erewhon, where there is a shortage of local, contemporary research. Lawmakers in the present will then be in position to judge whether these background, generative conditions apply in their jurisdiction (sharp-eyed readers may already have already picked up some clues in respect of a, b and c in our pocket description of Erewhon). Such propositional advice sits more realistically with modern portrayals of the nature of knowledge transfer (Nutley *et al.* 2007). Evidence rarely directs the policy decision; rather it provides one base on which to inform the decision.

Lastly, a word on the scientific certainty, or otherwise, of the strategy we outline here. With Popper, we have argued that the legitimization of a hypothesis (in our case a policy proposal) rests on the testing of a web of main and subsidiary propositions. There is nothing absolute about the evidence brought to test any theory; there are no bedrock facts awaiting discovery that will seal the fate of an explanation. Rather, science proceeds by erecting a wide body of evidence upon which a theory stands, and explanations gain ground as supplementary theories are articulated and tested. This process, as Popper sees it, is perpetual. As with all hypotheses, any supplementary theory is provisional, falsifiable and thus open to further revision. Theories, rather like policies, only hold 'for the time being' (Popper 1992).

Applying this logic to our case studies, we note that some research suggests that one of our secondary theories about denormalization may have limited application. In some circumstances, being an outcast may foster resistance. Thompson *et al.*, for instance, argue that smoking-related stigma combined with the inability to quit may lead to 'smoking islands' – micro-environments in which truculent opposition to bans and to quitting may fester (Thompson *et al.* 2007). Whilst acknowledging the truth of this, witnessed by many such gatherings outside workplaces and pubs (at least here in Erewhon), we would propose the counter-hypothesis that there is little reason to suppose that cars might provide a collective focus for such resistance. We may be right, we may be wrong on this particular matter. Our point is that the process of building conjectures and seeking evidence, for or against, is how evidence-based policy will approach (but never reach) certainty.

Acknowledgement

This work was supported by the UK Economic and Social Science Research Council (ESRC – RES-189-25-0007).

Notes

1. http://womenshistory.about.com/od/cattworks/Carrie_Chapman_Catt_Writings_Speeches.htm (accessed 19 March 2013).
2. We have borrowed the name of Samuel Butler's fictional country for this chapter. It approximates to *Nowhere* spelt backwards. The author, incidentally, wished it to be pronounced as a word of three syllables, all short – thus, E-re-whon.

References

Adams, S., Matthews, C., Ebbeling, C., Moore, C., Cunningham, J., Fulton, J. *et al.* (2005), The Effect of Social Desirability and Social Approval on Self-Reports of Physical Activity, *American Journal of Epidemiology*, 161: 389–98.

Aquilion, W. (1994), Interview Mode Effects in Surveys of Drug and Alcohol Use, *Public Opinion Quarterly*, 58: 210–40.

Balmford, J. and Borland, R. (2008), What does it mean to want to quit? *Drug and Alcohol Review*, 27: 21–7.

Bauman, A., Chen, X. C. and Chapman, S. (1995), Protecting children in cars from tobacco smoke, *BMJ*, 311, 7013: 1164.

Binns, H. J., O'Neil, J., Benuck, I., Ariza, A. J. and Pediatric Practice Research Group (2009), Influences on parents' decisions for home and automobile smoking bans in households with smokers, *Patient Education & Counseling*, 74, 2: 272–6.

Burstein, P. (1998), Bringing the Public Back In: Should Sociologists Consider the Impact of Public Opinion on Public Policy? *Social Forces*, 77: 27–62.

Converse, J. and Presser, S. (1986), *Survey Questions: Handcrafting the Standardized Questionnaire*, Thousand Oaks, CA: Sage.

Dunn, J., Greenbank, S., McDowell, M., Mahoney, C., Mazerolle, P., Occhipinti, S. *et al.* (2008), Community knowledge, attitudes and behaviours about environmental tobacco smoke in homes and cars, *Health Promotion Journal of Australia*, 19: 113–17.

Fong, G., Hammond, D., Laux, F., Zanna, M., Cummings, K., Borland, R. *et al.* (2004), The near-universal experience of regret among smokers in four countries: Findings from the International Tobacco Control Policy Evaluation Survey, *Nicotine & Tobacco Research*, 6: S341–S351.

Fong, G., Hyland, A., Borland, R., Hammond, D., Hastings, G., McNeill, A. *et al.* (2006), Reductions in tobacco smoke pollution and increases in support for smoke-free public places following the implementation of comprehensive smoke-free workplace legislation in the Republic of Ireland: findings from the ITC Ireland/UK Survey, *Tobacco Control*, 15: 51–8.

Freeman, B., Chapman, S. and Storey, P. (2008), Banning smoking in cars carrying children: an analytical history of a public health advocacy campaign, *Australian & New Zealand Journal of Public Health*, 32: 60–5.

Froberg, D. and Kane, R. (1989), Methodology for measuring health-state preferences-II: scaling methods, *Journal of Clinical Epidemiology*, 42: 459–71.

Gillespie, J., Milne, K. and Wilson, N. (2005), Secondhand smoke in New Zealand homes and cars: Exposure, attitudes, and behaviours in 2004, *New Zealand Medical Journal*, 118, http://journal.nzma.org.nz/journal/118-1227/1782 (accessed 19 March 2013).

Gonzales, M., Malcoe, L. H., Kegler, M. C. and Espinoza, J. (2006), Prevalence and predictors of home and automobile smoking bans and child environmental tobacco smoke exposure: a cross-sectional study of U.S.- and Mexico-born Hispanic women with young children, *BMC Public Health*, 6: 265.

Hammond, D., Fong, G., Zanna, M., Thrasher, J. and Borland, R. (2006), Tobacco Denormalization and Industry Beliefs Among Smokers from Four Countries, *American Journal of Preventive Medicine*, 31: 225–32.

Hill, A. and Roberts, J. (1997), Non-response bias in a lifestyle survey, *Journal of Public Health in Medicine*, 19: 203–7.

Holbrook, A., Green, M. and Krosnik, J. (2003), Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires, *Public Opinion Quarterly*, 67: 79–125.

Hyland, A., Travers, M., Higbee, C., Cummings, K., Dresler, C., Carpenter, C. *et al.* (2006), *A MultiCountry Comparison of Levels of Indoor Air Pollution in Different Workplaces*, Buffalo, NY: Roswell Park Cancer Institute, International Agency for Research on Cancer, Harvard School of Public Health.

Jalleh, G., Donovan, R., Stewart, S., and Sullivan, D. (2006), Is there public support for banning smoking in motor vehicles? *Tobacco Control*, 15: 71.

Kegler, M. C., Escoffery, C. and Butler, S. (2008), A qualitative study on establishing and enforcing smoking rules in family cars, *Nicotine & Tobacco Research*, 10, 3: 493–7.

Kegler, M. C. and Malcoe, L. H. (2002), Smoking restrictions in the home and car among rural Native American and white families with young children, *Preventive Medicine*, 35, 4: 334–42.

King, G., Mallett, R., Kozlowski, L., Bendel, R. B. and Nahata, S. (2005), Personal space smoking restrictions among African Americans, *American Journal of Preventive Medicine*, 28, 1: 33–40.

Kingdon, J. (1995), *Agendas, Alternatives and Public Policies* (2nd edn), New York, NY: Harper Collins.

Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Rautava, P., Helenius, H. *et al.* (2001), Non-response and related factors in a nationwide health survey, *European Journal of Epidemiology*, 17: 991–9.

Leatherdale, S. T. and Ahmed, R. (2009), Second-hand smoke exposure in homes and in cars among Canadian youth: current prevalence, beliefs about exposure, and changes between 2004 and 2006, *Cancer Causes & Control*, 20: 855–65.

Leatherdale, S. T., Smith, P. and Ahmed, R. (2008), Youth exposure to smoking in the home and in cars: how often does it happen and what do youth think about it? *Tobacco Control*, 17: 86–92.

Lee, C.-W. and Kahende, J. (2007), Factors associated with successful smoking cessation in the United States, *Am J Public Health*, 97: 1503–9.

Linblom, C. (1959), The Science of 'Muddling Through', *Public Administration Review*, 19: 79–88.

McColl, E., Jacoby, A., Thomas, L., Soutter, J., Banford, C., Steen, N. *et al.* (2001), *Design and Use of Questionnaires: A review of best practice applicable to health staff and patients*, Report No. 5, Norwich: HMSO, http://www.hta.ac.uk/fullmono/mon531.pdf (accessed 19 March 2013).

McMillen, R., Breen, J. and Cosby, A. G. (2004), Rural-urban differences in the social climate surrounding environmental tobacco smoke: a report from the 2002 Social Climate Survey of Tobacco Control, *Journal of Rural Health*, 20: 7–16.

Mintzberg, H., Raisinghani, D. and Theoret, A. (1976), The structure of 'unstructured' decision processes, *Administrative Science Quarterly*, 21: 246–75.

Monroe, A. (1998), Public Opinion and Public Policy 1980–1993, *Public Opinion Quarterly*, 62: 6–28.

Moulton, D. (2008), Stop that smoking car! *CMAJ Canadian Medical Association Journal*, 178: 19.

Mullins, R. and Borland, R. (1996), Do smokers want to quit? *Australian and New Zealand Journal of Public Health*, 20: 426–7.

Nederhof, A. (1985), Methods of coping with social desirability bias: A review, *European Journal of Social Psychology*, 15: 263–80.
Norman, G. J., Ribisl, K. M., Howard-Pitney, B. and Howard, K. A. (1999), Smoking bans in the home and car: Do those who really need them have them? *Preventive Medicine*, 29, 6, Pt 1: 581–9.
Nutley, S., Walter, S. and Davies, H. (2007), *Using Evidence: How research can inform public services*, Bristol: Policy Press.
Page, B. (1994), Democratic Responsiveness? Untangling the Links between Public Opinion and Policy, *Political Science and Politics*, 21: 25–9.
Pawson, R. (2006), *Evidence-based Policy. A Realist Perspective*, London: Sage.
Pawson, R. (2013), *The Science of Evaluation: A Realist Manifesto*, London: Sage.
Petticrew, M. and Roberts, H. (2006), *Systematic Reviews in the Social Sciences: A practical guide*, Oxford: Blackwell.
Phillips, D. and Clancy, K. (1972), Some Effects of 'Social Desirability' in Survey Studies, *American Journal of Sociology*, 77: 921–40.
Phillips, R., Amos, A., Ritchie, D., Cunningham-Burley, S. and Martin, C. (2007), Smoking in the home after the smoke-free legislation in Scotland: qualitative study, *BMJ*, 335: 553, http://www.bmj.com/content/335/7619/553 (accessed 19 March 2013).
Podsakoff, P., MacKenzie, S., Lee, J. and Podsakoff, N. (2003), Common method biases in behavioral research: A critical review of the literature and recommended remedies, *Journal of Applied Psychology*, 88: 879–903.
Popper, K. (1992), *Conjectures and Refutations: The Growth of Scientific Knowledge* (5th edn), London: Routledge.
Ross, C. and Mironsky, M. (1984), Socially-Desirable Response and Acquiescence in a Cross-Cultural Survey of Mental Health, *Journal of Health and Social Behavior*, 25: 189–97.
Samuelson, W. and Zeckhauser, R. (1998), Status Quo Bias in Decision Making, *Journal of Risk and Uncertainty*, 1: 7–59.
Thompson, L., Pearce, J. and Barnette, J. (2007), Moralising geographies: stigma, smoking island and responsible subjects, *Area*, 39: 508–17.
Thomson, G., Wilson, N., Weerasekera, D. and Edwards, R. (2008), Ninety-six percent of New Zealand smokers support smokefree cars containing preschool children, *New Zealand Medical Journal*, 121: 139–40.
Tieman, J., Abernethy, A., and Currow, D. (2010), Not Published, Not Indexed: Issues in Generating and Finding Hospice and Palliative Care Literature, *Journal of Palliative Medicine*, 13: 669–75.
Tripepi, G., Jager, K., Dekker, F. and Zoccali, C. (2010), Selection Bias and Information Bias in Clinical Research, *Nephron Clinical Practice*, 115: c94–c99.
Walsh, R. A., Paul, C. L., Tzelepis, F., Stojanovski, E. and Tang, A. (2008), Is government action out-of-step with public opinion on tobacco control? Results of a New South Wales population survey, *Australian & New Zealand Journal of Public Health*, 32: 482–8.
Walsh, R. A., Tzelepis, F., Paul, C. L. and McKenzie, J. (2002), Environmental tobacco smoke in homes, motor vehicles and licensed premises: community attitudes and practices, *Australian & New Zealand Journal of Public Health*, 26, 6: 536–42.
Weiss, J. (1980), Dilemmas of evaluating a balancing act: Policies for prevention of child abuse, *Policy Studies Journal*, 8: 1222–8.
Wong, G., Pawson, R. and Owen, L. (2011), Policy guidance on threats to legislative interventions in public health: a realist synthesis, *BMC Public Health*, 11: 222, http://www.biomedcentral.com/1471-2458/11/222 (accessed 19 March 2013).

6

Obstacles to Evidence-based Policy-making in the EU Enlargement Countries: The Case of Skills Policies

Will Bartlett

Introduction

The EU enlargement countries[1] have been severely hit by the global economic crisis (Bartlett and Prica 2011). Over the previous decade they experienced rapid growth based on external borrowing and inflows of credit and foreign direct investment, while another major source of external finance was remittance income, which reflected an exodus of skilled workers from many countries of the region. Several countries face an ageing population and an increasing burden of pension costs and health costs although some, such as Albania and Kosovo, have a youthful and rapidly growing population. In the face of these pressures, on the one side financial and on the other side demographic, it is likely that public expenditure on education will experience increasing constraints in the future. The steady inflows of foreign funds are unlikely to return over the next few years, implying a need for a new growth model that relies more on internal sources of finance and growth (European Commission 2010). The Eurozone debt crisis has put into doubt the accepted wisdom of slow but steady progress with EU enlargement. The enlargement countries are likely to find that the inflows of development funds to support EU-friendly reforms will also become less generous in the years to come. Although this may have negative effects on the pace of reform, it may also open up some new space for independent policy-making in the potential candidate countries. These countries may find that the need for an exclusive focus on the implementation of the *acquis communautaire* is less pressing than for the countries further developed in their accession process. In all countries, however, there is a need for better policy-making in all areas of social and economic policy response to these developments. In the past, the enlargement countries relied heavily on outside advice on the direction of policy, whether from international institutions or individual consultants (Deacon and Stubbs 2007). This has led to a situation in which the administrative capacity for independent policy-making has been weakened. Nevertheless, under the new circumstances, governments will need to develop a greater capacity for better policy-making that enables an improved performance of public services, given the constraints imposed by reduced funds. Under such circumstances, the

Evidence and Evaluation in Social Policy, First Edition. Edited by Ian Greener and Bent Greve. Book compilation

claims of the proponents of evidence-based policy as means to deliver improved public policy seem to offer a new way forward for policymakers in the region and also for their international advisers and donors.

The Origins of Evidence-based Policy-making

Evidence based policy-making has been around in various forms for a long time, under various names. Howlett (2009) locates its origins in the US experience with large-scale planning in areas such as defence, urban redevelopment and budgeting in the 1960s. The modern use of the term seems to date from experience in the health sector with the development of evidence-based medicine in the early 1990s (Black 2001). This led to the call by clinicians to extend the concept to the wider policy field, although it is still in the area of health policy that the procedures of evidence-based policy-making seem to have found their most extensive practical application, especially through the work of the Cochrane Foundation. The New Labour government that came to power in the UK in 1997 proclaimed that policy would in future not be decided by ideology but would be based on the best scientific evidence. The vision was to modernize government and to ensure improved public services by focusing on 'what works' irrespective of political preferences.

However, the evidence-based policy-making approach has come under criticism on a number of grounds. In a classic early article on the policy process, Lindblom (1959) argued that for complex policy problems the 'rational-comprehensive' method of policy-making is impractical and a process of 'successive limited comparisons' is all that can be achieved, making marginal adjustments from 'where we are now' rather than radical adjustments based on a complete survey of all possible means to achieve a variety of potential ends. It is held to rely on a linear relationship between evidence and policy and on an assumption of rationality in the policy-making process (Black 2001; Clarence 2002). It has also been pointed out that governments often use evidence selectively as a source of legitimacy (Walker 2000). For example, in deciding on introducing performance-related pay in schools in England and Wales, the Labour government perversely failed to follow its own prescriptions on evidence-based policy despite the lack of evidence on the effectiveness of the change and negative feedback from consultations. It did so for pragmatic reasons, in order to attract new teachers without breaking public sector pay inflation targets (Farrell and Morris 2009).

It has also been pointed out that governments may also lack the 'policy analytical capacity' to carry out effective evidence-based policy studies, and that that the attempt to do so may reduce resources available for the normal operational activities (Howlett 2009). This criticism was originally voiced by Lindblom (1959) in his analysis of 'muddling through' as a rational and realistic approach to the policy-making process.

Evidence-based policy and systematic review

According to the UK Department for International Development (DfID):

> evidence-informed policy is about decisions based on careful use of the most up-to-date evidence. Making policies and decisions in this way increases the success of policies, their value for money and their impact by basing decisions on what we know. This is important in international development, where limited funds are targeted at some of the world's most pressing problems.

While there is often much primary evidence available about various policy interventions, it is rarely systematically and neutrally laid out and disseminated to decision-makers. Policymakers and practitioners rarely have the time to assess the evidence base for each policy or practice question. They rely on single studies, advice from well-placed experts or traditional and unsystematic scoping studies or literature reviews. In other words, policy is typically made on an incremental basis, in which the 'means' and 'ends' are considered jointly rather than the best policy (means) being separately assessed in the light of well-defined objectives (ends). Policy in practice is typically carried out by a process of 'muddling through' rather than being a rational decision based upon the best available scientific evidence (Lindblom 1959).

But individual studies, no matter how rigorous or scientific, are not a sufficient evidence base from which to make informed policy and practice decisions. The purpose of a systematic review is to sum up the best available research on a specific question using transparent procedures to find, evaluate and synthesize the results of relevant research. Procedures are explicitly defined in advance, in order to ensure that the exercise is transparent, can be replicated and to minimize bias. Studies included in a review are screened for quality, so that the findings of a large number of studies can be combined. Peer review is a key part of the process; qualified independent researchers control the author's methods and results. Systematic reviews have been used in health, education and social policy to meet this need. They are a well-established and rigorous method to map the evidence base in as unbiased a way as possible, assess the quality of the evidence and synthesize it. They can be presented in a way to make it easier for policymakers and practitioners to understand the body of evidence and as a strong foundation on which to base policy and practice decisions.

Policy evaluation and pilot projects

Once policies have been implemented, they should be evaluated for effectiveness. Increasingly, policies are tried out on a pilot basis and subjected to evaluation before being rolled out on a national level. For example, in the late 1990s the Labour government in the UK began to pilot a number of policies in areas such as crime prevention, employment and welfare policy, health, education and local government. Innovative policies for the labour market introduced under the New Deal programme for disadvantaged groups (young people, lone parents, disabled, long-term unemployed) were all subject to initial piloting and evaluation before being extended on a national level (Hasluck 2000).

However, this approach may encounter a number of difficulties (Sanderson 2002). The first is the length of time for the effects for the policy intervention to become apparent and for pilots to give a good insight into how the policy might work when rolled out nationally. In addition, there have been cases in which policy has been changed during the period of the pilot project so that evaluation becomes impossible to fulfil. This may happen due to changes in political priorities of the government. Politicians may not be able to wait long enough for the pilot programme to be implemented due to electoral considerations. Pilot programmes may also be difficult to evaluate due to the effects of other exogenous changes in the environment that cloud the identification of impact. Cases have been noted in which several pilot programmes have been carried out in the same area through initiatives to target deprived populations, such as Health Action Zones and Education Action Zones, that were developed under the Labour government in the UK in the 1990s. The simultaneous implementation of several interventions may make it difficult to isolate the separate impact of any one of them. There may also be 'Hawthorn' effects through which the subjects of the pilot programme enjoy a 'pioneering spirit' among the professionals involved and so receive a better service than under normal conditions; in addition, the pilot intervention may be better funded than a national programme would be, giving spurious positive results to an evaluation. Lastly, in cases in which policymakers have already decided on a course of action, a pilot programme may be used more for exemplifying rather than experimenting, in order to demonstrate how a programme would work and develop instances of 'good practice' for other frontline workers to follow. In this context, piloting is more akin to 'prototyping', to ask questions about 'how it works' rather than questions about 'whether it works'.

Despite the critique of pilot project evaluation, in some cases pilot programmes have provided sufficient data for a robust impact analysis. In other cases 'natural experiments' can be identified in which some groups have been subject of a policy while others have not. This can happen when comparing experience across countries, for example. In this case a methodology has been developed to assess the effectiveness of policies known as the 'difference-in-difference' approach (DID). The key idea is drawn from medical control trials in which one group of patients is treated with a specific drug and another with a placebo. The difference in health outcomes in the treated versus the non-treated groups is measured and compared to the initial difference in health condition. The difference in these differences is measured and tested for statistical significance. The idea has been taken up in the evaluation of the effects of policies in the social field. Individuals involved in a policy intervention are viewed as a treated group, while those not covered by the policy are viewed as non-treated. The difference in 'before' and 'after' differences is the measure of the effectiveness of the policy. Hanushek and Woessmann (2006) use the DID approach to evaluate the effect of early tracking in schools on the level and distribution of student performance. Comparing across countries, they conclude that early tracking increases inequality in achievement, while there is little evidence of efficiency gains. A rare DID evaluation study in the enlargement countries has been carried out in Serbia. Based on the findings of a survey carried out by the United Nations Development Programme

(UNDP), the evaluation study shows the active labour market programme 'Beautiful Serbia' had only a small impact in generating new long-term employment (Bonin and Rinne 2006).

The Policy Process in Transition Economies

In transition economies, including those in the enlargement region, there has been an enormous debate on the type and nature of institutional reform policies that have underpinned the transition from a state-owned and managed economy to a market-based economy. Institutional reforms in transition economies can be seen in the outcome of a policy process which involves a political struggle between pro-reform and anti-reform elite groups and political coalitions which have specific interests in the outcome. In the transition literature there has been a long debate about the relative influence of 'winners' and 'losers' on the transition process. According to one account, the potential losers from the transition process are likely to resist reform, and present the reform process with severe political constraints (Roland 2002). Members of the old elites – including managers of state-owned enterprises and the top echelons of the security establishment who prefer the status quo to radical reform – may mobilize the losers – including workers thrown out of their jobs as a consequence of the privatization and restructuring of state-owned enterprises – into opposition to reform. In order to minimize this opposition, pro-reform leaders should establish a social safety net to compensate vulnerable groups for their losses (Kramer 1997). Another view holds that it is the *winners* from reform who are the most dangerous opponents of the reform progress (Hellman 1998). These winners are the new elites who prosper from the early stages of reform. They include managers of large privatized enterprises, politically well-connected tycoons who gained privatized assets at bargain prices, media barons and directors of public institutions who owe their positions to political connections, and political leaders who represent these groups. According to this view, in a partially reformed economy, new elites establish monopoly positions that provide opportunities for rent-seeking and strive to prevent further reforms that would undermine their new privileges.

In a more general version of this type of approach, Sabatier (1988, 2000) identifies 'policy subsystems' as a unit of analysis to understand the policy-making process. A policy subsystem is a group of actors from a variety of public and private organizations actively concerned with a particular policy issue. Within the subsystem, the theory assumes that actors can be aggregated into a number of 'advocacy coalitions' whose members share a set of normative and causal beliefs. Advocacy coalitions develop strategies and policy programmes in accordance with their beliefs in order to advance their policy objectives. Conflicting strategies can arise between advocacy coalitions, for example in relation to specific reforms. Political conflicts about whether to introduce reforms or resist them are represented through such advocacy coalitions. Sabatier's main insight is that such policy conflicts are often mediated by a third group of actors known as 'policy brokers'. Each advocacy coalition brings different forms of evidence to bear in the debate over a policy.

The policy broker is an agent or institution with the legitimacy to decide which body of 'evidence' is valid in legitimizing the adopted policy.

The Policy Process in the EU Enlargement Countries

Political parties in the EU enlargement countries in the Western Balkans are often closely related to the business sector. In some cases this amounts to the capture of the state by business interests so that policies are driven strongly by business interests rather than by any substantial evidence of the effectiveness of policy in achieving social goals. It also leads to the phenomenon of political patronage or 'partyisation' of the public sector in which jobs are handed out to political supporters and party members (Kopecky and Spirova 2011). Rarely do political parties have a comprehensive programme on which they fight elections and subsequently implement a consistent set of policies informed by available evidence. Often, a position as a government minister is seen as a way to promote one's own business interests. In some countries, pervasive corruption at high levels distorts the policy-making process (Vachudova 2009). In most Western Balkan countries there is no permanent civil service that could carry policy lessons from one administration to the next. Bureaucrats at a high level are often replaced when a new political party comes to power. Furthermore, poor pay and conditions have also contributed to high turnover among civil servants, disrupting project implementation and reducing the effectiveness of the public administration. This makes the policy-making process vulnerable to the imposition of policy advice from external advisers and donor agencies (of which more below) and weakens the ability of the domestic civil service to engage in a long-term policy learning process.

The example of skills policies

Skills have emerged as a central concern of transition countries in their attempts to develop more competitive market economies (Bartlett 2007; Commander and Kollo 2008). Skill shortages and surpluses of various types have appeared in these countries as a consequence of economic restructuring since newly created jobs typically require different types of skills to those that have been destroyed. This process of restructuring and the subsequent expansion of the demand for new skills has often taken place more rapidly than the education and training systems have been able to adapt. The skill mismatches that have subsequently emerged have often been an obstacle to labour reallocation from low to high productivity sectors and have slowed down the rate of economic growth (Brixiova *et al.* 2009). Moreover, it seems that skill mismatch is a more permanent phenomenon in transition countries than in developed ones, resulting in high levels of long-term unemployment. In addition, skill mismatches increase with the age of workers, rather than falling with age as typically occurs in more developed market economies. Long-term unemployment is consequently a serious problem in transition economies, especially affecting older workers with obsolete skills, while youth unemployment is generally extremely high (Kolev and Saget 2005).

The main challenges to the development of skills policies include the weak capacities of government institutions including public employment agencies, underfunding of provision of training services by the state, inadequate certification and accreditation of skills, slow reforms of the vocational education and training (VET) systems and qualification frameworks, and a low level of in-house training by employers. The VET sector, in particular, seems to be an example of a policy field in which it is difficult to introduce significant change. In the education sector, trade unions are relatively strong and form a powerful advocacy coalition resisting policy reform. Despite many donor-funded projects, the VET curricula in most countries are out-dated. While transition has involved large-scale structural change, the VET system has not adapted to provide the trained graduates that are needed by business. Moreover, there is a dearth of research-based evidence on which relevant policies could be formulated (Murthi and Sondergaard 2010).

For example, in Montenegro the '[l]ack of adequate research on the skills for which there is a widespread interest in the market, is a constraint for the education system in providing a good offer or labour supply for the labour market' (Golubovic 2011). Despite this evidential vacuum, the Employment Agency has introduced a programme ('A Job for You') to help people find jobs in northern Montenegro where unemployment is a severe problem. The programme includes several specific projects for different segments of the labour market and involves job subsidies and public works. Various programmes of vocational training, retraining and specialization are offered by the Employment Agency in collaboration with local employment offices and VET centres, mainly for disadvantaged people, but these are limited in scope. Despite the importance of this measure, no policy evaluations have been carried out to determine the success or otherwise of the effectiveness of this policy. Although the Employment Agency carries out employer surveys to identify labour market vacancies, no wider evaluation of research evidence is used to inform policy-making.

Weak capacities and underfunding

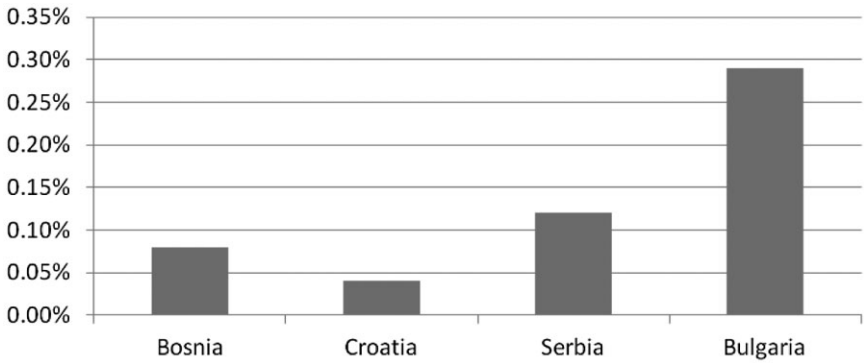
Civil service reforms in the enlargement countries have generally been slow and subject to significant delays (Nunberg 1999). It has been argued that the legacy of the communist past has led to weakness in the public administration in post-communist countries, associated with institutional instability and discretionary modes of governance (Dimitrov *et al.* 2006). Political parties have often intervened in personnel management decisions (Meyer-Sahling 2008). The extent of such legacies differs across countries, with those countries with a legacy of strong administrations under the pre-communist period having possibly better performance also in post-communism (Meyer-Sahling 2009). In this respect, the countries of the Western Balkans appear to have especially weak public administrations with high degrees of discretionary behaviour and political interference. The claim that the EU enlargement countries are held back in their reform progress by weak administrations has been so widely accepted that it has become one of the key drivers of the EU pre-accession assistance effort. Institution building is the first component of the current IPA

programme[2] through which financial and technical assistance is channelled to the enlargement countries. The weak capacity of public employment services in these countries to carry out policy evaluations of active labour market programmes has been noted by Johnstone (2006). The economic crisis measures that have been introduced throughout the region have involved caps on public sector employment in several countries. This has affected the ability of the institutions to carry out skills gaps analysis. In relation to policy towards VET more generally, a recent evaluation has concluded that 'the main problem is that too little use is made of the available data for policy making . . . [and] . . . there is no culture of evaluating policies and reforms' (ETF 2012: 17).

In addition to current restrictions on budgets, the field of employment and skills policy has been significantly underfunded for many years. As shown in figure 6.1, the expenditure on active employment measures is lower in the Western Balkans than among neighbouring countries in the EU. There are therefore very limited funds available for retraining, and skills gap analysis is therefore needed to better target scarce funds towards the creation of skills where gaps are most severe. For example, in Serbia the National Employment Service (NES) has 2,000 employees of whom only five work in the Department of Statistics and Analysis. In 2009, the NES was required to dismiss 200 people due to the government's austerity programme. The Department for Additional Education, which deals with matching needs on the labour market, has insufficient expert staff available to carry out skills gap analyses,[3] while the NES is unable to employ more analysts who could specialize in this area. All of this, in the context of weak administrative capacity at the best of times, has led in most of the EU enlargement coun-

Figure 6.1

Proportion of gross domestic product spent on active employment measures (%, 2008)



Source: Statistical Bulletin No. 2, Centre of Public Employment Services of Southeast European Countries.

tries to great difficulties in learning from evidence and taking research-based evidence into account in designing policies to improve policy effectiveness. At least, this is so in relation to domestically generated evidence. However, in this vacuum, other agents have entered with a far greater capacity to generate research evidence, namely the international organizations such as the World Bank, the UNDP and the European Commission, which consequently have a disproportionate influence on the design of skills policies in the enlargement countries.

International Agencies and Skills Policy in the Western Balkans

Governments frequently draw lessons from experiences in other countries in developing their own policies (Rose 1993). In addition, policy may be transferred across national settings either voluntarily or on a more coercive basis. An analysis of this 'policy transfer' approach has been made by Dolowitz and Marsh, who argue that policy transfer takes place through a variety of channels, at various levels of government, and through a variety of different agents of transfer, including central government, local government, international aid agencies and non-governmental organizations (Dolowitz and Marsh 1996, 2000). While earlier studies emphasized the role of government-to-government transfers, Stone (2004) has emphasized the importance of policy transfers from international agencies to national governments, and the role of transnational policy networks. The concept of policy transfer is highly relevant in the context of the complex systems of multi-level governance that can be observed in the enlargement countries. The debate centres on the identification of different patterns of policy transfer, and the agents and networks through which the transfer takes effect.

The policy transfer literature has identified three issues that are of relevance to the issue of effectiveness of international assistance in the context of endogenous policy formation. The first is the extent to which policy transfer is voluntary or coercive. This issue has been debated in research into conditionality surrounding the EU enlargement process. An important question is how far the appearance of voluntary transfer based upon the mechanisms of participation, ownership and consultation in fact disguise more or less effectively the practice of coercive policy transfer, and how aid effectiveness differs when associated policy transfers are based on coercive or voluntary transfer processes.

The second issue concerns the sources of policy convergence and divergence. As Stone (2004) points out, much of the institutionalist literature suggests that policy convergence is likely to take place where structural factors such as globalization, industrialization or – as in the case of the EU enlargement countries – Europeanization, are driving the adoption of similar policies in different national settings. However, Stone warns against an overly deterministic approach and argues that there may also be significant scope for agency in mediating the nature and effects of policy transfer. Indeed, if policy is endogenous, and the policy transfer process is influenced by the principal-agent interaction between donors and recipients, then there is likely to be

substantial scope for agency to affect the outcome of the process (Bartlett 2010). In such a case, structural path dependencies may be overcome by agents' conscious decisions so that policy transfer processes may lead to policy divergence rather than convergence.

The third issue concerns the circumstances under which policy transfer leads to policy failure. Dolowitz and Marsh (2000) focus attention on three potential sources of policy failure. Policy transfer may fail because policies are uninformed by effective assessment and understanding of the local context, a situation which is likely to occur if policy design is too heavily influenced by the external donor organization and if the donor has interests which are not aligned with those of the recipient agency. Second, policy transfer may fail if the transfer process is incomplete, which can occur if there are conflicts among donor agencies concerning the correct policies to be transferred along with the financial assistance package. Third, policy transfer may fail if transfer is inappropriate, which can occur for similar reasons to the case of uninformed transfer.

Policy-making in the enlargement countries of the Western Balkans is heavily influenced by policy transfer arrangements from numerous international organizations in the area of skills policy as in other policy areas. As a recent report by the European Training Foundation has identified:

> The donor community now greatly influences whether education and training sector reform is taken up as a policy priority, what focus such reform should take and how it is to be developed and implemented. . . . Coping with this and avoiding jumping into automatic policy taking and policy copying under the pressures from many international discourses and EU processes of education and training policy is problematic for policymakers. (Nikolovska 2007: 105–6)

Prominent among the international organizations which influence domestic policy-making in the region is the European Commission, which has adopted increasingly stringent conditionality in relation to the European integration processes. EU conditionality applies to all the stages through which Western Balkan countries must pass in their process of EU accession. The standard accession conditionality, known as the Copenhagen criteria, includes provisions on the creation of functioning market economies, the ability to take on the obligations of membership, on democratization, and having sufficient administrative capacity to implement the *acquis communautaire*. Most recently, specific short-term and medium-term reform goals leading towards harmonization with the *acquis communautaire* have been set out in the European Partnerships devised for each country by the European Commission. A central component is public administration reform and administrative capacity building. However, as Dimitrova (2002) has noted, this may lead to impermanent and opportunistic reforms that are readily reversed once accession has taken place. The leverage that the EU is able to bring to bear to ensure that the desired policies are adopted is therefore substantial. It is monitored through its diplomatic presence in the EU Delegation offices in each country, and enhanced through its control over

the disbursement of the IPA assistance programme. The effectiveness of EU policy transfer is further limited by domestic administrative and cultural traditions, political leadership and domestic reform management (Eriksen 2007). Moreover, as Noutcheva (2009) argues, domestic actors have indulged in fake compliance, partial compliance or non-compliance with the EU's conditions. Such responses can readily be seen through the creation of various new EU-model government agencies throughout the region, such as the Agency for Adult Education and the Agency for Vocational Education in Croatia and other similar agencies elsewhere. These often have an essentially self-referential basis and consequently have little impact on the real needs of their prospective clients. Often, such agencies were disbanded following accession (Randma-Liiv *et al.* 2011). Thus, the real influence of EU policy transfer on policy practice should not be exaggerated (Fagan 2010).

Another major external actor is the World Bank, which has resident representatives installed in each of the Western Balkan countries in country offices, with large research staffs involved in advising on policy in a wide range of policy fields guided by Country Assistance Strategies. The World Bank carries out a large research programme in various social policy fields including employment, education, health, pensions and social assistance. Most recently, it has produced a volume of research findings on vocational education and skills that is based on research carried out in Macedonia and other transition countries in the region (World Bank 2011).

A third international actor in the region is the Organisation for Economic Co-operation and Development (OECD) Investment Compact (IC) which sees its role as contributing to setting policy priorities in the enlargement countries. It addresses mainly economic issues but has an interest in education and skills policies. It produces a regular research report known as the *Investment Reform Index* that tries to benchmark countries according to peer performance on a range of issues related to the economic environment. In early 2011, the Secretariat of the *Investment Reform Index* requested 'country economic team leaders' to identify four to five policy priorities resulting from its most recent policy assessment. Project proposals related to these policies were debated at round tables in each country and policy priorities were ranked. Governments were then invited to select a policy that would be funded. However, governments did not necessarily select the projects for which there was much evidence of effectiveness, relying instead on the basis of subjective criteria. This example illustrates on the one hand how international organizations intervene in the policy process and, on the other hand, how domestic actors are nevertheless sometimes able to bring their own opaque interests to bear on policy implementation.

A fourth major international policy actor in the region is the UNDP with country offices throughout the region. The UNDP promotes a social agenda with respect to poverty alleviation and support for disadvantaged groups. It carries out a number of policy related research projects through which it attempts to influence government policy in a large number of social policy fields. The UNDP has carried out numerous surveys on social conditions and quality of life. One of the most recent is a survey of young people's transitions

from education to work (UNDP 2011). However, surveys such as these tend to be one-off affairs and there appears to be little attempt to conduct systematic reviews of the research literature.

All these international agencies are engaged in various processes of policy transfer. These sometimes conflict, although there is a surprising degree of informal coordination in which each agency takes responsibility for a specific policy area. Nevertheless, the need to respond to several different 'principals' in the process of policy design backed by substantial donor resource transfer often enables local administrations to engage in gaming of the various donors. Moreover, the attempt to transfer policies from the outside may be less effective than expected due to resistance to reform by the bureaucracies which have their own entrenched interests (Eriksen 2007). This sometimes gives the appearance of policy failure in the form of inconsistency in policy design and weakness in policy implementation. It may also, to some extent, underlie the widely noted weak administrative capacity to which no amount of 'capacity building' or externally funded training programmes will provide a solution, since their origin is in more fundamental structural causes as described above – namely, the multiple principals and the rational response of a domestic administration to their attempts at policy transfer.

As one practitioner put it to me:

> . . . where the donors want to put money, that is the priority. The Qualifications Framework is one good example. These issues have profound impact on the bleak reality of evidence-based policy making in Enlargement countries, but also on entire policy process. Everything becomes a priority – there are countries with six strategies for the sector of education, all to be implemented from 2012 to 2014 and all done together with different donors. To name few of the Strategies: Strategy for LLL, Strategy for VET, Strategy for Career guidance, than entrepreneurial learning, etc . . . I am wondering how this is possible – keeping in mind that all policies 'end' in the schools.

Conclusions

The global economic crisis has had a significant impact in the EU enlargement region through a sharp reduction in the inflows of external finance. Inflows of credit and foreign direct investment have dropped sharply and are unlikely to return to previous levels in the short to medium term. 'Euroisation' of debt, especially private debt, has created a situation in which it is difficult for countries to adjust exchange rates, and international competitiveness has suffered as a result leading to low prospects for export growth. Unemployment has begun to rise again, in a context in which high unemployment levels, long-term unemployment and youth unemployment were already significantly serious structural problems. Governments have sought to restrict their government budget deficits with varying degrees of success, and it is clear that for years to come there will be little scope for large increases in government expenditure in the region. Turkey is an exception to some extent as it has

weathered the economic crisis in better shape but, even there, significant structural reforms are needed to support continued economic growth. These effects of economic crisis have highlighted the need for better policy-making in the region, drawing far more than hitherto on a better understanding of the causes of economic and social problems and a better appreciation of the range of policy options and their relative chances of success or failure. This understanding can only be gained through a far greater use of evidence on 'what works and why' drawn from an engagement with the findings of policy-relevant research studies. It must be acknowledged that such studies are relatively rare in the EU enlargement region and so there is not an enormous evidence base on which to draw. Nevertheless, valuable lessons can be learnt from experiences in other countries, especially in the EU, provided care is taken to avoid the major pitfalls of such policy transfer processes. However, there remains a substantial local knowledge gap that can only be filled by additional resources devoted to well-designed research studies with research questions that are relevant to and informed by the needs of policymakers.

In recent years, the case for greater attention to evidence-based policy-making has been accepted in many European countries and internationally. Although having a longer lineage, the use of evidence-based policy-making has accelerated since the late 1990s. Although widely applied since then, the approach has come under a concerted critique from policy analysts and specialists in policy studies. These have pointed out that policy-making is not a rational linear process going from the definition of ends, the gathering of evidence, the formulation of a solution and to its implementation. Rather, policy takes place in a context of bounded rationality, in which political decisions are subject to conflicting pressures from advocacy coalitions and policy transfer from external agencies and institutions. Consequently, researchers and other 'policy brokers' need to better understand the policy process and engage in an informed dialogue with policymakers in order to form a realistic appreciation of what might be feasible. In this context, policymakers are likely to find it easier to implement small incremental changes rather than root and branch reform that might sometimes be suggested by disinterested and abstract reviews of the available research evidence (the cases in which 'shock therapy' has been tried in the transition countries mostly led to drastic negative unintended consequences).

Nevertheless, evidence-based policy-making techniques have a valuable role to play in improving the policy process in most countries and, as argued above, could have an especially important role in the context of the EU enlargement countries in the aftermath and continuing reverberation of the global economic crisis. Whichever of the various techniques of evidence-based policy-making is chosen, there needs to be a commitment from policymakers to provide sufficient funding to gather the evidence, and a sufficiently long-term perspective to take the findings into account in designing and implementing policy.

The chapter has discussed the specific nature of the policy process in transition countries and the difficulties of formulating rational policy during periods of rapid structural change in which administrations have become politicized, in which state capture by big business interests is common

(Hellman *et al.* 2003) and in which the alternative scenario of 'partyisation' of the business sector produces specific biases and interests in the formulation of policy which might be inimical to the use of the best available evidence put forward by researchers. In addition, in the EU enlargement countries, pervasive policy transfer, often of a coercive nature, is an additional constraint on rational policy-making and the conflicting advice received from multiple donors and external advisers only provides an incentive for playing the system and producing inconsistence policy formulas. Furthermore, weak administrative capacity for the development of evidence-based policy-making is a significant problem, not confined only to those countries but also to developed countries where limited policy analytical capacity has also been observed. The widespread use of 'strategies' and 'action plans' derived from externally driven policy advice only underlines the weakness of the domestic policy community in the region in its ability to develop independent policy programmes based upon the best available research evidence. This suggests that there is significant scope for improvement in policy-making through the use of evidence-based policy-making techniques. Governments, therefore, encourage the use of systematic review and ex-post evaluation of policy programmes and analysis of natural experiments where possible, while at the same time maintain a realistic awareness of the dangers and distorting effects of the influence of advocacy coalitions, state capture and partyisation of economies. All these considerations apply equally in the case of VET policy, with the added element that reform resistance in this sector seems to be a widespread phenomenon, even while structural change has created an intense need for further reform in order to better align the supply of skills with the jobs which are available, in the context of an increased reliance on new service sectors and the pervasive effects of skill-biased technical change on the demand for highly skilled workers.

Acknowledgements

An early version of this chapter was presented at a regional workshop on 'Evidence Based Policy Making in VET: examples from practice in the Western Balkans and Turkey', organized by the European Training Foundation in Torino, from 8 to 9 November 2011. I am grateful to participants in the workshop for their useful comments. I am also grateful to two anonymous referees and to Margareta Nikolovska for their critical reading of the chapter and constructive suggestions for revisions.

Notes

1. I refer here to countries that are candidates or potential candidates for EU membership. The candidates include Croatia, Macedonia, Montenegro, Serbia and Turkey, while potential candidates are Albania, Bosnia and Herzegovina, and Kosovo.
2. The Instrument for Pre-Accession Assistance (IPA) is the main EU programme to support the accession process in the region, focusing on administrative reform and capacity building, human capital development, regional development and rural development.

3. According to information provided at an interview at the Ministry of Economy and Regional Development November 2010.

References

Bartlett, W. (2007), Economic restructuring, job creation and the changing demand for skills in the Western Balkans. In A. Fetsi (ed.), *Labour Markets in the Western Balkans: Challenges for the Future*, Turin: European Training Foundation, pp. 19–50.

Bartlett, W. (2010), Creating functioning market economies: aid effectiveness, policy transfer and advocacy coalitions in the Western Balkans. In B. Cerović and M. Uvalić (eds), *Western Balkans Accession to the European Union: Political and Economic Challenges*, Belgrade: Faculty of Economics, pp. 87–110.

Bartlett, W. and Prica, I. (2011), The variable impact of the global economic crisis in South East Europe, *Economic Annals*, LVI, 191: 7–34.

Black, N. (2001), Evidence based policy: proceed with care, *BMJ*, 323, 7307: 275–8.

Bonin, H. and Rinne, U. (2006), Evaluation of the Active Labor Market program 'Beautiful Serbia', *IZA Research Report No. 8*, Bonn: Institute for the Study of Labor.

Brixiova, Z., Li, W. and Yousef, T. (2009), Skill shortages and labor market outcomes in Central Europe, *Economic Systems*, 33, 1: 45–59.

Clarence, E. (2002), Technocracy reinvented: the new evidence-based policy movement, *Public Policy and Administration*, 17, 3: 1–11.

Commander, S. and Kollo, J. (2008), The changing demand for skills – Evidence from the transition, *Economics of Transition*, 16, 2: 199–221.

Deacon, B. and Stubbs, P. (2007), *Social Policy and International Interventions in South East Europe*, Cheltenham: Edward Elgar.

Dimitrov, V., Goetz, K. and Wollmann, H. (2006), *Governing After Communism*, Boulder, CO: Rowman and Littlefield.

Dimitrova, A. (2002), Enlargement, institution-building and the EU's administrative capacity requirement, *West European Politics*, 25, 4: 171–90.

Dolowitz, D. and Marsh, D. (1996), Who learns what from whom: A review of the policy transfer literature, *Political Studies*, 44, 2: 343–57.

Dolowitz, D. P. and Marsh, D. (2000), Learning from abroad: The role of policy transfer in contemporary policy-making, *Governance-an International Journal of Policy and Administration*, 13, 1: 5–24.

Eriksen, S. (2007), Institution Building in Central and Eastern Europe: foreign influences and domestic responses, *Review of Central and East European Law*, 32: 333–69.

European Commission (2010), The pre-accession economies in the global crisis: from exogenous to endogenous growth, *Ocassional Paper No. 62*, Brussels: Directorate General for Economic and Financial Affairs.

European Training Foundation (ETF) (2012), VET systems and policies in the Western Balkans and Turkey. In M. Serban (ed.), *The Torino Process: evidence-based policy making for vocational education and training*, Torino: ETF.

Fagan, A. (2010), The new kids on the block – Building environmental governance in the Western Balkans, *Acta Politica*, 45, 1–2: 203–28.

Farrell, C. and Morris, J. (2009), Still searching for the evidence? Evidence-based policy, performance pay and teachers, *Journal of Industrial Relations*, 51, 1: 75–94.

Golubovic, V. (2011), *Anticipating and matching demand and supply of skills in Montenegro*, Podgorica: Institute for Strategic Studies and Prognoses.

Hanushek, E. A. and Woessmann, L. (2006), Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries, *The Economic Journal*, 116: C63–C76.

Hasluck, C. (2000), Early lessons from the evaluation of the new deal programmes, *Research Report No. 49*, Sheffield: Employment Service.
Hellman, J. S. (1998), Winners take all – The politics of partial reform in postcommunist transitions, *World Politics*, 50, 2: 203–34.
Hellman, J. S., Jones, G. and Kaufmann, D. (2003), Seize the state, seize the day: state capture and influence in transition economies, *Journal of Comparative Economics*, 31, 4: 751–73.
Howlett, M. (2009), Policy analytical capacity and evidence-based policy-making: Lessons from Canada, *Canadian Public Administration-Administration Publique Du Canada*, 52, 2: 153–75.
Johnstone, D. (2006), Lessons to be learnt? The role of evaluations of active labour market programmes in evidence-based policy making, *Public Administration and Development*, 26: 329–37.
Kolev, A. and Saget, C. (2005), Understanding youth labour market disadvantage: Evidence from south-east Europe, *International Labour Review*, 144, 2: 161–87.
Kopecky, P. and Spirova, M. (2011), 'Jobs for the boys'? Patterns of party patronage in post-communist Europe, *West European Politics*, 34, 5: 897–921.
Kramer, M. (1997), Social protection policies and safety nets in East-Central Europe: Dilemmas of the post-communist transformation. In E. Kapstein and M. Mandelbaum (eds), *Sustaining the Transition: the Social Safety Net in Postcommunist Europe*, New York, NY: Council on Foreign Relations, pp. 46–123.
Lindblom, C. E. (1959), The science of muddling through, *Public Administration Review*, 19, 2: 79–88.
Meyer-Sahling, J.-H. (2008), The changing colours of the post-communist state: the politicisation of the senior civil service in hungary, *European Journal of Political Research*, 47, 1: 1–33.
Meyer-Sahling, J.-H. (2009), Varieties of legacies: a critical review of legacy explanations of public administration reforms in East Central Europe, *International Review of Administrative Sciences*, 75: 509–28.
Murthi, M. and Sondergaard, L. (2010), Skills, Not Just Diplomas: The Path for Education Reforms in ECA, *Knowledge Brief, Volume* 20, Washington, DC: World Bank.
Nikolovska, M. (2007), How to achieve educational change in ETF partner countries: between dreams and reality. In S. Nielsen and M. Nikolovska (eds), *Quality in Vocational Education and Training: Modern Vocational Training Policies and Learning Processes*, Turin: European Training Foundation, pp. 101–14.
Noutcheva, G. (2009), Fake, partial and imposed compliance: the limits of the EU's normative power in the Western Balkans, *Journal of European Public Policy*, 16, 7: 1065–84.
Nunberg, B. (ed.) (1999), *The State After Communism: Administrative Transitions in Central and Eastern Europe*, Washington, DC: World Bank.
Randma-Liiv, T., Nakrosis, V. and Hajnal, G. (2011), Public sector organisation in Central and Eastern Europe: from agencification to de-agencification, *Transylvanian Review of Administrative Sciences*, Special Issue (November): 160–75.
Roland, G. (2002), The political economy of transition, *Journal of Economic Perspectives*, 16, 1: 29–50.
Rose, R. (1993), *Lesson Drawing in Public Policy: A Guide to Learning Across Time and Space*, Chatham, NJ: Chatham House.
Sabatier, P. (2000), Theories of the policy process, *Journal of European Public Policy*, 7, 1: 135–40.
Sabatier, P. A. (1988), An advocacy coalition framework of policy change and the role of policy-oriented learning therein, *Policy Sciences*, 21: 129–68.

Sanderson, I. (2002), Evaluation, policy learning and evidence-based policy making, *Public Administration*, 80, 1: 1–22.
Stone, D. (2004), Transfer agents and global networks in the 'transnationalization' of policy, *Journal of European Public Policy*, 11, 3: 545–66.
United Nations Development Programme (UNDP) (2011), *Educational and Employment Careers of Youth in Croatia*, Zagreb: UNDP.
Vachudova, M. A. (2009), Corruption and Compliance in the EU's Post-Communist Members and Candidates, *Journal of Common Market Studies*, 47: 43–62.
Walker, R. (2000), Welfare policy: tendering for evidence. In H. T. O. Davies, S. M. Nutley and P. Smith (eds), *Evidence-based Policy and Practice in Public Services*, Bristol: Policy Press.
World Bank (2011), *Skills, Not Just Diplomas*, Washington, DC: World Bank.

7

Understanding Employment Barriers for Lone Parents in Great Britain: Research Gaps and Missed Opportunities

Tina Haux

Evidence and Policy-making

The previous Labour government (1997–2010) in Britain wanted its policies to be based on evidence rather than ideology. Hence, 'evidence-based policy-making', its interpretation, benefits and problems became one of the central themes of that government. The focus on evidence is said to have started when the incoming Labour government was preparing for the first Comprehensive Spending Review in 1998 only to find 'many gaps in existing knowledge about "what works"' (Nutley and Webb 2000: 20). This discovery then led to the *Modernising Government* White Paper, which, among other things, called for 'better use of evidence and research in policy-making' (HM Government 1999). Much has been written on the particular interpretation of evidence-based policy-making by the previous government and the problems thereof, such as the focus on indicators, piloting and evaluations of programmes (Sanderson 2002), and on issues of measurement and causality leading to a predominance of quantitative data (Sanderson 2002) at the expense of theoretical approaches (Pawson and Tilley 1997). In addition, the focus on 'what works' has meant that policymakers and researchers became too concerned with ex-post evaluations – thereby neglecting the transferable lessons from generic policy instruments, other countries and theories of human behaviour (Solesbury 2001) – mainly piloted rather than tested policy initiatives (Walker 2001) and tended to be reactive rather than proactive (Head 2010; Bochel and Duncan 2007). Reviews of the nature and success of evidence-based policy-making under the previous Labour government have been calling for greater collaboration between government and academia, more cross-departmental coordination and transfer of knowledge, more forward looking research activity and regular reviews of the quality of government-sponsored research (Head 2010; British Academy 2008; Bochel and Duncan 2007; Science and Technology Committee 2006).

Duncan and Harrop (2006) argue that the 'acid test of quality in research . . . should surely be whether it is useful to anyone'. This usefulness needs to extend beyond the immediate evaluation research, particularly given that research and policy implementation often work on different timescales (Head

Evidence and Evaluation in Social Policy, First Edition. Edited by Ian Greener and Bent Greve. Book compilation

2010; Bochel and Duncan 2007). With a new government in place, the time has come to look back at the period of the previous Labour government and its particular interpretation of evidence-based policy-making and assess its contribution to our knowledge of a particular policy problem. The policy problem that is being used as an example are lone parents not in employment.

The policy problem

The high levels of benefit receipt and poverty of lone parents and the potentially negative outcomes for children living in these families continues to challenge policymakers. When the previous Labour government came to power in 1997, fewer than half of lone parents were in paid employment (Kiernan *et al.* 1998). While the employment rate of lone parents in Britain increased between 1997 and 2010, it is low compared to that of mothers in couples in Britain and compared to that of lone parents in other countries (Chzhen and Bradshaw 2012; OECD 2007). The previous Labour government initially focused on enabling lone parents to move into employment as part of its intention to halve child poverty in 2010 and to abolish it by 2020 (Blair 1999). Paid work had become not only the 'best route out of poverty' but was a crucial part of the 'new contract of welfare' (e.g. Lister 2002; Deacon 2002).

The policies introduced to enable lone parents to move into work since 1997, such as the National Childcare Strategy, tax credits and the New Deal for Lone Parents, have been 'unprecedented' in their scope and ambition (Millar 2005), even if they have not always fully worked in practice (Lister 2002). Since 2005 the policy focus has shifted from enabling lone parents to move into work towards increasing conditionality.

Aims and Objectives

The combination of the policy interest in lone parents and the focus on evidence-based policy-making meant that a large number of studies have been carried out on lone parents not in work between 1997 and 2010. The aims of this chapter are to assess the progress that has been made in our understanding of the barriers to work for lone parents and to place the findings into the broader context of producing evidence for policymakers.

The objectives of the chapter are as follows:

- to briefly list the research on lone parents not in work undertaken during 1997 and 2010;
- to assess the contribution of that research and explore reasons for the apparent lack of progress; and
- to draw conclusions for evidence-based policy-making more widely.

Research on Lone Parents not in Work under the Previous Labour Government

The case study used here is research on lone parents not in work that has been commissioned by the now Department for Work and Pensions (DWP). The

review focuses on surveys rather than qualitative research because they reflect the focus of the particular interpretation of evidence-based policy-making of the previous Labour government and because surveys have the additional advantage of being used for secondary analysis. The reasons for the focus on research sponsored by the DWP are that the expansion of research budgets under the previous Labour government soaked up much of the research capacity outside government and that it was commissioned in dialogue with policymakers.

This chapter is building on previous reviews of research on lone parents by Holtermann *et al.* (1999), Millar and Ridge (2001) and the DWP (2005) by covering the whole period of the previous Labour government and placing the research into the wider context of evidence-based policy-making.

The research carried out on lone parents under the previous Labour government can be divided into two time periods. For a more detailed discussion of the research set out below (including qualitative studies and research carried out before 1997), please see Haux (2009). The first phase started with Labour coming to power in 1997 and lasted until 2003. This phase is marked by substantial data collection exercises as part of the evaluation of New Deal for Lone Parents (NDLP) (Hales *et al.* 2000; Lessof *et al.* 2001; Coleman *et al.* 2003), the evaluation of various pilots and elements of NDLP (Hales *et al.* 1998; Hasluck *et al.* 2000) and other programmes such as ONE (Cotton *et al.* 2000). In addition, the Families and Children Study (FACS), a cross-sectional and longitudinal study designed to monitor family change and evaluate a range of policies, such as the new tax credits, was started in 1999 (Marsh *et al.* 2001; McKay 2002; Vegeris and McKay 2002; Marsh and Rowlingson 2002; McKay 2002, 2003; Marsh and Perry 2003; Kasparova *et al.* 2003; Farrell and O'Connor 2003; Vegeris and Perry 2003). Lastly, a research review was commissioned (Millar and Ridge 2001).

The second phase started in 2003 and lasted until the change of government in 2010. During that time, the focus shifted away from commissioning new surveys to making use of a broader range of research methods, such as increased use of administrative data, qualitative longitudinal data, Randomized Control Trials and experiments (e.g. Dolton *et al.* 2006; Knight and Lissenburgh 2005; Hoggart *et al.* 2006; Dorsett *et al.* 2007; Riccio *et al.* 2008; Millar and Ridge 2009). In addition, more secondary analysis of existing survey data was commissioned on particular topics, such as profiling benefit claimants, lone parents working fewer than 16 hours and employment transitions (Bryson and Kasparova 2003; Rafferty 2003; McKay 2004; Knight and Kasparova 2004; Cebulla *et al.* 2008; Knight *et al.* 2006; Barnes *et al.* 2008; Hales *et al.* 2008; Evans *et al.* 2004; Lyon *et al.* 2008; Browne and Paull 2010). Important in the context of this chapter is the review of research questions on lone parents and barriers to work, which resulted in a revision of the questions asked in the FACS from 2006 onwards (Collins *et al.* 2006; D'Souza *et al.* 2008). This review will be discussed in more detail below.

Altogether, more than 50,000 lone parents took part in government-sponsored research during 1997–2010. Yet, the progress with regards to understanding the reasons lone parents remain out of work has not kept pace with the number of new studies.

What Do We Know about Lone Parents not in Work after 13 Years of Research?

In addition to evaluating the specifics of policies such as NDLP, research on lone parents not in work has tried to examine the policy problem more generally. In their review, Millar and Ridge (2001) argue that the research to date has tended to concentrate on two areas: characteristics of lone parents not in work and barriers to work. I would add that the third area has been that of temporal 'distance to work', i.e. the timeframe of when lone parents plan to move into work. In other words, the three main research questions have been *who* is not working, *why* not and *when* they are looking to move into work. The level of understanding gained as a result of the research reviewed here will be discussed for each of the three questions in turn.

Question 1: Who are the lone parents not in work?

Millar and Ridge continue that 'a very complete and generally consistent picture' exists with regards to the characteristics of lone parents not in work (Millar and Ridge 2001: 147). Indeed, a number of studies and reviews have identified the following characteristics as being linked to non-employment for lone parents: number and age of children, route into lone parenthood, level of qualifications, recent work experience, health status and work orientation (see Millar and Ridge 2001; Bradshaw and Millar 1991; Holtermann *et al.* 1999; Marsh *et al.* 2001; Kasparova *et al.* 2003).

Question 2: Why are lone parents not in work?

The reasons for lone parents not being in work have been conceptualized as 'barriers to work' in much of the research reviewed here. However, progress on 'barriers to work' was less pronounced than on the characteristics of lone parents not in employment. The main barriers to work for lone parents tend to be a preference to stay at home with the child, the availability and affordability of childcare and personal health (see reviews by Holtermann *et al.* 1999; Millar and Ridge 2001; DWP 2005; Collins *et al.* 2006). Other barriers such as financial issues, lack of skills, qualifications and work experience, employer prejudice and lack of confidence also appear regularly among the responses. However, there has been little consensus with regards to the ranking of barriers across surveys (Collins *et al.* 2006).

Millar and Ridge (2001) have argued that research on 'barriers to work' to date has failed to illuminate the complexity of lone parents' decisions about moving into work, for example by failing to establish the interconnectivity and duration of the so-called barriers. In acknowledgement of the lack of progress in understanding the barriers to employment for lone parents, the DWP commissioned a review of existing research on barriers (DWP 2005). One of its conclusions was that:

> an over reliance on the concept of 'barriers' could be seen to be limiting our potential to: understand the choice lone parents make

> about caring for their children at different stages of their and their child's life; and understand the constraints lone parents experience in exercising their choices and in making decisions about work (DWP 2005: 1)

The concept of 'barriers to work' was therefore replaced by that of 'choices and constraints' (Collins *et al.* 2006) when new survey questions were introduced into FACS as a result of the review. Preliminary analysis of the new survey questions suggests that six clusters of lone parents not in work can be identified, in terms of their attitudes towards work and caring (D'Souza *et al.* 2008). Further longitudinal analysis by Tomaszewski *et al.* on the influence of choices and constraints on work entry, retention and exit suggests that the main factors influencing lone parents' decisions about employment are 'attitudes towards parenting as a job, perceptions of personal and family constraints and intentions to work' (Tomaszewski *et al.* 2010: 29), informal childcare, receiving multiple benefits, the effect of benefit receipt on attitudes towards parenting as well as the perception of constraints. More analysis is required to test the contribution of the new questions more fully, for example whether the six clusters of lone parents not in work identified by D'Souza *et al.* (2008) are also indicative of temporal 'distance to work' and whether they face different choices and constraints. However, as the new questions were introduced in 2006, they are only available for three waves as FACS was stopped in 2008 and, therefore, the possibilities for analysis over a longer time period are limited.

Question 3: When are lone parents looking to move into work?

Within the concept of 'distance to work', employment and unemployment are thought of as a continuum, with full-time employment at one end and economic inactivity at the other. Individuals not in work are placed along this continuum according to their own timescale for moving into work with the aim to measure and even predict movement along the continuum. Applied to lone parents not in work, their position on the continuum tends to be summarised as three groups: 'work ready', i.e. those looking to move into work immediately or very soon, 'work postpone', those who would like to work at some point in the future, and 'never work', those who do not know when they will enter employment and those who do not see themselves doing so at all. However, there are a number of problems with this division. First, it is difficult to establish the relative size of the three groups as the figures differ substantially between all the surveys. For example, for the 'never work' group the estimate varies between 3 per cent and 13 per cent (Green *et al.* 2000; Kasparova *et al.* 2003). Second, the 'work ready' and the 'never work' group are relatively homogeneous in terms of their characteristics. Yet the largest group – 'work postpone' – has been difficult to define both in terms of its composition and its size. The concept also has not been useful in predicting the timing of lone parents returning to work (Kasparova *et al.* 2003; Bryson and Kasparova 2003). Only lone parents in the 'work ready' group were likely

to move into work according to their own timescale but there were no clear patterns for lone parents in the other two groups (Kasparova *et al.* 2003). Thus, sufficient information to divide lone parents not in work according to their distance to the labour market is not available (Gregg 2008) nor is it possible to predict their movements along the continuum despite the significant resources expended on this topic in the past decade and a half.

In summary then, while the research on lone parents not in work during the previous Labour government established the *who*, the research has not satisfactorily answered the *why* or the *when*. Thus, the quote by Marsh *et al.* that for lone parents 'the decision to go to work, when it comes, seems to come so suddenly . . . They seem to "flip" between identities' (Marsh *et al.* 2001: 20) still seems to be an apt summary of our understanding of the journeys to work for lone parents.

What Went Wrong?

The following section discusses the reasons for the lack of progress in our understanding of the barriers and distance to work of lone parents. The reasons fall into two broad categories: the research content and the interpretation of the role of research and researchers in the policy-making environment. With regards to the research content, there are three main issues, which have contributed to the lack of progress regarding the understanding of barriers to work of lone parents. These issues centre around the factors affecting lone parents' decisions, duplication and inconsistency of questions, and lack of further analysis.

What factors determine lone parents' decisions regarding work and parenting?

Arguably, one of the most important questions in the policy area is what are the factors that influence lone parents' decisions on how to combine work and parenting over time? This question has been addressed in a range of academic studies as well as the qualitative studies commissioned by the DWP and, between them, have suggested, among other things, that the role of family and friends both in terms of their actions and their views is important, as is the role of social capital more generally (Duncan and Edwards 1999). Furthermore, while the government-sponsored research has focus on barriers to *work* with views on parenting being treated as one of several barriers, the research around work orientation more generally as well as other research on lone parents (as early as Bradshaw and Millar 1991) suggests for lone parents the question is not when or whether to move into work but *how to combine work with parenting* (Finch *et al.* 1999; Dawson *et al.* 2000; Woodfield and Finch 1999). As it is, the questions focusing on barriers to *work* did not capture an essential part of lone parents' identity. Moreover, the quantitative research commissioned by DWP did not capture the role of planning either, which would have been a useful additional dimension to capture 'distance to work'. For example, questions have not been asked about how lone parents were planning to

combine work and parenting, whether they had particular timescales in mind, whether they related to particular events such as a child starting school and what those timescales were. Also, the influence of friends and family mentioned in the academic literature (e.g. Duncan and Edwards 1999; Houston and Marks 2003) has not been investigated. For example, missing from the government sponsored surveys but present in other surveys are questions on whether their own mother worked/is working now, whether their three closest friends were working and whether their friends, family and former partner think they should be working. Lastly, it would also have been useful to include at an early stage questions illuminating the employability of lone parents further, such as more detailed questions on work experience, informal networks with people in work and aspirations.

Duplication

Moving on to duplication, it is striking that the research reviewed here seems to have been commissioned in isolation of other research that was going on within the same department and outside. Not only has much of it been a-theoretical, as shown above, but it has also failed to take into account the findings from qualitative research (Collins *et al.* 2006) or indeed from previous surveys. For example, the survey by Bradshaw and Millar (1991) covered a broad range of reasons for not being in work and concluded that the three main concerns of lone parents were childcare, finding the right job and confidence. Other concerns, such as whether lone parents had unrealistic wage expectations, turned out not to be relevant. Yet, rather than focusing on those three concerns, the whole list of topics included in the Bradshaw and Millar survey (1991) was implemented in FACS and also asked in all the ad hoc surveys. This duplication meant that valuable survey time was lost that could have been spent on particular aspects of the evaluations, on asking more in-depth questions about particular issues or adding topics not covered in the original survey by Bradshaw and Millar (1991), for example around social capital. Moreover, until recently it was rare for surveys to be analyzed beyond the first headline report. More in-depth secondary analysis of particular areas of interest of existing data in the first place would have made a good preparation for the future survey. As mentioned in the introduction, the focus and methodological approaches changed in the second phase and became more sophisticated. However, by then few new surveys were commissioned, and so the opportunity to include different or better survey questions had passed.

Consistency

Lastly, the issue of how the questions have been asked: while the content of the questions did not change substantially between surveys, the wording of the questions, for example around barriers, did (DWP 2005). For example, while both the Bradshaw and Millar survey (1991) and the NDLP pilot survey (Hales *et al.* 2000) asked about the main reasons for not wanting a (regular) paid job, FACS asked about factors stopping lone parents from looking for a job of 16 or more hours. In some surveys, lone parents were asked both about the

reasons for not working as well as the perceived barriers. Comparing the answers to the two questions shows that responses to the question about perceived or potential barriers were fuller, more personal and more tangible than those on reasons for not working (Haux 2009), thereby highlighting the link between question wording and responses even within the same survey. This suggests that the answers to the questions on barriers are not strictly comparable across surveys.

In summary, a number of problems can be identified with the design of the research itself, namely that the political focus on getting lone parents into work skewed the content of the questions away from the perhaps more appropriate and fruitful line of questioning of how to combine work and parenting. In addition, the level of duplication and inconsistency of question wording means that the available data is less useful for secondary analysis as would otherwise have been the case.

The Role of Research and Researchers

Linking the research output back to the overall approach of evidence-based policy-making, the following observations can be made. The criticism by Duncan (2006) that too much attention has been paid to the collection rather than the analysis and assessment of data applies in particular to the first phase of research on lone parents (1997–2003). During that phase a number of surveys were commissioned to evaluate the different implementation stages of NDLP (Hales *et al.* 2000; Lessof *et al.* 2001; Coleman *et al.* 2003). The content of the surveys was tailored towards the questions at the heart of NDLP, namely how to get more lone parents into employment. This political focus is likely to have contributed to lone parents not being asked about how/whether they were planning to combine work and parenting but solely about barriers and distance to work. Perhaps researchers working in government got carried away in the excitement of working for a government that was keen to pilot and evaluate policies, and thereby lost sight of the opportunity to expand the understanding of the phenomenon in addition to the, usually more narrowly defined, policy problem.

Focusing not just on the questions of policy interest at that point in time is particularly important as the introduction of flagship policies often comes with additional resources for research that might not be available later on. When looking at the second phase of research on lone parents not in work (2003–2010), when there were fewer large-scale policy innovations, we can see that no new surveys were commissioned. Hence, the opportunity to gain a fuller understanding of the issues based on quantitative research around lone parents and work was lost during that first phase. As a result, there are no detailed questions examining the identity of lone parents in terms of how to combine work and parenting, nor on the role of networks and social capital in any of the surveys commissioned between 1997 and 2010. These questions could have been explored earlier in the form of rotating modules in FACS, for example, while leaving a core element that is comparable over time.

In addition to the above points, I would also argue that the forward-looking research activity, also referred to as horizon scanning, called for by reviews of

evidence-based policy-making under the previous Labour government should include a thorough review of previous research and policy agendas, as policy problems often re-appear in different guises. The question of whether lone parents should work and, therefore, the assessment of whether they are able to work goes back to the *Finer Report on One Parent Families* (HM Government 1974) and arguably as far back as the Poor Law Amendment Act 1834 (Lewis 1995). Researchers in government are seen as experts in research methods. Their role is to advise on the best design of research and then to commission and manage those studies (see GSR 2011). However, the research organizations carrying out the survey research are also usually research methods specialists. Hence, the absence of theoretical input and awareness of previous research is partly due to the institutional set-up where all the key players are experts in research methods rather than the subject matter.

Lessons for Evidence-based Policy-making

The conclusions by previous reviews of evidence-based policy-making under Labour, namely calling for greater collaboration between government and academia, more cross-departmental coordination and transfer of knowledge, more horizon scanning and regular reviews of the quality of government sponsored research (Head 2010; British Academy 2008; Bochel and Duncan 2007; Science and Technology Committee 2006) are supported by this analysis. Greater collaboration between government and academics could have prevented the continued narrow focus on barriers to work of lone parents. Furthermore, academic advisers are likely to have pointed towards the importance of the influence of family and friends on the decisions of lone parents around how to combine work and parenting (see Duncan and Edwards 1999). An earlier review of the progress made with regards to understanding the issues around lone parents and employment would have highlighted the duplication of research activity and, therefore, left more time and resources to improve the survey content. Furthermore, not only is horizon scanning important but it should also include a review of past policy problems and the nature and effectiveness of government responses, as many policy problems appear time and again. Lastly, however, returning to the usefulness of government research mentioned by Duncan and Harrop (2006), the previous Labour government spent substantial amounts of money and resources on the research programme on lone parents not in work. The surveys generally and the longitudinal survey, FACS in particular, could have been very useful resources for secondary analysis both at the time as well as to provide comparisons across decades with other longitudinal data. However, this is not the case due to the duplication, inconsistency and lack of theoretical input. Therefore, when commissioning research the longer-term usefulness should form part of the considerations alongside the immediate need to address a particular interpretation of a policy problem.

By Way of Conclusion: Why it Matters . . .

Towards the end of its time in office, the focus of the previous Labour government shifted from enabling lone parents to work to requiring (most)

lone parents with older children to be looking for work (Haux 2011). This reform has already been extended by the Coalition government, whose approach to welfare reform has been described as a continuation and 'intensification' of the approach of the previous government (Deacon and Patrick 2011; McKay and Rowlingson 2011).

When Gregg (2008) recommended profiling lone parents according to their readiness to work as an alternative to using the age of the youngest child as the main indicator for their ability to work, this was not possible due to the lack of robust data. Regardless of whether or not his recommendation would have been adopted by the government at the time, the fact remains that the research should and could have been available would have formed part of the evidence base to discuss options such as to tailor services, and perhaps even to tailor the level of conditionality to criteria and characteristics over and above the age of the youngest child (Haux 2011).

The lack of progress also matters as there is now neither the political nor the financial environment to commission large amounts of research. FACS, arguably the main vehicle for monitoring changes in the composition and outlook of lone parents, has come to an end after ten waves (1999–2008). The longitudinal element of the Lone Parent Obligation survey, containing the 'choices and constraints' questions, has also been curtailed. Therefore, data will only be available for waves one and two (Coleman and Lanceley 2011; Coleman and Riley 2012). The main longitudinal survey vehicle seems now to be *Understanding Society*, the successor of the British Household Panel Survey. However, *Understanding Society* is a general purpose panel study which covers a much broader remit than government policy, let alone government policies affecting lone parents. This means that it carries fewer relevant questions for the analysis of lone parents, e.g. it does not include questions on choices and constraints. Thus, there is not going to be any in-depth, high quality survey data available in terms of the sample size and question content on lone parents not in work in the near future.

In conclusion, it seems fair to say that, in the first half of Labour's time in office, a major opportunity was missed to use the political interest in evidence-based policy-making to produce high quality research shedding light on the phenomenon as such. Research that would, therefore, have been more useful to policymakers at the time and in the future.

References

Barnes, M., Conolly, A. and Tomaszewski, W. (2008), *Circumstances of Persistently Poor Families with Children: Evidence from the Families and Children Study (FACS)*, Department for Work and Pensions Research Report No. 487, Leeds: Corporate Document Services.

Blair, T. (1999), Beveridge revisited: a welfare state for the 21st century. In R. Walker (ed.), *Ending child poverty: Popular welfare for the 21st century*, Bristol: Policy Press, pp. 7–20.

Bochel, H. and Duncan, S. (2007), *Making Policy in Theory and Practice*, Bristol: Policy Press.

Bradshaw, J. and Millar, J. (1991), *Lone Parent Families in the UK*, Department of Social Security research report no. 6, London: HMSO.

British Academy (2008), *Punching our weight in the humanities and social sciences in public policy making*, London: British Academy, http://www.britac.ac.uk/policy/punching-our-weight.cfm (accessed 22 March 2013).
Browne, J. and Paull, G. (2010), *Parents' work entry, progression and retention, and child poverty*, Department for Work and Pensions Research Report No. 626, Leeds: Corporate Document Services.
Bryson, A. and Kasparova, D. (2003), *Profiling benefit claimants in Britain: A feasibility study*, Department for Work and Pensions Research Report No. 196, Leeds: Corporate Document Services.
Cebulla, A., Flore, G. and Greenberg, D. (2008), *The New Deal for Lone Parents, Lone Parent Work Focused Interviews and Working Families' Tax Credit: A review of impacts*, Department of Work and Pensions Research Report No. 484, Norwich: HMSO.
Coleman, N. and Lanceley, L. (2011), *Lone Parent Obligations: supporting the journey into work*, Department for Work and Pensions Research Report No. 736, London: Department for Work and Pensions.
Coleman, N. and Riley, T. (2012), *Lone Parent Obligations: following lone parents' journeys from benefits to work*, Department for Work and Pensions Research Report No. 818, London: Department for Work and Pensions.
Coleman, N., Rousseau, N. and Laycock, M. (2003), *National Evaluation of Lone Parent Adviser Meetings: Findings from longitudinal analysis*, Department for Work and Pensions Research Report W172, Leeds: Corporate Document Services.
Collins, D., Gray, M., Purdon, S. and McGee, A. (2006), *Lone parents and work: developing new survey measures of the choices and constraints*, Department for Work and Pensions Working Paper No. 34, Leeds: Corporate Document Services.
Cotton, D., Stone, V. and Thomas, A. (2000), *Why not ONE? Views of non-participants before full participation*, Department of Social Security Research Report No. 127, Leeds: Corporate Document Services.
Chzhen, Y. and Bradshaw, J. (2012), Lone parents, poverty and policy in the European Union, *Journal of European Social Policy*, 22, 5: 487–506.
Dawson, T., Dickens, S. and Finer-Cragg, S. (2000), *New Deal for Lone Parents: Report on qualitative studies with individuals*, Employment Service Research No. 55, Sheffield: Employment Service.
Deacon, A. J. (2002), *Perspectives on welfare: ideas, ideologies and policy debates*, Maidenhead: Open University Press.
Deacon, A. J. and Patrick, R. (2011), A new welfare settlement? The Coalition Government and Welfare to Work. In H. Bochel (ed.), *The Conservative Party and Social Policy*, Bristol: Policy Press, pp. 161–80.
Department for Work and Pensions (DWP) (2005), *DWP methodological review of lone parent parenthood and employment*, London: DWP.
Dolton, P., Azevedo, J. P. and Smith, J. (2006), *The econometric evaluation of the New Deal for Lone Parents*, Department for Work and Pensions Research Report No. 356, Leeds: Corporate Document Services.
Dorsett, R., Campbell-Barr, V., Hamilton, G., Hoggart, L., Marsh, A., Miller, C., Phillips, J., Ray, K., Riccio, J. A., Rich, S. and Vegeris, S. (2007), *Implementation and first year impacts of the UK Employment Retention and Advancement (ERA) demonstration*, Department for Work and Pensions Research Report No. 412, Leeds: Corporate Document Services.
D'Souza, J., Conolly, A. and Purdon, S. (2008), *Analysis of the Choices and Constraints: Questions on the Families and Children Study (2006)*, Department for Work and Pensions Research Report No. 481, Leeds: Centre for Disability Studies.
Duncan, S. (2006), Towards evidence inspired policy, *Social Sciences*, 61: 10–11.

Duncan, S. and Edwards, R. (1999), *Lone Mothers, Paid Work and Gendered Moral Rationalities*, Basingstoke: Macmillan.
Duncan, S. and Harrop, A. (2006), A User Perspective on Research Quality, *International Journal of Research Methodology*, 9, 2: 159–74.
Evans, M., Harkness, S. and Ortiz, R. A. (2004), *Lone parents cycling between work and benefits*, Department for Work and Pensions Research Report No. 217, Leeds: Corporate Document Services.
Farrell, C. and O'Connor, W. (2003), *Low-income Families and Household Spending*, Department for Work and Pensions Research Report No. 192, Leeds: Corporate Document Services.
Finch, H., O'Connor, W., Millar, J., Hales, J., Shaw, A. and Roth, W. (1999), *New Deal for Lone Parents: Learning from Prototype areas*, Department of Social Security Research Report No. 92, Leeds: Corporate Document Services.
Government Social Research Service (GSR) (2011), Government Social Research Service, http://www.civilservice.gov.uk/networks/gsr (accessed 22 March 2013), last updated on 14 March 2011.
Green, H., Smith, A., Lilly, R. and Marsh, A. (2000), Part One, Survey of Clients: Cohort One, Wave 1, and Johnson, C. and Fielding, S., Part Two, Qualitative Research with Clients: Stage 1, *First Effects of ONE: Findings from Survey and Qualitative Research with Clients*, Department of Social Security Research Report No. 126, Leeds: Corporate Document Services.
Gregg, P. (2008), *Realising Potential: A Vision for Personalised Conditionality and Support, An independent report to the Department for Work and Pensions*, London: HMSO.
Hales, J., Roth, W., Barnes, M., Millar, J., Lessof, C., Gloyer, M. and Shaw, A. (2000), *Evaluation of the New Deal for Lone Parents – Early lessons from the phase ONE prototype – Findings of the survey*, Department of Social Security Research Report No. 109, Leeds: Corporate Document Services.
Hales, J., Roth, W. and Shaw, A. (1998), *Evaluation of the New Deal for Lone Parents Preliminary estimates of the counterfactual*, Department of Social Security Research Report No. 42, Leeds: Corporate Document Services.
Hales, J., Tipping, S. and Lyon, N. (2008), *Mothers' participation in paid work: the role of 'mini-jobs'*, Department for Work and Pensions Research Report No. 467, Leeds: Corporate Document Services.
Hasluck, C., McKnight, A. and Elias, P. (2000), *Evaluation of the New Deal for Lone Parents – Early lessons from the phase ONE prototype – Cost benefit and econometric analysis*, Department of Social Security Research Report No. 110, Leeds: Corporate Document Services.
Haux, T. (2009), *Lone parents and welfare reform: A policy appraisal*, PhD thesis, University of Bath, http://ebookbrowse.com/univbath-phd-2009-t-haux-pdf-d78763541 (accessed 25 October 2012).
Haux, T. (2011), Activating Lone Parents: An Evidence-Based Appraisal of the Welfare-To-Work Reform in Britain, *Social Policy and Society*, 11, 1: 1–14.
Head, B. (2010), Reconsidering evidence-based policy: Key issues and challenges, *Policy and Society*, 29, 2: 77–94.
HM Government (1974), *Recommendations of the Finer Report on One Parent Families*, Cmnd 5629, London: HMSO.
HM Government (1999), *Modernising Government*, Cm 4310, White Paper, London: Cabinet Office.
Hoggart, L., Campbell-Barr, L., Ray, K. and Vegeris, S. (2006), *Staying in work and moving up: Evidence from the UK Employment Retention and Advancement (ERA) demonstration*, Department for Work and Pensions Research Report No. 381, Leeds: Corporate Document Services.

Holtermann, S., Brannen, J., Moss, P. and Owen, C. (1999), *Lone Parents and the Labour Market: Results from the 1997 Labour Force Survey and Review of Research*, Employment Service Research Report No. 23, Sheffield: Employment Service.
Houston, D. and Marks, G. (2003), The Role of Planning and Workplace Support in Returning to Work after Maternity Leave, *British Journal of Industrial Relations*, 41: 197–214.
Kasparova, D., Marsh, A., Vegeris, S. and Perry, J. (2003), *Families and children 2001: Work and Childcare*, Department for Work and Pensions Research Report No. 191, Leeds: Corporate Document Services.
Kiernan, K., Land, H. and Lewis, J. (1998), *Lone motherhood in twentieth-century Britain: from footnote to front page*, Oxford: Clarendon Press.
Knight, G. and Kasparova, D. (2004), *Lone parents: In Work Benefit Calculations – work and benefit outcomes*, Department for Work and Pensions Research Report No. 367, Leeds: Corporate Document Services.
Knight, G. and Lissenburgh, S. (2005), *Evaluation of the extension to Lone Parent Work Focused Interviews eligibility: administrative data analyses*, Department for Work and Pensions Research Report No. 237, Leeds: Corporate Document Services.
Knight, G., Speckesser, S., Smith, J., Dolton, P. and Azevedo, J. P. (2006), *Lone parents Work Focused Interviews/New Deal for Lone Parents: combined evaluation and further net impacts*, Department for Work and Pensions Research Report No. 368, Leeds: Corporate Document Services.
Lessof, C., Hales, J., Phillips, M., Pickering, K., Purdon, S. and Miller, M. (2001), *New Deal for Lone Parents Evaluation: A Quantitative Survey of Lone Parents on Income Support*, Employment Service Research Report ESR 101, Sheffield: Employment Service.
Lewis, J. (1995), *The Problem of Lone Mother Families in Twentieth Century Britain*, London: London School of Economics.
Lister, R. (2002), The dilemmas of pendulum politics: balancing paid work, care and citizenship, *Economy and Society*, 31, 4: 520–32.
Lyon, N., Barnes, M. and Millar, J. (2008), *Employment transitions and the changes in the economic circumstances of families with children: Evidence from the Families and Children Study (FACS)*, Department for Work and Pensions Research Report No. 506, Leeds: Corporate Document Services.
Marsh, A. and Perry, J. (2003), *Family Change 1999 to 2001*, Department for Work and Pensions Research Report No. 181, Leeds: Corporate Document Services.
Marsh, A. and Rowlingson, K. (2002), *Low/moderate-income families in Britain: Changes in Living Standards 1999–2000*, Department for Work and Pensions Research Report No. 165, Leeds: Corporate Document Services.
Marsh, A., McKay, S., Smith, A. and Stephenson, A. (2001), *Low-income families in Great Britain: work, welfare and social security in 1999*, Department of Social Security Research Report No. 138, Leeds: Corporate Document Services.
McKay, S. (2002), *Low/moderate-income families in Britain: Work, Working Families' Tax Credit and childcare in 2000*, Department for Work and Pensions Research Report No. 161, Leeds: Corporate Document Services.
McKay, S. (2003), *Working Families' Tax Credit 2001*, Department for Work and Pensions Research Report No. 180, Leeds: Corporate Document Services: Leeds.
McKay, S. (2004), *Lone Parents in London: Quantitative analysis of differences in paid work*, Department for Work and Pensions In-House Report No. 136, London: HMSO.
McKay, S. and Rowlingson, K. (2011), Social security and welfare reform. In H. Bochel (ed). *The Conservatives and Social Policy*, Bristol: Policy Press, pp. 145–60.
Millar, J. (2005), Work as welfare? Lone mothers, social security and employment. In P. Saunders (ed.), *Welfare-to-work in Practice: Social Security and Participation in Economic and Social Life*, Aldershot: Ashgate.

Millar, J. and Ridge, T. (2001), *Families, poverty, work and care: A review of the literature on lone parents and low income couple families with children*, Department for Work and Pensions Research Report No. 153, Leeds: Corporate Document Services.
Millar, J. and Ridge, T. (2009), Relationships of Care: Working Lone Mothers, their Children and Employment Sustainability, *Journal of Social Policy*, 38: 103–21.
Nutley, S. M. and Webb, J. (2000), Evidence and the Policy Process. In H. T. O. Davies, S. M. Nutley and P. C. Smith (eds), *What works? Evidence-based policy and practice in public services*, Bristol: Policy Press, pp. 13–42.
Organisation for Economic Co-operation and Development (OECD) (2007), *Babies and bosses: reconciling work and family life: A synthesis of findings for OECD countries*, Paris: OECD.
Pawson, R. and Tilley, N. (1997), *Realistic evaluation*, London: Sage.
Rafferty, A. (2003), *The Characteristics of Lone & Coupled Mothers Working Fewer Than 16 Hours per Week*, Department for Work and Pensions In-House Research Report No. 125, London: HMSO.
Riccio, J. A., Bewley, H., Campbell-Barr, V., Dorsett, R., Hamilton, G., Hoggart, L., Marsh, A., Miller, C., Ray, K. and Vegeris, S. (2008), *Implementation and second-year impacts for lone parents in the UK Employment Retention and Advancement (ERA) demonstration*, Department for Work and Pensions Research Report No. 489, Leeds: Corporate Document Services.
Sanderson, I. (2002), Evaluation, policy-learning and evidence-based policy-making, *Public Administration*, 80, 1: 1–22.
Science and Technology Committee (2006), *Scientific Advice, Risk and Evidence-Based Policy-making*, HC 900-1, London: HMSO.
Solesbury, W. (2001), *Evidence-based policy: Whence it came and where it's going*, UK Centre for Evidence-based Policy and Practice Working Paper No. 1, http://www.kcl.ac.uk/sspp/departments/politicaleconomy/research/cep/pubs/papers/paper-01.aspx (accessed 22 March 2013).
Tomaszewski, W., Chanfreau, J. and Barnes, M. (2010), *Lone parents and employment: an exploration of findings from the Families and Children Study 2006–08*, Department for Work and Pensions Working Paper No. 93, London: Department for Work and Pensions.
Vegeris, S. and McKay, S. (2002), *Low/moderate-income families in Britain: Changes in Living Standards 1999–2000*, Department for Work and Pensions Research Report No. 164, Leeds: Corporate Document Services.
Vegeris, S. and Perry, J. (2003), *Families and children 2001: Living standards and the children*, Department for Work and Pensions Research Report No. 190, Leeds: Corporate Document Services.
Walker, R. (2001), Great Expectations: Can Social Science Evaluate New Labour's Policies? *Evaluation*, 7, 3: 305–30.
Woodfield, K. and Finch, H. (1999), *New Deal for Lone Parents – Evaluation of innovative schemes*, Department of Social Security Research Report No. 89, Leeds: Corporate Document Services.

8

Putting the Research Boot on the Policymakers' Foot: Can Participatory Approaches Change the Relationship between Policymakers and Evaluation?

Liz Richardson

Introduction

The moral and instrumental arguments in favour of more participatory approaches to research suggest that they may be more likely to influence policy than conventional research (Durose *et al.* 2011). However, there is little evidence or evaluation of impact in participatory research (Catalani and Minkler 2009); as one commentator points out, simply 'intending to create social change is no assurance of actually doing so' (Riger 1992: 736). Perhaps more problematically, co-productive approaches have been seen as posing threats to 'the integrity of the research process' (Martin 2010: 213).

Participatory approaches have tended to be aligned with methodologies which de-privilege the idea of objective evidence of policy effectiveness. Participatory research has commonly been associated with a focus on multiple forms of knowledge and on the principles and values of empowering practice, rather than positivist scientific standards of evaluation, whereas traditional approaches to evaluation which privilege independent and objective methods have typically been non-participatory. One consequence has been limited use of evidence of policy effectiveness in decision-making processes. This artificial separation of positivist and non-positivist methods into non-participatory and participatory approaches is all the more troublesome when the political nature of policy-making is taken into account. Many competing pressures push policy decisions away from the evidence-base and towards ideologically driven choices. Deeper and more thoroughgoing engagement of policymakers with research could help to anchor decisions more firmly in evidence from evaluation. However, amongst policymakers there remains a strong assumption that evidence, particularly evaluation, is compromised if it is not undertaken by an external and 'independent' body, a concern which is shared by many academics.

While the impact and quality of co-produced or participatory research is unproven, the idea that academic independence *per se* is a better guarantor of objectivity or reliability is questionable. As a structural tool for generating

Evidence and Evaluation in Social Policy, First Edition. Edited by Ian Greener and Bent Greve. Book compilation

unbiased results, the separation of knowledge production from knowledge consumption or use is arguably a crude one, and biases can remain (e.g. see Bond *et al.* 2011). Acknowledging the inevitability of bias and value orientations in social science (Myrdal 1972) is easier to do on paper than it is to put in place checks and balances against bias. Methodologies exist to guarantee the robustness of findings, and also allow for peer scrutiny of research. However, even the strongest 'gold standard' methods are not perfect or perfectly applied.

Perhaps the strongest reason to re-visit non-participatory methods is the fact that, across disciplines in the social sciences, potential knowledge users fail to maximize the use of academic evidence in public policy-making (Walt 1994; Davies *et al.* 2000: 13; Palfrey *et al.* 2012), with faults identified on both sides (Blunkett 2000; Black 2001; DeLeon and Longobardi 2002; Campbell *et al.* 2007: 6; Flinders 2010: 313). Instrumental uses of social science are not its only function, but a body of pristine (or not so pristine) and unheard evidence seems like a lost opportunity for society, and for ambitions for the policy sciences and applied social science, as imagined by Lasswell, Lynd, Dewey and others (DeLeon 1994: 78; Parsons 2002: 43).

In much existing research, including some participatory approaches, the implicit assumption is that researchers and research users such as policymakers are separate entities with distinct roles and claims to knowledge, although the parties may be in a dynamic interactive relationship (Orr and Bennett 2009; Pohl *et al.* 2010). However, there is also the possibility that 'researcher' and 'policymaker' could be the same 'tribe'. Could this offer a different way of bridging the research-policy gap, while maintaining defensible research methods and generating reliable results? First, this chapter discusses the relationships between politics and science in decision-making. It then describes tensions in existing approaches between participatory research and positivist methods, before outlining a case study of a self-evaluation scheme by politicians in local government to explore what happens in practice when local policymakers become researchers.

The Policy Process, Social Science and Politics

Simplistic linear views of the role of social science in policy-making have been critiqued, and the relationship shown to be much messier and politically situated (Lindblom 1959; Cohen *et al.* 1972; Weiss, C. 1979; Gains and Stoker 2011). One driver for complexity and messiness is that policy-making is political, and party political. Many have noted political ideology, values and public opinion as influencing policy (Weiss 1993; Davies *et al.* 1999; Duncan 2005; Mulgan 2005). In the absence of effective coalitions, 'the voice of research' can be 'drowned out by the noise' of competing interests and other pressures on political policymakers (Wilensky 1997: 1248). Political naivety is a risk to policy influence in a pluralist process (Etzioni 1988). Political interests are overshadowed by short-term electoral incentives (Davies *et al.* 2000: 30).

Given, then, what is known in social science about messiness and the political nature of the policy process, there are gaps in mainstream academic approaches to managing complexity and politics. Tactics for the dissemina-

tion of evaluation research include a focus on better communication (Davies *et al.* 2000) such as 'writing in human' (Flinders 2012). Basic advice about more effective communication tools is unobjectionable. However, it is hard to see how these tools could have enough purchase on the political nature of decision-making or the rejection of externally produced knowledge. Tensions have been identified between political science's preoccupation with policy networks and their own adherence to professional boundaries (Smith and Joyce 2012). Some individuals in different disciplines have tried becoming policy entrepreneurs and engaging in policy networks. Boyer (1996: 15), John (2013 forthcoming) and others paint a positive picture in political and policy sciences of a long history of policy engagement. However, in social policy, political science and sociology, academics who have been close to policy present rueful descriptions of their encounters (Le Grand 2006; Newman 2011).

Policy use of evaluation is centred on evidence about 'what works' (Ekblom 2002). Given this focus, some solutions have been proposed, which could be seen as side-stepping these policy demands. For example, the 'realist' position that 'evidence-inspired' or 'evidence-informed' policy is more achievable (Duncan 2005) or proposals for a longer-run game to shape elite perceptions and conceptualizations of social and political reality (Wilensky 1997: 1257; Davies and Powell 2012: 213).

Participatory Research Approaches and Politics

If sidestepping the issue, playing the long game, avoiding policy engagement, becoming engaged in policy and more professional communication techniques are inadequate to bridge the evaluation evidence and public policy gap, then could participatory policy research be an alternative? There is now a growing body of work which argues for a wider role for research users, such as political decision-makers, in scientific research as co-producers of knowledge (Nutley *et al.* 2007; Armstrong and Alsop 2010; Martin 2010). Understanding the evidence-policy gap as not a problem of knowledge transfer but of knowledge production, offers a route to engaged scholarship (Boyer 1996; Van de Ven 2007). This builds on a longer tradition of engaged scholarship, following Lasswell, to overcome disconnection between analysis and politics.

Arguments for the involvement of practitioners and members of the community in research are not new (e.g. Bennett and Roberts 2004). There is a significant body of work loosely focused around participatory policy-orientated research and practice-based learning. This includes distinct approaches, such as participatory action research, participatory poverty assessments (Robb 2002), experiential learning (Kolb 1984), and reflective practice (Schön 1983; Boud *et al.* 1985). But what unites different approaches to participatory research and practice-based learning are attempts to value people's direct experience in knowledge. They also share a value-base in social justice and human rights (Lewin 1946; Heron and Reason 1997), with participatory approaches used to overcome marginalization and transform imbalances in power (Freire 1970; Minkler and Wallerstein 2003; Ledwith 2007).

However, a fundamental issue is whether those of us who wish to walk along the 'positivist yellow brick road' (Parsons 2002: 45) can also use a rainbow of participatory methods. Proponents, like Parsons, of self-transforming decision-making systems, and of participatory policy analysis (Torgerson 1986) largely identify themselves as post-positivists. In participatory research there has often been more focus on principles and values of empowering, transformative and reflective practice than on traditional concerns with standards of scientific integrity and research quality, or how the two might be achieved simultaneously.

Relying on direct experience is an unsatisfactory solution to the evaluation of policy impact and observable outcomes. Participatory policy analysis (DeLeon 1994: 88) privileges the common sense and 'practical reason' (Sanderson 2002: 19) of policymakers (and the decision-making competence of ordinary people) on a par with technical expertise. Therefore, the line between the 'conventional wisdom' preferred by policymakers (Leicester 1999: 6) and scientific knowledge begins to get blurred when looking at reliable evidence of policy outcomes. For politicians, a danger is that practice-based insights and common sense become a justification for the sorts of ideologically driven policy decisions described above.

What happens when politicians conduct their own evaluations? Does putting the research boot on the policymakers' foot achieve both participation and positivism? We now turn to a case study of a self-evaluation scheme for local councillors in the north west of England to explore these questions.

A Case Study Example: Self-evaluation by Politicians in Local Government

This chapter uses primary research from an award scheme to accredit self-evaluation undertaken by policymakers in local government in the north west of England. The scheme, called the North West Member Development Charter Level Two (the Charter) was started in 2007, using criteria, systems and a process designed by the author. It is sponsored and administered by North West Employers (NWE).[1] Given our questions about the political nature of policy-making in public policy, looking at local government bodies is useful. Local government is said to be democratically accountable through the electoral mandate given to its local councillors, and has a responsibility for setting local policies across a wide range of policy areas. It is a site where policy is explicitly made by politicians with ultimate power lying with a politically controlled Executive. It has several features identified as making it potentially more susceptible to influence by social science such as the decentralization of policy decisions,[2] and less susceptible, including the relatively weak 'professional' position of local councillors, strong institutional cultures, the complex nature of decisions and multiple alternative sources of knowledge (Weiss, J. 1979). Much of the focus in the existing literature has been on evidence-based policy-making by central government.

The selection of this case is because it was explicitly designed to test the questions posed in this chapter, that is, whether participatory self-evaluation could generate co-production of local government policymakers in research,

Table 8.1

Numbers participating in the Charter

Total no. of local government organizations participating	Discussions only (no written draft application)	Written draft applications	Withdrawn before full assessment completed	Assessment completed	of which, unsuccessful	of which, awarded
25	7	18	2	16	1	15

while maintaining rigorous evaluation methods. There are few other examples of such attempts. As a regional membership body, NWE had the authority to stipulate evaluation standards across a region, and was prepared to sponsor the research.

Description of the Charter scheme

The Charter required local authorities to collect evidence of the community outcomes created by investment in learning and development for local councillors. It specified local councillors being in a lead role in the research design, a minimum of a pre- and post-intervention research design, with a comparison or control group recommended. Accreditation was assessed by a panel of an academic, a local councillor as a peer assessor and a local government officer. The process of applying for the Award was a series of iterations of the application; applicants were offered comments ranging from major to minor revisions. After submission of a final written draft, an on-site verification visit was held with the applicants and a sample of their respondents.

Research methods for the case study

For each authority which expressed an interest in the Charter scheme, the following data was collected by the author: copies of all draft written applications, including comments on drafts; recorded discussions with potential applicants; verbatim notes of discussions on verification days; assessment reports synthesizing the written applications with the feedback from the verification visits. In addition, there were: a group discussion with six successful authorities after the first two rounds of applications to debrief on their experiences of the process; two workshops for authorities interested in applying; and a booklet about six of the successful applications produced independently of the author, and based on interviews with the authorities (NWE 2010).

Participation in the Charter scheme

As shown in table 8.1, a total of 25 local government organizations expressed a serious interest in the Charter scheme between 2007 and May 2012,[3] of

which 18 progressed to written draft applications. Of these, 15 to date have been awarded the Charter. None of the authorities' written draft applications were accepted without revisions. Some applicants withdrew from the process if they could not respond to the comments.

Table 8.2 summarizes the 15 successful Charter awards. Of the applicable authorities, six were politically controlled by Labour administrations, five by Conservatives, and one by a Liberal Democrat and Conservative coalition. There were a wide range of topics for the applications, however, all had the core intervention in common, which was work by local councillors on the specific topic which had been part of a learning and development programme.

Evaluation Done by Local Policymakers: Findings

Tables 8.3 and 8.4 show examples of research designs and methods used in the successful awards, and the unsuccessful or withdrawn written draft applications. In the successful applications, all used a mixed methods design, and all included pre- and post-measures. None used comparative (quasi-experimental) or experimental designs, although two used other local authority comparator averages for specific pieces of statistical data. As well as data from national and local surveys, and public consultation and customer survey feedback, successful applicants made good use of existing administrative data, for example appeals against licensing decisions. Given the nature of interventions, conventional data sources and administrative data were not always sensitive enough to track impacts. To deal with this, imaginative use was made of more sensitive and tailored measures, for example, the number of signatories to petitions opposing council decisions.

The qualitative methods were primarily mainstream methods of in-depth interviews, focus groups and case studies. In the Warrington Council example, focus groups with vulnerable children in care and care leavers were made accessible for the young people by structuring feedback on services around '*the good, the bad and the ugly*'. In addition, there were isolated examples of more innovative qualitative methods. Blackpool Council undertook recorded observations of meetings. Three authorities used content analysis to examine policy changes before and after the interventions. The written application from Wyre Council compared the marginality of 'call-ins', tracking changes in the decisions challenged from, '*the disposal of beach huts (small wooden chalets) on Fleetwood seafront to the physical masterplan for the whole Fleetwood area*'. All of the successful applications used documentary evidence, for example documentation of legal cases, photographs, websites and official reports.

The unsuccessful or withdrawn written applications suffered from a range of problems in the research designs and methods, including: unclear research methods or poor evaluation frameworks; insufficient range and depth of methods; insufficient evidence of community outcomes; impacts not clearly linked to the core intervention; weak baseline measures; weak evidence of role of intervention in creating outcomes. Discussion did not move to written applications where the research questions were not sufficiently formulated or crystallized, where outcome measures were inadequately operationalized.

Table 8.2

Description of successful Charter award

Authority	Date of award	Political control at application	Topic of application
Blackpool Council	2007	Labour	Liquor licensing/management of the night-time economy
Cheshire West and Chester Council	2012	Conservative	Corporate parenting/looked after children
Chorley Council	2009	Conservative	Neighbourhood working/councillors' roles in wards
Flyde Council	2009	Conservative	E-democracy
Gt. Manchester Fire & Rescue Authority	2011	n/a	Review of corporate plan/service re-structuring
Halton Council	2007	Labour	Councillors' roles in wards/community engagement
Hyndburn Council	2012	Labour	Housing market renewal/regeneration
Knowsley Council	2009	Labour	Household recycling
Lancashire Combined Fire Authority	2010	n/a	Review of emergency fire and rescue services
Merseyside Fire and Rescue Authority	2012	n/a	Prevention of youth anti-social behaviour
Rossendale Council	2011	Conservative	Review of leisure provision
St Helens Council	2011	Labour	Roles for councillors
Tameside Council	2010	Labour	Community chest grants
Warrington Council	2009	Lib Dem/Cons coalition	Corporate parenting/looked after children
Wyre Council	2010	Conservative	Review of budget setting/council tax

Table 8.3

Examples of research designs and methods used in successful Charter awards

Authority	Topic	Quantitative methods and measures	Qualitative methods	Documentary evidence
Flyde	E-democracy	– No. of online transactions overall – Public consultation responses – Customer feedback responses – No. of councillors with published emails, websites, and blogs – No. of hits on councillor websites and councillor-initiated online surveys – Correspondence conducted electronically – No. of new community websites	– Individual case studies – In-depth interviews with councillors and community organizations – Comments from online discussion on community forums	– Community websites and online discussion forums
Hyndburn	Housing market renewal/ regeneration	– No. of legal enforcement actions against landlords – No. of neighbourhood visits by councillors – Level of central government funding	– In-depth interviews with councillors, senior officers and front-line council staff – Case studies of community groups and neighbourhood work	– Constituent correspondence – Grounds given in legal cases – Presentations by councillors – Press reports – Pre- and post-photographs – Neighbourhood issues resolved
Knowsley	Household recycling	– % waste collected going to recycling – No. of households recycling – Levels of unacceptable litter and detritus – No. of complaints about bins – No. of residents satisfied with service, and data from comparator areas – Household attitudinal survey – No. of sickness absence days for refuse staff – Costs, staffing levels and working patterns	– Simulation exercises and documentation of demonstrations – In-depth interviews with councillors and senior officers – Feedback from community organizations	

Lancashire CFA	Review of emergency fire and rescue services/service restructuring	– Costs, staffing levels and working patterns – No. and location of incidents (fires/call outs) – Service standards – Response times to incidents/call outs – No. of consultation responses – Staff surveys, incl. agreement with outcome of review; satisfaction with new arrangements – No. of staff opting for new working patterns	– In-depth interviews with councillors and senior officers – Analysis of policy changes implemented compared to original proposals – Focus groups with representatives from Lancashire local authorities, and community representatives – Analysis of consultation feedback	– Official reports and documents – Review and committee reports and minutes of meetings – Submissions from trade unions – Press reports – External audit and inspection reports
St Helens	Review of roles for councillors	– Demographic data on councillors – Staffing levels for democratic services	– In-depth interviews with councillors, senior officers, and community organizations – Testimonials from public and voluntary sector partners, and senior officers – Feedback from public consultation – In-depth individual case studies – Case studies of community work	– Minutes of council meetings – Official reports and documents – Responsibilities for democratic services
Wyre	Review of budget setting/council tax	– No. of 'call-ins' (challenges to Executive decisions) – Costs of services, and data from comparator 'family group' councils – Survey of councillors	– In-depth interviews with councillors and senior officers – Focus groups with the public – Analysis of nature of call-ins and governance arrangements – Analysis of policy changes implemented compared to original proposals – Feedback from Parish Councils and community representatives – Testimony from local journalists	– External audit and inspection reports – Press reports – Council website

Table 8.4

Examples of research designs and methods used in unsuccessful or withdrawn written draft applications

Status of application	Topic	Proposed methods/measures	Issues with design and methods
Assessment completed Unsuccessful at verification stage	Local scrutiny/ working in neighbourhoods	*Quantitative* – Numbers of scrutiny committee meetings – Numbers of policy recommendations from Scrutiny to the Executive – % of recommendations adopted by the Executive – Numbers of residents attending meetings – Survey of councillors' perceptions of Scrutiny *Qualitative* – In-depth interviews with councillors – Case studies of neighbourhood work *Other* – Official reports and documents – External consultant reports	– Did not continue to meet minimum level one criteria – Lack of structured evaluation framework – Contradictory evidence on verification day – Increase in key quantitative outcome measure is not significant – Some perception measures got worse over period – Impacts and policy intervention not clearly linked to learning and training for councillors – Weak baseline on councillors' skills and policy issues – Weak evidence of role of intervention in creating outcomes – Gaps in documentary analysis – Insufficient range of respondents for qualitative evidence
Application withdrawn before verification	Roles for newly elected councillors	*Quantitative* – Turnover of councillors – Survey of councillors *Qualitative* – In-depth interviews with councillors – Analysis of policy recommendations – Case study of community work *Other* – External audit/inspection reports	– No baseline measures of community issues or negative impacts of previous policies on community – Insufficient evidence of community outcomes – Community outcomes different to descriptions of baseline problems (pre- and post-not the same indicators) – Assertions of community impact not clearly linked to interventions – Impact evaluation only covers learning impacts for councillors – Insufficient range and depth of methods – Rationale for case study selection not clear

Discussion

Our review of the literature identifies a series of problems when participatory research approaches are conflated with non-positivist methods, and vice versa. The Charter was an attempt to overcome an artificial separation of positivist and non-positivist methods into non-participatory and participatory approaches, and take account of the political nature of policy-making. We now describe and discuss findings from the Charter case study.

Engagement of policymakers in evaluation evidence

For many proponents of conventional non-participatory evaluation, social scientists are left bemoaning low levels of engagement of policymakers with hard outcome evidence. 'Enemies' of evidence-based policy include a distrust of information generated outside the system (Leicester 1999: 6). Others have pointed to cultural and institutional gaps between the 'two communities' (Caplan 1979) or 'two tribes' (Orr and Bennett 2010: 199) of researchers and policymakers which has hindered evidence-based policy.

In the Charter scheme, local politicians expressed a desire for scientific evidence and robust outcome evaluation, and to undertake this work themselves. A committee for the sponsoring organization which agreed the design and criteria was made up of local councillors across the region, representing their political groups. During the same period, other regions in England were developing similar schemes. Doing evaluation was a departure from normal practice, and required significant extra effort.[4] In a 2009 review meeting, successful applicants described the high standards as '*justifiably tough*'. Serious interest from around half of the region's authorities was a validation of the value of evaluation, although take-up was gradual over the life of the scheme. Our findings echo those of Caplan, who found 'government executives do not need to be convinced of the potential usefulness of scientific information' (Caplan 1979: 461) for policy formulation.

Re-politicizing evidence-based policy

We have discussed how research can have an ancillary role, with policymakers ignoring evidence in favour of conventional wisdom or ideologically driven choices in line with dominant values or consensus, and biased towards electoral gain. The Charter is a rare counter-example to research which has found politics and 'rational' science to be in opposition (Davies *et al.* 2000: 14). This scheme suggests ways in which evidence-based policy could be re-politicized and brought more in line with 'the Lasswellian approach [which] is about ensuring that knowledge is politicised', and value based (Parsons 2002: 56).

The Charter played a role in forcing local politicians to address the ideological biases and pressures on them towards political or electoral considerations. In order to provide adequate baseline data, politicians needed to overcome their reluctance to admit failures in the democratic process, and accept 'inconvenient truth' over 'comfortable spin' (Bond *et al.* 2011: 90) about their own previous practice. For example, in Rossendale, the changed

relationships between councillors and the public towards more positive collaborative arrangements could not be proven without first evidencing negative relationships before the intervention. The application included evidence of a high profile oppositional campaign conducted by a residents' organisation through social networking websites and in public meetings. Hyndburn Borough Council's application took a politically bold approach to acknowledging community anger and resentment over demolition of homes and previous resource allocation decisions. In only one successful application was the identification of previous democratic failures related to the policies of a previous administration from a different political group. In all of the other successful applications, the baseline data was from a period under the same political party's leadership, although – as many politicians are aware – judicious use of tactical honesty can also be a way of making political capital. Admissions of failure were in the context of positive evaluations which attempted to demonstrate how the administrations had improved.

Organizational capacity for learning

Parsons has argued that 'government and [public] policy-making [is] a process of learning . . .' (Parsons 2002: 48). In contrast to the more anti-positivist and pessimistic stances in Parson's and Schön's original works, authorities were keen to develop their capacity for learning and improve their evaluation techniques, but were intimidated. As one authority said, '*When . . . I [thought about] how to evidence everything . . . and I was mind boggled, I haven't got the foggiest*' (NWE 2008: 7). The Charter was a 'steep learning curve that forced them to assess how they properly evaluated training' (NWE 2010: 6). Authorities in the 2009 review meeting compared the Charter criteria to other accreditation they had undergone and described it as being like '*a jump from "O" to "A" levels*'. 'All without exception agreed that working for Level Two had not been easy, but the outcomes had made the hard work worth it in the end . . . they all faced stiff challenges when identifying and collecting evidence' (NWE 2010: 3).

A focus on high salience policy issues

While some academic research is esoteric, irrelevant or over-theoretical, practitioners are said not to address challenging or important policy problems with sufficient technical rigour (Schön 1983: 42–3). Assigning a value or level of priority to the policy topics in the evaluations is extremely complex. However, none of the applications covered policy areas which could be said to be irrelevant, marginal or low importance and priority, although one implication of this is that traditionally neglected policy areas remained neglected. Several covered controversial policy changes, for example, both applications from the Fire Authorities concerned significant service restructuring following cuts in central government funding, which had been similarly negatively received by the public, and from staff and unions as a '*return to Victorian working practices*', and likely to cause deaths by fire.

Integrity of self-evaluation research

A dominant tradition in participatory research has engaged potential research users and subjects, but is critiqued for its over-reliance on practitioners' insights and practical reason rather than defensible or gold standard evaluation designs. The research designs and methods used in the case study scheme suggest that a wide range of research methods, some more innovative than others, were used within defensible but restrained research designs. Applications which did not progress were a positive sign of recognition of gaps in research designs. Many of the applicants and potential applicants demonstrated an intuitive understanding of some aspects of evaluation. For example, (unresolved) problems with potential applicants who did not submit written applications including finding pragmatic methods to get a suitable baseline measures, and questions about reasonable timescales over which to assess outcomes. Some gaps remained in understandings of evaluation methods for the successful applicants. At least two of the successful applications included survey or statistical data which was not plausibly related to their particular intervention. Ironically, one of the organizations reflected that the experience had taught them that 'some of the best evidence is not statistical or survey based' (NWE 2010: 9).

In many cases, conventional data sources and administrative data were not fully suitable for the applications, and applicants shifted from their initial over-reliance on quantitative data towards mixed methods designs and 'diverse evidence gathering techniques' (NWE 2010: 3) including 'what Councillors, residents and partners have to say' (NWE 2010: 9). One successful application developed a recording system for qualitative material, offering one lesson for other applicants as, 'record all feedback, including evaluation forms, emails, casual conversations' (NWE 2010: 39). Having done this, applicants managed to correct over-reliance on self-reported outcomes solely from councillors, by bringing in other qualitative data. Cohen and Levinthal identify as part of critical knowledge awareness of, and ability to assimilate and apply, 'useful complementary expertise . . . within and outside the organization' (Cohen and Levinthal 1990: 133). As the research designs and methods indicate, applications used data from other departments and from partner organizations, and one lesson identified by applicants was the role of 'constructive communication with other departments' (NWE 2010: 7).

Objectivity of findings

Self-evaluation is seen as questionable as it lacks integrity and independence which is said to guarantee objective results, and is therefore biased towards positive results. Partly because of the community impact criteria, the applications and potential understandably focused on interventions which had had positive effects. A bias towards publication of positive findings and against confirmation of null hypotheses is also common in academic journal publishing, and the release of results by commissioners of external research. Possible applications to the Charter scheme did not choose policy areas which were likely to show negative or null effects. Indeed, in one example, the authority

developed a credible research design and collected data, but did not proceed to a written application when the evidence began to indicate change in the wrong direction, showing citizens' unhappiness with decisions made. However, some applications managed to achieve a degree of subtlety in their positive portrayals. In one example, a swimming pool still closed; residents remained opposed to the decision. Their application documented positive improvements to citizens' perception of the transparency of the decision, and increased two-way dialogue between policymakers and citizens.

Conclusion

Barriers to full and effective use of social science in public policy-making are related to a separation of knowledge production from its consumption and utilization, and by a lack of mechanisms to allow for the political nature of knowledge transfer and decision-making processes. Avoiding policy engagement is a denial of some core ambitions for policy sciences and applied social science. One implication is that more participatory and co-productive approaches to evaluation could be used by researchers and policymakers, particularly politicians, working together. A dominant tradition in much participatory research has focused on values of empowering practice and recognition of different but equal claims to knowledge, more than it has on the quality of evaluation methodologies and objective reliable standards of evidence.

A case study was used of an attempt to reconcile tensions between participatory approaches and positivist methods, and close the divide between evidence and politically situated policymakers. A self-evaluation scheme by local government policymakers specified rigorous evaluation methods and criteria. We found positive levels of interest in outcome evaluation evidence. Undertaking self-evaluation to academic standards forced local politicians in the case study to address some of their ideological biases, although in politics even honesty can be used for political gain. Successful evaluations by local government in the case study used 'hard' data on policy outcomes in mixed methods pre- and post-intervention designs. This enhanced organizational capacity for learning. Policymakers in the case study did not all achieve robust research designs and methods, and some were defeated by lack of suitable data or sufficiently crystallized research questions. A bias towards positive results was in-built in the scheme, and also remained in the choice of policy intervention for evaluation. One conclusion from the case study is that the participation of policymakers in conducting research may simply replicate difficulties faced by academic researchers using more conventional methods. The Charter gave policymakers more appreciation of the value of academic expertise, with reflections from participating authorities urging potential applicants to get academic support (NWE 2010).

Overall, the Charter case study was a way of testing the proposition that a belief in positivism and a commitment to participatory approaches does not need to lead to a schizophrenic research life. There are other ways to achieve participatory positivism, for example, field experiments conducted with

citizen and local government partners (Cotterill and Richardson 2010; John *et al.* 2011). One hope for the Charter is that appreciation of evaluation methods will be more deeply embedded as a result of the depth of processing of the material (Cohen and Levinthal 1990: 131). Dewey advocated 'a kind of thinking that consists in turning a subject over in the mind and giving it serious . . . consideration' (Dewey 1933: 3), and reflection which was 'a conscious and voluntary effort to establish belief upon a firm basis of evidence and rationality' (Dewey 1933: 9). Placing politicians in a dual role, as decision-makers and as evaluators, is an attempt to demonstrate that the research boot of evidence-based policy can be worn on a policymaker's foot, even if there are a few blisters along the way.

Notes

1. A regional organization for local government which provides services on executive recruitment, personnel management, workforce development, learning and development for elected members. NWE is one of nine similar regional employers' organizations across England, with each regional body operating independently.
2. Although decentralization in the fragmented US context is seen as a hindrance to the use of research in policy (Wilensky 1997).
3. There were 52 local government organizations in the region at the start of the scheme, dropping to 47 in 2009 due to local government reorganization.
4. Authorities anticipated spending around ten to 20 days in total on each completed application, which represented around 5 to 10 per cent of one full-time officer, in addition to the additional work for the councillors.

References

Armstrong, F. and Alsop, A. (2010), Debate: co-production can contribute to research impact in the social sciences, *Public Money and Management*, 30, 4: 208–10.

Bennett, F. and Roberts, M. (2004), *From input to influence: participatory approaches to research and inquiry into poverty*, York: Joseph Rowntree Foundation.

Black, N. (2001), Evidence based policy: proceed with care, *BMJ*, 323, 4: 275–9.

Blunkett, D. (2000), *Influence or irrelevance: can social science improve government?* Swindon: Economic and Social Research Council and Department for Education and Employment.

Bond, L., Sautkina, E. and Kearns, A. (2011), Mixed messages about mixed tenure: do reviews tell the real story? *Housing Studies*, 26, 1: 69–94.

Boud, D., Keogh, R. and Walker, D. (eds) (1985), *Reflection: turning experience into learning*, London: Kogan Page.

Boyer, E. L. (1996), The scholarship of engagement, *Journal of Public Outreach*, 1, 1: 11–20.

Campbell, S., Benita, S., Coates, E., Davies, P. and Penn, G. (2007), *Analysis for policy: evidence-based policy in practice*, London: HM Treasury.

Caplan, N. (1979), The two-communities theory and knowledge utilization, *American Behavioral Scientist*, 22, 3: 459–70.

Catalani, C. and Minkler, M. (2009), Photovoice: a review of the literature in health and public health, *Health Education and Behavior*, 37, 3: 424–51.

Cohen, M. D., March, J. G. and Olsen, J. P. (1972), A garbage can model of organizational choice, *Administrative Science Quarterly*, 17, 1: 1–25.
Cohen, W. M. and Levinthal, D. A. (1990), Absorptive capacity: a new perspective on learning and innovation, *Administrative Science Quarterly*, 35, 1: 128–52.
Cotterill, S. and Richardson, L. (2010), Expanding the use of experiments on civic behavior: experiments with local government as research partners, *Annals of the American Academy of Political and Social Science*, 628: 148–64.
Davies, H. T. O. and Powell, A. E. (2012), Communicating social research findings more effectively: what can we learn from other fields? *Evidence and Policy*, 8, 2: 213–33.
Davies, H. T. O., Nutley, S. M. and Smith, P. C. (1999), What works? the role of evidence in public sector policy and practice, *Public Money and Management*, 19, 1: 3–5.
Davies, H. T. O., Nutley, S. M. and Smith, P. C. (2000), *What works: evidence-based policy and practice in public services*, Bristol: Policy Press.
DeLeon, P. (1994), Reinventing the policy sciences: three steps back to the future, *Policy Sciences*, 27: 77–95.
DeLeon, P. and Longobardi, R. C. (2002), Policy analysis in the good society, *The Good Society*, 11, 1: 37–41.
Dewey, J. (1933), *How we think*, Boston, MA: D C Heath and Co.
Duncan, S. (2005), Towards evidence-inspired policy making, *Social Sciences*, 61: 10–11.
Durose, C., Beebeejaun, Y., Rees, J., Richardson, J. and Richardson, L. (2011), *Towards co-production in research with communities*, London: Arts & Humanities Research Council.
Ekblom, P. (2002), From the source to the mainstream is uphill: the challenge of transferring knowledge of crime prevention through replication, innovation and anticipation. In N. Tilley (ed.), *Analysis for crime prevention: Crime prevention studies, volume 13*, Monsey, NY: Criminal Justice Press.
Etzioni, A. (1988), *The moral dimension: towards a new economics*, New York, NY: Free Press.
Freire, P. (1970), *Pedagogy of the oppressed*, London: Continuum.
Flinders, M. (2010), In defence of politics, *Political Quarterly*, 81, 3: 309–26.
Flinders, M. (2012), Keynote speech to Political Studies Association, Belfast, 12–14 April.
Gains, F. and Stoker, G. (2011), Special advisers and the transmission of ideas from the policy primeval soup, *Policy and Politics*, 39, 4: 485–98.
Heron, J. and Reason, P. (1997), A participatory inquiry paradigm, *Qualitative Inquiry*, 3, 3: 274–94.
John, P. (2013 forthcoming), Political science, impact and evidence, *Political Studies Review*, 11, 2.
John, P., Cotterill, S., Moseley, A., Richardson, L., Smith, G., Stoker, G. and Wales, C. (2011), *Nudge, nudge, think, think: experimenting with ways to change civic behaviour*, London: Bloomsbury Academic.
Kolb, D. A. (1984), *Experiential learning: experience as the source of learning and development*, Englewood Cliffs, NJ: Prentice Hall.
Ledwith, M. (2007), On being critical: uniting theory and practice through emancipatory action research, *Educational Action Research*, 15, 4: 597–611.
Le Grand, J. (2006), Academia, policy and politics, *Health Economics, Policy and Law*, 1: 319–22.
Leicester, G. (1999), The seven enemies of evidence-based policy, *Public Money and Management*, 19, 1: 5–7.
Lewin, K. (1946), Action research and minority problems, *Journal of Social Issues*, 2, 4: 34–46.

Lindblom, C. E. (1959), The science of 'muddling through', *Public Administration Review*, 19, 2: 79–88.
Martin, S. (2010), Co-production of social research: strategies for engaged scholarship, *Public Money and Management*, 30, 4: 211–18.
Minkler, M. and Wallerstein, N. (eds) (2003), *Community-based participatory research for health*, San Francisco, CA: Jossey-Bass.
Mulgan, G. (2005), Government, knowledge and the business of policy making: the potential and limits of evidence-based policy, *Evidence and Policy*, 1, 2: 215–26.
Myrdal, G. (1972), How scientific are the social sciences, *Journal of Social Issues*, 28, 4: 151–70.
Newman, J. (2011), Boundary troubles: working the academic-policy interface, *Policy and Politics*, 39, 4: 473–84.
North West Employers (NWE) (2008), *Making a difference: a toolkit to help councils demonstrate where and how elected member development has community impact*, Manchester: NWE.
North West Employers (NWE) (2010), *North West Member Charter Case Studies*, Manchester: NWE.
Nutley, S. M., Walter, I. and Davies, H. T. O. (2007), *Using evidence: how research can inform public services*, Bristol: Policy Press.
Orr, K. M. and Bennett, M. (2009), Reflexivity in the co-production of academic-practitioner research, *Qualitative Research in Organisations and Management*, 4: 85–102.
Orr, K. M. and Bennett, M. (2010), Editorial, *Public Money & Management*, 30, 4: 199–203.
Palfrey, C., Thomas, P. and Phillips, C. (2012), *Evaluation for the real world: the impact of evidence in policy making*, Bristol: Policy Press.
Parsons, W. (2002), From muddling through to muddling up – evidence based policy making and the modernisation of British government, *Public Policy and Administration*, 17, 3: 43–60.
Pohl, C., Rist, R., Zimmerman, A., Fry, P., Gurung, G. S., Schneider, F., Speranza, C. I., Kiteme, B., Boillat, S., Serrano, E., Hirsch Hadorn, G. and Weismann, U. (2010), Researchers' roles in knowledge co-production: experience from sustainability research in Kenya, Switzerland, Bolivia and Nepal, *Science and Public Policy*, 37, 4: 267–81.
Riger, S. (1992), Epistemological debates, feminist voices: science, social values and the study of women, *American Psychologist*, 47, 6: 730–40.
Robb, C. M. (2002), *Can the poor influence policy? Participatory Poverty Assessments in the developing world* (2nd edn), Washington, DC: International Monetary Fund and World Bank.
Sanderson, I. (2002), Evaluation, policy learning and evidence-based policy making, *Public Administration*, 80, 1: 1–22.
Schön, D. A. (1983), *The reflective practitioner*, New York, NY: Basic Books.
Smith, K. E. and Joyce, K. E. (2012), Capturing complex realities: understanding efforts to achieve evidence-based policy and practice in public health, *Evidence and Policy*, 8, 1: 57–78.
Torgerson, D. (1986), Between knowledge and politics: three faces of policy analysis, *Policy Sciences*, 19: 33–60.
Van de Ven, A. H. (2007), *Engaged scholarship: a guide for organizational and social research*, Oxford: Oxford University Press.
Walt, G. (1994), How far does research influence policy?, *European Journal of Public Health*, 4: 233–5.
Weiss, C. (1979), The many meanings of research utilization, *Public Administration Review*, 39: 426–31.

Weiss, C. (1993), When politics and evaluation research meet, *American Journal of Evaluation*, 14, 1: 93–106.
Weiss, J. A. (1979), Access to influence: some effects of policy sector on the use of social science, *American Behavioral Scientist*, 22, 3: 437–58.
Wilensky, H. L. (1997), Social science and the public agenda: reflections on the relation of knowledge to policy in the United States and abroad, *Journal of Health Politics, Policy and Law*, 22, 5: 1241–65.

Index

aboriginal communities, 2, 47–58
  alternative strategies, 57–8
  critical framework, 55–7
  evidence as knowledge, 50–2
  evidence as politics, 52–5
  evidence-based policy discourse, 49–50
acquis communautaire, 97, 106
Action on Smoking and Health (ASH), 29
activation programmes, 14–15, 16
active labour market policies (ALMP), 14
Actor–Network Theory (ANT), 36–7, 40
advertising and tobacco, 29, 31
advocacy coalition framework (ACF), 33–6, 39, 40, 41
advocacy coalitions, 3, 101, 109, 110
Africa, 10
Albania, 97, 110
alcohol, 19, 30, 47, 85
ante-natal care, 10
Attitude Survey Study, 68–74
Australia, 2, 68, 47–58
  alternative strategies, 57–8
  critical framework, 55–7
  evidence as knowledge, 50–2
  evidence as politics, 52–5
  evidence-based policy discourse, 49–50
  income management, 19–21, 48, 51–2, 54–5, 58
  tobacco and smoking, 32, 82–4, 86, 89, 90

background information, 8
Bangladesh, 12
Bartlett, Will, 2–3, 97–113
Berridge, Virginia, 30, 31, 32, 33, 38, 39
bias, 1, 7, 17, 74–6
  participatory research, 130, 139, 141, 142
  public opinion and smoking, 84–8
  systematic reviews, 2, 63, 65, 67, 70–3, 74–6, 99
Blackpool Council, 134, 135
Blunkett, David, 64
Boolean operators, 69, 78
Bosnia, 104, 110
Brazil, 9, 10, 13, 14
Bulgaria, 104

Cambodia, 12
Campbell Collaboration, 22, 64
Canada, 34, 68, 83, 89, 90
case-controlled studies, 8
Centre for Reviewers and Dissemination (CRD), 64–5, 70
Cheshire West and Chester Council, 135
child labour, 16
childcare, 15, 20, 116, 118, 119, 121
Chile, 11, 12
China, 16, 17
Chorley Council, 135
Coalition government, 3, 20, 124
Cochrane Foundation, 64, 70, 98
cohort studies, 8, 91
Colombia, 10, 11, 12, 13
conditional cash transfer (CCT) programmes, 5, 8–13, 16, 18, 20, 21
conditionality, 5, 10, 14, 18–19, 20–1
  EU accession, 1–6
Conservative party, 134, 135
context of systematic reviews, 65, 66, 67, 72
Copenhagen criteria, 106
cost-benefits, 10, 56
  tobacco and smoking policies, 32, 39
  welfare policy, 15, 18, 22
Country Assistance Strategies, 107
Cox, E, 56
critically-appraised topics, 9
Croatia, 104, 107, 110
cross-national surveys, 63–4, 74, 76, 77

debt, 17, 108
Deeming, Christopher, 2, 5–27
Department for International Development (DfID), 98–9
Department for Work and Pensions (DWP), 116–17, 118–19, 120
development policy, 6–10, 11, 12, 13, 14
difference-in-differences (DID), 11, 12, 100
distance to work, 118, 119, 120, 122
duplication of research, 120, 121, 122, 123

Evidence and Evaluation in Social Policy, First Edition. Edited by Ian Greener and Bent Greve. Book compilation © 2014 John Wiley & Sons Ltd.

Economic and Social Research Council (ESRC), 64
Ecuador, 10, 11, 12, 13, 14
education, 15, 16, 18, 103
  aboriginal communities, 47–8, 50, 53–4
  development policy, 6, 9–10, 12
  EU enlargement countries, 97, 105–10
  evidence-based policy-making, 97–100
  performance-related pay, 98
  systematic reviews, 63, 64, 70, 75, 99
Education Action Zones, 100
employment, 14–15, 16, 19, 20, 21
  aims and objectives, 116
  EU enlargement countries, 102–10
  evidence-based policy-making, 99, 101, 115–17, 122–3, 124
  knowledge after 13 years of research, 118–20
  lone parents, 3, 115–24
  research under Labour government, 116–17
  role of research, 122–3
  skills, 102–3, 105–8
  Western Balkans, 105–8
  what went wrong, 120–2
England and Wales, 98, 132, 135–9, 143
enlightenment model, 38–9
ethics, 2, 16, 19, 20, 37
Euro-Barometer, 74, 77
Europe, 4, 20, 21, 29
  enlargement areas, 2–3, 97–111
  evidence-based policy making, 3, 21, 29, 97–111
  policy process, 101–5
  systematic reviews, 68, 71, 73, 74, 76
  transition economies, 101–2
  Western Balkans, 105–8
European Commission, 105, 106
European Social Survey (ESS), 74, 77
European Training Foundation, 106
European Values Survey (EVS), 74, 77
evidence as knowledge, 50–2
evidence as politics, 52–5
evidence-based policy-making, 1, 5–6, 19, 21, 98–101, 139–40
  alternative strategies, 57–8
  critical framework, 55–7
  employment, 99, 101, 115–17, 122–3, 124
  EU enlargement countries, 3, 21, 29, 97–111
  evidence as knowledge, 50–2
  evidence as politics, 52–5
  evidence-based policy discourse, 49–50
  Labour government, 29, 123
  Northern Territory, 2, 47–58
  origins, 3, 98–101
  participatory research, 129, 130, 132, 139–40, 143
  public opinion, 84, 93
  systematic reviews, 64, 98–9
Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), 64, 70
evidence synthesis, 8
expert opinion, 8
explanation building, 88

Families and Children Study (FACS), 117, 119, 121, 122, 123, 124
Finer Report on One Parent Families, 123
Fisher, Ronald, 31
Fixed Effects statistical model, 12
Flyde Council, 135, 136
food and diet, 6, 10, 13, 17
frames and framing, 33, 36, 37, 38–9, 40
France, 71, 74

Germany, 71
Gillard, Julia, 47, 48, 52
Great Britain and lone parents' employment, 115–24
  aims and objectives, 116
  evidence-based policy making, 115–17, 122–3, 124
  knowledge after 13 years research, 118–20
  research under Labour government, 116–17
  role of research, 122–3
  what went wrong, 120–2
Greater Manchester Fire and Rescue Authority, 135
Greener, Ian, 1–4
Greve, Brent, 1–4
Guatemala, 10

Halton Council, 135
Haux, Tina, 2, 115–28
Hawthorn effects, 17–18, 100
Health Action Zones, 100
health and healthcare, 1, 15, 17, 18, 118
  development policy, 6, 9–11
  evidence-based policy-making, 99–100
  smoking, 2, 29–41, 82, 85
  systematic reviews, 63, 64, 99
  see also medicine
Herzegovina, 110
heterogeneity, 16, 65, 66, 67, 75–6
Hicks, D, 71–2, 73
hierarchy of evidence, 6, 7, 8, 65–6, 74
Honduras, 11, 12, 13
horizon scanning, 122–3
housing development policy, 6, 10
Howard, John, 47, 58
Hyndburn Council, 135, 136, 140

ideational approaches, 37–9, 40–1
immunization, 10

income management, 19–21, 48, 51, 52, 54–5, 58
India, 17
indigenous policy, 2, 19, 48–9, 54, 57–8
Instrument for Pre-Accession Assistance (IPA), 103–4, 107, 110
integrity, 141
International Bibliography of the Social Sciences (IBSS), 68–9, 71–4
International Monetary Fund, 7
International Social Survey Programme (ISSP), 74, 77
Investment Reform Index, 107
Ireland, 91
ISEA, 77
ISJP, 77

Jamaica, 11, 12, 14
JISC Academic Database Assessment Tool, 71, 72
judgement, 65, 66, 67, 70, 72, 74

Knowsley Council, 135, 136
Kosovo, 97, 110

Labour party (including New Labour), 1, 29, 98, 99–100
  employment and lone parents, 115–17, 120, 123, 124
  participatory research, 134, 135
Lancashire Combined Fire Authority, 135, 137
Liberal Democrat party, 134, 135
Lindblom, C E, 98–9
living standards development policy, 6
LMICs, 5, 10, 16
  welfare policy, 15, 21–2
local government, 3–4, 132–43
Lone Parent Obligation Survey, 124
lone parents, 3, 19, 20, 115–24
  aims and objectives, 116
  evidence-based policy making, 115–17, 122–3, 124
  knowledge after 13 years research, 118–20
  research under Labour government, 116–17
  role of research, 122–3
  what went wrong, 120–2

Macedonia, 107, 110
Macklin, Jenny, 48, 51, 52–4, 55
malaria, 10
Malawi, 6–7, 9
medicine, 98, 100
  systematic reviews, 63, 66, 70, 74
  see also health and healthcare
Merseyside Fire and Rescue Authority, 135
meta-analysis, 22, 64, 65
methodology, 2, 3, 16, 17
  employment and lone parents, 117, 121, 123
  participatory research, 129, 130, 136–8, 141
  public opinion and policy-making, 81, 84–5, 90, 92
  systematic reviews, 65–8, 70, 71, 74, 75
Mexico, 9, 10, 11, 12, 14, 18
Millennium Development Goals (MDGs), 6, 21
Modernising Government White Paper, 115
Montenegro, 103, 110

National Centre for Research Methods (NCRM), 64
National Childcare Strategy, 116
National Health Service Centre for Reviewers and Dissemination (CRD), 64–5, 70
National Institute for Clinical Excellence (NICE), 64
natural sciences, 65, 66, 67, 68–9, 70
network-based approaches, 36–7, 39, 40, 41
New Deal for Lone Parents (NDLP), 116, 117, 118, 121, 122
New Deal programme, 99
New South Wales, 82, 83
New Zealand, 68
smoking, 83, 84, 86, 88
Nicaragua, 10, 11, 12, 13
nicotine replacement therapy, 34, 35
non-governmental organizations (NGOs), 7, 10
North West Employers (NEW), 132–3, 143
North West Member Development Charter Level Two (The Charter), 132–43
Northern Territory, Australia, 2, 47–58
  alternative strategies, 57–8
  critical framework, 55–7
  evidence as knowledge, 50–2
  evidence as politics, 52–5
  evidence-based policy discourse, 49–50
Northern Territory Emergency Response (NTER), 47–58

objectivity, 141–2
ONE, 117
Organisation for Economic Co-operation and Development (OECD), 14, 107
  Investment Compact (IC), 107

Pakistan, 12
participatory research, 129–43
  case study, 132–42
  politics, 131–2

Partridge, Emma, 2, 47–62
passive smoking or second hand smoke, 30, 39
  public opinion, 84, 88, 89
Pawson, Ray, 2, 65, 66, 81–96
Philip Morris (tobacco company), 36, 38–9
pilot studies, 99–101, 115, 117, 121
Poland, 71
policy transfer, 3, 105–8, 109–10
Political Action, 77
Poor Law Amendment Act (1834), 123
positivism, 3, 63, 66, 67
  participatory research, 129, 130, 132, 139, 140, 142
post-positivism, 49–50, 132
poverty, 16, 17, 19, 22, 35, 116
  development policy, 7–8, 9, 10, 14
power relations, 19, 50, 51, 55, 57
public health policy, 29–41
  UK "Tobacco Wars", 30–9
public opinion, 2, 15, 16, 19–20
  Attitude Survey Study, 68–74
  dealing with uncertainty, 88–92
  evidence for policy, 92–3
  overview of evidence, 82–4
  quality and bias, 74–6
  social policy, 63–78
  systematic review, 64–8
  tobacco and smoking, 38, 39, 81–94
  uncertainties in data, 84–7

quality appraisal, 84
question wording, 117, 119, 121–2
  public opinion and smoking, 85, 86, 88, 90

Racial Discrimination Act (1975) of Australia, 47–8
randomized controlled trials (RCTs), 2, 3, 5–23, 31, 65, 117
  development policy, 6–9, 11, 12
randomistas, 7–8
rational evidence-based approach, 30–2, 39
realist synthesis, 66, 81
refutation, 66
regression discontinuity design (RDD), 11, 12
remittance income, 97
replication, 65, 66, 67, 99
research evidence hierarchy, 6, 7, 8, 65–6, 74
Richardson, Liz, 3–4, 129–46
rights, 18, 19, 33, 131
Rossendale Council, 135,139–40
Rudd, Kevin, 47, 48, 52

St Helens Council, 135, 137
scepticism, 66, 67
Scopus database, 68–9
Scotland, 88
SEAM programme, 48, 58
self-evaluation, 130, 132–43
  discussion, 139–42
  local policy makers, 134–8
  politicians in local government, 132–4
Serbia, 100–1, 104, 110
skills, 2–3, 14, 102–3, 105–8
  EU enlargement countries, 102–3, 104–10
Skinner, B F, 5, 21
Smith, Katherine, 2, 29–45
smoking and tobacco, 2, 19, 29–41
  dealing with uncertainty, 88–92
  evidence for policy, 92–3
  overview of evidence, 82–4
  public opinion, 81–93
  taxation, 30, 31–2, 34–6, 40
  UK "Tobacco Wars", 30–9
  uncertainties in data, 84–7
social assistance, 5, 6–7, 10, 19, 22, 107
Social Care Institute for Excellence, 67
social desirability effect, 85–7, 88
social science, 20, 74–6, 130–1
  participatory research, 130–2, 139, 142
  systematic reviews, 63–4, 65–71, 73–7
social security, 5, 14, 17–18, 19, 22, 23
Social Services Citation Index (SSCI), 71, 73
Somalia, 17
stigmatization, 19, 35, 85, 93
Sundberg, Trude, 2, 63–80
systematic reviews, 2, 3, 8, 10, 15, 64–8
  Attitude Survey Study, 68–74
  attitudes to social policy, 63–78
  evidence-based policy, 98–9
  quality and bias, 74–6

Tameside Council, 135
tax credits, 116, 117
taxation and tobacco, 30, 31–2, 34–6, 40
Taylor-Gooby, Peter, 2, 63–80
tobacco *see* smoking and tobacco
Tobacco Control, 30, 33
transition economies, 101–2
transparency, 142
  systematic reviews, 64, 65, 66, 67–8, 99
Turkey, 12, 108–9, 110

unconditional cash transfers, 17, 18
Understanding Society, 124
United Kingdom, 1, 3, 14, 15, 19, 20, 21
  evidence-based policy-making, 98–9
  smoking, 29, 30–9, 40, 41, 90
  systematic reviews, 70–1, 73
  "Tobacco Wars", 30–9
United Nations, 17
  Developmental Programme (UNDP), 100–1, 105, 107–8
  Universal Declaration of Human Rights (1948), 18

United States of America (USA), 10, 14, 98, 143
  smoking, 34, 39, 83, 89, 90
  systematic reviews, 68, 70–1, 73, 74, 77

value-orientated approaches, 33–6, 39, 40
vocational education and training (VET), 103–4, 110

Warrington Council, 134, 135
Washington Consensus Policies, 16, 21
Web of Knowledge (WoK), 72
Web of Science (WoS), 68–9, 70–4
  dominance, 72–3
Weiss, C, 38
welfare policy, 5–6, 10, 14–15, 18, 19, 20–2
  aboriginal communities, 47–8, 50, 53–4
  employment and lone parents, 116
  smoking, 88
  systematic reviews, 69, 74, 75–6
welfare-to-work, 14–15, 20
Western Balkans, 102, 103, 104, 105–8
Wong, Geoff, 2, 81–96
workfare, 10, 14, 21
World Bank, 7, 10, 39, 105, 107
World Health Organization (WHO), 29
World Trade Organization, 7
World Values Survey (WVS), 74, 77
Wyre Council, 134, 135, 137