Forecasting the Effectiveness of Policy Implementation Strategies

A thesis submitted to the University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Humanities

2010

Nicolas D. Savio

Manchester Business School

Contents

Word Count: 57,525

List of Figures

List of Tables

Forecasting the Effectiveness of Policy Implementation Strategies

Nicolas D. Savio

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Humanities

2010

Abstract

An important stage in the policy process involves deciding what strategy is to be adopted for implementation so that the objectives of the policy are met in the best way possible. A Policy Implementation Strategy (PIS) adopts a broad view of implementation, which is argued to transcend formulation and decision-making, thereby offering a more realistic view of the policy process. Governmental decision-makers are often faced with having to choose one PIS amongst several possible alternatives, at varying cost levels. In order to aid in such a decision-making process, PIS effectiveness forecasts are proposed as a decision-support tool.

Current methods for such a purpose are found to include ex-ante evaluative techniques such as Impact Assessment (IA) and Cost-Benefit Analysis (CBA). However, these approaches are often resource-intensive and such an investment is not always rewarded with accurate predictions. Hence, a judgmental forecasting approach for making PIS effectiveness predictions is proposed as a means for screening the different PIS under contention to provide a shortlist of candidates with particular potential. The selected few can then be further analysed via the quantitative evaluative techniques such as IA and CBA. Judgmental approaches to forecasting are considered ideal for such a role because they are relatively quick and inexpensive to implement. More specifically, a structured analogies approach is proposed as information about analogous PIS is believed to be useful for such a purpose.

The proposed structured analogies approach is tested over a series of experiments and the evidence suggests that a structured analogies approach is more accurate when compared to unaided judgment and the more support given to the expert the better. Furthermore, experts were seen to produce considerably more accurate predictions than non-experts. Level of experience and number of analogies recalled did not seem to affect accuracy. The expert forecasts were also comparable to those produced by governments. The thesis concludes with suggestions for future research in the area.

Declaration

The author of this thesis hereby states that no portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university of institute of learning.

Copyright Statement

(i)   The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

(ii)   Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

(iii)   The ownership of certain Copyright, patents, designs, trade-marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

(iv)   Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on Presentation of Theses.

Acknowledgements

I would first of all like to thank my supervisor, Professor Kostas Nikolopoulos, not only for the faith he showed in me but also for his continued guidance, support and encouragement throughout the whole process and without whom, such an investigation would not have been possible.

I would also like to thank my second supervisor, Dr. Michael Keenan, for the advice and suggestions he gave to help direct the investigation in the best possible line.

I am also grateful to the Economic and Social Research Council as well as the Engineering and Physical Sciences Research Council for their continued funding.

My thanks also extend to the individuals from the academic and governmental organisations which partook in the numerous interviews and experiments conducted throughout the research process.

I would lastly, but by no means least, like to thank my parents, for their moral and financial support for not only my PhD (which is only the latest step) but also for their investment in the educational and academic project that has been the last 20 years of my life.

Chapter 1: Introduction

1.1 Background

1.1.1 The Field of Forecasting

The Forecasting Dictionary defines forecasting as,

Estimating in unknown situations. Predicting is a more general term and connotes estimating for any time series, cross-sectional, or longitudinal data. Forecasting is commonly used when discussing time series.

(Forecastingprinciples.com, 2009a)

The activity of 'forecasting' is an integral part of everyday life. Many individuals do not realise that they are constantly making predictions about the future, albeit most of the time subconsciously. What's more, whether it be about what route to take to the office or what clothes to wear that day, strategy is almost always based upon estimates of future events, and rightly so. The same occurs in the professional world, in the private and public sector alike, whether it is about predicting the demand of a product or formulating a policy to solve a problem, forecasting is everywhere.

All the more surprising then for such an important activity, that forecasting only became a formal independent field 30 years ago[1]. In the early 1980s, a group of the field's pioneers[2] joined to form the International Institute of Forecasters (IIF), which was quickly followed by the creation of the two main existing journals, first the Journal of Forecasting (JoF) in 1982 and then the International Journal of Forecasting (IJF) in 1985. The IIF's aim was the

---

[1] Of course, before then, forecasting was thriving as a sub-discipline in various domains.
[2] Spyros Makridakis, Scott J Armstrong, Robert Fildes, Robert Carbone.

"…developing and furthering the generation, distribution, and use of knowledge on forecasting through the following objectives:

- Develop and unify forecasting as a **multidisciplinary field of research** drawing on management, behavioural sciences, social sciences, engineering, and other fields.
- Contribute to the **professional development** of analysts, managers, and policy makers with responsibilities for making and using forecasts in business and government.
- Bridge the gap between **theory and practice**, with practice helping to set the research agenda and research providing useful results.
- Bring together decision-makers, forecasters, and researchers from **all nations** to improve the quality and usefulness of forecasting".

(Forecasters.org, 2009)

With the creation of the IIF and the two journals, the field finally had the infrastructure it needed for development. As the word spread, much research became published in journals and books, resulting in the body of forecasting knowledge which exists today (Fildes and Nikolopoulos, 2006).

Forecasting is multidisciplinary in nature as it lies at the junction of several different fields such as statistics, mathematics, psychology, information sciences, management, policy and so on. In any given forecasting situation the field(s) which provides the basis for the approach to be used will necessarily intersect with the field where it is being applied, resulting in a large number of possible combinations. In other words, any forecasting approach (quantitative or qualitative) can theoretically be applied to any field (strategy, finance, policy, weather, conflict, etc).

Being such a broad field, few books have been successful in enveloping all this knowledge into a single source. The difficulty of such a task lies in that because there are so many application areas for forecasting and there is such a vast range of methods available for use, the book would have to cover from advanced econometric techniques used in finance to purely judgmental approaches used in long range

planning, and everything in between (Fildes and Nikolopoulos, 2006). One which has succeeded and is regarded as the most complete forecasting textbook is Makridakis et al. (1998). Makridakis et al. (1998) includes the theory on virtually all the major forecasting approaches known to date as well as the scenarios in which they have been applied and gives advice about their proper use to practitioners. Another core text that has seen significant success is Armstrong (2001a). This book supports the notion of 'evidence-based forecasting' and is an amalgamation of expert knowledge and empirical results in forecasting research, culminating in 139 principles the author claims can be used to summarise forecasting knowledge. Both texts are considered fundamental in the study of forecasting for researchers, practitioners and students alike.

Forecasting approaches fall naturally into two broad categories (Forecastingprinciples.com, 2009c) depending on the nature of the available information, qualitative or quantitative (also known as judgmental). In situations of sufficient numerical data, quantitative methods are preferred. In the absence of such a luxury, judgmental approaches are used (Makridakis et al., 1998; De Gooijer and Hyndman, 2006; Forecastingprinciples.com, 2009b). Table 1.1 summarises the main advantages and disadvantages of each type of method.

	Judgmental	Quantitative
Advantages	- Ability to incorporate valuable domain knowledge and so predict unsystematic changes in data - Can function without the need for past numerical data - Often inexpensive	- Can be very accurate - More objective than judgmental approaches - Can be used to predict over multiple forecast horizons
Disadvantages	- Prone to biases - Can require the need for experts in a field	- Unable to incorporate domain knowledge and hence predict unsystematic changes in data - Can be expensive

Table 1.1: Advantages and Disadvantages of Judgmental and Quantitative Forecasting Approaches

Although quantitative approaches have been subject to a vast amount of research (De Gooijer and Hyndman, 2006), judgmental approaches have arguably seen the greatest evolution (Armstrong, 1986; Goodwin and Wright, 1993; Lawrence et al. 2006; Parackal et al., 2007). Makridakis et al. (1998) come to the consensus that judgmental approaches can be very useful and have a lot to offer forecasters but that their successful implementation will depend on understanding the biases and limitations together with the advantages such methods have to offer (chapter 10, p. 496).

A particularly important result that has emerged from forecasting research is that judgmental and quantitative forecasts often work well in combination with each other (Blattberg and Hoch, 1990; Bunn and Wright, 1991; Armstrong and Collopy, 1998). The argument here is that through combination, the advantages of each can be complemented whilst the effects of their limitations, minimised.

1.1.2 Policy Implementation Strategies and Decision-Making

'Policy analysis' can be considered as the approach to public policy that attempts to 'integrate and contextualise models and research from those disciplines which have a problem and policy orientation' (Parsons, 1995, p.XV)

A popular model for such a study, which gains merit in its value as a heuristic for the policy process but which is nonetheless criticised for not being an accurate representation of reality, says that policy can viewed as a five-stage sequential process, which can be compared to the stages of a problem-solving process, as seen in table 1.2.

Stages of the policy cycle	Stages in applied problem solving
Agenda – setting	Recognition of the problem
Policy formulation	Proposal of the solution
Decision-making	Choice of the solution
Policy Implementation	Putting solution into effect
Policy Evaluation	Monitoring results

Table 1.2: Five Stages of the Policy Cycle and Their Relationship to Applied Problem-solving (Howlett and Ramesh, 2003, p.13).

Two fundamental stages where the decisions made will ultimately give shape to the policy and how it is implemented are its formulation and implementation. One of the most important steps in the policy process is the formulation stage where the solution to a public problem is selected, amongst various possible alternatives, for implementation (Parsons, 1995; Hillman and Hitt, 1999; Howlett and Ramesh, 2003; Moran et al., 2008). It also at this stage that it is decided what strategy will be adopted for such an implementation. Such implementation strategies will necessarily make use of the different instruments or tools that governments have at their disposal for such a purpose (Howlett and Ramesh, 2003). They are the actual means governments have for attaining the goals or objectives set out by the policy. The successful delivery of these instruments will ultimately determine whether or not an implementation strategy has been effective in attaining the goals set out by the policy.

Thus, such implementation strategies play a central role in the decision-making process within formulation where governments must decide which will provide the best way to approach the problem within any constraints imposed upon them (Parsons, 1995; Howlett and Ramesh, 2003).

One particular type of strategy which much of this research will centre around is a Policy Implementation Strategy (PIS) (Savio and Nikolopoulos, 2009a, 2009b, 2009c). Policy Implementation Strategies are essentially very similar to policy instruments in that they are used by governments to attain the objectives or targets set out by policy but are broader in their scope. A Policy Implementation Strategy (PIS) adopts a broad view of implementation, which is argued to transcend formulation and decision-making to offer a more realistic view of the policy process. They are the strategy or plan of action that governmental decision-makers have chosen to implement a policy with the added difference that they use financial incentives (provided and funded by governments through tax rebates, credits, VED, scrappage schemes, etc) to promote the adoption of new forms of environmentally clean technology[1] (ECT).

The distinction is made between these strategies and 'policies' because they do not define what the policy goals are, as these have already been established, nor are they

---

[1] Although it will be argued that they are not necessarily confined to such a definition but can refer to any policy instrument in any domain area.

'instruments', as they include an added strategic element not found in these. Moreover, they transcend these artificial formulation and implementation stages set out by the cyclical policy model above, presenting a more realistic perspective of how such a process actually develops, as illustrated in figure 1.1.

Policy Process

Agenda-Setting

Formulation

Decision-making

Policy Strategy and Objectives

Policy Implementation Strategy

Implementation

Evaluation
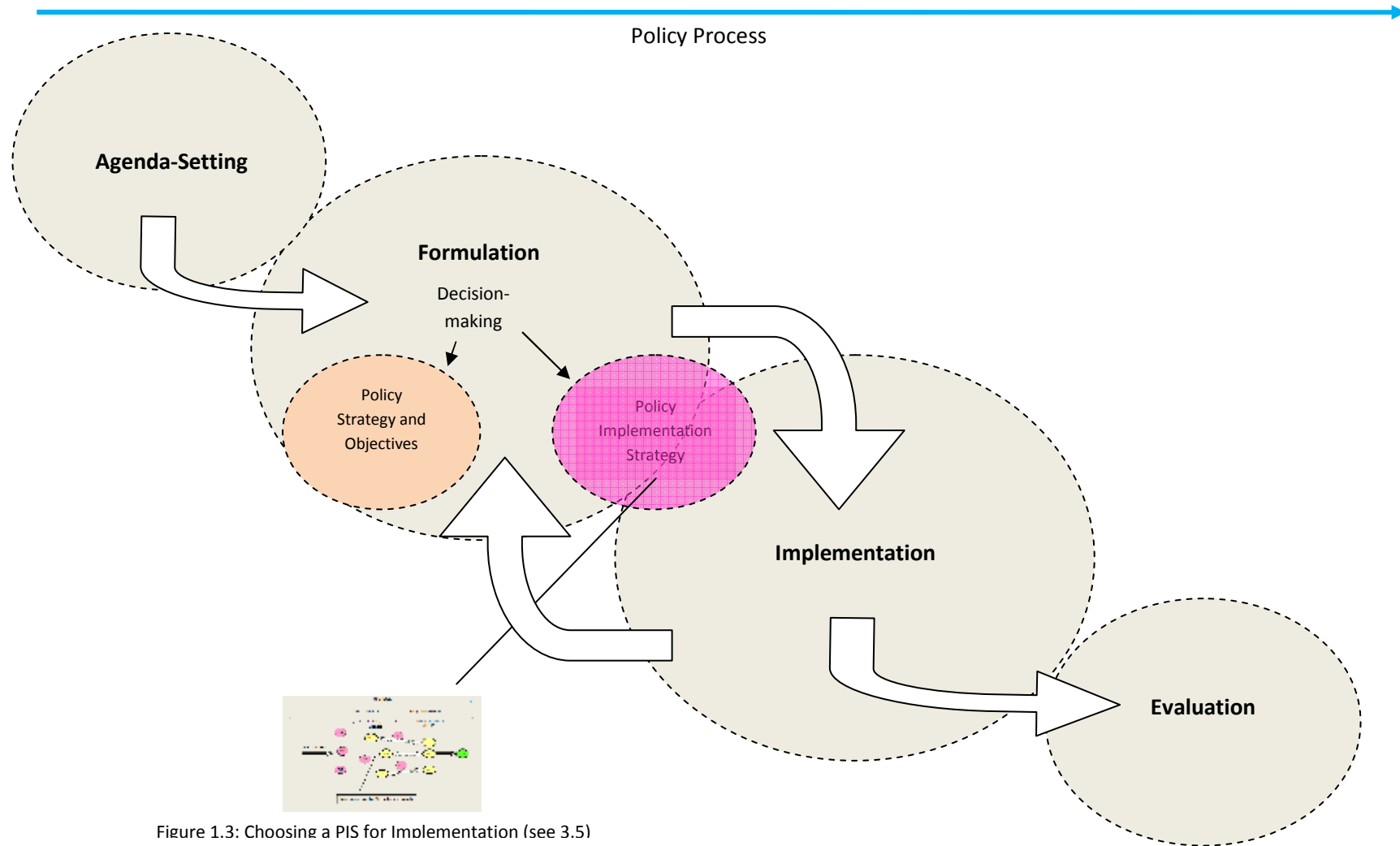
Figure 1.3: Choosing a PIS for Implementation (see 3.5)

Figure 1.1: Stages of the Policy Process and the Positioning of the PIS

Figure 1.1 shows that the positioning of the PIS to be intersecting the formulation and implementation processes.

Clearly these strategies are important because they define how a policy will be delivered. Like for any implementation strategy, a successful PIS is a big step towards a successful policy. Moreover, what might have been a successful and effective policy can be ruined by a poorly designed and ineffective strategy.

With such a broad variety of shapes a PIS can take, with differing costs and suitability, the decision as to which to select is not so straightforward and has been the subject of extensive study (Peters and Van Nispen, 1998; Bemelmans-Videc et al., 1998). Such a decision is illustrated graphically in figure 1.2.
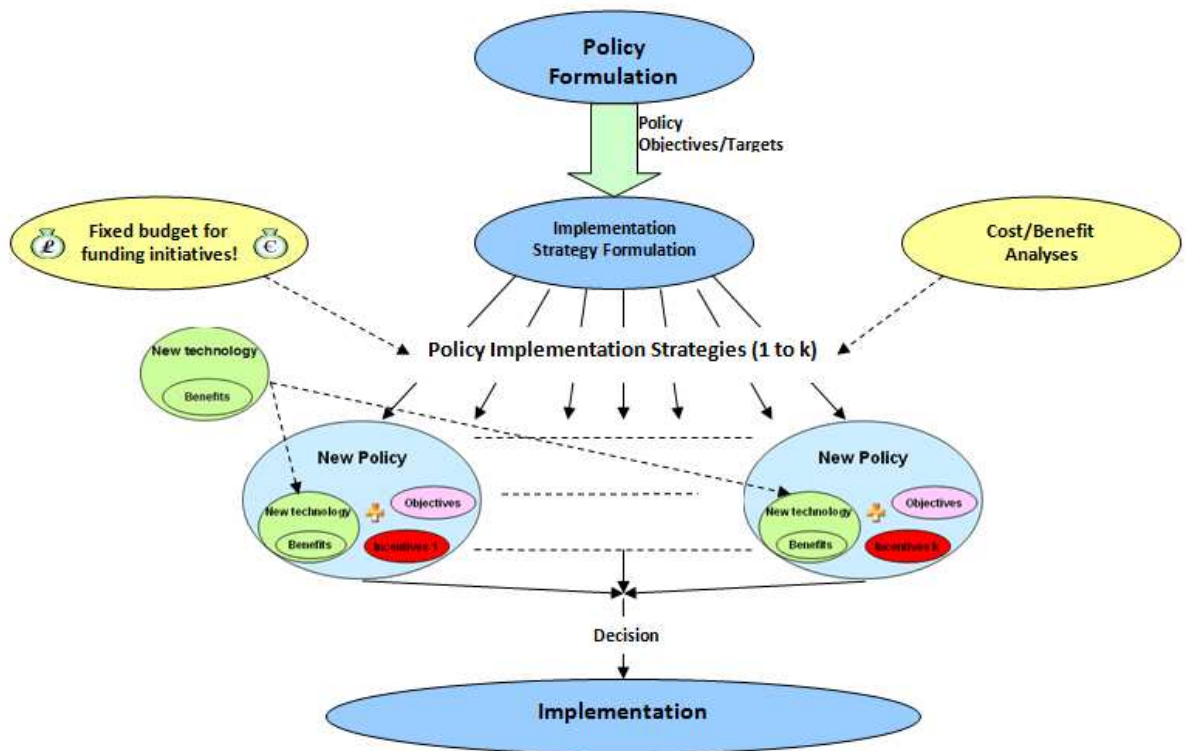
Figure 1.2: Graphical Illustration of the PIS Decision-Making Process (adapted from Savio and Nikolopoulos, 2009a, p. 9)

1.2 Scope and Aims of this Thesis

If then, PIS effectiveness is defined as 'the extent by which the strategy moves the status-quo towards the desired target set by the policy' (Savio and Nikolopoulos, 2009b), it is easy to imagine that a priori information about the effectiveness of a PIS

could be an important criterion for choosing a strategy for implementation as described in figure 1.2. What's more, the recent Performance Measurement[1] (PM) and evidence-based policy-making movements in the public sector have placed an accentuated emphasis on decision-making based on rational criteria such as efficiency and effectiveness[2] (Savio and Nikolopoulos, 2009b). Hence any decision made during the formulation of a policy, whether it be about a general policy objective, or what instrument to use or the exact specifications of the strategy designed for implementation should be supported by evidence of this kind.

Thus, after a qualitative exploration into the forecasting practices currently undertaken for such a purpose, this thesis will propose that PIS effectiveness predictions be made for the rivalling strategies and that these forecasts then be used as a decision-support tool for choosing a PIS for implementation. That is, these forecasts could potentially be used to shortlist potentially effective PIS and the selected few could then be further analysed by the more resource-intensive ex-ante evaluative approaches already in place for such a purpose, such as Impact Assessment (IA) or Cost-Benefit Analysis (CBA). Based on this further analysis, the decision as to which PIS to implement can be made. This would ensure that an extensive analysis of all possible PIS is made without the considerable investment of conducting this analysis with more costly approaches like IA or CBA. This idea is illustrated in figure 1.3.

---

[1] Performance Measurement (PM) places accentuated importance on evaluative techniques to measure performance in public and private sector organisations. PM puts an emphasis on meeting objectives and targets in an effective and efficient manner.

[2] Despite this being an optimistic, but unrealistic view on how decisions are actually made, as will be seen in chapter 2.
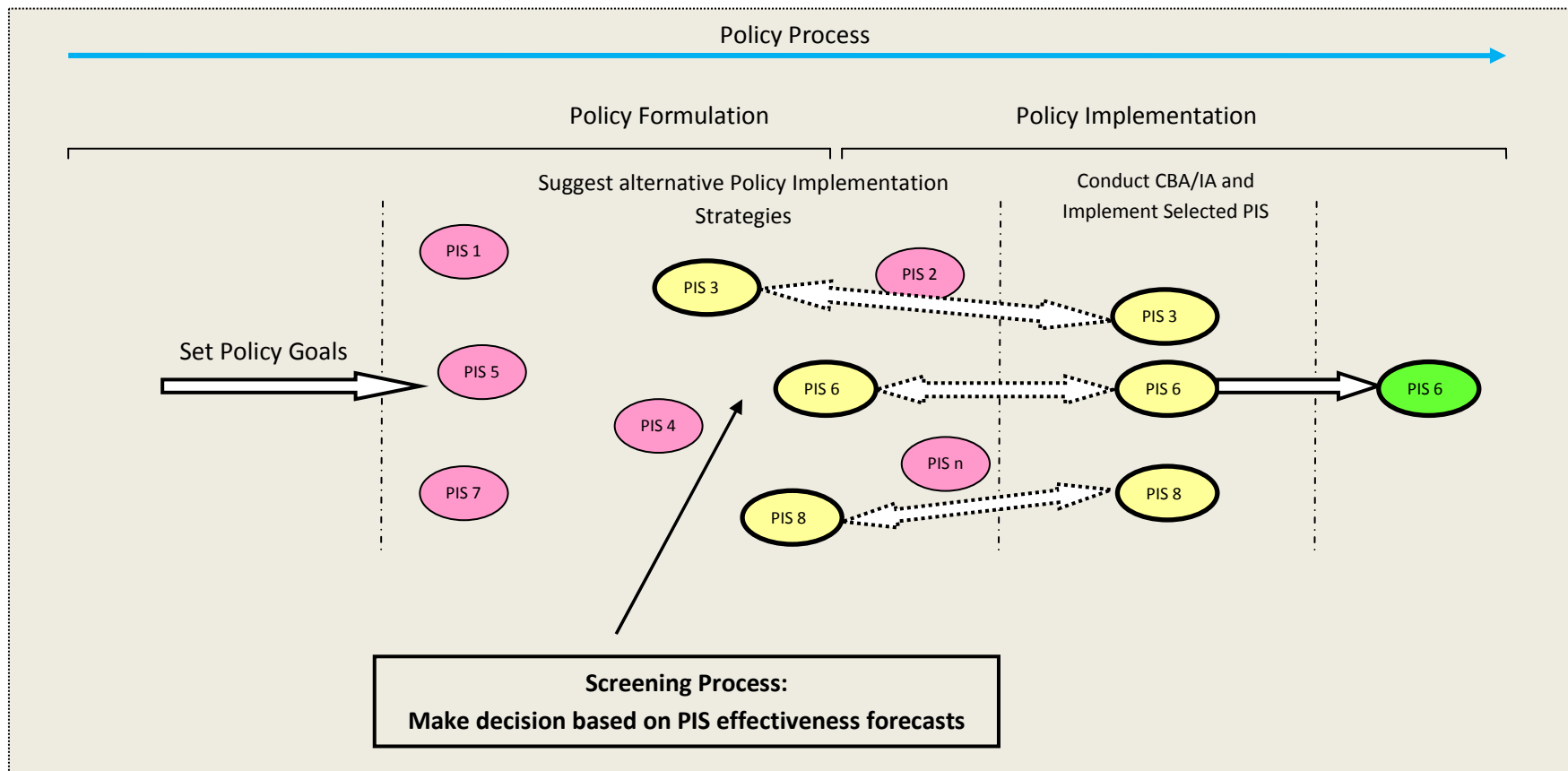
Figure 1.3: Choosing a PIS for Implementation

Furthermore, it will be proposed that a judgmental approach be used for producing such PIS effectiveness predictions. The reason for this is that they are quick and inexpensive to implement (Makridakis et al., 1998). More specifically, it will be proposed that this judgmental approach should be based on the use of structured analogies. It will be seen that forecasting by analogies has seen success for variables of similar characteristics to PIS effectiveness and that by structuring this use, several of the biases associated with their use can be minimised.

Hence, this thesis aims to contribute to research in the following three ways:

1. To propose that effectiveness forecasts can be made for a range of alternative PIS and that these forecasts can be used for short-listing the best strategies ahead of more resource intensive evaluations, such as Impact Assessment or Cost-Benefit Analysis. In this way such forecasts can provide support in the decision-making process when defining implementation strategy.

2. That such forecasts should be based on a judgmental approach due to their quick and inexpensive nature.

3. More specifically, these judgments should be based on a semi-structured analogies approach.

1.3 Preview of Subsequent Chapters

Figure 1.4 gives a graphical overview of the structure of this thesis. A more detailed outline of each chapter is given below.

Chapter 2 reviews the pertinent literature needed to position this research within what is currently known about the topic. It will be seen that this research is positioned at an intersection between the fields of forecasting and policy analysis. As a result, the review commences with a general overview of the field of forecasting and then moves on into an examination of policy analysis, policy formulation, implementation strategy and then policy evaluation. Focus will then be returned to forecasting by examining the current practices found to exist in public policy. Then, the role of expert judgment in public policy is assessed and a more in-depth view of judgmental forecasting

approaches and structured analogies is given. The literature review ends with an exploration into working with experts, including how to identify them, how to elicit such data and the difficulties that arise. As a result, little will be found about the forecasting practices to support decision-making at an implementation strategy level and that structured analogies will be found to be a particularly suitable approach for forecasting in situations of low availability of quantitative data, high uncertainty and where the use of analogous information comes naturally.

Chapter 3 introduces the notion of a Policy Implementation Strategy and discusses its importance. After stressing the importance of the notion of PIS effectiveness, the formal research questions put forward in this thesis are presented. It turns out that decisions are made regarding implementation strategy without the use of formal effectiveness predictions. In light of this, such forecasts are proposed as a support tool for such a decision and a structured analogies approach is proposed for such a purpose.

Chapter 4 initially discusses the philosophical position that the investigation assumes, including the method of scientific reasoning used for investigating the research questions. Then, the research process adopted for investigating the research questions is presented and discussed. Details of how the theory was constructed, the strategy and method of the investigation and an overview of the empirical procedure undertaken are given.

Chapter 5 is concerned with presenting in detail the series of steps taken to explore the research questions set out in chapter 3. Detailed information is given on the case studies used, the preparation of the cases studies and design of the five experiments.

Chapters 6 and 7 present and discuss the results of the experiments respectively. The evidence found goes a considerable way in supporting the propositions made.

Finally, chapter 8 outlines the conclusions of the thesis, along with its limitations and suggestions for where the research should go from here.

```
                    ┌─────────────────────────────┐
                    │       1. Introduction       │
                    └──────────────┬──────────────┘
                                   │
                                   ▼
                    ┌─────────────────────────────┐
                    │  2. Review of judgmental     │
                    │  forecasting approaches and  │
                    │  relevant policy literature  │
                    └──────────────┬──────────────┘
                                   │
                                   ▼
              ┌──────────────────────────────────────────┐
              │  3. Presentation and discussion of        │
              │  importance PIS effectiveness forecasts   │
              │  for governmental decision-making and     │
              │  presentation of research questions        │
              └──────────────────┬───────────────────────┘
```

4. Development of philosophical positioning and presentation of research process for investigating research questions

5. Details on how the cases were prepared and the experiments designed for investigating the research questions

6. Presentation of the results of the experiments

7. Discussion of the results

8. Conclusions and limitations of this thesis as well as suggestions for further research

Figure 1.4: The Structure of the Thesis

Chapter 2: Literature Review

2.1 Introduction

Forecasting has come a long way since its beginnings as an independent, multidisciplinary field in the early 1980s. In these 30 years, much of the research has centred on the quality of forecasts as well as the suitability of different approaches in different scenarios, resulting in a substantial body of knowledge for practitioners as well as a solid foundation for further research. An application area which has received notably less attention up until very recently has been that of public policy. Little research exists on the forecasting practices for the purpose of decision-making for governmental strategy and planning. More specifically, literature on predicting the effectiveness of incentive-based policy instruments designed as implementation strategies, known as Policy Implementation Strategies (PIS), has been virtually non-existent. The purpose of this literature review is to explore and analyse what has been written on such an explicit intersection of research bodies.

The second section of the review will give a broad introduction to the different methods for forecasting. Different approaches will be mentioned with more detail going into judgmental methods, along with some discussion on where the field has been and where it is going.

With the application field of the thesis being that of policy, this will be the subject of the third section. Here, a broad introduction into the policy analysis field will be given along with an examination into the cyclic model of the policy process and a view of the different schools of thought behind decision-making in policy.

The fourth section will look in more detail into policy formulation and implementation strategy. An analysis of the different literature attempting to classify instruments used

by policy makers for implementation as well as investigating the relevant literature on instrument selection and the implications this has on implementation strategy will be carried out.

The section after that will look at the all-important literature on policy evaluation, the field of systematically evaluating policies in the aim of improvement, including the different tools for this purpose as well as the *ex-ante* or *ex-post* debate. This section will also include an analysis of the literature documenting how such evaluations are actually carried out in government.

Section six will cover the role forecasting plays in public policy by examining the evidence of its presence and the outlook it is given by the field. The new found importance given to forecasting by the Performance Measurement (PM) regime as well as the role forecasting plays in policy implementation and formulation will be explored.

The section following that will provide more detail in the idea of expert judgment and explore how expertise is used in policy.

This will lead on to section eight, which will look at how expert judgment can be used in forecasting. Particular attention will be given to the forecasting by analogies approach as this will ultimately provide the rationale for the use of a structured analogies method for predicting policy instrument effectiveness.

Section nine presents the difficulties of working with experts. Firstly, it provides a discussion on the importance of being able to identify an expert and then of being able to measure their level of expertise. This section also gives an overview of the elicitation process and difficulties that can arise during such an exercise.

Finally, section ten will attempt to synthesise all the key points covered in the review and thereby provide a solid base for the construction of the thesis.

2.2 Forecasting

The nature of forecasting methods allows them to be classed into two broad categories: quantitative and qualitative (which are also known as judgmental).

Quantitative methods are used when there is enough numerical data available to allow for such an analysis and often judgmental methods are best when little or no numerical data is available but enough qualitative information is on hand to generate forecasts (Makridakis et al., 1998; De Gooijer and Hyndman, 2006; Forecastingprinciples.com, 2009b). A third, less common area, is distinguished by Makridakis et al. (1998) which concerns situations in which little or no information (of any nature) is available, deeming them essentially 'unpredictable'.

2.2.1 Judgmental Approaches

With much forecasting research having been published even before the formalisation of the field in the 1980s, it is the judgmental area of forecasting methodology which has perhaps seen the greatest progression (Armstrong, 1986; Goodwin and Wright, 1993; Lawrence et al. 2006; Parackal et al., 2007). In reviewing the literature on judgmental approaches published between 1981 and 2006, Lawrence et al. (2006) found that the attitude towards these kinds of approaches had evolved. After initially opposing the use of these methods, researchers had begun to understand the advantages that they could offer, to the point where now, the benefits from using judgment in forecasting are deemed indispensable (Lawrence et al. 2006; Parackal et al., 2007). Research had shown that judgmental forecasting could offer many advantages over quantitative approaches but had nevertheless the risk of several biases[1], which is the current consensus. Since then, much research has been done in trying to make sense of and manage these strengths and weaknesses (Lawrence, M. et al. 2006; Parackal et al., 2007).

Makridakis and Gaba (1998) highlight the importance of judgment in forecasting. The authors advocate that the very reason why models that best fit past data are not the ones that best predict the future is why judgment is important. Such a correction or 'adjustment' made by human judgment is what will allow for an improved prediction, they argued. Nevertheless, the authors warn of the multiple biases involved with such a procedure. They argue that it is best to just accept this contradiction and use

---

[1] The main sources of these biases are cognitive, related to an individual's ability to process information. Other sources can include motivational, which affect data collection during the elicitation process. A more detailed discussion of these two sources of bias can be found in sub-sections 2.2.3 and 2.9.2.2 respectively.

judgment, but in such a way to identify and minimise the negative effects of these biases. The biases associated with the use of expert judgment in forecasting will be discussed further in sub-section 2.2.3 of this chapter.

As their name would suggest, judgmental forecasting methods are based on qualitative past data but also on subjective information that the forecaster might have from experience or training. The usual scenarios for the use of a judgmental approach would be ones where little numerical data was available (leaving no other choice), the level of uncertainty is high[1] (Makridakis and Gaba, 1998; Makridakis et al., 2009), and/or domain knowledge is deemed beneficial in improving forecast accuracy[2] (Goodwin and Wright, 1993; O'Connor and Lawrence, 1998).

The choice of which judgmental approach to use in what scenario will factor on aspects such as the nature and quantity of the qualitative data, the training and level of experience of the forecaster and the specific forecasting requirements for the task at hand (Makridakis et al., 1998; Lawrence et al., 2006; Parackal et al., 2007). Makridakis et al. (1998) claims that comparison between the different judgmental methods is made difficult by their qualitative nature and so determining which is the most suitable to a given scenario is a difficult task. The 'Selection Tree' on Forecastingprinciples.com (2009b) gives advice on how to choose an appropriate method based on the data available and the requirements of the forecasting situation.

In addition to its use in making and improving predictions, Bolger and Harvey (1998) argue that judgment plays another role in forecasting. That is, judgment plays a big part in every stage of the forecasting process; implementation, choice of method, application of method, comparison and combination of forecasts, adjustment of forecasts and evaluation. So the forecasting process is inherently subjective and the decisions made by the forecasters will affect the way in which forecasts are produced (Armstrong, 1985; Bolger and Harvey, 1998). Subjectivity often translates into bias even if the forecasts are produced in a quantitative way, and so research warns that forecasters must be aware of this danger and strive to minimise it (Armstrong, 1985;

---

[1] When, for example, needing to forecast over a large time horizon.
[2] When, for example, having to predict an un-systematic change in past data, to anticipate the impact of a special event or to incorporate contextual domain knowledge.

Bolger and Harvey, 1998; Makridakis et al., 1998; Harvey, 2001; Lawrence et al., 2006; Parackal et al., 2007).

Judgmental approaches have gained merit due to their unique ability to anticipate un-systematic changes in established patterns and causal relationships in past data (Goodwin and Wright, 1993; Makridakis et al., 1998; Lawrence et al. 2006). The key to the successful use of judgmental methods, according to Makridakis et al. (1998, chapter 10, p.496), is to understand the biases and limitations along with the advantages of using judgment in a given situation. Furthermore, judgmental approaches are often an attractive alternative because they are relatively inexpensive and quick to implement (Makridakis et al., 1998).

2.2.2 Quantitative Approaches

Of less relevance to this research but still worth covering is the other main category of forecasting methods available, quantitative. In similar fashion to Lawrence et al. (2006) for judgmental approaches, De Gooijer and Hyndman (2006) offer a review of the progress of time series forecasting from 1981 to 2006. These methods have unquestionably seen a lot of research dedicated to them over the past years with over one third of all papers in IJF and JoF focussing on time series research (De Gooijer and Hyndman, 2006). The review documents the progress seen for each individual model type: exponential smoothing, ARIMA[1], non-linear, ARCH/GARCH[2], etc as well as for various subjects of interest such as seasonality, accuracy measures, combining and so on. The conclusions are that although enormous progress has been made in some areas, others need further development.

Quantitative forecasting methods can be subdivided into two major classes based on their principal model; time series or explanatory. Quantitative methods are generally considered as having more history than qualitative ones, with time series and explanatory models being branches of statistics and econometrics, which are 'older sciences', respectively. As a consequence, there are quite a lot of books which cover time series analysis and econometrics (Hamilton, 1994; Chatfield, 2003; Mills and

---

[1] Autoregressive Integrated Moving Average (ARIMA)
[2] Autoregressive Conditional Heteroskedasticity (ARCH) / Generalised Autoregressive Conditional Heteroskedasticity (GARCH)

Markellos, 2008), but few which take a forecasting 'evidence-based' perspective (Fildes and Nikolopoulos, 2006).

Explanatory models are based on the causal relationships between the dependant variable and the multiple independent variables that are assumed to exist. For this reason, these kinds of models need a certain level of information about the forecasting situation, the domain. Time series models however, do not need such information as they take more of a 'black box' approach to the system, thereby discarding the environmental information that could affect the behaviour of the variable (Makridakis et al., 1998).

The appeal for the use of time series models is two-fold. Firstly, such models are objective, inexpensive and replicable, making them perfect for short-term forecasts. Secondly, they are suitable when the objective is to make a prediction about what will happen, not why it happens, or similarly when a system is difficult to understand properly[1] (Makridakis et al., 1998; De Gooijer and Hyndman, 2006)

Mechanically, all time series methods are similar; they are all designed to study the systematic pattern of the past data, model it and use this model to project into the future. Every model has its own way of 'studying' the past data and has its own characteristics making each suitable to different kinds of data type.

When it comes to performance of explanatory models, two studies stand out; Armstrong (1985) and Fildes (1985), with slightly differing conclusions. Armstrong (1985) argues that explanatory models will forecast well when three conditions are met; (1) that the causal relationship between variables can be estimated accurately, (2) that the causal variables change significantly over time and (3) that this change can be predicted accurately. The results showed that explanatory models were not useful in all situations, rather extrapolative methods were more accurate in the short term (under small environmental changes) and vice versa in the long term (under large environmental changes). Fildes (1985) on the other hand, concluded that, no matter what the horizon, causal methods performed better than extrapolative ones. When these two results were later reconsidered in Allen and Fildes (2001), it was concluded

---

[1] Or that it is understood but the relationships between the causal variables are not.

that econometric methods are more accurate than extrapolative about as often for the short term as for the long term. Furthermore, although Allen and Fildes (2001) agree with Armstrong (1985)'s second condition, the first and third are more questionable according to the review. It suggests that problems in performance are due to a poor model (referring to the first condition) or poorly forecasted causal variables (referring to the third condition) or both. The paper concludes by arguing that an econometric model will perform well if it is 'well specified': that is, that if it was built in the right way using the principles of model building.

De Gooijer and Hyndman (2006) write that since the 1960s, there was a rise in new, more complex extrapolation models (including the Box and Jenkins (1976) model) and significant literature was published prophesising the improved accuracy of these new models, a claim which was challenged by various papers. One of the first of these papers was Makridakis et al. (1979), where the authors analysed 111 time series to examine the performance of a multitude of forecasting methods. The results not only showed why some models perform better than others for different types of data, but controversially concluded that simpler methods performed well against the more sophisticated, 'cutting-edge' ARIMA models of the time. Despite stiff opposition from the field (Fildes and Nikolopoulos, 2006), notably statisticians, the authors insisted upon their results to the point of conducting another study, this time comparing 24 methods, implemented by seven experts, forecasting up to 1001 series for six to eighteen different horizons. The results of this study, known as the M-competition, were presented in Makridakis et al. (1982) seminal paper. The results of the competition supported the findings of the previous study and concluded that statistically sophisticated methods did not perform better than relatively simpler methods.

In reaction to these results, Armstrong (1984) conducted a research review to find all empirical evidence comparing the performance of simple against sophisticated methods. Of the 39 papers found, 21 showed a negligible difference in performance, 5 favoured sophisticated models and 7 favoured simple models (6 cases had to be removed for various reasons (see p.2 Armstrong, 1984)).

Despite persistent objection from the field, these two studies provided undeniable evidence that 'simpler is not necessarily worse'. Specifically, the M-competition is seen as having marked the change in the paradigm of model performance which meant that models best fitting past data could no longer be presumed to make the best predictions. Since then several similar studies have been conducted, notably the M2-Competition (Makridakis et al., 1993), and the M3-Competition (Makridakis and Hibon, 2000), and they have largely found to support these results. The latest study in this line of research, the M4-Competition (Makridakis et al., 2010), is scheduled to start in 2010.

So, having reviewed the key literature on quantitative and qualitative forecast performance, it has been seen that both categories of approaches have their strengths and weaknesses. Judgmental approaches have the unique ability to predict un-systematic changes in past data as well as allowing for the incorporation of expertise into the forecast but they nevertheless suffer the danger of producing biased results due to the subjectivity and limitations of the forecaster's judgment. Quantitative approaches have the capacity to accurately model the behaviour of past data and extrapolate this into the future, which can result in highly accurate forecasts, particularly in the short term. However, their limitation comes from their rigidness in holding the continuity assumption to be true. If a change in pattern occurs, this kind of model will fail to detect it, resulting in possibly quite large prediction errors.

An obvious solution to these individual shortcomings has been to combine approaches to maximise their strengths and attempt to minimise their weaknesses. The idea of combining methods was first presented in Reid (1968) and Bates and Granger (1969) and since then much research has been published in the area (Armstrong, 2001b) with combinations between quantitative and judgmental approaches, known as 'judgmental adjustment' seeing particular success in improving accuracy (Goodwin, 2000; Fildes et al., 2009; Syntetos et al., 2009).

2.2.3 Bias

Bias is an important issue in forecasting as it can be the source of considerable prediction error if not handled appropriately. Unfortunately, sources of bias are

sometimes inevitable and rather than trying to eliminate them altogether (which is arguably an impossible task), researchers advise rather, to concentrate efforts on identifying the potential sources of bias and attempting to minimise their negative effects on forecast accuracy.

Bias can generally be defined as a skewing from some reference point, the un-biased, which is defined as the 'reality or truth' (Meyer and Booker, 2001). Hence, bias is always considered with a negative connotation[1].

Bias in forecasting is predominantly associated with the use of expert judgment in judgmental approaches rather than quantitative approaches. The reason for this is that quantitative methods are inherently objective because they all use the same past (numerical) data to develop the models and then to make predictions[2].

Judgmental approaches however, rely on the use of expert judgment, which is inherently subjective in nature. Meyer and Booker (2001) reason that expert judgment is the result of complex thought-processing known as knowledge-based cognition.

One of the most crucial stages in the use of expert judgment is the data extraction, or elicitation as it is known, from the experts (Meyer and Booker, 2001; Ayyub, 2001). The authors argue that this is a very delicate procedure; poor elicitation techniques can lead to biased results, which often translate into losses in accuracy. Research has shown that if careful consideration is not taken in the elicitation process, the experts will display the typical biases common in all humans. The two main categorisations of bias present in the way information is extracted from experts are *cognitive* and *motivational* (Meyer and Booker, 2001). The former will be dealt with now and the latter will be covered in sub-section 2.9.2.2 when elicitation is discussed.

Cognition refers to the mental exercise that takes place when an individual processes information. Hence knowledge-based cognition is a complex interpretive or analytical

---

[1] Such a result is unfortunate seeing as the reference point is always some theoretical and unachievable ideal. Thus, it is better to try and understand these sources of bias rather than to waste time trying to completely eliminate it.

[2] Although as explained in sub-section 2.2.1, judgment is argued to play a part in every stage of the forecasting process so essentially the whole process can be considered subjective, even if the mechanics of quantitative methods are in themselves objective.

activity that an individual performs when faced with a new or uncertain decision-making situation (Meyer and Booker, 2001).

In an investigation aimed at understanding the errors originating from survey respondents and why these errors occur, Mathiowetz (1987) identified four main cognitive processes that respondents perform during elicitation. It is clear that due to the subjective nature of such cognitive processes, each of these four steps can result in a source of bias.

- *Comprehension of the question*. Here it is easy to see that interpretation of a question is highly subjective. A series of characteristic traits in the respondent such as background, culture, education, etc will affect the way in which the problem is approached.

- *Retrieval of relevant information*. Meyer and Booker (2001) write that humans do not perceive and store all the information that is ever available to them (this would be a seemingly impossible task!) but rather tend to selectively notice data that support the opinion that they already maintain. In other words, humans do not objectively analyse new data but rather form their opinion on how the new data fits in with what they already know. Armstrong (1985) showed that humans generally give more attention to data that support one's hypothesis and tend to ignore information that causes conflict. In addition, of the information that is stored, the cognitive process of recollection is far from perfect as will be discussed later on.

- *Integrate information and make judgments*. This involves processing the information recalled in the previous stage. Research has shown that humans naturally use shortcuts for doing so called heuristics, which in the process of simplifying the process, also skew the final answer (Tversky and Kahneman, 1974; Hogarth, 1980). Hence, this step is similar to the previous one in that it is limited by the cognitive capabilities of individuals.

- *Formulating and reporting a response*. This step might too be governed by set decision rules distinctive to each respondent so may also generate bias.

The ability to carry out these four steps is limited by how much information an individual can remember and process (Hogarth, 1980). These limitations are the main source of bias present when using expert judgment, known as *cognitive*. This kind of bias leads experts to subconscious tendencies which have a detrimental effect on the data such as anchoring, inconsistency, availability and underestimation of uncertainty (Meyer and Booker, 2001). Through consideration of two proposed models of memory and information processing, one can begin to understand more fully the origins of cognitive bias.

The first, the traditional fixed image model, is based on the idea that humans store memory according to capacity as well as capability and use a central processor to move between each (Ericsson and Simon, 1980). It works in a similar way to that of a modern day computer. Short term memory (STM) contains information of limited capacity and intermediate duration (such as the cache memory in a computer) whereas long term memory (LTM) contains information of large capacity and fairly permanent duration (the hard-drive). The LTM contains information about past experiences and insight gained. When solving a problem, the central processor will take information from the LTM and move it to the STM for further processing. In this way the central processor will move continuously between the LTM and the STM until a solution has been found.

The second, known as the re-categorisation model, proposes that recollection should be viewed as a re-categorisation of groups of brain cells, called neurons, with temporarily strengthened connections. This approach, presented by Rosenfield (1988), posits that the way individuals perceive stimuli depends on how these have been previously organised or categorised. In other words, the way humans approach new situations is by re-conceptualising them in accordance to existing beliefs, needs or desires.

Meyer and Booker (2001) believe that both models are useful and should be taken into consideration during the elicitation process. For example, the fixed image model is helpful in timing when best to ask a respondent about his or her problem solving techniques, which according to the theory would be when it was still in the STM.

Similarly, the re-categorisation can help explain and comprehend the evolution in experts' thinking processes over time as well as the discrepancies between expert and non-expert thinking processes.

The other source of bias, *motivational*, stems from the experts' approach and attitude to the elicitation procedure and whose dynamics can also be detrimental to forecast accuracy if not handled appropriately. This category of bias will be discussed in more detail in 2.9.2.2.

2.3 Policy Analysis

As this thesis is essentially concerned with evaluating the suitability of a new forecasting approach in a policy context, a review of the background theory on policy analysis was deemed necessary. First, the concept of policy analysis will be defined. Secondly, the different stages of the policy cycle will be explored. Lastly, the different theories of decision-making that are said to transcend the stages will be reviewed.

2.3.1 Definition

Before attempting to define 'policy analysis', the definition of 'public policy' must be understood. An early and particularly notable definition, for its brevity and simplicity was given in Dye (1976, p. 1) as 'what governments do, why they do it and what difference it makes'.

Several more complex definitions for public policy exist in the vast number of existing published texts and journals about the subject (e.g. Parsons, 1995; Richards and Smith, 2002; Howlett and Ramesh, 2003; Dorey, 2005; Hill, 2005; Moran et al., 2008). Most of these definitions are rather similar so the one adopted here, which is fairly typical, will be 'the measures taken by governmental bodies to maintain the status quo or to alter it when faced with a particular problem or set of problems' (Howlett and Ramesh, 2003). Hence, it can be considered as the action (or inaction) a government takes to address a single or multitude of problematic issues.

'Policy analysis' is simply the analysis of the policy process and can be either descriptive or prescriptive (Howlett and Ramesh, 2003). Such a study can be

considered as the approach to public policy that attempts to 'integrate and contextualise models and research from those disciplines which have a problem and policy orientation' (Parsons, 1995, p.XV). In other words, the term 'policy analysis' brings together all the models designed and research conducted with the intention of solving a problem in a policy context.

Most texts (Parsons, 1995; Howlett and Ramesh, 2003; Hill, 2005) on public policy or policy analysis itself, define two broad categories of policy analysis, which are

- Analysis *of* policy (descriptive) – improving the understanding of policy (i.e. how problems are defined, agendas set, policy formulated, decisions made and policy evaluated and implemented.

- Analysis *for* policy (prescriptive) – improving the quality of policy (i.e. the use of analytical techniques, research and advocacy in problem definition, decision-making, evaluation and implementation.

Hill (2005) describes the relationship between both categories by describing analysis of policy to be concerned with the *ends* and analysis for policy with the *means*. While most books attempt to keep these two approaches separate, Parsons (1995) does not and attempts to formalise the kinds of policy analysis as comprising a variety of activities on a continuum of 'knowledge in the policy process' (Parsons, 1995, p.55).

Analysis of Policy			Analysis for Policy	
1 Analysis of policy determination	2 Analysis of policy content	3 Policy monitoring and evaluation	4 Information for policy	5 Policy advocacy

Figure 2.1: Varieties of Policy Analysis (Parsons, 1995, p. 55)

The research presented here would comprise sections in stages 3, 4 and 5 of this spectrum. By studying current methods of policy evaluation (stage 3), the intention is to provide an alternative decision-making tool for instrument selection (stage 4), which

will ultimately suggest a new approach for choosing the strategy to be adopted for policy implementation (stage 5).

2.3.2 The Policy Cycle

One popular way of describing the policy process is to consider it as a sequence of stages, which can be considered to be cyclical in nature (Parsons, 1995; Howlett and Ramesh, 2003; Hill, 2005; Moran et al., 2008). Authors generally agree upon five main stages, which can in turn be compared to the stages of applied problem solving, as can be seen in table 2.1 below.

Stages of the policy cycle	Stages in applied problem solving
Agenda – setting	Recognition of the problem
Policy formulation	Proposal of the solution
Decision-making	Choice of the solution
Policy Implementation	Putting solution into effect
Policy Evaluation	Monitoring results

Table 2.1: Five Stages of the Policy Cycle and Their Relationship to Applied Problem-solving (Howlett and Ramesh, 2003, p.13)

A logical starting point in the cycle is *agenda-setting*, where the problems, to which a solution will eventually be required, first become recognised. *Policy formulation* is the stage at which the problem is considered and several possible solutions are studied and proposed[1]. Once several solutions to the problem have been proposed by the policy developers, the *decision-making* stage will determine which of these solutions is to be carried forward and implemented. The next stage is concerned with the execution of the proposed solution, or in other words, the *implementation* of the policy[2]. Once the implementation strategy has been defined, the policy is put into practice. Finally, the last stage in the policy process is *evaluation*, which is a monitoring or quality control process in which the results and impacts of the policy are

---

[1] This stage in the policy cycle will be given further emphasis in section 2.4.
[2] More emphasis on the policy implementation process will be given in sub-section 2.4.2.

evaluated[1].   Any information and insight gained in this last stage can be used for improving the design and implementation of future policies, thereby completing the cycle.

Although such a cyclical model is popular in that such a heuristic appeals to many human and natural processes[2], John (1998) argues policy analysis is far more complex than what such a sequential model would suggest.

Policy is by definition complex and changeable...The linear model is more relevant in elucidating the presentation of policy than in detecting the reality of bargaining.

(John, 1998, p. 27)

In other words, the sequential model is useful on a conceptual level as a way of presenting the policy process but for serious analysis of the policy process, it must be acknowledged that such a process is far more complex in reality.

2.3.3 Theories of Decision-making

Decisions are made by government officials[3] throughout the whole policy process, beyond the selection process of choosing the best alternative, of those proposed in the formulation stage, for addressing the problem. Parsons (1995) writes

Decision-making...extends throughout the policy cycle; for example: decisions about what to make into a 'problem'; what information to choose; choices about strategies to influence the policy agenda; choices about what policy options to consider; choices about what option to select; choices about ends and means; choices in how a policy is implemented; choices about how policies may be evaluated. At each of these points decision-making is taking place.

Parsons (1995, p.245)

In other words, decision-making theories in policy look beyond the sequential nature of the model described in sub-section 2.3.2 and treat policy as a less rigid and continuous process. Several perspectives of decision-making have been seen through the evolution of policy analysis. Of the early theories, the most prominent was the rationalist approach.

---

[1] Policy evaluation will be the topic of section 2.5.
[2] Such as representing the policy process as an 'input/output' mechanism, or allowing for links to be made between the different governing institutions involved, etc.
[3] And increasingly by other actors, e.g. business and NGOs, as policy agenda-setting and implementation are increasingly carried out in multi-actor policy networks.

Rationalism

Such a rational approach to decision-making has come from two sources: economic rationality as developed in economic theory and bureaucratic rationality as developed by organisational and industrial society theory (Parsons, 1995). This theory postulates that the alternative to be chosen should be the one that will lead to the most efficient means of achieving the policy goals. In other words, having analysed the different alternatives, policy decision-makers will strive to choose the one which maximises their utility. In this way, from the rationalist's view, policy is a 'logical, reasoned and neutral way organisations assess problems' (John, 1998, p. 33).

The basis of this approach is attributed to German sociologist and political economist, Max Weber and is presented in a book of his major work (Weber, 1946). This was analysed, developed and critiqued by the work of Herbert Simon, acclaimed Nobel Prize winner and polymath. Simon (1957) identified three major difficulties in the rational theory:

- It is difficult to define a rational process of decision-making if it is not clear whose objectives are being considered, the decision-maker's or the organisation's, as these may be different.

- Referring to organisational goals may not make sense as these are often prone to dispute and modification.

- Decision-making is bounded by the multitude of human and contextual limitations.

That is, due to the complexity of political or bureaucratic roles, it might not be possible for organisations to reach the most rational decision, meaning that policy decision-making as a whole cannot be considered rational (John, 1998).

Simon was not against the idea of rationality in decision-making, rather, he believed that rationality existed in the 'process' of decision-making and that individuals will make elections which will 'satisfice' rather than 'maximise' value (Simon, 1957). In other words, because it is impossible for an actor to collect all the information needed

for a rational selection due to the limitations presented above, Simon (1957) argues that decision-makers will be forced to come to a satisfactory outcome given the constraints imposed on them. This was Simon (1957)'s idea of 'bounded rationality'.

Incrementalism

This rationalist model met opposition through the work of Charles Lindblom (Lindblom, 1959), who was unconvinced by the idea that rational analytical techniques could somehow replace the need for political agreement and consensus when making a decision. Rather Lindblom (1959) believed that the idea of 'muddling through', the notion of making many small, step-by-step changes (decisions) over time to create an eventual larger, broader change, was indeed 'scientific' and was a better representation of how decisions in policy were actually made. The power of such an approach comes in that it challenges the idea of policy being a multi-stage process as was previously thought (John, 1998) but rather a gradual, continuous process in which policy decision-making is a collective rather than an individual exercise. Lindblom (1959) argued that decision-makers rarely attempt to act rationally and that most decisions are made based on very little information.

Another advantage according to Lindblom (1959) is that such a cautious, incremental approach would allow the evasion of serious mistakes and allow for a flexible, dynamic approach to decision-making (Hill, 2005). What's more, this approach, which would be termed 'incrementalist', argued that the decision-making process should be governed by what is politically feasible rather than what is desirable and what is possible rather than maximal (Parsons, 1995; Howlett and Ramesh, 2003). Lindblom believed incrementalism to be a good depiction of the policy making process and a good model for how the decision-making process is and should be carried out.

New Models

More recently, several newer models have appeared including mixed scanning (Etzioni, 1967), which attempts to bridge the shortcomings of the rationalist and incrementalist approaches, and the garbage-can (March and Olsen, 1976).

On the mixed scanning model, Etzioni (1967) writes:

A rationalistic approach to decision-making requires greater resources than decision-makers command. The incremental strategy which takes into account the limited capacity of actors, fosters decisions which neglect basic societal innovations. Mixed scanning reduces the unrealistic aspects of rationalism by limiting the details required in fundamental decisions and helps to overcome the conservative slant of incrementalism by exploring long-run alternatives (Incremental decisions tend to imply fundamental ones, anyway). The mixed-scanning model makes this dualism explicit by combining (a) high-order, fundamental policy-making processes which set basic directions and (b) incremental ones which prepare for fundamental decisions and work them out after they have been reached.

(Etzioni, 1967, p. 385)

With its origins in weather forecasting, the mixed scanning model would involve a close, incremental look at some areas of the problem, with a more general, fundamental overlook at the problem as a whole. This would render the mixed scanning model flexible, a key characteristic for a decision-making model according to Etzioni (1967).

In Parsons (1995)'s interpretation of the 'garbage can'[1] model, the author argues that

…there is essentially a condition in which some issues will have solutions attached to them, others will not, other solutions may be roaming around looking for an issue to which to attach themselves. Decision-makers may well dumb a problem or solution into whatever can they have to hand, or whatever can is empty enough to contain the problem/solution.

Parsons (1995, p. 302)

What the model tries to convey is that problems and solutions are untidy and messy in nature whose manner of identification is dependent on when it was picked up and the availability of a can to place them in. As such, this approach views decision-making as an irrational activity. That is, the garbage can model proposes that decisions are made in an 'act now/think later' process.

Most texts on policy analysis tend to present the above models, and in particular the rationalist and incrementalist views, as being mutually exclusive. John (1998), however argues that this should not be the case as decision-making can be rational in some aspects and incrementalist in others.

Hence, the view taken in this thesis is that the policy process can indeed be viewed as being a multi-staged, cyclical process. However, these stages serve primarily as

---

[1]This model gets its name from the use of such an analogy in its description in March and Olsen (1976) whereby problems and solutions are dumped into a can, like garbage.

constructs for heuristic value, as in reality, the process is rather more complex. Moreover, it is believed that whilst decision-makers strive to be rational, political forces limit this and the result is a more incrementalist model in which decision-makers rational approaches are bounded by inherent information constraints which ultimately lead them to make more 'garbage can' like decisions.

2.4 Policy Formulation and Implementation Strategy

As was seen in section 2.3, although decision-making is believed to transcend all the stages of the policy process (Parsons, 1995), two fundamental stages where the decisions made will ultimately give shape to the policy and how it is implemented are its formulation and implementation. Such is the importance of the decision-making task during these stages that the model in table 2.1 attributes it its own stage. Although as an incrementalist would argue, these three stages are very much intertwined and can arguably considered as one. Thus, this section of the literature review is dedicated to what is known about these two crucial stages of the policy process.

This initial part of the section will give some general insight into the policy formulation stage and why it is important. Sub-section 2.4.2 will give more focus to the formulation of the implementation strategy of a policy and more specifically on the tools or instruments used by governments for such a purpose. The literature attempting to classify these instruments will also be reviewed, followed by the literature dealing with different views on instrument selection.

2.4.1 Policy Formulation

The policy formulation stage in the policy cycle is the crucial stage where governments develop the course of action to be taken to address a public issue which has been identified (Parsons, 1995; Hillman and Hitt, 1999; Howlett and Ramesh, 2003; Moran et al., 2008). As was seen in section 2.3, policy formulation is a complex, multi-layered, analytic decision-making process where various organisational and political forces are seen to act, resulting in defeating efforts to carry out such a process in a rational manner (Parsons, 1995; Howlett and Ramesh, 2003). During such a stage, input from

all the different stakeholders[1] will be taken and several alternative strategies will be considered (Dorey, 2005).

The policy formulation process will vary from deciding upon a solution to tackle a public issue to deciding on the strategy that will be used for the implementation of this solution (Parsons, 1995; Howlett and Ramesh, 2003; Hill, 2005; Dorey, 2005). As will be seen in sub-section 2.4.3, implementation strategies often take the form of policy instruments, used for defining the exact program specifications to be used in the delivery. For this reason, the stages of policy formulation, decision-making and implementation are considered to be particularly closely linked (Parsons, 1995; Howlett and Ramesh, 2003; Hill, 2005).

Table 2.2 below represents a model proposed by Howlett and Ramesh (2003), adapted from Hall (1993), which is evidence of the decisions that must be made during the formulation stage with respect to the different types of policy options. In such a model, it is clear how the different policy components can be expressed in terms of their level of generality and the policy element they affect.

		Level of Generality of Policy Content	
		Conceptual/Policy	Practical/Program
Affected Policy Element	**Ends**	Policy Goals	Program Specifications
	Means	Instrument Types	Instrument Components

Table 2.2: Types of Policy Options by Level of Generality and Policy Component Affected (Howlett and Ramesh, 2003, p.147)

In other words, the selections made in the formulation of a policy will centre on decisions made on a conceptual, policy level in addition to decisions made on a more practical, program level. No matter what level of generality is being considered, the decisions made during the formulation stage will necessarily shape the strategy adopted in its implementation.

---

[1] Regular meetings are held between policy officials and all interested parties concerned with or affected by the development and implementation of the new policy.

- Decisions on a policy level will focus on aspects such as what the policy goals or objectives are (the *ends*) as well as which policy instrument type is to be used for attaining these objectives (the *means*).

- Decisions on a program level will be more detailed and focus on the program specifications as well as any targets to be hit (the *ends*) and the instrument components which define the parameters of the implementation program (the *means*).

Hence, together, these kinds of decisions will ultimately define how a policy will be delivered, a very important factor in determining its success. A potentially successful policy strategy can become not so if a bad decision is made with regards to the implementation strategy and the same is true for the vice-versa (Parsons, 1995; Howlett and Ramesh, 2003). In this way, deciding upon the implementation strategy becomes all the more important.

2.4.2 Implementation Strategy

Policy instruments, also referred to as policy tools or governing instruments, will help shape the strategy used by governments for the purpose of implementation (Howlett and Ramesh, 2003). They are the actual means governments have for attaining the goals or objectives set out by the policy. The successful delivery of these instruments will ultimately determine whether or not an implementation strategy has been effective in attaining the goals set out by the policy. In other words,

Instrument choice...*is* public policy-making, and the role of the policy analyst is one of assisting 'in constructing an inventory of potential public capabilities and resources that might be pertinent in any problem-solving situation' (Anderson, 1971).

(Howlett and Ramesh, 2003, p.89)

Thus, such implementation strategies form an integral part of the decision-making process within the policy formulation stage in that governments have to decide which of them will offer the best way to approach the problem within any constraints imposed upon them[1] (Parsons, 1995; Howlett and Ramesh, 2003). With a wide range of such tools available for this process, at differing costs and suitability, such a decision

---

[1] Such as could be political, financial or time constraints.

is no easy task and has been the subject of considerable study (Peters and Van Nispen, 1998; Bemelmans-Videc et al., 1998).

One of the major figures involved with the research of government instruments is Lester Salamon who argued in Salamon (1981) that one of the biggest limitations in early implementation research was that focus was being given to the wrong unit of analysis when assessment was being considered. Rather than focusing on individual programs, or groupings of them according to objectives, he argued that focus should be given to the tools of government action, on the *means* of intervention. More detail on the evolution of implementation research will be given in section 2.4.4.

In his early research, Salamon (1981) also asked two important research questions which would make a considerable impact in implementation research; (1) how does instrument selection affect the effectiveness and operation of a governmental policy and (2) what factors influence instrument selection.

These questions triggered a wave of classification research (Rondinelli, 1983; Lowi, 1985; McDonnell and Elmore, 1987; Salamon and Lund, 1989; Hahn and Stavins, 1992; Acutt and Dodgson, 1996; Bemelmans-Videc et al., 1998; Perrels, 2001; Howlett and Ramesh, 2003) which led to an attempt to group similar instruments into categories, which could then be analysed to see what determined their effectiveness (Balch, 1980; Bardach, 1980, Rondinelli, 1983; Perrels, 2001).

Hence, the core of the research has focussed on two main points; attempting to build and classify the array of instruments available to policy makers, which will be the focus of the next section and theories of instrument selection, which will be the focus of sub-section 2.4.5. Before that, sub-section 2.4.4 will present the evolution of implementation research as this is considered important in the context of instrument selection theories.

2.4.3 Instrument Classification

The variety of instruments for addressing a given public problem is limited only by the policy maker's imagination.

(Howlett and Ramesh, 2003, p.88)

The work of Salamon (1981) sparked a new line in implementation research which led researchers to study these tools in more detail. The first natural step in this line of research was classifying these instruments, which could then be studied and from which conclusions about their effectiveness could be drawn. However, these ideas of policy instruments as strategies for implementation originates from the work of Laswell (1958) who argued that governments had developed a limited number of strategies which they used to manipulate policy outcomes. This notion led to the first attempts to classify such instruments as early as the 1950s and regularly since then (Dahl and Lindblom, 1953; Lasswell, 1958; Lowi, 1966, 1972, 1985; McDonnell and Elmore, 1987; Salamon and Lund, 1989; Hahn and Stavins, 1992; Acutt and Dodgson, 1996; Bemelmans-Videc et al., 1998; Perrels, 2001; Howlett and Ramesh, 2003). The rationale behind this kind of research and such a classification desire is fairly straightforward in that it attempts to ultimately establish what instruments work best in a given scenario (McDonnell and Elmore, 1987): information whose value to governmental decision-makers is indisputable. In a review of such past research, Howlett and Ramesh, (2003) found that generally such attempts were of limited success due to the complexity of the task and the sheer number of different combinations of tools and contexts available. Nonetheless, a few of these classifications are presented here.

- In one such attempt, McDonnell and Elmore (1987) identify four categories in which instruments can be classed:

mandates are directives aimed at the public and agencies and are intended to force a desired objective (e.g. regulation);

inducements involve the transfer of money to the public or agencies in exchange for certain actions (e.g. grants);

capacity-building involves the transfer of money to the public or agencies for the purpose of investment in material, intellectual or human resources (e.g. contractual research);

system changing is the transfer of official authority among individuals and agencies (e.g. vouchers).

- In more of an environmental policy focus, Hahn and Stavins (1992) argue that economists usually divide policy instruments for achieving environmental objectives into two categories; the classic 'command and control' providing little flexibility in the means by which an objective is to be reached and 'market-based or incentive-based mechanisms' providing greater flexibility. The authors find that conventional command and control approaches often fail to reach the objectives in the cheapest way possible but Stavins (1997) adds that these can be effective at achieving established environmental goals. Economists have shown a preference for incentive based instruments because of their greater level of efficiency and the encouraging effect they have on the adoption and diffusion of environmentally clean technologies (ECT, Eder and Leone, 1999; Stavins, 1997). As a result, economic incentives such as environmental taxes and marketable permits[1] are becoming increasingly popular in the policy arena (Hahn and Stavins, 1992).

- In a paper exploring this rise of policy instruments in an environmental context, termed new environmental policy instruments (NEPIs), Jordan et al. (2003) find an explosion in these kinds of tools not only in the EU but worldwide, a movement away from the traditional 'command and control' approach. This paper offers yet another typology which concentrates on a more narrowly defined concept of policy instruments consisting of four categories (as opposed to five categories suggested by other authors):

  (1) regulatory instruments,

  (2) market based instruments (MBIs),

  (3) voluntary agreements (VAs) and

  (4) informational devices.

---

[1] Marketable permits are a type of economic incentive scheme that allow the trading of any credits or allowances obtained through reduced pollution (in an environmental context).

The paper also classes these different instruments according to whether or not the regulator specifies; the goal to be achieved and how the goal is to be achieved.

- In another classification attempt, Perrels (2001) categorises instruments into four classes,

    Regulating instruments, as the command and control style instrument described above.

    Instruments that imply *deregulation*, as they attempt to change behaviour without legislation, through voluntary agreements for example.

    Fiscal and financial instruments, these include instruments that try to change behaviour through financial incentives such as taxes, scrappage schemes, etc.

    Supportive actions, refer to the instruments that attempt to improve knowledge levels and market transparency, such as research grants, etc.

- In the last of the classifications, Hood (1986) proposes to classify all policy tools by categorising them by the governing resources that they employ; (1) *nodality* or 'central policy actor', (2) *authority* or 'their legal powers', (3) *treasure* or 'money' and (4) *organisation* or 'the formal organisation available to them'.

Instruments from different classes can of course be mixed and matched to ensure their maximum effectiveness the author argues. This, in contrast to previous attempts, is more detailed and classes instruments to a higher level of specification according to their characteristics.

According to Howlett and Ramesh (2003), many of the classification schemes, including the ones described above, unfortunately are either

...pitched at a high level of abstraction, making them difficult to apply in practical circumstances, or dwell on the idiosyncrasies of particular tools, thereby limiting the range of the descriptions and explanations they provide.

(Howlett and Ramesh, 2003, p.88)

The authors argue that a scheme is required which is abstract enough to include all the possibilities but tangible enough to correspond with the way governments actually

interpret their selections. Thus, with a strong influence from Hood (1986)'s model presented above, Howlett and Ramesh (2003) propose their own basic taxonomy of instruments. Howlett and Ramesh (2003) borrow and enhance Hood (1986)'s model by adding illustrative examples for each category as can be seen in table 2.3.

Nodality	Authority	Treasure	Organisation
Information Monitoring and Release	Command and Control Regulation	Grants and Loans	Direct Provision of Goods and Services and Public Enterprises
Advice and Exhortation	Self-Regulation	User Charges	Use of Family, Community, and Voluntary Organisations
Advertising	Standard-Setting and Delegated Regulation	Taxes and Tax Expenditures	Market Creation
Commissions and Inquiries	Advisory Committees and Consultations	Interest Group Creation and Funding	Government Reorganisation

Table 2.3: Policy Instruments by Principal Governing Resource (Howlett and Ramesh, p.92)

Of particular interest to this thesis is the third category in the taxonomy in table 2.3, *treasure*, which concerns instruments based on financial resources, also known as subsidisation. The reason for this being that the schemes being used to test the approach in the experiments carried out, are based on the use of such financial resources.

The effectiveness of such instruments tends to rely more on these resources than on government personnel or government authority. These instruments entail the financial transfers between individuals, firms and organisations and the government or between themselves under government direction. Usually, these transfers are used as incentives or disincentives issued by the government to change public behaviour. Naturally, an incentive attempts to reward a desired activity whilst a disincentive attempts to penalise an undesired one (Howlett and Ramesh, 2003).

The three main kinds of subsidies are grants, tax incentives and loans. Grants are considered one of the most prominent and 'are usually offered … with the objective of making them provide more of a desired good or service than they would otherwise'

(Howlett and Ramesh, 2003, p.108). Another popular form of subsidy is the tax incentive in which the government agrees to forgive either all or part of a tax that would have otherwise been collected contingent on some act (or its omission). Loans from the government are also considered as subsidies with such being the difference between the interest charged and the market rate, not the entire amount (Lund, 1989). The authors argue that these kinds of financial instruments are useful to governments as they offer several advantages. Tax incentives for example are particularly interesting because they do not need budgetary approval (as no money is actually spent, the revenue from tax is merely forgiven (Maslove, 1994), they are not constrained by availability of funds (Howlett and Ramesh, 2003) and they are fairly easy to implement as they can be hidden in complex tax codes (McDaniel, 1989; Leeuw, 1998)).

Howlett and Ramesh (2003) offer five main advantages and disadvantages that subsidies as a policy instrument can offer, they are summarised in the table 2.4 below.

Advantages	Disadvantages
They can be used to encourage a behavioural change that the government would like to see and the public would like to make, but needs the means for doing so	Since they need financing (except tax incentives), subsidies are often too difficult to include in the formal budgetary process and they must compete with government programs needing funds, backed by their own respective political forces
They offer flexibility in implementation in that the individual participants choose themselves how to respond to the subsidy in light of the changing circumstances (this emphasises the importance of predicting the take up rate of the scheme in order to assess if the program is viable)	The cost of establishing the right size of the subsidy through investigation may be high. The usual trial and error method can be an expensive means of implementation that won't necessarily ensure effectiveness
They encourage innovation by offering far more flexibility in their use by the public than for example, directives	Subsidies work indirectly so there is often a time lag before the desired effects become noticeable
They offer potentially low final costs as these will depend on the take up from individuals	In cases where the desired behaviour would have occurred without governmental intervention, subsidies may be redundant
They are often easier to justify politically as they will be supported strongly by the beneficiaries and opposed weakly by their opponents	Some subsidies could be banned by international agreements because of the negative effect they might have on local industries and employment

Table 2.4: Advantages and Disadvantages that Subsidies can Offer as Policy Instruments.

Thus, as was explained at the beginning of the section, this instrument classification research was considered important so much as the first step towards understanding how and why these instruments are selected. Before this is done however, a brief synopsis of the evolution of implementation will be given as this is considered important in placing such selection theories in the appropriate context.

2.4.4 History of Implementation Research

Until the early 1970s, little attention was given to implementation as an activity, simply because it was not considered to be problematic in a policy sense (Hargrove, 1975; Sabatier, 1986). However, as research started to show that policy goals were not being met as a result of poor implementation, and not poor policy, this began to change (Derthick, 1972; Pressman and Wildavsky, 1973; Bardach, 1977; Sabatier, 1986; Hill, 2005). By the 1980s, Public Policy began to see an emerging line of research, with a focus on understanding the issues that dictated the success of policy implementation (Van Meter and Van Horn, 1975; Hood, 1976; Mayntz, 1979; Sabatier and Mazmanian, 1981; Hjern, 1982).

This second wave of implementation research saw the emergence of a disagreement in what should be the focus of such research. A 'top-down' versus 'bottom-up' debate divided the field and caused an inevitable stalemate in the progression of its research (Hill, 2005). On one side, it was argued that effective policy implementation should adopt a 'top-down' approach which reasoned that the policy process should be viewed as a chain of command in which the policy preferences of political leaders are passed down through the different levels of the governmental organisation (Van Meter and Van Horn, 1975; Hood, 1976; Dunsire, 1978a, 1978b, 1990; Sabatier and Mazmanian, 1981). In other words effective implementation would only be achieved if a good, hierarchical chain of command with the ability to co-ordinate and control was set-up (Pressman and Wildavsky, 1973). A good amount of research stemmed from Pressman and Wildavsky (1973) in the quest for finding a definition for 'perfect implementation'. One such book, Hood (1976), proposed five conditions for perfect implementation as

- ideal implementation is a product of a unitary 'army'-like organisation, with clear lines of authority;

- that norms would be enforced and objectives given;

- that people would do what they were told and asked;

- that there should be perfect communication in and between units of organisation;

- that there would be no pressure of time.

In contrast, their opposers believed that for implementation to be effective, a 'bottom-up' approach must be adopted, which involved studying the actions of those affected by and involved in policy implementation (Lipsky, 1971; Wetherley and Lipsky, 1977; Elmore, 1978, 1979). Parsons (1995) argued that if the 'street-level' implementers had prerogative in how policy is applied, how could they not have an input in formulation? Such implementers will have gained knowledge and experience from their profession, which could prove valuable to future policy. That is, such civil servants working at street-level to deliver the policy, will gain important insight into such a process, including how it can be improved, and such insight should be taken into account by officials at the 'top' of the process in a view for policy learning.

On the other hand, the level of such 'input' should be controlled, as a scenario in which street-level professionals such as policemen and teachers shape future policy could be criticised by he who claims such officials, although being knowledgeable in the task, lack training for its purpose (Parsons, 1995; Howlett and Ramesh, 2003). In other words, any insight provided by street-level officials should be considered and analysed as just that because any such feedback has ultimately been received from an individual untrained for such a task. Hence, although these implementers might hold the necessary tools for carrying out effective implementation, they are oblivious to the existing political and resource constraints acting on the policy, which only policy designers higher up the scale will be aware of. Furthermore, these policy designers will have training and experience in the necessary tools for implementation design such as strategy, resource-planning and political manoeuvring, which the policy implementers fall short of.

In light of this debate, research was being published offering analysis and prescriptions from both schools of thought, resulting in a negative effect on the development of knowledge of this stage of the policy cycle (Howlett and Ramesh, 2003). That is to say, researchers were concentrating their efforts on siding within the conflict rather than producing research which would enable the progression of the field in a positive direction. This led to a call in the late 1980s and early 1990s for a third generation of new implementation research (Sabatier, 1986; Parsons, 1995; Howlett and Ramesh, 2003; Hill, 2005).

This call was answered by researchers and the 1990s proved a prosperous decade for implementation research (Sabatier, 1986; Parsons, 1995; Howlett and Ramesh, 2003; Hill, 2005). This new research boom saw the emergence of a third approach with a different focus. Instead of studying the strictly administrative concerns of implementation, this new research focussed on the process of implementation as one in which the range of policy instruments applied to cases through a process of policy design (Salamon, 1981; Mayntz, 1983). Efforts were made to investigate the characteristics of policy instruments and the reasons for their selection by governmental decision-makers, with the aim of improving implementation (Hood, 1986; Linder and Peters, 1989). These studies proved successful and very positive in that they tended to focus on the reasons for the selection of a particular tool and their potential use in future scenarios. This laid the foundation for more recent research which has gone beyond the question of individual instrument choice and has entered into the description and application of instrument 'mixes' and 'implementation styles' (Lowi, 1985; Kagan and Axelrad, 1997; Knill, 1998). Research had now moved on from the 'top-down' vs. 'bottom-up' debate to theories of hybrid administrative systems allowing policy leaders to control street-level officials whilst granting these enough independence to carry out their functions effectively (McCubbins and Schwartz, 1984; Sabatier, 1986; McCubbins and McCubbins, 1994).

Moreover, these ideas would eventually give light to new views on public administration, including the ideas of New Public Management [1] (NPM) and

---

[1] New Public Management (NPM) is a philosophy to public sector management whose main hypothesis is that a greater market orientation will lead to an improved cost-efficiency ratio for governments.

Performance Measurement (PM) (Parsons, 1995; Howlett and Ramesh, 2003; Hill, 2005).

One very important consequence of this results-based movement was the conception of *policy evaluation.* The idea of policy evaluation, which is the final stage of the policy cycle, refers to the phase in which the success of the policy is judged in light of the problem it was meant to address. It is a very broad topic which can entail a multitude of different evaluation criteria and approaches and will be covered in section 2.5.

2.4.5 Instrument Selection

As a result of the surge in research to improve the administrative structure of governments and the advent of public administration doctrines such as NPM and PM, the desire to study the characteristics of policy instruments and their reasons for selection by governmental decision-makers was seen vital in improving the implementation process (Sabatier, 1986; Howlett and Ramesh, 2003).

In their paper on policy instruments, McDonnell and Elmore (1987), attempted to explore the conditions under which instruments are most likely to produce their intended effects. They offered one of the early theories on instrument choice. The authors believed that instrument choice is based on two factors; the way in which the policy problem is defined and the resources and constraints policy-makers face when confronted with a policy issue. In other words, instrument choice is governed by the way the policy maker frames the problem as well as the limiting effects of resource constraints such as time and money.

Stavins (1997) proposed a rather more direct approach to instrument choice in which the author says that the instrument to be chosen is the one that maximises efficiency. Admittedly, as this is often hard to evaluate (quantitatively), cost-effectiveness should also be used (Stavins, 1997). Regardless of which is to be used however, the cost must be measured correctly. This is not a very simple task however. In the author's view, costs are multi-faceted and entail not just the on-budget costs to the government and the administrative costs, as many think, other non-direct costs (such as environmental), have also got to be considered. Other criteria can also be considered in instrument

selection. Each nation, based on their socio-economic and cultural contexts, can consider many different and varying criteria (Stavins, 1997).

More recently, Jordan et al. (2003) propose two theories of instrument selection. The first, which it refers to as 'ideational', is based on the idea that instruments must be chosen to fit new policy requirements and policy paradigms. This theory would encourage the adoption of new, innovative instruments (provided they are convincing) as policy learning is considered important in driving instrument selection. The second theory, termed 'institutional', is limited by the effect of national institutional forms. That is, instrument selection is governed by the collection of ideologies and practices of the policy maker's own national government.

The three theories mentioned above suggest that instrument selection is based on: conceptual factors, such as consistency with national political forces as well as consistency with the policy paradigms in place, in addition to more technical factors, such as making a selection based on efficiency criteria, thereby considering the importance of any resource constraints.

Such a view is supported by the three different doctrines for instrument choice summarised by Howlett and Ramesh (2003). Although the authors claim that efforts are made to incorporate notions and ideas from the other schools of thought, they can be easily distinguished. They are,

- *Technical*: proposed by economists who believe that instrument is, at least in part, a technical exercise in which the attributes of specific tools must be matched to the requirements of the task at hand (Posner, 1974; Breyer, 1979, 1982; Mitnick, 1980; Utton, 1986; Wolf, 1987).

- *Political*: proposed by political scientists who propose that instruments are more or less interchangeable from a technical point of view and that instrument choice is governed primarily by political forces (Doern, 1981; Tupper and Doern, 1981; Phidd and Doern, 1983).

- *Subsystem*: which believes that instrument choice is governed by principles from both the technical and political doctrines. This theory proposes that some

instruments are more suited to certain tasks than others and so a technical analysis of the features of the instruments is important. However, it also proposes that a nation's policy style and culture will have a firm bearing on instrument selection (Linder and Peters, 1989).

The first two approaches raise concerns. In a perfect policy arena, one would like to think that if developers had the political freedom to choose the most suitable instrument for a task based on their attributes and a rational analysis, they would. However, policy arenas are arguably the most susceptible to political forces which do not always share the same policy objectives as the developers (Doern, 1981; Tupper and Doern, 1981; Phidd and Doern, 1983). Similarly, any sort of decision based purely on political reasons with no ex-ante evaluation into the optimal choice based on resource constraints, as the second theory suggests, would necessarily be subjective, leading to a biased outcome. Hence the realistic option would be the third position. The subsystem view suggests that rationality could be maintained through such a technical approach (using criteria such as efficiency or cost-effectiveness) whilst acknowledging that such an approach will necessarily be influenced by the political forces present in such a decision-making process.

In sum, rational techniques such as efficiency or cost-effectiveness criteria act as resources which decision-makers use as support in their bargaining games when selecting an instrument and hence shaping implementation strategy. But such a decision will be subject to the same political constraints present in any decision-making process in the policy process, as was seen in section 2.3.

Thus, this section has highlighted:

- The importance of the decisions made during the formulation stage regarding how a policy will be shaped as a solution when faced with a public problem. Such decisions not only include defining the policy objectives but also the strategy that will be adopted for implementation. For the latter, it was seen that policy instruments are considered as the principal means for such a purpose and hence are an area of interest.

- Instrument selection plays a vital role in ultimately defining implementation strategy so it comes as no surprise that policy research has shown interest in how these instruments can be classified but particularly how these should be selected.

- The rationality of instrument selection is constrained by political forces, as is any decision-making process. Rational techniques such as efficiency or cost-effectiveness are used by decision-makers as bargaining tools during such a process.

2.5 Policy Evaluation

In section 2.4, the emphasis was on a strategic decision-making level with an exploration into how different government tools are selected and how this ultimately shapes implementation strategy. Attention will now be turned to the practice of *evaluation* and its intent on assessing policy effectiveness for the purpose of improvement as this will prove to be important in later developments of the thesis.

Policy evaluation, or sometimes referred to as simply *evaluation*, refers to the investigation into the effectiveness of a policy or program (Suchman, 1967; Nachmias, 1979; Parsons, 1995; Howlett and Ramesh, 2003; Rossi et al., 2004). Analysts often place evaluation as the last stage of the policy cycle as such an investigation can be performed after a policy has been implemented (*ex-post*), as a policy learning exercise, or before implementation (*ex-ante*) as a tool for policy assessment. More formally, evaluation draws on concepts and techniques of social science disciplines and has as a primary goal to improve policy programs (Suchman, 1967; Nachmias, 1979; Parsons, 1995; Howlett and Ramesh, 2003; Rossi et al., 2004).

2.5.1 Origins and Exercises

The beginnings of evaluation research can be traced back to post-World War II when several public administration movements saw the rationalisation and professionalization of policy analysis (Rossi et al., 2004). By the 1970s and 1980s, several formal evaluation systems had surfaced in the United States (Reid, 1979; Krane, 2001). A steady evolution in this line of research has seen the importance of such a practice escalate considerably to the point where it is now more essential than ever to

evaluate effectiveness in light of the present-day concerns regarding the allotment of scarce resources (Howlett and Ramesh, 2003). The idea of evaluation currently represents a prominent role in the ideas of results based management in the Performance Measurement (PM) doctrine of public administration (Krane, 2001; Howlett and Ramesh, 2003; Rossi et al., 2004).

Evaluation is just as important as a policy learning exercise as it is for assessing the effectiveness of a policy. In other words, evaluation helps to assess the strengths as well as weaknesses of a policy and this knowledge can then be used for the future design and formulation of new prospective policies.

In terms of the different dimensions of a policy which can be exposed to such an evaluative exercise, although opinions from one source to another differ slightly (Parsons, 1995; Howlett and Ramesh, 2003; Rossi et al., 2004), the core policy axioms are summarised in Rossi et al. (2004). They are,

(1) *need for the policy*, this is concerned with evaluating if there is an actual need for a policy to be put in place to address an issue. Such an exercise necessarily involves reviewing the initial conditions and rationale which led to a policy being developed (Soriano, 1995; Reviere et al., 1996; Rossi et al., 2004).

(2) *design of the policy*, this is the assessment of the assumptions which relate the policy to the issue it is designed to address, including the strategy the policy has adopted to achieve its objectives (Weiss, 1972; Wholey, 1979; Chen and Rossi, 1980; Howlett and Ramesh, 2003). Such a stage will necessarily involve decision analysis techniques when selecting which policy alternative and implementation strategy will meet the policy objectives in the most cost-effective way (Palumbo, 1987). The policy theory explains why the policy does what it does and provides the rationale for expecting that doing so will achieve the desired results (Rossi et al., 2004).

(3) *implementation of the policy*, also known as the policy's process. This area of evaluation looks at the performance of the processes used to put the policy into place. That is, the policy activities that actually take place and the services that are actually delivered in routine policy operation. This kind of evaluation is designed to describe

how a policy is operating and assess how well it performs its intended functions (Howlett and Ramesh, 2003; Rossi et al., 2004).

(4) *impacts or outcomes of the policy*, is concerned with assessing how and the extent to which a policy produces (or will produce) the outcomes it is intended to produce. Albeit difficult, evaluators attempt to assess the effect of a policy by measuring its outcome: the state of the target population that it is expected to have changed (Parsons, 1995; Howlett and Ramesh, 2003; Rossi et al., 2004).

The principal tool for evaluating policy impact is called an Impact Assessment (IA) (Mohr, 1995; Parsons, 1995; Howlett and Ramesh, 2003; Rossi et al., 2004). The primary function of an IA is to explore whether a policy actually produces its intended effects. Admittedly, an IA is used to produce estimates of impacts, and so are never 100% accurate (Mohr, 1995; Rossi et al., 2004; UK Gov, 2009a; European Commission, 2009a). Generally, the principle of proportionality is applied to the funding of such assessments (Rossi et al., 2004; UK Gov, 2009a; European Commission, 2009a). That is, the amount to be invested will be determined by the potential value the results could bring. What's more, the greater the investment, generally, the more valid the results (Rossi et al., 2004). One of the governing principles in IA is the idea that the outcome for variables under the effect of the policy should be estimated and compared with estimates for the same variables which have not been exposed to the policy (Parsons, 1995; Mohr, 1995; Shadish et al., 2002; Howlett and Ramesh, 2003; Rossi et al., 2004). The leading approach for doing so is the randomised field experiment, as these have been shown to exhibit the most credible and statistically valid results (Mohr, 1995; Shadish et al., 2002; Rossi, Lipsey et al., 2004). An IA can be carried out at any stage in the life of a program or policy but because they are usually costly, the principle of proportionality is again used to judge if and when one is necessary (UK Gov, 2009a; European Commission, 2009a). It must be noted that IA is not a policy formulation tool (UK Gov, 2009b). These are only conducted once a policy has been formulated and are often used as support for a new policy proposal.

(5) *efficiency of the policy*, in this kind of assessment, evaluators attempt to relate the results of a policy to their costs. The two main types of efficiency analysis approaches

are cost-benefit and cost-effectiveness analyses (Gramblin, 1990; Parsons, 1995; Nas, 1996; Howlett and Ramesh, 2003; Rossi et al., 2004). Generally when policy impacts (including benefits and costs) can be assigned a monetary value then the former is preferred but more often than not, all impacts cannot be quantified in such a way so cost-effectiveness analysis, which quantifies only costs and expresses benefits in outcome units, is used (Rossi et al., 2004).

Cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA) are examples of the rational techniques to improve the effectiveness of public decision-making which policy analysis is said to study (John, 1998). But, although they are widely used as they address real and continuing problems, they are the source of considerable controversy (Simpson and Walker, 1987; Eddy, 1992; Zerbe, 1998). Simpson and Walker (1987) provide an example of why the approach has often been called into question. The review identifies three main problems with the method; (1) its one-dimensional nature because it attempts to reduce every impact/effect to a monetary value, which just may not be appropriate in many cases, (2) uncertainty and risk because it attempts to represent an outcome, which is necessarily prone to uncertainty by a single parameter, which is risky and (3) intergenerational bias, which is bias arising from the selection of discount rates, which CBA argues should be held constant for competing projects.

Essentially, these are choices made by relevant analysts and are subjective decisions, which can hence result in biased results (Simpson and Walker, 1987; Eddy, 1992; Zerbe, 1998). So, the subjective nature of such techniques lead them to being as politically influenced as the original techniques for supporting decision-making that they try to avoid (John, 1998).

Efficiency analysis is often considered an extension of IA and is generally carried out (1) prospectively (ex-ante) as part of the planning and design phase or (2) retrospectively (ex-post) as a policy learning exercise whose results will benefit the possible future status of the policy (Gramblin, 1990; Nas, 1996; Howlett and Ramesh, 2003; Rossi et al., 2004). The exact timing of an efficiency analysis is subject to some dispute with ex-ante and ex-post providing differing advantages but also suffering from a number of limitations (Cabinet-Office, 2003; Davies, 2004; Rossi et al., 2004).

What is certain is that through the appraisal of a policy that has recently been implemented ex-post, any form of evaluation exercise generates valuable insight for the strategy to be taken for analogous policies in the future (Gramblin, 1990; Nas, 1996; Howlett and Ramesh, 2003; Rossi et al., 2004). The exercises described in points (3), (4) and (5) above, in particular, can be used to assess the quality of the implementation of a policy on a strategic level, and such information can eventually be used to make predictions into the effectiveness of future implementation strategies (Rossi et al., 2004).

2.5.2 Important Considerations

An important stage before the initiation of the evaluation exercise consists of identifying and formulating the question that the evaluation will address (Gramblin, 1990; Nas, 1996; Howlett and Ramesh, 2003; Rossi et al., 2004). In order for an evaluation question to be useful, it must be reasonable, appropriate and most of all, answerable (Rossi et al., 2004). Similarly, the importance of all stakeholders and evaluation sponsors in having an input into such an exercise so that all parties are up-to-date and in agreement about the process to be adopted cannot be stressed enough (Howlett and Ramesh, 2003; Rossi et al., 2004).

This is never a straightforward task as there could be many stakeholders, with different views and interests, often considerably geographically separated. A useful and popular consensus building tool for such a purpose is the Delphi method[1] (Geist, 2010). One of the main advantages of such a method is that it can be applied without the need of congregating all participants in the same location; it can be done via several methods of interaction. Geist (2010) showed that the data collection process, whether it is correspondence based or web based, makes no difference in the quantity and quality of data collected. Despite such a success level, the Delphi method is not used for such a purpose in this thesis, rather its use is proposed as a line of potential future research.

---

[1] The Delphi method is a structured, interactive judgmental data collection process which is most commonly used in forecasting. Conceived by the RAND Corporation over 50 years ago, it consists of an administrator conducting multiple Delphi 'rounds' in which information is gathered from experts and then anonymously presented back to them until a consensus is usually arrived at.

Any evaluation, irrespective of the policy stage it is intended to address, must be tailored to the political and organisational context of the policy being evaluated (Howlett and Ramesh, 2003; Rossi et al., 2004). In other words, when planning an evaluation, evaluators must not only strive to find an approach which is in accordance with the type of question to be addressed but they must also tailor the specifics of the evaluation design, by acknowledging and incorporating the political and organisational forces that could affect the exercise, to fit the circumstances of the situation (Weiss, 1993).

Practically speaking, the evaluation of a policy is far from an exact science and can often prove to be a very difficult task (Howlett and Ramesh, 2003; Rossi et al., 2004). Because evaluation is dynamic in nature, causing potential changes in circumstances and activities during an evaluation, a suitable trade-off must be made between the scientific and pragmatic considerations in the evaluation design (Rossi et al., 2004).

In an investigation into the past, present and future of evaluation literature, Krane (2001) recalls the progression of the field since its beginnings to the modern era with the different controversies along the way. With the main debates occurring over differences of opinion in the 'methods' and 'purposes' of evaluation, the author argues that despite the turmoil, the field has not only progressed substantially, but it has been diverse and disorderly in nature, creating a strong and flexible toolbox for evaluators, which has made evaluation itself an indispensable instrument in policy analysis and program management.

2.5.3 Evaluation Applied

The UK government uses a large variety of evaluation approaches (Davies, 1999, 2004; Cabinet-Office, 2003; HM Treasury, 2009), including those that have been mentioned already, to guarantee that policies, programs and public services are designed and implemented as effectively and efficiently as possible.

Examples of such evaluation exercises include (Davies, 2004):

- systematic reviews of existing evidence,

- policy pilots,

- demonstration projects,

- ex-ante and post-hoc evaluations of specific intervention,

- economic appraisals and evaluation methods,

- strategic audit and international benchmarking,

- regulatory impact assessment,

- performance management mechanisms.

The basis of such research is to aid in strategic planning and development as well as the operational management and implementation of public services (Davies, 2004).

In a typical evaluation report of this kind, DEFRA (2006) synthesises the results from multiple evaluations of climate change policies. The evaluations were carried out by experts across the relevant governmental departments, other organisations with policy responsibility or by consultants, who have followed standard Green book (HM Treasury, 2009) procedure and have been subject to a peer review process by the Interdepartmental Analysts Group (IAG). In this particular illustration, evaluation comprised of both cost-benefit and cost-effectiveness analyses of the policies intended to reduce green-house-gases (GHGs) across seven sectors (Agriculture, Business, Domestic, Energy Supply, Public Sector, Transport and Waste) with the results showing that Domestic measures to be by far the most cost-effective.

These types of reports, summarising and comparing different policy measures for achieving the same objective across different sectors are exactly what evaluation is all about. The idea behind reports like these is to learn why some measures are working better in other sectors and attempt to use this insight to improve measures in other sectors that might not be working so well.

In a review of the policy evaluation practices in the UK, Davies (2004) argues that the principal driving force for such an investment of resources to ensure high quality policy evaluation in the UK is the government's commitment to evidence-based policy making. Such a commitment necessarily requires policy makers and implementation

officials to call upon the valuable evidence available to them, from sources such as national statistics, academic research, economic theory, pilots, evaluation of past policies, commissioned research, consultation with delivery officials, etc.

The UK government provides a guide on evaluation for 'policy evaluators and analysts, and people who use and commission policy evaluation', with a strong focus on such an exercise in government, with it calls the *Magenta Book* (Cabinet-Office, 2003).

The Magenta Book gathers information from core texts in the subject, as well as governmental literature to develop an information source with a view on the demands of 'evidence-based policy making and the changing needs of analysis in and for government'. Past governmental research has highlighted the importance of sound evidence, proper evaluation and good analysis as the pillars of effective policy making (Cabinet-Office, 2003). Consequently, a rise in demand in support on how to approach and undertake quality evaluation, appraisal and analysis for policy making, has been seen, which has as a central protagonist, the *Green Book* (HM-Treasury, 2009). The latest source in this wave of evaluation guidance, the Magenta Book, is aimed at specialists and generalists alike, and provides information and assistance on the status quo of evaluation research, data collection and statistical analysis, issues determining policy effectiveness, without forgetting assistance into the broad range of methods used in policy evaluation (Government-Social-Research-Unit, 2009).

The Magenta book (Government-Social-Research-Unit, 2009) specifies two types of evaluation used in government; summative and formative. The first, is what was referred to as Impact Assessment earlier on and is intended to explore *whether* a policy was effective or not. The second, is intended to explore for *whom*, *why* and under *what conditions* a policy was effective. The book also covers other types of evaluation that exist in government, such as theory-based evaluation, goals-based evaluation, goals-free evaluation, experimental and quasi-experimental evaluation, qualitative evaluation and lastly, economic appraisal and evaluation. Despite being mentioned briefly already, further discussion of most of these is beyond the scope of this research.

So in this section, it has been seen that evaluation plays an important part after a policy has been implemented, as a policy learning exercise, but also before implementation, as a tool for policy assessment. This will have important implications later on when the different forecasting practices in public policy are explored.

2.6 Forecasting in Public Policy

Forecasting activities exist everywhere in public policy, although not always apparent or in a formalised manner (Publicpolicyforecasting.com, 2009). Like in the private sector, any attempt to predict future events will result in better planning and strategy (Makridakis, 1996), which allows for better policy making (Publicpolicyforecasting.com, 2009). The recent move towards evidence-based policy making (Government-Social-Research-Unit, 2009), performance measurement and the key concept of evaluation in public administration has called upon the need for better forecasts to provide decision-makers with a sound rationale for their choices (Government-Social-Research-Unit, 2009; Publicpolicyforecasting.com, 2009).

In this section, an attempt will be made to explore the different forecasting practices in public policy and the purpose they serve. The aim is to try and assess how forecasting is used in public policy. Due to such an application of forecasting having a relatively low-profile in forecasting research, this search extends over a broad range of sources and comprises insight mainly from policy texts, governmental documents, web-pages and independent studies. The search commences in quite a broad nature with a look at how changes in public administration doctrines have put a new emphasised importance on forecasts. Then, the forecasting practices found to be in use will be examined together with the purpose they serve.

2.6.1 Public Administration: A New Outlook

The advent and rise of the Performance Measurement (PM) philosophy, also known as results based management or managing for results, in the UK can be traced back to the Margaret Thatcher and John Major administrations and more recently with Tony Blair (Hughes, 2003; Davies, 2004; Talbot, 2008). This philosophy, regarded as a key component of the New Public Management (NPM) doctrine in public administration,

refers to the idea that governing should be centred on ensuring objectives are met in an effective and efficient manner (Hughes, 2003; Talbot, 2008). The concepts that define PM are believed to be, along with experience, expertise and judgment of policy makers, the influencing factors in policy making and service delivery in the UK (Davies, 2004).

In a study comparing the principles of PM with those of evaluation, Nielsen and Ejler (2008) find strong correlations between the two. This is no surprise as both philosophies, although different in the way they manifest themselves, share the fundamental objectives of policy learning and improvement. What's more, it is no coincidence that the rise of the NPM paradigm has been matched by the surge in evaluation research (Nielsen and Ejler, 2008). Nielsen and Ejler (2008) also argue that evaluative knowledge should be applied during all stages of the policy cycle and that any decisions made along the way should be supported by sound rationale. For example, the selection of a policy tool should be based on a prediction of its suitability for the policy problem. This prediction should be based on evidence of the effectiveness or efficiency of the tool for the particular policy problem in question. This evidence can be gathered through the evaluation of past, similar tools in similar contexts, if possible, and applied to the target policy. In this way, the various instruments of evaluation should be used to improve the application of the PM philosophy (Nielsen and Ejler, 2008).

It is such a philosophy that has been the cause of the latest push by the UK Government to drastically improve the management and delivery of public policy to make it evidence-based, cross-cutting and innovative (Government-Social-Research-Unit, 2009). Davies (1999) defines evidence-based policy as the integration of experience, judgment and expertise with the best available external evidence from systematic research. Government-Social-Research-Unit (2009) believes that evidence-based decision-making relies heavily on the findings of scientific research gathered through the scientific process.

A result of this new paradigm saw the creation of the Policy Hub website (Government-Social-Research-Unit, 2009). This website, developed by the

Governmental Social Research Unit was recommended by the report of the Better Regulation Task force published in July 2002 (Better-Regulation-Task-Force, 2002) as the key gateway for promoting best practice, guidance and case studies to policy makers. Policy Hub (Government-Social-Research-Unit, 2009) is a web-based platform that provides support and guidance within the context of evidence based policy and has an aim of improving the way public policy is designed and implemented. The developers claim that it provides

- tailored access to initiatives, projects and tools that support better policy making and delivery

- extensive guidance on the use of research and evidence in the evaluation of policy

- links to a wide range of research resources and tools from the UK and around the world

(Government-Social-Research-Unit, 2009)

In a fairly recent attempt to centralise the area of policy forecasting, the principal global forecasting page, ForecastingPrinciples.com (2009d), established a 'special interest group' (SIG[1]) in Public Policy Forecasting (Publicpolicyforecasting.com, 2009). No books or journals specific to the forecasting practices in policy exist, simply a multitude of studies and government documents spread across several research contexts, so the developers saw a chance to unify what is known on the subject. The SIG was established to provide a platform for the rational analysis of government policies. In the words of the developers,

Forecasting is more important for the public sector than for the private sector because public policy involved coercion, can result in large changes, and is not guided by prices. The injunction to "first, do no harm" is therefore appropriate for public policy decision-making. Scientific forecasting can help decision-makers to choose the best policies.

(Publicpolicyforecasting.com, 2009)

The intention of the Public Policy Forecasting SIG was to include evidence-based assessments of forecasting procedures behind major policy initiatives in areas such gun control, capital punishment, climate change, immigration, public construction projects and minimum wage laws, via forecasting audits. Unfortunately, only two such audits have been produced as of yet: one on Polar Bear Populations (Armstrong et al., 2008) and the other on Global Warming (Green and Armstrong, 2007a). But there will

---

[1] A special interest group (SIG) is a group with a common interest in the development of a specific field of knowledge.

undoubtedly be more in future as the need for a formalised way of looking at forecasting procedures in public policy gains momentum.

The only one of relevance to this research is that of Global Warming. Green and Armstrong (2007a) offer a review of the forecasts and forecasting procedures prepared in the IPCC[1] report. The authors claim that these forecasts are "opinions of scientists, transformed by mathematics and obscured by complex writing", and not "scientific forecasts", as they claim to be. The importance of the paper lies in that it provides evidence of a very high-profile case with significant global implications where bad forecasts are being used for decision-making in public policy.

Hence, the recent PM doctrine has given precedence to the idea of evaluation and evidence-based policy making. Greater importance has been given to evaluative measures such as efficiency and effectiveness to support rational decision-making in light of such a results-based, Performance Measurement paradigm.

2.6.2 Forecasting Practices

It was seen in section 2.5 that tools such as ex-ante efficiency analyses and economic appraisals (including impact assessments and cost-benefit analyses) can not only be used for predicting the impact of the policy but also for the purpose of planning and designing the implementation process itself (Howlett and Ramesh, 2003). That is, these tools can be used as means for making forecasts about the implementation on an operational level. There is also evidence however, that these techniques are used to make strategic decisions.

As was explained in section 2.5, such exercises can be carried out using a policy's anticipated outcomes and costs (Gramblin, 1990; Nas, 1996; Howlett and Ramesh, 2003; Rossi et al., 2004). Hence, these kinds of evaluative analyses will inevitably involve predictions of policy impact as well as the costs of providing and delivering the intervention policy, which aid in the design and planning of the implementation process (Howlett and Ramesh, 2003).

---

[1] Intergovernmental Panel on Climate Change (IPCC)

Rossi et al. (2004) state that in some cases, estimates of these inputs and magnitude of impact can be made with considerable confidence either because there has been a pilot policy or because the policy is fairly straightforward in its implementation. Logically, in order to determine whether or not a policy has achieved its anticipated outcome, an ex-post evaluation must also take place. A comparison between the forecasts made ex-ante with the observed impacts in the ex-post evaluation can be used to ultimately improve the forecasting process for future ex-ante appraisals (Rossi et al., 2004).

Two reports, investigating the accuracy and applicability of CBA in ex-ante and ex-post assessments are Harrington et al. (2000) and Boardman et al. (1994). Boardman et al. (1994) argue that CBA is intended to help in public sector decision-making. The 'help' it claims, depends on the timing within the policy when it is performed.

- An ex-ante CBA can help decision-makers decide upon the characteristics of the policy and whether it should be proceeded with. This implies that CBA also plays a role in defining the strategy of implementation.

- An ex-post evaluation will assess the quality of the policy itself but also that of the ex-ante CBA predictions. That is, an ex-post evaluation will provide detailed feedback and insight into the policy (in a direct way) to help decisions for future policies and into the original ex-ante CBA, so that the CBA process can be improved.

For example, Harrington et al. (2000) compares ex-ante and ex-post estimates of the direct costs of the same policies. It found that for total costs, the ex-ante assessments tended to overestimate, which is fairly common. This is particularly true when the policies use economic incentives for implementation. So an ex-post evaluation has identified a shortcoming of an ex-ante assessment which can be attempted to be remedied for the future.

Despite the clear advantages of forecasting the potential benefits and costs of a policy for the reasons described above, CBA as an exercise has been called into doubt for several reasons. Indeed, as Rossi et al. (2004) argues, the issue of accuracy with

regarding policy input and output predictions is one of the main areas of concern in ex-ante analysis.

Rossi et al. (2004) believes that ex-ante CBAs are particularly important when either a policy will require extensive investment of resources to be realised or when a policy will be difficult to abandon once it has been put into place. Hence, CBA would provide heavy influence in the decision-making process when the proposed policy would require considerable expenditure. This, in similar fashion to Boardman et al. (1994)'s example above, suggests that CBA plays a part in strategic decision-making for implementation.

In the social arena, many social policies are initiated and modified without the use of CBAs. There seems to be an existing sentiment that because of the disbelief in the accuracy of such ex-ante predictions, either the exercise is directly not undertaken or the predictions that are made are not considered when making decisions (Rossi et al., 2004). Instead of avoiding their use, time and money should be invested in attempting to make these predictions more accurate. A start could be by formalising the forecasting procedure and applying various methods instead of just one or a few.

Thus, the ex-ante evaluation tools covered in section 2.5, such as economic appraisals, impact assessments and cost-benefit analyses are presented in the literature as the important tools involved in forecasting the impact of a policy and in designing and planning the implementation process. There is also some evidence however, that they are also used for strategic purposes. This then, is further evidence to the claim made in section 2.3 that to attempt to segment the policy process into stages in which certain methods are applied for certain phases is dangerous. It is not a good representation of what happens in reality as this example has shown.

With these techniques being presented in the literature as primarily intended for forecasting on more of an operational level (despite evidence that they are also used for strategic purposes), it was of interest to see what other forecasting processes could be found for making predictions on more of a strategic level. That is, for instrument selection or for shaping implementation strategy.

Hence, attention was drawn to exploring the literature for further evidence of the formal use of forecasting approaches when making decisions on instrument and implementation strategy. After an extensive examination of the public policy and forecasting literature, no such evidence was found. In fact, aside from what was found on instrument selection in sub-section 2.4.5 and the evidence of the evaluative techniques above, very little information was found on how strategic decisions are made during the policy formulation stage and what role forecasts play in this. Furthermore, no evidence of the use of formal judgmental approaches, such as structured analogies, was found.

In sum, this section has shown that the recent PM doctrine has given primacy to the concepts of evaluation and evidence-based policy making. Greater significance has been given to evaluative measures such as efficiency and effectiveness to support rational decision-making in light of such a results-based, Performance Measurement paradigm.

Furthermore, it has been shown that ex-ante evaluation tools are the important approaches used in predicting the impact of a policy and in designing and planning the implementation process. They are forecasting procedures for making predictions about implementation at an operational level. There is also evidence that these techniques are also used for deciding implementation strategy, although to what extent, is unclear. When a further examination was conducted into the formal forecasting approaches used specifically in decision-making for instrument selection and implementation strategy, no evidence was found. No evidence of the use of formal judgmental approaches, such as structured analogies, was found either.

2.7 Expert Judgment

In the previous section, the review of the literature on the forecasting practices in public policy revealed that forecasting approaches are readily used in implementation for operational reasons but it also revealed some evidence for their use in deciding implementation strategy. Of these forecasting approaches however, none were found to be judgmental, that is, based on expert judgment. The use of expert judgment in

forecasting, judgmental approaches, were introduced in sub-section 2.2.1 and will be covered further in more detail in section 2.8.

This section will focus on the underlying concept behind judgmental forecasting, the notion of expert judgment, or expertise. Attention will first be given to the theory behind this concept and then the role it plays in public policy will be examined.

2.7.1 Theory

How expert judgment can best be used to make better predictions is a key research area in the study of judgmental approaches (Lawrence et al., 2006). Meyer and Booker (2001) define expert judgment to be data given by an expert in response to a technical problem. The authors believe that expert judgment can be considered as a 'snapshot' representation of an expert's knowledge or opinion on a particular issue at a given point in time. This view was originally defined as such by Keeney and Von Winterfeldt (1989, p.6). Ayyub (2001) provides a similar definition in that an expert opinion is seen to be a formal judgment of an expert on a matter, in which his or her advice is needed. By nature, an opinion or judgment is a 'subjective assessment, evaluation, impression or estimation of the quality or quantity of something of interest that seems valid or probable to the expert's own mind' (Ayyub, 2001, p.98). Moreover, an expert opinion is more than just a guess (Meyer and Booker, 2001).

Naturally, any discussion on expert judgment necessitates a definition of what an 'expert' really is. For example, Meyer and Booker (2001) believe that an expert is a person with training and experience in the subject area and is furthermore acknowledged by the field or any relevant stakeholder, as qualified to answer the question at hand. According to Shanteau et al. (2002), expertise is formed through personal training and experience and hence can never really be objectively measured (making comparisons between levels of expertise difficult), which the authors argue leads to obvious complications when having to select experts for a particular task.

The problem of 'when' the use of expert judgment is appropriate is one that has been the topic of much discussion in any and all of the areas in which such an approach is used. Helmer and Rescher (1958) believed that the predictive usage of expert

judgment is acceptable in any field that has not yet developed to the point of having scientific laws which govern it. Generically speaking, experts are needed when either there is no other way of addressing a problem or if the use of expertise is deemed the best way to address the problem (Meyer and Booker, 2001). The choice of when expert opinions are to be used, and to what extent, is a delicate one whose outcome will vary depending on the characteristics of the situation (Meyer and Booker, 2001; Ayyub, 2001; Lawrence et al. 2006; Parackal et al., 2007).

Some situations can only be unlocked by the appropriate use of expert judgment (Makridakis et al., 1998). In sub-section 2.2.1, the use of judgment for forecasting was briefly discussed, a greater discussion will follow in the section 2.8. The different advantages and disadvantages of such an approach in forecasting will be examined along with how these determine how and when they are to be used.

Generally speaking however, five common areas where expert judgment is popular can be identified (Meyer and Booker, 2001);

1) To provide estimates on new, rare, complex, or otherwise poorly understood phenomena.

2) To forecast future events: In general, when good data are unavailable, predicting future events or actions requires use of expert judgment.

3) To integrate or interpret existing data.

4) To learn an expert's problem solving process or a group's decision-making processes.

5) To determine what is currently known, what is not known, and what is worth learning in a field of knowledge.

Despite encompassing a significant proportion of the situations in which expert judgment is used successfully, this list is not entirely comprehensive[1] (Lawrence et al. 2006). One aspect that the list fails to include is that expert judgment is often used because sometimes it provides the means to a kind of information that no other approach can. To forecast future events for example, expert judgment is not only used

---

[1] Moreover, one can argue that it is impossible to redact a completely comprehensive list of such cases as new situations are always being discovered and investigated.

when no other approach is possible, but also to incorporate domain knowledge which can prove vital in predicting unsystematic changes in data patterns, something which other approaches fail to do (Makridakis et al., 1998; Lawrence et al. 2006; Parackal et al., 2007).

The performance of expert judgment approaches is mixed and depends largely on the scenario in which they are applied (Lawrence et al. 2006). When it comes to forecasting, expert predictions are particularly useful and should not be ignored because they are cheap and can be quite accurate due to the fact they represent the most updated consensus on the core assumptions behind a given decision-making situation (Ascher, 1978). The beauty of this is that an expert can continuously update his or her opinion as new information is received.

The quality of expert judgment will vary upon how the data are gathered (Meyer and Booker, 1987, 2001; Booker and Meyer, 1988; Ayyub, 2001), which can range from the subconscious to the deliberate. Within deliberate data collection, such a process can range from informal to formal. On the informal end of the continuum, experts are asked to produce forecasts 'off the top of their heads', in an approach termed unaided judgment. On the formal end, experts are 'walked through' the process with a whole range of different techniques for doing so (Meyer and Booker, 2001). The authors believe that the key to proper elicitation is in understanding the cognitive limitations (mentioned in sub-section 2.2.3) of experts and using these to tailor and enhance such techniques. This is supported by Armstrong (1985, chapter 6), who found that forecast accuracy is improved when judgmental approaches are structured.

2.7.2 In Policy

In previous sections it has been established that a common, but inherently flawed, view of the policy process is one with multiple stages, known as the policy cycle, which starts with the problem and ends with the evaluation of the solution. This model then suggests that policy making is not an instantaneous event defined by a single decision made by a single decision-maker but rather it is a series of events or decisions over a lengthy period of time involving many analysts and decision-makers.

Rich (2004) centres on how think tanks[1], organisations offering services in research, and experts become influential at different points in the policy process. Expertise is argued to play an active and important role in each stage of the policy process (Rich, 2004). For example, expertise can be useful during agenda setting as warning to policy developers of forthcoming problems and as support to decision-makers on how to revise policy. The use of this expertise can vary from long-range, high-level strategic decisions to low-level, operational decisions (Rich, 2004).

According to Rochefort and Cobb (1994) expertise can help establish a framework for problems and intervention before issues are considered for debate. Typically, during the debate process in which different policy officials defend their particular viewpoint on a decision that is to be made, expertise through research can serve as valuable ammunition for such a cause. Then, once a policy has been deliberated, expertise becomes useful again for the officials responsible for its implementation (Rich, 2004).

Evidence exists of the big part expertise plays in EU policy, to the point where many believe such a technocratic philosophy leads to considerable bias in governmental decision-making and policy (Radaelli, 1999; Rayner, 2003). The biases referred to by Radaelli (1999) and Rayner (2003) were not the same cognitive and motivational biases which experts are prone to when giving their opinions but rather political biases which skew the democratic factor in governmental decision-making and policy. Nevertheless, whatever their nature, the question here, is, could these biases be avoided, or at least minimised, by a formal approach to the use of expertise in public policy.

So, the presence of expertise in public policy exists mainly as an influential force in decision-making, throughout all stages of the policy cycle, and primarily as a support for decision-making in situations where specialised information is deemed necessary and beneficial. However, as was the case in section 2.6, no evidence was found in the literature of any formal judgmental approaches based on expert opinions being used to make predictions.

---

[1] A think tank (also called a policy institute) is an individual or organization that provides research and advocacy in a given area.

2.8 Judgmental Forecasting and Structured Analogies

So, section 2.7 showed that although expertise is used in decision-making (how so, is unclear), there is no evidence to suggest that any expert-based judgmental approaches are used when making predictions.

Having been introduced briefly in sub-section 2.2.1, attention will now return to judgmental forecasting approaches, with details of their advantages and shortcomings. Furthermore, particular attention will be given to the idea of forecasting by analogies as this is seen as particularly useful in the context of PIS effectiveness predictions. Finally, the approach proposed in this research, a structured approach to the use of analogies will be presented and explored.

2.8.1 Judgmental Approaches

In sub-section 2.2.1, it was mentioned that judgmental approaches to forecasting hold particular value

- (1) because of their ability to predict unsystematic changes in patterns of past behaviour making them useful in situations in which it is thought that the future will not be simply the continuation of the past,

- (2) when no past quantitative data is available making it the only remaining alternative,

- (3) when there is a high level of uncertainty surrounding the forecasting exercise

- (4) to integrate information, known as domain knowledge, which is believed can improve prediction accuracy.

However, judgmental forecasting implies making predictions based on expert judgment (Makridakis et al., 1998; Lawrence et al. 2006; Parackal et al., 2007). As has been mentioned already and will be investigated in more depth, this can lead to a score of biases stemming from a human's cognitive limitations (Meyer and Booker, 2001). Hence, expert judgment is best used in forecasting when done so in a formal, structured way, in order to minimise these biases (Meyer and Booker, 2001; Ayyub, 2001; Armstrong, 1985, chapter 6).

The Forecasting Dictionary (Forecastingprinciples.com, 2009a) defines judgmental forecasting as,

A subjective integration of information to produce a forecast. Such methods can vary from unstructured to highly structured.

The Forecasting Dictionary (Forecastingprinciples.com, 2009a)

On the unstructured side of the spectrum lies unaided judgment. Perhaps surprisingly, but unaided judgment is the most common forecasting method (Makridakis et al., 1998; Lawrence et al. 2006), and occurs when the forecasters has good knowledge of the situation resulting in a great confidence in the forecasts. When a situation arises where unaided judgment is not suitable, in one of high uncertainty for example, or when not enough quantitative information is available for the use of statistical methods, a structured approach to the use of judgmental forecasting is warranted (Armstrong, 1986). Such structured approaches include intentions and expectations, conjoint analysis, judgmental bootstrapping decomposition, structured analogies, simulated interaction, and expert forecasting[1] (Forecastingprinciples.com, 2009c).

An examination of the Methodology Tree (Forecastingprinciples.com, 2009b) in figure 2.2 places judgmental approaches on the left hand side with approaches such as expert systems, quantitative analogies, rule-based forecasting and causal methods classed as hybrids and placed between the judgmental and quantitative branches.

---

[1] E.g. Nominal Group Technique, Delphi, and Prediction Markets.
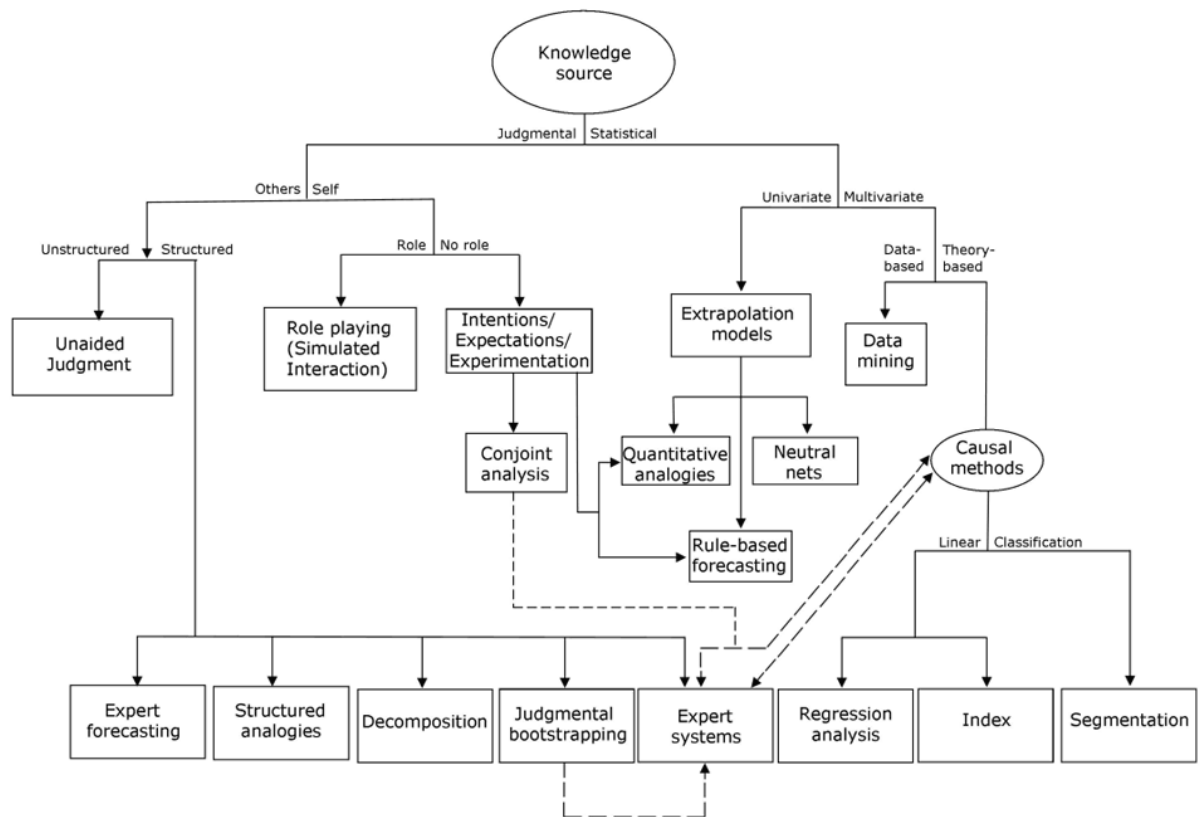
Figure 2.2: Methodology Tree (Forecastingprinciples.com, 2009b)

The Forecasting Principles (Armstrong, 2001a) summarises what is known on the performance of these methods in given situations. Practically speaking, judgment is actually required and important for all forecasting methods but arguably even more so, in those that rely purely on judgment. The concept of combining quantitative approaches with judgment is a popular topic in forecasting research which has shown promising results (Armstrong, 2001b).

2.8.2 Forecasting by Analogies

One important objective for future research in judgmental forecasting is to develop ways of supporting judgmental forecasters in their task (Lawrence et al., 2006). One such method for support is the use of analogies (Lawrence et al., 2006; Green and Armstrong, 2007b; Lee et al, 2007). In situations of little availability of quantitative data and high uncertainty where a prediction of the outcome of an event must be made, it seems logical that information about similar, analogous events from the past

or other contexts along with their outcome will be of use (Kokinov, 2003). This idea is nothing new and has appeared in several other fields.

Green and Armstrong (2007b) conducted an extensive review for performance of analogies in forecasting and found very little evidence. Nonetheless, some studies do exist, with differing levels of success. Some cases that have shown the use of analogies to good effect are as follows.

- Kokinov (2003) defines analogy making as a "process of perceiving one situation (target) in terms of another" (p.168). In one of the first comprehensive books on the uses of analogies, Markman and Moreau (2001) found this idea was found to play a fundamental part in decision-making. Within the context of decision-making in game theory for example, if a player chooses a certain strategy in a previous game, one would expect him or her to behave similarly in an analogous situation (Kokinov, 2003). This leads Kokinov (2003) to conclude that human behaviour can hence be explained by assuming decisions are made by analogy with previous cases.

- The use of analogies for economic and business forecasting dates back to the 1930s with analysts using explicit analogies from previous business cycles to predict things such as the end of the depression (Goldfarb et al., 2005). So in this case, the cyclical property that business is believed to exhibit is capitalised on by using information from past cycles to predict existing or future ones.

- Software cost estimation is an area where analogies are quite an attractive prospect. The reason for this is that outcomes and costs of past projects can be stored in historical databases and used to predict the cost of future projects (Angelis and Stamelos, 2000). In a paper reviewing the state of the art in software cost estimation, Heemstra (1992) finds that the majority (365 of 598) of organisations who forecast the cost of software projects did so by building a database of previous projects and their costs, and used these as analogies for their target case.

- Forecasting by analogy has been quite popular in the area of technology (Schnaars, 2009). In such a context, analogies are formed when two similar technologies are

diffusing in a similar fashion or are predicted to diffuse in the same manner due to their similarities. The logic behind this is that if two events resemble each other enough, their outcome is presumed to be similar as well. Schnaars (2009) found that despite their popularity, unfortunately little research has centred on how well or how poorly such a technique has performed.

- In an early investigation in a Marketing context, Easingwood (1989) looked at predicting the diffusion of a new product. The paper argues that any product can fall into what he calls a product 'class' and this 'class' has a known diffusion shaper or path. Hence, through comparison between the new product and its class diffusion shaper, predictions about the new product's own diffusion can be made.

- Finally, the use of analogies has also been seen in scenario planning (Dortmans and Eiffe, 2004). Dortmans and Eiffe (2004) propose a new way of predicting the likeliness of any future analysis based on the use of analogies taken from the past. The paper proposes that the model to be used to make the scenario likeliness prediction be tested, by being made to predict and establish links with recent historical environments.

So the above examples show that analogies have been seen to be useful in several different decision-making and forecasting situations. These studies all share two fundamental common traits which Green and Armstrong (2007b) believe are the conditions that make analogies useful for forecasting; low levels of past quantitative data and a high level of uncertainty surrounding the situation.

However, evidence also exists of cases where the use of analogies has shown to be unsuccessful.

- Schrodt (2002) reviewed the empirical evidence of analogy-based approaches for forecasts of decisions made in conflict situations and found no evidence for the superiority of analogy based methods compared to other methods. Furthermore, Schnaars (2009) found that analogy forecasts served more to misdirect than to improve demand forecasts.

- Similarly, in a study to forecast sales promotions, McIntyre et al. (1993) found no difference in the accuracy of predictions made using analogies and that of an expert buyer.

Nonetheless, Green and Armstrong (2007b) maintain that from the existing literature in the area, analogies show that they can provide some improvements in accuracy relative to other methods. Hence, if the common traits present in the successful use of analogies for forecasting as proposed by Green and Armstrong (2007b) are combined with Kokinov (2003)'s idea that analogies will work well in situations whose nature suggest that the use of analogous information is instinctive and will hence be beneficial, a set of conditions for the use of analogies can be drawn. This idea is illustrated in table 2.5.

Conditions in which analogies will be useful in forecasting
- Low level of available quantitative data
- High uncertainty levels surrounding the situation
- Analogous information is considered beneficial when the nature of the situation is such that the use of analogous information comes instinctively

Table 2.5: Evidence of Suitable Conditions for the Use of Analogies in Forecasting

So, although the existing literature has revealed some preliminary conditions under which analogies seem to be useful, much remains to be done to fully understand the exact conditions under which such an approach is recommendable. What's more, aside to what is presented above, little has been done in trying to investigate the issues surrounding the use of analogies and the possible shortcomings of the approach (Green and Armstrong, 2007b). Moreover, past research has not yet fully explored ideas such as the effect of structuring the use of analogies so as to offer the forecaster support when using them.

Hence, this section has shown that forecasting by analogies seems to be an attractive prospect when there is a low level of quantitative data available, a high level of uncertainty surrounding the forecasting situation and in situations where it seems inherent that analogous information can be positively applied to the target situation.

For these reasons, the use of analogies for predicting implementation strategy effectiveness seems a plausible and interesting prospect.

2.8.3 Structured Analogies

Two recent studies which have tackled the use of analogies in forecasting are those of Lee et al. (2007) and Green and Armstrong (2007b). These studies have recognised that the unstructured use of analogies can be hindered by the various cognitive limitations of the human mind which affect the use of any unstructured judgmental approach. Past research in judgmental decision-making and forecasting has shown that when judgmental processes are structured, they tend to capitalise more on the information individuals possess (Armstrong, 1985, chapter 6). Both Lee et al. (2007) and Green and Armstrong (2007b) posit that a forecaster requires such support when using analogies and as a result propose different approaches of structuring their use.

In an examination of the process of using analogies to make predictions, Lee et al. (2007) identify three key steps in such an approach

- *Recall* – the process of actually retrieving similar, past cases from memory

- *Similarity judgments* – the process of evaluating the suitability of the retrieved cases by rating their similarities

- *Adaptation judgments* – the process of making alterations to the analogies to suit the target situation in order to make predictions

The authors acknowledge the restrictions of the human mind and the detrimental effect this can have when using analogies. Without any kind of support, the forecaster will experience difficulties in each one of the above stages. Firstly, the *recall* of past cases can be bounded by recollection restrictions in the form of: capacity (only a small number of recent cases may be recalled), specificity (details of the case may be recalled incorrectly) and method (the way in which the cases are recalled might be defective). As a solution, the authors propose that the formation of a formal database of past cases will provide support for the recall process. Secondly, similarity judgments can prove problematic due to cognitive limitations so the paper suggests support for

this too should be provided. Lastly, the paper calls for support when making adaptation judgments and hypothesises that providing support for all three of these stages will improve the use of analogies when making predictions.

These notions of forecaster support form the basis for a forecasting support system (FSS) proposed by the paper in the context of predicting the impact of sales promotions on demand. The purpose of the FSS is to support users with similarity and adaptation judgments when recalling past information. The results of the study show encouraging signs for the support of analogies as the FSS was found to improve forecast accuracy (see Lee et al., 2007, pp.10-12).

Nonetheless, despite being a step in the right direction when it comes to using analogies for forecasting, the development and use of such an FSS will be relatively costly and unsuitable in all situations (Savio and Nikolopoulos, 2009a). In addition to being costly, the creation of such a database would necessarily involve a certain degree of subjectivity when providing similarity and adaptation support. Someone has to decide how and the manner such support will be given. Lastly, such an FSS presumes that past cases will be readily available for building such a database, which might not necessarily be the case.

In a similar investigation, Green and Armstrong (2007b) also believed that a formalisation and structuring of the use of analogies will ultimately improve accuracy. To this, the authors propose a formal method for using analogies which attempts to minimise the biases caused by a human interface. The paper recognises the benefits of analogies but fears that experts will be prone to choosing easy to recall analogies, which would simply serve to confirm their current opinion, not help form a new one. In other words, if analogies are used in an unstructured way, biased predictions will follow.

In earlier work, Armstrong (1985, chapter 6) found evidence to support the structuring of judgmental approaches as this was shown to improve accuracy. It was thought that the subjectivity that led to biases was the reason for the large errors and so the objective was to make the process more objective. On the basis of this, Green and Armstrong (2007b) posit that a structured approach to the use of analogies would

encourage experts "to consider more information from analogies, and to process it in a more effective way" (p.366). This leads the authors to the hypothesis that if a way of using analogies in a more objective manner could be found, biases would be minimised and accuracy surely improved. Thus the paper proposes a five step approach for structuring the use of analogies which the authors suitably call 'structured analogies' (SA).

- *Description of the target situation*

The first stage involves the administrator (responsible for the implementation of the approach) preparing a concise but detailed description of the forecasting situation through consultation with either expected unbiased experts[1] or from experts with expected differing biases[2] (Savio and Nikolopoulos, 2009a).

- *Selection of experts*

In the second stage, the administrator gathers a set of experts who are likely to have information about cases analogous to the target. The experts should ideally be chosen on the premise of how knowledgeable they are on analogous situations, the variability in their responses and the importance of obtaining accurate forecasts.

- *The experts each identify and describe analogies*

The third stage involves the experts being asked to describe as many analogous cases as possible and to match their analogies' outcomes with the target outcomes.

- *The experts each rate the similarity of the situations*

In the fourth stage, the experts are asked to list similarities and differences between the analogies and the target situation and then give a similarity rating.

- *Derivation of the forecasts*

In the final stage, the forecaster or administrator transcribes this information and uses it to make a final forecast. The final forecast would be generated using a pre-defined mechanical rule in order to promote logical consistency and replicability. E.g. the final forecast of the expert would be the outcome of the analogy rated as most similar to the target case. Data could be elicitated through questionnaires, interviews, or panel groups and collected either personally or via correspondence.

---

[1] In the case of EU policy, these can be found in academia for example.
[2] In the case of EU policy, these can be found in the different stakeholder organisations, lobby groups, etc.

Green and Armstrong (2007b) decided to test this approach by asking a large panel of experts to individually produce forecasts for the outcome of eight different conflict situations and compared the SA approach with unaided judgment (UJ) and chance. The results show that when predicting decisions made in eight conflict situations, 46% of structured analogies forecasts were accurate, compared to the 32% accuracy of unaided experts' forecasts, which were little better than chance. In addition, when experts recalled more than one analogy, hence demonstrating greater domain knowledge, accuracy of the SA method improved further.

Despite being the pioneering approach to structuring the use of analogies, the SA approach does not so without critique. The developers claim the method to be objective, a very bold and precarious statement when dealing with a judgmental approach. A better statement would probably have been 'less subjective...than the unstructured use of analogies'. The reason for this being that the method can never be entirely objective as there are steps in the procedure which cannot avoid subjectivity. Although the approach does well in minimising the biases associated with the cognitive limitations of experts when recalling the analogies, the administrator will be responsible for selecting, eliciting and analysing the expertise, and will inevitably involve subjective judgment on his or her part. Nonetheless, Green and Armstrong (2007b)'s method makes excellent progress in making a highly subjective procedure as objective as possible and can hence be seen as substantial advance in the evolutionary line of forecasting by analogies.

So, both Lee et al. (2007) and Green and Armstrong (2007b) have provided means of support when using analogies to forecast, the latter with a formalised and structured approach. Both studies report that this support leads to an improvement in accuracy.

A review of the existing literature on the use of analogies in forecasting has shown that these can be helpful in situations which are seemingly difficult to forecast, such as the outcome of a conflict or the diffusion of a new technology (amongst others), and in which very limited past quantitative data is available, there is high level of uncertainty and when the use of analogies comes instinctively. For this reason, it would be reasonable to find FBA as an attractive prospect for predicting implementation

strategy effectiveness. The above review also showed that a structured approach to use of analogies can improve forecast accuracy yet further.

2.9 Working with Experts

In sub-section 2.7.2 it was seen that expertise exists within public policy mainly as an influential force in decision-making when such specialised information is considered necessary and beneficial. No evidence was found however, to suggest that formal expert opinions are used as a means for making predictions of any sort.

Whenever experts are needed for a particular purpose, several issues can arise which can ultimately affect the quality of the data that is obtained from them (Howlett and Ramesh, 2003; Ayyub, 2001). This section will focus on some of the issues associated with the use of expertise. These issues are important because they translate directly to judgmental forecasting, which is quintessentially based on the use of expertise, and can hence help understand some of the issues with these, as raised in sub-section 2.8.1.

The first of these issues is the dilemma of being able to identify experts. This is no easy task and can be the difference in the quality of the data collected and the validity of the results. The second important question when working with experts is whether or not an individual's level of expertise can actually be objectively measured. Lastly, provided one has managed to identify true experts for the study, the difficult task of extracting or 'eliciting' the required information from an expert is no easy task and should not be underestimated. Such a process is extremely important when one considers that experts are prone to the same biases as normal individuals and that if care is not taken in accounting for them, these can translate into biased results.

2.9.1 Identifying Experts

An 'expert' was defined in sub-section 2.7.1 as an individual with advanced training and experience in a subject who is additionally recognised by his or her field for being so. Although such a definition is fairly easy to understand, the notion of what expertise actually is and more importantly how experts can be identified is rather more abstract and not so straightforward. Furthermore, is it possible to measure levels of expertise? And how then can experts be selected? These ideas are of course very

important in studies such as this one, which attempt to explore judgmental forecasting approaches, which centre on the use of expert knowledge. In this section, more thought will be given to these ideas as such a discussion will form the basis of the strategy adopted in this research to select experts, which will be covered in subsequent chapters.

2.9.1.1 How Can an Expert be Identified?

Several studies exist which attempt to define the properties that true experts exhibit and subsequently use these as a means for demarcating between experts and non-experts (Johnson, 1983; Shanteau, 1987). The focus here however will be on a more recent study of this kind, Shanteau et al. (2002).

The authors of this study present the dilemma of describing who and who is not an expert in a very simple way. The authors argue that if there is some external criterion (which they compare to a 'gold standard'), against which experts judgments could be compared, one could simply call everyone whose answers come within a certain range of this criterion, an expert. The problem however with such a 'validity-based approach' the authors argue is that paradoxically, experts are needed in the very situations where 'correct answers' or such an external criterion just does not exist, making such comparisons impossible. Moreover, if these answers were available, experts would not be needed in the first place. Quite, the identification of experts is no easy task and there is no full proof method available for doing so to date.

What is certain is that the identification of experts is more than just a case of labelling those who know the most about a topic, as an 'expert'. Although this may be one way of identification, in a review of research on the topic, Shanteau et al. (2002) present nine ways of identifying experts that have been suggested in the past.

Experience

This is one of the most classic approaches to identifying experts. The idea is that an individual with experience must be an expert in comparison to someone with little experience. Although this seems logical, it does not account for the fact that it is possible for someone to be working on a subject for a long time and never become an

expert; perhaps due to other reasons (e.g. behavioural) meaning that his or her performance does not reflect the years of experience. Moreover, as Shanteau et al. (2002) explains, there is no doubt that there are situations in which a positive correlation exists between experience and performance, but it would be dangerous to hold this universally true (Trumbo et al., 1962; Goldberg, 1968). In some instances, experience merely reflects seniority (Shanteau et al., 2002).

Certification

This is another classic way which is used to judge level of expertise, whether it is in industry, academia, the public sector, the military, etc. In all of these professions individuals receive official recognitions based on skill which then determines their level within the organisation. Unfortunately, these promotions are in part attributed because of experience and not just skill. This then brings back the problems of the previous approach. What's more, the other problem with certification is what Shanteau et al. (2002) calls the 'ratchet-up effect' in which an accreditation last forever. One never goes down on the scale. So, even if performance went down for any particular reason, one's title or job position would remain unchanged and so would be representative of performance level.

Social acclamation

Another fairly popular method involves 'experts' in a field being identified by their peers for being so. Such a process of social recognition involves a degree of agreement amongst professionals about who exactly should be labelled an expert.

Shanteau et al. (2002) believe that this is a fairly good method in the absence of any other means and pretty safe as it would be unlikely that several experts would name the same unqualified person as an expert. This approach however suffers two elementary flaws. The first, identified by Shanteau et al. (2002), is the 'popularity effect' by which someone who is better known within a social group (more popular) is more likely to be identified as an expert. Conversely, someone outside the peer group is less likely to be considered an expert despite being at the cutting edge of the field. Quite, it is commonplace for individuals proposing a shift in paradigm, causing what

Thomas Kuhn referred to as a 'scientific revolution' (Kuhn, 1962), to be unpopular figures at the time of the breakthrough. Secondly, there are often strong limitations to this kind of approach attributed to group dynamics. That is to say, the opinion of one popular professional in the field could influence the opinions of other professionals who would rather agree with this person instead of causing disruption by disagreeing. In this way, an unqualified individual could be labelled an expert through this strategy simply as a result of their acclamation from a select few of influential professionals, leading to a chain reaction of acceptance through the field.

Consistency (within) reliability

The fourth strategy of identifying an expert proposes to examine a person's consistency. This approach argues that for an individual to be an expert he or she must be consistent, just as inconsistency would be a trait of a non-expert (Einhorn, 1972, 1974). The problem with this according to Shanteau et al. (2002) is that an individual can exhibit consistency simply by following some basic, but incorrect, rule. Indeed, one could behave consistently but in an incorrect manner. This strategy will only stand a chance of being useful if an individual is behaving consistently correct.

Consensus

This approach argues that a necessary condition for expertise is the agreement between individuals on a certain subject (Einhorn, 1972, 1974). If there is any disagreement then this suggests that at least one, if not all, are not experts as they claim to be. Although initially this seems to be quite a logical property in experts, Shanteau et al. (2002) believe that the problem with such an approach is that 'agreement can result in premature closure' (Janis, 1972). In other words, a final solution could be incorrect despite being the consensus within a group because it was based on an initial incorrect assumption. But because consensus has been reached, the group will fail to consider better, alternative solutions. So although experts all agree, they could all be wrong.

Moreover, this strategy goes against Kuhn's idea of how new scientific knowledge is created, in which scientific revolutions occur when an individual challenges the ideas

of the current paradigm, thereby going against the consensus, causing a shift. Kuhn would strongly disagree with an answer being considered correct simply because consensus about it exists. So such an approach of identifying experts can similarly be thrown into dispute.

Discrimination ability

In this approach, it is argued that a defining property of experts when compared to non-experts is that the former are able to perform more rigorous analysis, perceive and act on subtle differences that non-experts might not pick up on (Hammond, 1996). Shanteau et al. (2002) believe that this ability is 'necessary, but not sufficient' in identifying experts because this ability might not always be of relevance or an advantage when having to give an answer in every situation.

Behavioural characteristics

This idea proposes that all experts exhibit common behavioural characteristics such as self-confidence, creativity, perceptiveness, communication skills, stress tolerance, etc (Abdolmohammadi and Shanteau, 1992). Using these characteristics, it is proposed that an ideal profile could be established for an expert. Tests could be developed to assess the suitability of prospective experts through comparison with the ideal profile.

According to Shanteau et al. (2002), this approach has three problems. Firstly, such tests do not exist for many of these characteristics. Second, even if they did they would have to be adapted to the domain being tested. Third and most importantly, the theory that only experts exhibit these traits is still uncertain. Although encouraging, before a strategy identifying experts using such an approach is used, it should be investigated further.

Knowledge tests

This approach posits that expertise can be verified through knowledge tests. In other words, an individual's knowledge, a premise for expertise, is examined through a test (Shanteau et al., 2002). The problem with this is that expertise is more than just knowledge. Knowledge is necessary but how this knowledge is used is just as important. An individual can be knowledgeable but if they cannot apply this

knowledge in a given situation, this knowledge is useless and they cannot be considered an expert.

Creation of experts

Finally, instead of proposing a way to identify experts, the last approach suggests that experts should be created through extensive training (Chase and Ericsson, 1981). Such a way would ensure the individual had the knowledge and was trained in how to behave in certain situations. The two fundamental problems with this is (1) it ever really possible to fully train someone, as expertise also involves experience and other characteristics argued to be ingrained in an individual's personality and (2) such a training process would be very lengthy and costly.

So in reviewing the different approaches for identifying experts from past research, Shanteau et al. (2002) have shown that identifying an expert goes beyond the knowledge an individual might have and transcends social as well as behavioural traits in humans. As has been shown however, these all suffer from one or multiple limitations.

2.9.1.2 Can Expertise be Objectively Measured?

An issue that is often considered in studies surrounding expert identification is can the level of expertise in an individual ever really be measured? As the review in Shanteau et al. (2002) is evidence, expertise is more than just knowledge and as a result is rather more intangible. Several of the approaches discussed above attempt to assign values to expertise level on an ordinal or even cardinal scale, with the exception of social acclamation, consensus and creation of experts. For example, *consistency* and *discrimination* attempt to give an estimate of these values to experts and use these to make comparisons between them. An expert with a higher estimated value in each will be considered 'more of an expert'.

Whatever the case may be, any individual attempt to quantify expertise by the approaches above would necessarily be biased, as only one dimension would be considered at any one time. The only possible logical solution would be to devise an approach combining all of the axioms of the methods described above into one

objective strategy. This is of course assuming that all of the axioms defining expertise are included in the above nine methods, which is possible, but unlikely.

With this in mind, Shanteau et al. (2002) propose their own approach which uses both the *discrimination* and *consistency* approaches. The authors propose to combine these two to form a ratio (discrimination estimate/inconsistency estimate), which they call the Cochran-Weiss-Shanteau (CWS) ratio (p. 258). An individual will be deemed more of an expert if his or her CWS ratio is high, indicating he or she discriminates consistently, which the authors believe true experts do. After testing their approach on past studies, the authors find that the CWS approach appears to work better than every other abovementioned approach for identifying experts. Nevertheless, because the approach is still very new the authors warn against using it as the sole approach and recommend rather using it as a complement to the other ones.

An important aspect which was overlooked by the authors in Shanteau et al. (2002) is the effect resource limitations can play on the process of identifying and selecting experts for a study. Time, money and practicality issues can play significant roles in constraining such an approach for a particular task. This was indeed the case for the studies involving experts conducted in this research as will be seen in later chapters when the methodology of these experiments is described.

In sum, the task of identifying experts for a task is a difficult exercise and by no means is there a single, infallible method for such a purpose. It has been seen that the difficulty lies largely in the complex nature of the notion of expertise and the resulting multitude of different characteristics that can be seen to classify an individual as an expert. In addition very little literature has been published on this topic of expertise measurement and aside from the CWS ratio, few methods have been proposed for such a purpose.

2.9.2 Elicitation of Expertise

In sub-section 2.7.1, it was explained that expert judgment is the result of complex thought-processing, known as knowledge-based cognition and that cognition refers to the mental exercise that takes place when an individual processes information. It was explained that when expert judgment is used in the face of a new or uncertain

situation, it is often inhibited by the cognitive and motivational limitations present in all humans. These limitations often manifest themselves as biases if careful consideration is not taken during the process of extracting the information, known as the elicitation procedure. These biases in judgment often translate into loss in quality so designing the elicitation method to minimise (if possible, or to account for when not) these two sources is an important challenge for decision-makers. In this section, greater focus will be given to this elicitation process with emphasis on the different available approaches as well as the different strategies of minimising bias.

2.9.2.1 Elicitation Process

Elicitation can be formally regarded as 'the process of gathering the expert judgment through specially designed methods of verbal or written communication' (Meyer and Booker, 2001). The authors structure the whole process into different, well defined stages. The main stages in the process are as follows.

Selecting and refining the questions: The first step involves defining the overall purpose of the project, the general question areas and then the more specific questions to be asked. Once this has been done, the cognitive limitations of individuals are taken into account and as a result the questions are then refined through a structuring process which allows experts to assimilate and process the questions more easily.

Selecting and motivating the experts: This step involves deciding upon the experts to be used in the study, depending on what kind of data is to be elicited. Furthermore, once these have been selected, these must be put in the appropriate frame of mind through motivational techniques.

Selecting the components of the elicitation: This is arguably the most important stage in the process and is where the specifications of the procedure are defined. The authors identify seven elicitation components which must be defined;

- **Situation** – which refers to the general methodology of the procedure and can be one of three possibilities. (1) Individual interviews; which consist of one on one collection process between the interviewer and the respondent. The approach

ensures in-depth elicitation without any distractions or detrimental effects caused by group dynamics. (2) Interactive groups; in which all the respondents are gathered in one place and a moderator chairs the meeting to a highly structured or completely unstructured standard of interaction. (3) Delphi situations; as described in chapter 1 of the literature review which involve a structured, multiple feedback rounds until consensus is reached.

- **Modes of communication** – The way the data is transferred between the moderator and the respondents, i.e. face to face, telephone or mail (or email).

- **Techniques** – The means for transferring the data, i.e. verbal report, verbal probe, ethnographic technique.

- **Response mode** – The format in which the answer is to be given, i.e. estimate of physical quantity, probability estimate, probability distribution, etc.

- **Dispersion measures** – These will be applicable of the respondents are asked to give an estimate of the variation or uncertainty in their answers and can involve ranges, percentiles, confidence intervals, etc.

- **Aggregation types** – The way in which the answers from multiple experts will be combined in the end to form one single, final answer. Approaches include behaviour and mathematical.

- **Documentation methods** – This is the final component to be defined and refers to what is to be recorded from the respondent, just the answer or the problem solving process used as well.

Designing and tailoring the elicitation. Once the components have been defined then these can be assimilated to form the elicitation design. This design is then tailored to the requirements of the situation. Particularly important aspects to be considered, amongst others, during this phase are logistics, costs and handling bias.

Having reviewed the main steps of the elicitation process, it is clear that much care and attention must be put into such an exercise to ensure that the information experts can provide is fully capitalised on. In chapter 5, the way in which this process was

undertaken for the studies conducted in this research will be examined closely. Details about the elicitation along with the rationale behind the decisions taken will be given.

2.9.2.2 Motivational Bias

In sub-section 2.2.3, the notion of bias in forecasting was introduced. It was explained that this bias can have a detrimental effect on forecast accuracy, particularly when judgmental approaches are used. The first kind of bias, termed *cognitive* was said to arise from the cognitive limitations present in all humans when having to process information, or expertise, and in sub-section 2.8.1 it was seen that a structured approach to judgmental methods was seen to minimise such a bias. The second source of bias was said to arise from the elicitation process of this expertise, and will be the subject of this sub-section.

Meyer and Booker (2001) argue that motivational bias is said to occur in one or more of three possible scenarios; firstly, when experts do not report their true solutions or thought processes because of external factors such as social pressure, group dynamics, wishful thinking, etc; secondly, when the data collector misinterprets the experts' answers; and thirdly, when the data collector misrepresents the experts' answers.

It is no surprise that these sources of bias degrade the quality of the data and hence must be monitored, or controlled as well as accounted for when interpreting results, so as to ensure credibility and validity.

Handling such cognitive and motivational biases is extremely difficult because they are so difficult to study and so ways of doing so are rare. Meyer and Booker (2001) for instance, describe them as much as an art as they are science. One six step approach which the authors suggest can help, but which is not foolproof, is,

(1) Anticipate any potential biases

(2) Reformulate the planned elicitation to account for the suspected biases

(3) Inform the experts of the potential incursion of the biases and familiarise them with the elicitation procedure

(4) Monitor the elicitation for the occurrence of bias

(5) Analyse the data for the occurrence of particular biases

Such an approach will be considered further in chapter 5, when the research design will be covered in detail.

In conclusion, this sub-section has shown the difficulties in identifying experts for a study. Past research has shown that such a process goes beyond the knowledge an individual might have and transcends social as well as behavioural traits in humans. These different approaches nonetheless suffer from one or multiple limitations. In addition, the concept of measuring expertise has been shown to be another area of difficulty. Shanteau et al. (2002) propose an approach but this is argued to be quite resource-intensive.

Furthermore, this sub-section has reviewed the main steps involved in the elicitation of expert knowledge. The design of such a process was found to be a very meticulous activity in which care must be taken in accounting for, and attempting to minimise the different possible sources of cognitive and motivation biases that experts are prone to.

2.10 Conclusions

After an extensive review of the pertinent literature needed to position the research presented in this thesis within existing knowledge, a few conclusions can be drawn.

- Judgmental forecasting can provide much positive insight into a forecasting situation where such information is considered of value, which can translate into improved accuracy. These methods are particularly useful when little quantitative data is available or when this data is unreliable. Furthermore, simple methods should not be overlooked because of their unsophisticated nature as they have been found to produce surprising results on several occasions. Finally, judgmental methods are attractive because they are relatively quick to use and inexpensive to implement.

- The position taken in this thesis is that the policy process can indeed be considered as a multi-staged, cyclical process. However, the stages serve principally as constructs for heuristic purposes, as in reality, the process is rather more complex.

In addition, it is believed that whilst governmental decision-makers strive to be rational, political forces limit this and the resulting model is more incrementalist in which decision-makers rational attempts are bounded by inherent information constraints which eventually lead them to make more 'garbage can' like decisions.

- The formulation stage of the policy process is when the decision is made as to what shape a policy will take as a solution to a public problem. Such a decision was said not only to involve defining the policy objectives but also the strategy chosen for implementation. For the latter, it was seen that policy instruments are considered as the primary means for such a purpose and hence are an area of interest.

- Instrument selection plays a fundamental role in shaping implementation strategy so it is unsurprising that policy research has shown an interest in the different theories behind instrument classification and selection.

- As for any decision-making process in policy, the rational selection of a policy instrument was seen to be constrained by political forces. Rational criteria such as efficiency or cost-effectiveness were seen to be used by decision-makers as bargaining tools during such a process.

- Evaluation was seen to play an important role after a policy has been implemented, as an ex-post policy learning exercise, but also before implementation, as a tool for ex-ante policy assessment.

- The recent PM doctrine has put extra emphasis on the ideas of evaluation and evidence-based policy making. Particular precedence has been given to evaluative measures such as efficiency and effectiveness to support rational decision-making in government.

- Ex-ante evaluation techniques are not only important and valuable exercises for assessing the impact of a policy but whose predictions are also used for the design and planning of the implementation process. They are forecasting approaches for making predictions about implementation at an operational level. In addition, there is also evidence that these approaches are also used for deciding implementation strategy, although to what extent, is still unclear. A further

exploration into the presence of formal forecasting approaches used specifically in decision-making for the purpose of instrument selection and implementation strategy revealed no evidence of such. Moreover, no evidence for the use of formal judgmental approaches, such as structured analogies, was found either.

- The presence of expertise in government exists primarily as an influential force in to aid in decision-making, throughout all stages of the policy cycle, particularly in scenarios where such expert knowledge is considered necessary or beneficial. No evidence was found of expertise being used in either a structured or formal way to make predictions.

- Past studies have shown that the use of analogies in forecasting has been shown to be particularly suitable in situations which are difficult to forecast, in which very limited past quantitative data is available, there is a high level of uncertainty and when the use of analogies is instinctive.

- Forecasting by analogies, as for any judgmental approach, suffers from the limitations and biases present in all individuals. A structured approach to the use of analogies has been shown to improve accuracy further.

In summary, the literature review has shown that forecasting practices in the form of evaluation techniques are used particularly for purposes such as anticipating the impact of a policy or as an aid in the planning and design of the implementation process. That is, their use is mainly seen as one for providing forecasts related to implementation on an operational level. Despite some evidence showing that these techniques are also used for implementation strategy this area remains unclear. Aside from these evaluative techniques, no evidence was found of the role that formal forecasting approaches play in decision-making for instrument selection and implementation strategy.

In addition, no evidence of the use of structured analogies was found for making such predictions despite it being seen that such an approach is suitable in situations of low availability of quantitative data, high uncertainty and where the use of analogous information comes naturally.

Chapter 3: Policy Implementation Strategy Effectiveness Forecasts

3.1 Introduction

The idea of policy instruments or tools, used by governments as implementation strategies, was explored in the literature review. Section 2.4 showed that these implementation tools were seen to be the means governments have for attaining the policy objectives. It was also seen that selection theory for choosing an instrument will ultimately define implementation strategy and that technical criteria such as effectiveness are used, among others, for such a decision-making process during policy formulation. However, the literature review revealed some evidence that ex-ante evaluative techniques also play a part in defining implementation strategy, at least at a UK level, but this area remains unclear. These results also revealed the lack of formal judgmental approaches, in particular structured analogies, for making such predictions.

In order to shed light on the strategic forecasting practices in government, it was decided to arrange an informal interview with an EU Policy Official. The results of this interview and the insight gained will be presented first in this chapter. Then, the notion of a Policy Implementation Strategy (PIS) will be presented. A discussion into why PIS effectiveness forecasts are suitable for supporting the decision of implementation strategy will follow. Finally, the last section of this chapter will present the research questions, based on the evidence presented, that this thesis will investigate.

3.2 Insight into the Strategic Forecasting Practices of the EU through an Interview with a Policy Official

One of the main results of the literature review was the evidence of forecasting processes in the form of evaluative exercises for the purpose of anticipating policy

impact or as an aid in the planning and design of the implementation process. In other words, forecasts were seen to be used for making decisions about implementation on more of an operational level. Despite evidence that these techniques are also used for shaping implementation strategy, this area remained unclear. Moreover, apart from these evaluative approaches, no evidence was found for the role that formal forecasting approaches play in instrument selection and decision-making for implementation strategy. Hence, this gap in knowledge led to the arrangement of an interview with an EU official in order to provide first-hand insight into this area at an EU level. The aim of this sub-section is hence to shed some light on this area as this will ultimately provide a foundation for the propositions made later on in the chapter.

The reason for the selection of an interview as a data collection method is that they are a flexible and adaptable way extracting the information required (Robson, 2002). Although such a data collection method would necessarily be based on information from a smaller sample of respondents than say, a questionnaire-based study, the greater depth of information associated with interviews (Denscombe, 2003) was considered a determining factor.

The interview process was semi-structured with the interviewee being allowed to speak freely around three questions.

- What are the steps involved in policy formulation?

- How is a solution decided upon?

- What role do forecasts play in deciding implementation strategy?

3.2.1 Summary of Interview Results

The main conclusions from the interview with an EU Policy Official are summarised in the points below.

- During the policy formulation stage, input is welcomed and encouraged from all stakeholders concerned with the policy proposal. The policy developer acts as an administrator, whose job it is to make sense of all this information. Naturally, each of these stakeholders proceeds with their own interests at mind and this is often evident in the solutions that they propose. Needless to say then, that each

interested parties' individual input can be considered biased. Furthermore, if there are any areas of incertitude with any particular proposed solution, it is common for experts to be brought in to investigate and report their opinions. Usual figures for the number of experts used vary between 5 and 10.

- The policy developer then has the task of integrating all of this information into a final solution to the problem, which naturally must try to please the stakeholders as best as possible. When agreement has been reached on a potential solution, an Impact Assessment (IA) is carried out to estimate the anticipated social, economic and environmental impacts the proposed policy will have. This IA is then attached to the policy proposal and all of this then makes its way to the other executive branches of the EU.

- No formal forecasts of the effectiveness of alternative solutions to the problem are made, and if they are, they are almost always done in an unstructured way. Such predictions are simply not used as a tool for supporting any decisions made at implementation strategy level.

3.2.2 Conclusions

The results of this interview show a lack of support for decision-making at the policy formulation stage at an EU level. Unlike what was found in sub-section 2.6.2, these results suggest that forecasts are not used when deciding on implementation strategy. Rather, these results suggest that their forecasting activities only start once the policy has been developed. The predictions about the effectiveness of the proposed policy made during the cost-benefit analysis of the IA are made to assess the impact of an already selected strategy. So such predictions are merely done to support the proposal which will aid in deciding whether or not the policy will be passed and put into effect. These results are in line with the garbage can model used to explain decision-making presented in sub-section 2.3.3 of the literature review, in that decisions are made and then justified post-hoc using technical analysis.

3.3 Policy Implementation Strategy

Many kinds of policy instruments are available to policy developers. They all have a common goal, that is to attain any objectives or targets set out by the policy. Policy Implementation Strategies (PIS) are essentially very similar to policy instruments in that they are used by governments to attain the objectives or targets set out by policy but are broader in their scope. A PIS adopts a broad view of implementation, which is argued to transcend formulation and decision-making to offer a more realistic view of the policy process. They are the strategy or plan of action that governmental decision-makers have chosen to implement a policy.

Conversely, not all policy instruments are Policy Implementation Strategies. PIS are more detailed than policy instruments as well as being more strategy orientated. Often, a PIS will include the specified parameters of the implementation strategy.

Although PIS can essentially refer to any policy instrument in any domain area, a particular type of PIS will be focused on in this thesis. That is, as defined in Savio and Nikolopoulos (2009a, 2009b, 2009c), the PIS considered in this research are policy instruments which promote the adoption of new forms of environmentally clean technology (ECT) through financial incentives (provided and funded by governments through tax rebates, credits, VED, scrappage schemes, etc). With such a definition, a PIS will include information such as the size of the incentive, qualification criteria, duration, etc. For illustrative purposes, an example of a policy with a possible PIS is given in table 3.1.

Policy	Combat the negative effects of climate change
Target set out by policy	Reduce household carbon emissions by 33% by 2020
Policy Implementation Strategy	Governmental initiative to promote the adoption of solar panels as main source of household power supply through subsidisation of panel cost as well as discounted installation

Table 3.1: Policy Implementation Strategy Illustration (Savio and Nikolopoulos, 2009b, p.89)

The above example shows a clear target set out by an overarching policy and a PIS based on the promotion of solar panels to contribute towards meeting this target.

Recent high profile policy implementation strategies have included the Vehicle Scrappage Scheme which instigated the renewal of the current fleet into a greener, more efficient one, and had dual objectives including economic and environmental. Such an initiative was implemented by several nations worldwide and will be used as a case study later on. Similarly, a Boiler Scrappage Scheme ended a short time ago in the UK promoting the purchase of a new grade A (highest efficiency level) by offering a discount on its purchase price if it replaced a working grade G (lowest efficiency level) boiler.

Policy implementation strategies are similar to the 'treasure' category of policy instruments defined by Howlett and Ramesh (2003) that was seen in sub-section 2.4.3 of the literature review, in that they use monetary incentives to attempt to change behaviour.

It must clear that these strategies are not 'policies' because they do not define what the policy goals are, these have already been established, nor are they strictly instruments, as they include a strategic element not found in instruments. Moreover, they transcend the artificial policy formulation and implementation constructs set out by the cyclical policy process model, offering a more realistic view of how such a process actually occurs, as illustrated in figure 3.1.

Policy Process

Agenda-Setting

Formulation

Decision-making

Policy Strategy and Objectives

Policy Implementation Strategy

Implementation

Evaluation
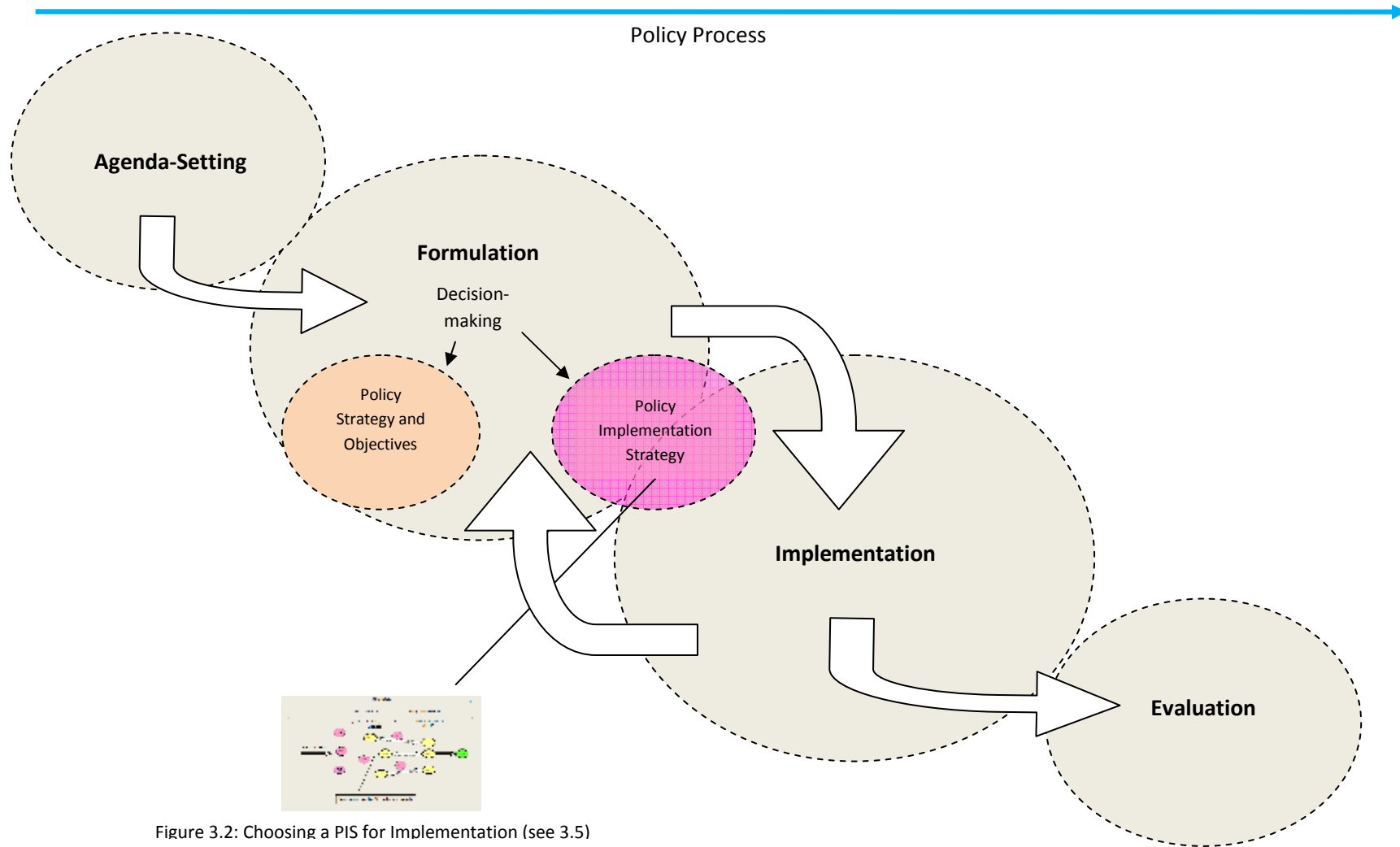
Figure 3.2: Choosing a PIS for Implementation (see 3.5)

Figure 3.1: Stages of the Policy Process and the Positioning of the PIS

Figure 3.1 shows that the positioning of the PIS is between the policy formulation process, once the policy has been decided upon and its objectives have been defined, and the implementation stage.

Clearly these strategies are important because they define how a policy will be delivered. Like for any implementation strategy, a successful PIS is a big step towards a successful policy. Moreover, what might have been a successful and effective policy can be ruined by a poorly designed and ineffective strategy.

3.4 PIS Effectiveness

In sub-section 2.4.1 and from the interview in section 3.2, it was shown that during policy formulation, policies are developed in part through input from various experts and stakeholders. At such a stage, all of the different implementation strategies and instruments possible for attaining the objectives set out by the policy are considered and a decision must be made as to which one to proceed with. Many alternative, rivalling policy implementation strategies with different characteristics, incentives, and costs, but all intended to arrive at the same goal, are considered. This idea is illustrated in table 3.2.

Policy Target	Reduce household carbon emissions by 33% by 2020	
		Cost (£)
	A:	
	- 45% return on insulation materials	L
	- 25% purchase price return on new generation efficient boilers	M
	- Differentiated council tax scheme with greater returns for greener households	N
Possible policy implementation strategies	Total	L+M+N
	B:	X
	- 80% return on insulation materials	
	- 35% purchase price return on new generation efficient boilers	Y
	Total	X+Y

Table 3.2: Alternative Policy Implementation Strategies (Savio and Nikolopoulos, 2009b, p.90)

The above example only shows two possible strategies under consideration but evidently there will usually be many more alternative strategies for the same policy.

Another of the results of literature review in chapter 2 was that there is evidence to suggest that ex-ante evaluative criteria, such as effectiveness, are used for deciding implementation strategy. What's more, sub-section 2.6.1 also showed that the PM doctrine has given precedence to such evaluative criteria such as effectiveness, among others, in an attempt to support rational decision-making in government (despite it being shown that in reality decisions follow more of a garbage can model). It may also be recalled from sub-section 2.4.5 that technical criteria such as effectiveness are

actually used when deciding between rivalling implementation tools. Then, in light of these results, it is reasonable to imagine that in the same way, effectiveness can be used as a criterion for assessing PIS and hence be used as a means for deciding which is to be implemented.

PIS effectiveness was defined in Savio and Nikolopoulos (2009b) as the extent by which the strategy moves the current situation towards the desired target set by the policy. Such a definition is illustrated with three examples, found in table 3.3.

Policy Objectives (Set Out During Policy Formulation)	Policy Implementation Strategy (PIS)	Measurement of PIS Effectiveness
Reduce average vehicle CO2 emissions by 2012 to 120g/km	PIS launching new type of eco-friendly engines (e.g. hybrid), with promotion of change over via incentives	Impact on Hybrid sales (in number of vehicles)
Reduce overall household energy consumption	PIS promoting the change-over/adoption of new energy efficient light bulbs though subsidised pricing	Change in percentage of new energy efficient light bulbs present in households
Reduce average vehicle CO2 emissions by 2012 to 120g/km	PIS offering a £2000 discount on the purchase price of a new car if a car of 10+ years is traded in for scrapping	Impact on sales of new cars (in number of vehicles)

Table 3.3: Examples of PISs and their Effectiveness Measurement (Savio and Nikolopoulos, 2009b, p.91)

The examples in table 3.3 show that such a definition allows for a quantitative measurement of effectiveness. Ekins et al. (2002) used a similar criterion for measuring effectiveness in a number of energy efficiency schemes of UK households.

3.5 Ideas and Research Questions

3.5.1 Ideas

It is known from the results of the literature review (section 2.6) that implementation strategy is decided upon using ex-ante evaluative techniques. The literature review (section 2.4) and section 3.4 showed that there are many possible strategies available

for attaining the same objectives and only one can go forward to be implemented. Section 3.4 also argued that effectiveness can be used as a criterion for such a decision. This leads to the first idea put forward by this thesis.

That is, this thesis proposes that PIS effectiveness predictions for the rivalling strategies be used as a decision-support tool for selecting potentially effective strategies for implementation.

In other words, these judgmental PIS effectiveness predictions could allow for a screening process of the alternative PIS proposed and the selected few could then be further analysed via the ex-ante evaluative approaches such as CBA or IA. Based on these results, it can be decided which PIS to implement. Such a decision-making process ensures that an extensive analysis of all possible PIS is made without the considerable investment of conducting this analysis with much more expensive approaches such as IA or CBA. This idea is illustrated in figure 3.2
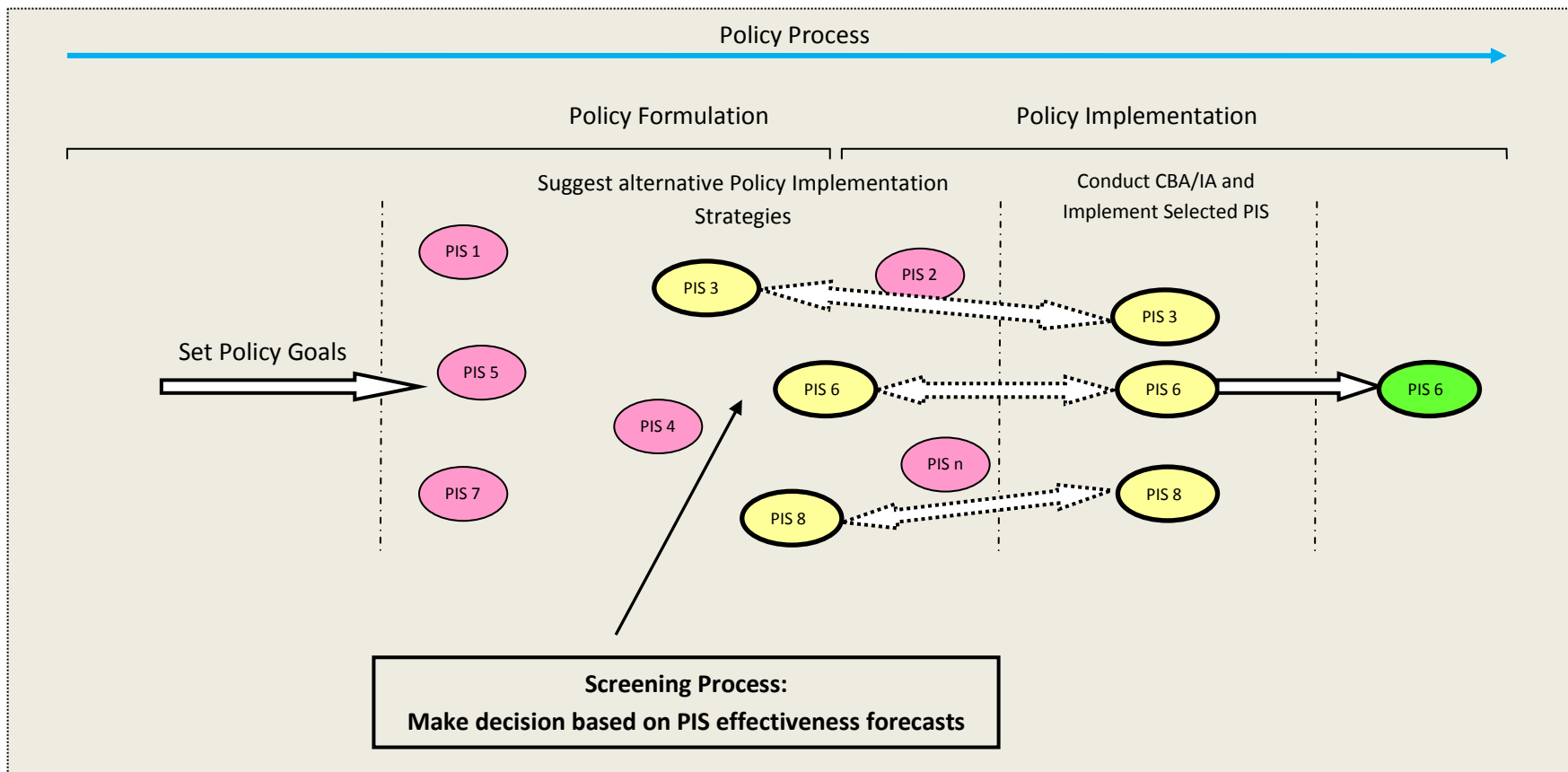
Figure 3.2: Choosing a PIS for Implementation

As can be seen in figure 3.2, the PIS effectiveness forecasts can be used to select among the rivalling PIS (pink). The most promising PIS according to the effectiveness predictions (yellow) are then further analysed using CBA/IA and based on this a final PIS is chosen for implementation (green). The double-headed arrows show that if the CBA/IA results are unfavourable for a PIS, it can be discarded and the screening process can restart.

Then, if this is combined with the claim that judgmental forecasting approaches are relatively quick and inexpensive to implement (sub-section 2.2.1), there is a lack judgmental procedures for forecasting in public policy (sub-section 2.7.1) and that judgmental approaches would be a welcome addition in government (section 3.2), then this leads to the second idea put forward by this thesis.

That is, it is proposed that a judgmental forecasting approach be used to produce the PIS effectiveness forecasts described above.

Two of the important attributes of judgmental forecasting approaches is that they are quick and inexpensive (Makridakis et al., 1998). So, such an approach would be a relatively quick and inexpensive way to shortlist potentially effective strategies prior to the investment of valuable time and money in assessing all competing strategies using the more resource-intensive ex-ante evaluative techniques. This, then places considerable importance on the accuracy of the PIS effectiveness forecasts as such a screening must be able to make sound predictions as to which PIS will actually be the most effective so that policy-makers are confident that the correct ones are being selected. The issue of the accuracy of the forecasts forms one of the main research questions of this thesis, RQ3, as will be seen later on.

Both intuition and evidence point towards the idea of analogies being useful in predicting PIS effectiveness. It is difficult to refute the logic that knowledge of the outcomes of similar, past cases of a certain type of PIS will not be useful in trying to predict the outcome of a new PIS of similar characteristics. One only has to look as far as the policy learning objective of evaluation exercises to see that the use of analogies is inherent in policy formulation. Evaluations of current policies are used to gain

insight about them and this insight can then be applied to the design of new future policies.

Sub-section 2.8.2 of the literature review provides evidence that analogies are indeed an attractive prospect for such a purpose for these very reasons. In particular, past research showed that forecasting by analogies was seen as a good solution to forecasting situations of low past quantitative data, high levels of uncertainty and when the use of analogous information comes instinctively, which in most cases forecasting PIS effectiveness is.

In addition, with cost minimisation clearly being a priority in any organisation including those in the public sector, the approach used for producing these PIS effectiveness forecasts must be inexpensive. Moreover, because of the evidence supporting the structured use of judgmental approaches to minimise subjectivity bias, a structured approach to the use of analogies, similar to the one proposed by Green and Armstrong (2007b), is preferred.

Therefore, the third idea put forward by this research is that the judgmental approach used for making PIS effectiveness predictions should be based on the use of structured analogies.

3.5.2 Research Questions

3.5.2.1 Research Question 1
In order to explore this third idea, a series of research questions were put forward for investigation. The first of these, which attempts to compare the accuracy of predictions made using the structured analogies method with those made using unaided judgment, is as follows.

RQ1.0: The structured use of analogies improves forecast accuracy compared to an unaided approach when predicting the effectiveness of a Policy Implementation Strategy.

There is not one unique way of structuring the use of analogies. Different techniques and levels of structuring are possible and really depend on the context of the forecasting situation, as the two studies reviewed in sub-section 2.8.3 of the literature

review are evidence. The level and shape of the support given to the forecaster should be based on the requirements of the forecasting exercise. In this thesis, two approaches to the structured use of analogies will be investigated, with differing levels of structure, leading to the second research question posed.

RQ1.1: The performance of the structured analogies approach improves in comparison to the unaided judgment approach when the level of structuring is increased.

3.5.2.2 Research Question 2

Bolger and Wright (1994) found that experts will only perform better than non-experts if the forecasting situations are 'ecologically valid' (the extent to which the individual is experienced at making such forecasts) and 'learnable' (the degree to which it is possible to master such a forecasting task in a useful way). If either or both are low, the authors argue, there is no reason to believe that experts will perform better than non-experts. In the case of forecasting PIS effectiveness, it is believed that the task is both ecologically valid (the experts have the necessary knowledge for such a task, collected on a daily basis in their professions) and learnable (such a task can be mastered through the study of existing and new strategies along with their levels of effectiveness). In this way, it is believed that the experts will perform better than non-experts when predicting PIS effectiveness, which leads to the RQ2.0.

RQ2.0: The higher the level of expertise of the respondent, the more accurate the forecasts will be.

However, it is also believed that within these experts, those with a greater level of experience will benefit more greatly from the use of a method like structured analogies and hence provide more accurate forecasts. So, as will be explained in section 5.9, an experience rating (which will be based on the responses to questions about experience and suitability for the task) will be given to each respondent. Such a rating does not, by any means, claim to be objective. Nonetheless, these ratings will help explore whether or not there is a correlation between such a figure and forecast accuracy. That is,

RQ2.1: The higher the experience rating, the more accurate the forecasts will be.

In addition, the work of Green and Armstrong (2007b) found that if several analogies were recalled by the respondent, the better his or her predictions would be. It can be argued that the ability of recalling several analogies is an indication of a respondent's level of expertise, although expertise can also be argued be more than that (see sub-section 2.9.1). This leads to RQ2.2,

RQ2.2: The greater the number of analogies recalled, the more accurate the forecasts will be.

3.5.2.3 Research Question 3

Finally, in order for such PIS effectiveness forecasts made using a structured analogies approach to be a reliable screening method for potential PIS, policy-makers must be confident in their ability. In order to test the accuracy of the structured analogies forecasts, these were to be compared against the forecasts found from the governments who implemented the PIS. It is believed that despite being much quicker, less expensive and less sophisticated than the approaches used by governments, the expert predictions will be no worse. Hence, the final research question explore in this thesis is

RQ3: The PIS effectiveness forecasts generated by the experts will be no worse than those produced by the government.

3.6 Conclusions

Hence, this chapter has proposed that PIS are a more realistic approach to how implementation strategy is formulated in government. Like other decisions which define implementation strategy, technical criteria such as effectiveness can be used for PIS. Hence PIS effectiveness forecasts are proposed as a decision-support tool which will allow for a screening of all possible PIS to identify the most promising, which can be submitted to a more rigorous analysis through ex-ante evaluative techniques such as IA and CBA. Furthermore, a structured analogies approach is proposed as the method for producing such PIS effectiveness forecasts because it is quick and

inexpensive to implement and is suited to the nature of such a forecasting task. This chapter also presented the research questions associated with the use of structured analogies for such a purpose that will be investigated in this thesis.

Chapter 4: Methodology

4.1 Introduction

In order to understand any methodology behind a piece of scientific research (whether it is natural or social), it is useful to understand the philosophical positioning behind the investigation. That is to say, any decisions made regarding the methodology or research process of the study will be better understood if the positioning, amongst the multitude of different philosophies for conducting scientific research, is clear. Thus, the position of this research within the different philosophies of social science will be presented and explained. Based on this positioning, attention will then be given to the chosen research strategy and method that make up the research process of this thesis.

4.2 Philosophical Positioning

Regardless of which position is taken within those in social science research, it is important that any research conducted be scientific. This idea of the 'scientific method' and the exact form it should take is one that causes considerable dispute amongst the rivalling views on how research should be conducted in a social science arena. Robson (2002) argues that in order to seek the 'truth' about whatever subject, research should be carried out 'systematically, sceptically and ethically';

- *systematically* means giving serious thought to what you are doing, and how and why you are doing it; in particular, being explicit about the nature of the observations that are made, the circumstances in which they are made and the role you take in making them;

- *sceptically* means subjecting your ideas to possible disconfirmation, and also subjecting your observations and conclusions to scrutiny;

- *ethically* means that you follow a code of conduct for the research which ensures that the interests and concerns of those taking part in, or possibly affected by, the research are safeguarded.

(Robson, 2002, p. 18)

This definition is not too far off from the traditional scientific method used in the natural sciences which says

The principles and procedures for the systematic pursuit of knowledge involving the recognition and formulation of a problem, the collection of data through observation and experiment, and the formulation and testing of hypotheses.

(Merriam-Webster, 2009)

This definition is the cornerstone of the traditional positivist philosophy of science (Robson, 2002). Positivists always try to find links between events through this scientific method in order to build theories within the subject, something which simply does not occur in the social sciences, which is one of the more important critiques of such an approach for such a purpose (Robson, 2002, see Blaikie, 1993, p.101 for details of other critiques).

Post-positivism, which as its name suggests is derived from positivism, on the other hand, is much more suited to social research. This view is the result of positivists taking on the criticism laid out by rivalling philosophies to somewhat 'loosen' their views on scientific research (Robson, 2002).

One important change is that although like in positivism, the researcher and the research subject are considered independent of each other; post-positivists believe that observation is theory-laden. In other words, the subjective characteristics of the researcher, such as his theories, hypotheses, background knowledge and values can ultimately have an effect on the outcome of an experiment, or what is observed. Hence, despite committing to the positivist belief of a sole objective truth, post-positivists acknowledge that such a truth can be known only in an imperfect way due to the limitations of the researcher.

Another view to social research is critical realism. Critical realists believe that the outcome of an action depends largely on the context in which is applied. This means that in the critical realist way of conducting experiments, any results that are found are not considered as facts beyond dispute; they merely provide evidence which if found again through replication, can lead to the generalisation of a theory.

This idea leads to an important conjecture amongst critical realists which says that there is no indisputable foundation for science (Robson, 2002). In other words, any theory is open to dispute and it is the task of science to not only build the potential theory for explaining the real world, but also to rigorously test these theories by rational criteria.

As has been seen, post-positivism and critical realism have some clear differences in some aspects but importantly, they share the view that knowledge is objective and it is through science that this objective knowledge can be built. Thus, the position taken by this research will be one situated between the post-positivistic and critical realist approaches.

The reason for this is straightforward. It is believed that common to both schools of thought, there exists an objective reality in that expert forecasts will or will not be improved when these are given structured support in their use of analogies. But it is also acknowledged that the results of any study undertaken to explore this truth will be affected by the theories, hypotheses, background knowledge and values of the author, a premise of the post-positivistic view. These very same factors will also define the context in which any such study will take place and thus in maintaining with the critical realist view, will only be found to hold true in. Finally, as postulated by the critical realist approach, whatever the results of this study, they will be considered as either positive or negative evidence towards the verification of the proposed theory. In other words, regardless of the outcome of this investigation, it will be considered as the first step towards verifying if expert forecasts are indeed improved when such individuals are given structured support in the use of their analogies. Such an investigation must be replicated and verified before any sort of general theory can be confidently established. This of course, also being the case for the other research questions put forward.

4.3 Method of Scientific Reasoning

The method of scientific reasoning adopted in this research can be summarised in figure 4.1 with RQ1.0 being used as an illustration.

Premises

- The use of analogies improves expert forecast accuracy in contexts of low past quantitative data and high levels of uncertainty, comparable to that of forecasting PIS effectiveness.

- The structured use of analogies reduces the effect of cognitive biases, which improves forecast accuracy.

Generalisation

There is evidence to suggest that either the structured use of analogies _does_ or _does not_ improve forecast accuracy when predicting PIS effectiveness.

Deduction

Induction

RQ1.0: The structured use of analogies improves forecast accuracy when predicting the effectiveness of a policy implementation strategy.

Figure 4.1: Method of Scientific Reasoning in this Research with RQ1.0 as an example.

In the example above, RQ1.0 is deduced from the premises in the top-left box and then based on the results of the experiments to test RQ1.0, inferences are made about general theories through induction.

Thus, the research questions of this research have been proposed through a deductive process of existing theories in forecasting research. Then, through experimentation, these research questions will be explored. Based on these results, inferences will be made about general theories in forecasting though inductive logic.

4.4 Research Process

Three key ingredients are required for any good piece of social research;

1. Construction of theory

2. Collection of data

3. Design of methods for gathering data

(Gilbert, 2001)

124

4.4.1 Construction of Theory

In order to test a theory then, as Gilbert (2001) argues, it must be well constructed. For this to be the case there must be a good way of measuring the truth that is being explored, known as the indicator. In this research, the indicator variable will be the errors[1] of the experts' forecasts. By analysing the errors of expert's forecasts, it will be possible to analyse the accuracy of their predictions as required by the three research questions presented in the previous section.

An indicator variable will only be considered as valid if it is actually good at representing and measuring the concept in question. Similarly, an indicator variable will only be considered reliable if it is consistent from one measurement to the next. In the case of this research, it is believed that the indicator variable chosen is indeed both valid and reliable but is only constrained as far as the contextual and theory-laden limitations held by the post-positivist and critical realist philosophies of social science discussed above.

Moreover, measurement theory, which is said to link the truth under investigation with observable facts, determines the validity and reliability of the indicator variable (Gilbert, 2001). What's more, Gilbert (2001) argues that a good way of justifying measurement theory is by the use and reference to previous research which has employed the same principle. In the case of this research, the measurement theory says that the performance of experts (the observable facts), measured as the errors of their forecasts (the indicator variable), when predicting PIS effectiveness in controlled experiments will be a good representation of their ability to predict PIS effectiveness (the truth under investigation). This same measurement theory is the standard practice in forecasting research and so it is considered appropriate in these circumstances (Armstrong, 2001a).

4.4.2 Research Strategy

The strategy adopted in this research to explore the research questions set out in section 3.5 will be of a flexible design. It will be composed of lab-based experiments

---

[1] Although several different error measures will be used, see section 6.2.

and experiments conducted via alternative platforms. The reason for this mixed design is straightforward in that the type of experiment depended on the accessibility and availability of participants. In addition, the design of the experiments were modelled on the multitude of past experiments, designed to investigate very similar research questions (but in other contexts), published in forecasting research.

All experiments were based on exploring the above research questions for differing levels of expertise. By running experiments in which half the participants would be asked to produce a forecast using an analogies approach in a structured way and half being asked to use unaided judgment, the accuracy of the two approaches could be compared and RQ1.0 could be explored. Then, by using structured analogies approaches with different levels of structuring (one for experiments I, II and III and another for experiments IV and V), RQ1.1 could be investigated.

RQ2.0 could then be explored by comparing the accuracy of forecasts made by experts and non-experts. In addition, comparison of the s-SA forecasts with the experience score of the respondent as per table 5.4 in section 5.9, RQ2.1 could be investigated. Finally, by comparing the s-SA forecast accuracy with the number of analogies recalled, RQ2.2 could be explored.

As for RQ3, a comparison would be made between expert forecasts and the same predictions made by government, which are known. In this way, any results obtained from these experiments could be used as evidence to support a potential generalisation of any of the theories put forward in these three research questions via inductive logic.

The idea behind each experiment was to mimic the situation during the policy formulation stage which required a prediction of the effectiveness of a given PIS. Thus, case studies were prepared to simulate one of the many alternative implementation strategies proposed at such a stage.

4.4.3 Research Method

The selection of data collection methods also depended very much on the conditions of the experiment. Whether the data was collected through questionnaires,

interviews or email-based questionnaires depended on the accessibility and availability of participants. Details of which was applied in each experiment is given in sub-section 4.4.4.

Common to all experiments however was not only the strategy as described above but also the design. Each participant was presented with a case study, which was taken from a real life PIS. This would include a brief description of the proposed strategy, which would be disguised to avoid recognition, followed by a questionnaire asking for a PIS effectiveness prediction. For those participants required to use the structured analogies approach, an analogy table would be provided after the strategy description which they were asked to complete before moving on to the questionnaire.

Since it is also this analogy table that would specify the level of structure to be used in the structured analogies approach, different tables were used for experiments I, II and III and experiments IV and V. For the analogy table in experiments I, II and III, respondents were asked to recall, describe, rate similarity (to the target strategy) and provide the outcome of any analogies they might know. In this way, the analogy based approach was similar, but not identical to the one developed in Green and Armstrong (2007b). Instead of asking respondents to match analogy outcomes with target outcomes, they were simply asked to give an outcome if they could. The reason for this was that it was uncertain how well or easily analogies of implementation strategies and their outcomes could be recalled.

As a result of the nature of the analogy table, it was not possible to derive a structured analogies (SA) forecast using a mechanical rule in the way that Green and Armstrong (2007b) did. This approach was thus labelled as semi-structured analogies (s-SA), to differentiate itself from the original SA approach developed by Green and Armstrong (2007b). To differentiate between the two s-SA approaches used for the different experiments, the s-SA approach used in experiments I, II and III will be called s-SA1 and as will be seen next, the approach used in experiments IV and V will be labelled s-SA2.

Hence, in slight contrast to the original SA approach presented in sub-section 2.8.3 by Green and Armstrong (2007b), the steps of the s-SA1 are as per table 4.1.

Step 1	Description of target situation
Step 2	Analogy recall through the analogy table which asks for descriptions and similarity ratings
Step 3	Forecast PIS effectiveness

Table 4.1: s-SA1 Approach Used for Experiment I, II and III

Although the s-SA1 approach does not permit the administrator to derive a final objective prediction based on the analogies provided and their similarities, the rationale behind it is that through the use of the analogy table, respondents will be aided in framing the situation and structured support for recalling and utilising analogous information is provided, thus minimising cognitive biases. This idea is illustrated with in figure 4.2.
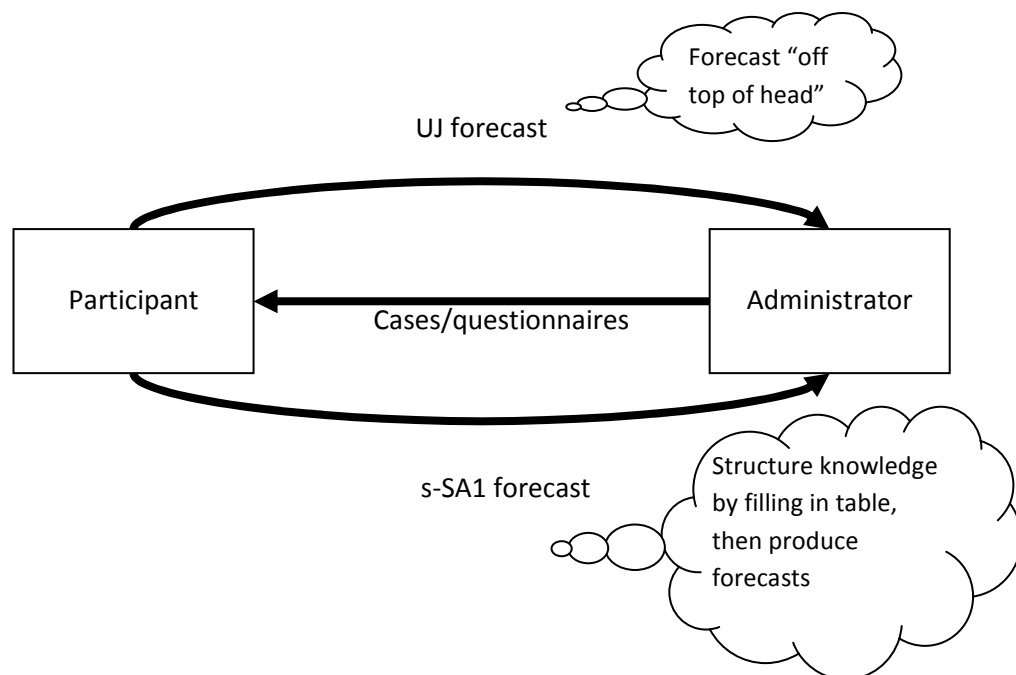
Figure 4.2: Soft Systems View of the Data Collection Process (Savio and Nikolopoulos, 2009c, p.89).

After the third experiment, feedback from respondents revealed that more support was needed when recalling the outcomes of the analogies. For this reason, it was decided that for the final experiments, a further level of structure would be added to the structured analogies approach, giving s-SA2.

In s-SA2, the possibility of matching analogy outcomes with target outcomes would be given in an attempt to provide such support. Hence, in addition to the original point PIS effectiveness forecast, the possibility of giving a PIS effectiveness interval prediction by selecting of one of the available interval outcomes was given. The rationale behind this was that it would provide insight into whether participants would 'obey' their analogies and choose the outcome interval that these suggested. Thus, the s-SA2 approach used in experiments IV and V was as per table 4.2:

Step 1	Description of target situation
Step 2	Analogy recall through the analogy table which asks for descriptions and similarity ratings
Step 3	Match analogy outcomes with target outcomes
Step 4	Forecast PIS effectiveness (interval)
Step 5	Forecast PIS effectiveness (point)

Table 4.2: s-SA2 Approach Used for Experiment IV and V

Although producing forecasts using the original SA approach would now be possible, it was not done so. The reason for this is that a purely SA method could only possibly result in an interval forecast being produced (if for example the mechanical rule used would define the SA forecast to be the outcome of the analogy which was rated most similar to the target case). When deciding which of several alternative implementation strategies should be implemented, a point forecast of effectiveness would be more precise, and of more use to decision-makers, than an interval prediction. For this reason, it was decided to carry on with the s-SA2 method, which would result in a point forecast. The details of the s-SA2 approach as well as the analogy section and all of these case studies, including how they were prepared, will be covered in chapter 5.

A summary of the different experiments along with the structured analogies approach and cases used is found in table 4.3

Structured Analogies Method		Experiment	Cases	
s-SA1	s-SA2		Original Three	Vehicle Scrappage Schemes
X		I (Non-Experts)	X	
X		II (semi Experts)	X	
X		III (Non-Experts)		X
	X	IV (Non-Experts)		X
	X	V (Experts)		X

Table 4.3: Summary of Experiments and s-SA Approach and Cases Used

Once the data of the experiments had been collected, they were analysed in the view of the research questions put forward without the use of statistical significance testing. The reason for this is two-fold.

- Evidence in the literature showed that the typical number of experts involved during the implementation strategy decision-making process during the policy formulation stage is between 5 and 10. Furthermore, considering the time and availability constraints and low response rates for studies such as these ones involving experts, it is anticipated that it will be difficult to obtain more than 10 responses for the final experiment (see below and section 5.9). Such a number would be insufficient for any sort of statistical significance test.

- Armstrong (2007) argues that over a review of past studies reporting statistical significance tests in their results, no evidence was found to support the use of such methods under any conditions. It is argued that such tests are often incorrectly interpreted and important results of smaller studies, failing to attain statistical

significance are discarded for this reason. For this reason, Armstrong (2007) argues that statistical significance tests should never be used when reporting scientific results and so practitioners should ignore them and journals discourage them. Rather, Armstrong (2007) believes that focus should be placed, amongst other criteria, on effect size of the results (e.g. an estimate of practical significance), which is what will be done in this research.

4.4.4 Overview of Experimental Procedure

In order to obtain a general view of the experimental process, an outline of the logical progression of the experiments undertaken for investigating the research questions will be given.

Experiment I

The first experiment was intended as a study to test the proposed s-SA1 approach in the hands of non-experts but also to test the elicitation procedure ahead of future experiments involving real experts. The experiment involved the use of 31 Manchester Business School students with training in forecasting techniques. They were each presented with three case studies in the space of two hours and asked to make predictions according to their assigned approach. Once participants had completed the case studies individually, the experiment was repeated with the participants working in small groups of three or four. Experiment I will be covered in more detail in section 5.3.

Experiment II

The second experiment was a further study but which involved participants from the Environment Directorate of a large European governmental organisation. The participants were chosen by a contact within the organisation and were all policy developers, spanning six different nationalities and different levels of experience, within this department. The design of the experiment was similar to the first, with the participants being presented with the same three cases, with the same three questionnaires and being asked to make the same three effectiveness predictions. However, due to availability constraints, the data was collected through one-to-one

interviews with each participant of about 1 hour. In addition, due to time and availability constraints, it was not possible to have the participants work in groups as was done in the first experiment. The details of this experiment are covered in section 5.4.

Experiment III

With the then launch of the UK Vehicle Scrappage Scheme, it was decided that using such scrappage schemes as case studies would be topical and interesting. Such schemes are good, popular examples of PIS with clear policy objectives and well-defined incentive schemes which have been implemented in different ways across the globe, considered ideal for testing the new approach. Hence, three new cases were designed using three real life vehicle scrappage schemes in an identical structure to the first three, with a disguised description, an analogy table (for those using the analogy-based approach) and a questionnaire asking for an effectiveness prediction.

With these new cases, another experiment was run with non-experts. However, the aim of the experiment was to test the appropriateness of the new cases through feedback from the respondents to ensure their quality, ahead of experiments IV and V. In terms of set-up and design, the third experiment was very similar to the first, using trained forecasting Manchester Business School students. Unlike the first experiment however, the participants were not asked to produce forecasts in groups. It was decided that because such results could not be collected from experts due to time and availability constraints, it was better to focus on testing the approach on an individual basis and leave the group predictions for further research.

At the end of the session, time was taken to gain feedback from the participants about the cases. This feedback was used to refine the cases in preparation for the two final experiments, IV and V. Details of the new cases, experiment III and the refining of the new cases can be found in sections 5.5, 5.6 and 5.7 respectively.

Experiment IV

One of the main results from the refining of the cases was the added support to respondents when having to assess the outcome of their analogies. Thus, in the

analogy table, along with all the previous requirements, respondents were now being asked to match the outcomes of their analogies with five possible, categorical, target outcomes (s-SA2).

Such a modification would theoretically allow for an SA forecast to be produced. Although the SA method as proposed by Green and Armstrong (2007b) was deemed inapplicable when having to predict PIS effectiveness, it was thought of interest to see how such a method performed. Naturally, because respondents gave outcomes of their analogies in the intervals provided, any such SA forecast would be an interval forecast, not a point one (which is not much use when considering that to make a decision between alternative strategies, more precision is required, hence the need for a point forecast). More detail regarding the use of the SA method will be given in section 5.7.

With the refined cases ready, experiment IV was the final one to test the s-SA2 approach in the hands of non-experts. Once again, a new set of Manchester Business School trained forecasting students were used. The design of the experiment was similar to that of experiments I and III with the important difference that participants were now given one week to complete the questionnaire. The idea behind this experiment was to provide a benchmark for the s-SA2 performance against which the performance of the experts using s-SA2 in experiment V could be compared. More details on the new analogy section of the cases and the experiment IV can be found in sections 5.7 and 5.8.

Experiment V

In the final experiment, the s-SA2 approach was evaluated in the hands of experts. Particular individuals in various sectors such as Academia, Government, Non-Profit Organisations (NPOs), Media, and Consultancies who through their job title, were considered to be experts on environmental policy and vehicle scrappage schemes if possible, were targeted. Through an extensive correspondence procedure, over 90 individuals were contacted, whom on request, were sent the three case studies (the very same as those used in experiment IV) and given one week to return these

completed. Only 10 provided valid results. Full details of this experiment can be found in section 5.9.

4.5 Conclusions

To recap, the philosophical position taken to this research is situated between the post-positivistic and the critical realist approaches to social research. The reason for this is that it is believed that there exists an objective reality, common to both schools of thought, that the expert forecasts will or will not be improved when these are given structured support in their use of analogies. However, it is also believed that any results will be theory-laden, which is a post-positivist belief. In addition, as postulated by the critical realist approach, whatever the results of this study, they will be considered as either positive or negative evidence towards the verification of the proposed theory. In other words, the outcome of this study will be considered as the first step in verifying the theories behind the research questions proposed and that such an investigation will have to replicated and verified before a general theory can be confidently established.

In view of this, the research strategy adopted for exploring the research questions set out in sub-section 3.5.2 will be a flexible design. It will consist of lab-based experiments and experiments conducted via alternative platforms. The reason for this design was partly determined by the accessibility and availability of participants. Further validation for the strategy adopted can be found in the considerable body of forecasting research, in which experiments of this kind are readily used. The data collection methods depended largely on the conditions and constraints of the experiment but these varied from questionnaires, interviews and email-based questionnaires.

In sum, the empirical research for investigating the research questions set out in sub-section 3.5.2 consisted of five experiments carried out under different conditions, depending on the requirements. Experiments I and IV tested the approach in the hands of non-experts. Experiments II and V tested the approach in the hands of experts. In addition to generating results for the performance of the proposed

approach, experiments I to III provided useful feedback which was used to optimise the design of experiments IV and V.

Chapter 5: Testing the Approach

Having established the research questions to be investigated by this research and the general methodology by which these propositions were to be explored, details from each of the experiments will now be given.

5.1 The Case Studies

In order to test the structured analogies approach for predicting PIS effectiveness, it was decided that the best and most practical way of doing this was to provide the participant with a short but comprehensive description of a real life strategy intended for policy implementation and ask the participant to produce a forecast as to its anticipated effectiveness, as explained in sub-section 4.4.4 Details on how these descriptions were written will be included in section 5.2 (for the cases used in experiments I and II) and section 5.5 (for the cases used in experiments III, IV and V). The idea however is to simulate one of the possible alternatives that are proposed as PIS during formulation. This chapter will provide the rationale for the selection of the case studies chosen for all the experiments undertaken throughout the research.

What was needed then were cases suitable for such kinds of experiments. In order for a case to be deemed suitable, three conditions would have to be met, they were

1) the policy proposed led to the development and implementation of a governmental PIS aimed at hitting a policy target
2) a *forecast* of the effectiveness of this PIS was produced during the ex-ante cost/benefit analysis (or other)
3) a measure of the *actual* PIS effectiveness is available and accessible

As illustrated by the diagram in figure 5.1 below

Policy
Implementation

Time
X Years

Forecasted PIS
effectiveness

PIS
implementation

Actual effectiveness of
PIS
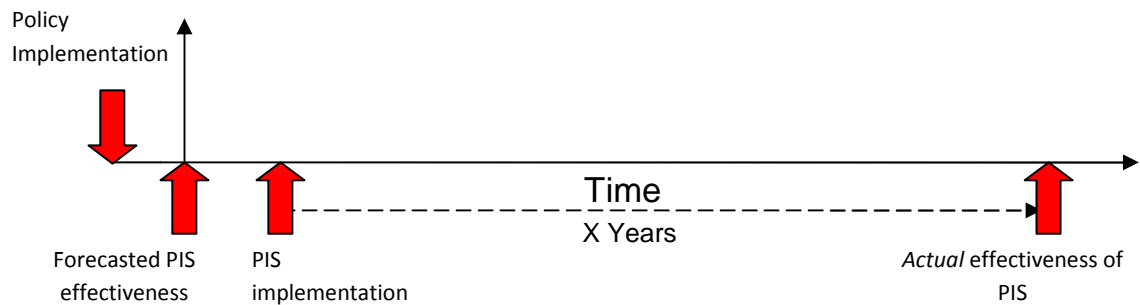
Figure 5.1: Evaluation PIS (Savio and Nikolopoulos, 2009b, p.93)

As will be seen, all the cases used were taken from different global contexts.

5.1.1 The Original Three

For experiments I and II, the three case studies used were the United States' *New Eco-Friendly Technology Adoption*, the European Union's *New CO$_2$ Differentiated Tax Scheme* and the Greek *Get Digital*. The first two cases were selected because of their political importance and to compare differing strategies for similar objectives in two World powers. The third case was chosen to offer variety in the experiment as it is a more local, low key and less complex strategy compared to the others.

New Eco-Friendly Technology Adoption

This first case is based on the 'Alternative Motor Vehicle Credit' scheme which is a result of the 'Energy Policy Act of 2005' in the United States. The scheme entailed a tax credit on any new qualifying hybrid vehicle purchased on or after the 1st January 2006 starting at a maximum of 15% of the original price, depending on its fuel economy. However, after the sale of 60,000 units of any particular make and model, the % of the price returned phases out over the next 5 quarters (i.e. once 60,000 are sold with a ± 15% return, the % returned gradually decreases over time to 0%). The IRS website explains

Taxpayers may claim the full amount of the allowable credit up to the end of the first calendar quarter after the quarter in which the manufacturer records its sale of the 60,000th hybrid passenger automobile or light truck or advance lean burn technology motor vehicle. For the second and third calendar quarters after the quarter in which the 60,000th vehicle is sold, taxpayers may claim 50 percent of the credit. For the fourth and fifth calendar quarters, taxpayers may claim 25 percent of the credit. No credit is allowed after the fifth quarter.

(IRS, 2010)

New CO_2 Differentiated Tax Scheme

This scheme was based on the Community Strategy in the EU to reduce CO_2 emissions from passenger cars and improve fuel economy. The strategy is based on three pillars (voluntary commitments from manufacturers, improvement in consumer information and promotion of 'greener' vehicle types via fiscal measures) implemented in 1995 to reduce such emissions to the required level by 2007. The first two pillars are expected to reduce CO_2 emissions the majority of the way. The last pillar, which consists of a tax scheme which differentiates between vehicle types based on energy efficiency and favours the "greenest" ones, has the aim of reducing CO_2 emissions the rest of the way.

Get Digital

The third case is based on a governmental scheme in Greece which aimed at supporting the use of computers by students. The initiative consisted of subsidising the purchase of new laptops by first year university students. Any such student purchasing a computer would be entitled to an 80% refund with a maximum fixed at 400 €.

5.1.2 The Vehicle Scrappage Schemes

For experiments III, IV and V, it was decided to focus on one particular type of PIS, with both economic and environmental objectives, the vehicle scrappage scheme. Such was chosen because of its popularity (present in over 16 countries around the world), and to compare different strategies across the world for the same objective. These schemes are good examples of PIS because they attempt to promote a change in behaviour through financial incentives. It is up to the policy developers to specify the parameters of the scheme, the size of the incentive, the duration of the scheme, qualification criteria, etc. The case studies are as follows.

Operation Green

This is based on the 'Retire Your Ride' campaign in Canada and is a scheme with an environmental objective intended to encourage owners of inefficient vehicles to get them off the road and offer a reward in exchange for doing so. Funded by the Canadian Government and implemented by the Clean Air Foundation (Summerhillgroup.ca, 2009a) the program is intended to run from February 2009 to February 2013.

To be eligible for a reward, vehicles must be at least fourteen years old, in running condition, registered and insured for the last 6 months prior to application. Rewards for such surrender include a public transit pass or a membership to a car-pooling scheme, £200 (approximately), or a discount on the purchase of a vehicle made in the last five years.

Forecast: It is unclear if the prediction has come from the Canadian Government or the Clean Air Foundation but Summerhillgroup.ca (2009b) reports a prediction of 0.729% (29, 166 of 4M) of old, polluting vehicles to be taken off the road in the first seven months. Neither the Canadian Government nor the Clean Air Foundation report how this figure was generated.

Actual Effectiveness: After seven months, the number of vehicles scrapped recorded was 15,000 of 4M (0.375%) (Retireyourride.ca, 2009).

Reducing Average CO_2 Emissions

This strategy was the scrappage scheme from France, called 'La Prime a la Casse'. The proposed strategy is designed to encourage owners of dated, inefficient vehicles to purchase newer, more efficient vehicles via a rebate incentive. The strategy promises a reward of £950 when a vehicle emitting less than 160g of CO_2/km is purchased and a (functioning) vehicle of more than 10 years is handed in for destruction. This strategy was due to commence in December 2008 and last for twelve months.

Forecast: The French Government estimated that 11.6% (104,000 of 900,000) of all new vehicles registered in the first five months would have done so through the

strategy (Lexpansion.com, 2009). Details on exactly how this prediction was arrived at were not found.

Actual Effectiveness: After five months, the number of new vehicles registered through the scheme was 175,000 of 900,000 (19.5%) (Lexpansion.com, 2009).

Killing Two Birds with One Stone

This is based on the 'Cars Allowance Rebate System' (or more commonly known as the 'Cash for Clunkers' initiative) in the United States, which had dual objectives; environmental and economic. The strategy was implemented to not only improve the fuel efficiency of the current fleet but also to inject a much needed financial boost into the failing automotive sector. This strategy rewarded the purchase of a new vehicle (with a retail price of less than $45,000) when an older vehicle (of less than 25 years) was traded in. The size of this reward, which would come in the form of credit towards the purchase of the new vehicle, ranged from $3,500 to $4,500 depending on the difference in efficiency between the vehicle traded in and the one purchased. The overall available budget for such a strategy was $3 billion. The incentive offered by the strategy commenced 1st July 2009 and lasted till 1st November 2009 (or until the budget was exhausted).

Forecast: The American Government estimated that 33.3% of the ($1 Billion of the $3 Billion budget) would be spent in 4 months (NHTSA, 2009). As was explained in sub-section 1.6.5, no insight into how these figures were produced was provided.

Actual Effectiveness: After 4 months, 2,800M of the 3,000M budget had been consumed (93%) (Wikipedia.org, 2009).

5.2 Preparing the Cases

Once it had been decided which strategies were to be used as case studies for the experiments, these had to be appropriately converted into a format for the respondents that would ensure the quality and validity of the data collected. The chosen format was a verbal report with a case description and questionnaire section. Particular attention was paid to the suggested elicitation guidelines from Ayyub (2001)

and Meyer and Booker (2001) when designing the cases. The steps involved in the design of these case studies and their questionnaires are described in this chapter.

It is important to note that a determining factor in the case study design was the severe time constraints imposed on such a data collection process. For non-experts taking part, (which were university students) the whole exercise had to fit into a two-hour slot, including breaks and inevitable delays. For the experts willing to partake, whose time is even more stringent, the duration of each case study had to minimised as much as possible to avoid frustration and loss of interest on their part.

Hence each case study was designed to take about 15-20 minutes. Furthermore, because these case studies are meant to represent one of many possible alternatives for an implementation strategy, proposed at the policy formulation stage, the decision as to which one to adopt must not be delayed too much and so these predictions are required fairly quickly.

The three cases, all followed a similar design. They consisted of two main sections: (1) case description and (2) questionnaire. Only for respondents being asked to use the structured analogies method, an analogy table[1] was provided between the two sections.

At the top of each case, instructions for the participants were clearly laid out. These were

1/ Read the description and

2/ try to think of several analogous situations and

3/ about how similar your analogies are to the case.

4/ Fill-in the questionnaire

For respondents using UJ, steps 2 and 3 were omitted in their documents. Clear instructions such as these were considered an important measure towards avoiding confusion as to what is required from the respondent, something which is vital in reducing motivational biases.

---

[1] A table which allowed respondents to list describe and rate similarity of analogies.

Case Description

For each of the three cases, the formulation of the PIS description was carried out carefully. It was reasoned that enough information had to be given to allow an informed participant to be able to recall analogies from his long term memory (LTM), which he or she could use to benefit the forecast. However, not too much information could be given as it would run the risk of the strategy being identifiable to the expert, making his or her prediction ultimately biased, and hence invalid.

For each case, details about the policy context, the objective and details of the incentive were given (including its format and duration, where applicable). Bearing in mind that due to the serious time constraints meaning each case could only really take 15-20 minutes, such a description had to be fairly concise and easily digestible.

The different measures used to mask the three cases used in experiments I and II (as presented in section 5.1) when it was thought that revealing this information would make identification too obvious, included:

- Omitting the names of the nations in question and labelling them merely with an X.

- Not giving specific details of the technology being proposed.

- Omitting dates and using letters to denote a year and numbers to represent elapsed time

Details of the exact descriptions given can be found in Appendix 1.

Analogy Table

An analogy table was provided after the description of the PIS for respondents who were assigned the structured analogies approach. Such a table (figure 5.2) provided a structured means for the use of analogies by guiding the respondents through the recollection, description, analysis, similarity rating and outcome of analogous strategies.

The support given to experts in the use of analogies is the cornerstone of the structured analogies approach developed by Green and Armstrong (2007b). As was

seen in the literature review, the authors argued that experts will be limited by their cognitive abilities which can result in biases, so a 'structured' approach must be taken in order minimise such a risk. It is precisely for this reason then that the analogy table is provided; as an aid to allow the respondent to structure his or her analogous information, which is thought to improve recollection and processing.

It was seen in sub-section 4.4.3 that for the first three experiments, a semi-Structured Analogies (s-SA1) approach was used, wherein respondents were aided in recalling, describing, rating the similarity and recalling the outcomes of the analogies to the target strategy. However, these outcomes, if provided, were not to be used to derive a forecast as in the SA method, hence the prefix 'semi'. The reason for this that it was not known how well respondents would be able to recall the outcomes of analogies, as this was considered more difficult than recalling their nature and rating their similarity.

(A)	(i) Description	(ii) Source	(iii) Similarities and Differences	(B) Similarity Rating	(C) Outcome
a.					
b.					
c.					
d.					
e.					
e.g.	New technology promoted by Gov. to combat climate change	Media	Similarities: new technology introduced for improvement purposes Differences: Different context, not necessarily for directly reducing CO2	8	Reduced waste (any kind) by 24%

Figure 5.2: Analogy Table Given to Respondents for Experiments I, II and III

Questionnaire

The case studies all ended with the same questionnaire that had the important role of allowing the respondent to generate a prediction for PIS effectiveness, which

incidentally was the first question. Depending on each particular case, a different type of forecast was asked for. For all three however, the forecast required was an estimate of the strategy's effectiveness, as per defined in the case.

Case 1

The sales of the new technology were given for the six years leading up to the date of the strategy's implementation and respondents are asked to forecast the sales of the first three years of the PIS (in other words the impact of the PIS on sales).

Case 2

Respondents were told that the PIS was implemented in year Y and were asked to predict the percentage decrease in average CO_2 over a 13 year period.

Case 3

Respondents were simply required to predict the percentage of students taking up the government's offer in the next academic year.

The other questions in the questionnaire (figure 5.3) were aimed at circumstantial aspects of the exercise such as the time spent on the task; the likeliness to change the prediction if given more time: if the case was recognised, the number of individuals spoken to about the task and then two questions about the respondents experience with issues similar to the ones presented in the descriptions.

2) Roughly, how long did you spend on this task?

{include the time you spent reading the description and instructions}   [____] mins

3) How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) ... 10=practically certain (99/100)}   [__] 0–10

4) Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have as a environmental issues management specialist?

[ ] years

7) Please rate your experience (out of 10) with cases similar to this one   [____] 0–10

Figure 5.3: Extract from the Questionnaire for Cases 1, 2 and 3

Questions 2 and 3 were intended to provide feedback on whether enough time was being given to complete the task. This information could eventually be used to refine the cases. Question 4 was intended to identify if any participant recognised the case, in which case the forecast would be biased because the true outcome would be known. In such an event, the results would be considered void. Questions 6 and 7 were aimed at providing feedback on the respondents' level of expertise in the field as well as experience with cases such as the one described. As was mentioned in the methodology, it was this feedback that led to the refining of the case design for experiment IV and V. The exact format of the original three cases as was presented to participants in experiments I and II can be found in Appendix 1.

5.3 Experiment I

The first experiment was intended as a study to not only test the proposed approach in the hands of non-experts, but to test the elicitation procedure in light of future experiments involving real experts.

A total of 31 students from the Manchester Business School were selected as participants in the study. All the participants were Undergraduate students who had

nearly completed a 'forecasting and applications' module and so were already familiar with judgmental forecasting approaches such as unaided judgment and structured analogies. In order to minimise the main source of motivational bias from respondents, it was thought that the best way of doing this would be to offer an attractive incentive for the students. This was the prospect of being able to apply the methods they had learnt in addition to the more tangible incentive of extra credit on their final module grade for the top three forecasters.

The participants were to be presented with the first three cases as discussed in the previous section and asked to predict the effectiveness of the PIS using either of the two methods already discussed, Unaided Judgment (UJ) or the semi-Structured Analogies (s-SA1) approach (depending on their assigned group number), individually and then in groups.

Participants were split into two random groups, 1 and 2.

Group 1 size: n = 15

Group 2 size: n = 16

Each of these two groups was then split into a further three groups (1a, 1b, 1c and 2a, 2b, 2c) of 3/4 participants in each for the group exercises.

With the participants split into groups 1 and 2, every participant was individually presented with case 1. Participants in group 1 were asked to use UJ and participants in group 2 were asked to use s-SA1 to produce individual forecasts of PIS effectiveness. Participants were given 15 minutes to complete the task. Such a time span was set in accordance with the time constraint of the double period in which the students were available for the exercise. What's more, this time span was deemed as sufficiently long enough for the participant to understand the case and generate the forecasts.

Participants were then asked to get into their subgroups and each was presented with case 1 again. Subgroups 1a, 1b and 1c were asked to produce group forecasts using UJ and subgroups 2a, 2b and 2c were asked to produce group forecasts using s-SA1. Again, 15 minutes was given for this task. This process was repeated for case 2 and case 3, as

shown in table 5.1. Ten minute breaks were given (for coffee, toilet, etc) between changing cases.

Group	Case 1	Case 2	Case 3
1	UJ	UJ	UJ
2	s-SA1	s-SA1	s-SA1
1a, 1b, 1c	UJ	UJ	UJ
2a, 2b, 2c	s-SA1	s-SA1	s-SA1

Table 5.1: Treatments for Experiment I

5.4 Experiment II

An experiment was carried to test the performance of the s-SA1 approach in the hands of individuals whom it was not known what their level of expertise would be, albeit that it would be greater than that of students. The experiment was seen as a good opportunity to perform a study of what would be another experiment in the future, with real experts. As the results will show, these individuals showed a mid-level of expertise and so were labelled 'semi-experts'.

Participants were from the Environment Directorate of a large European governmental organisation. Most participants (if not all) were policy developers in this department. Participants were chosen non-randomly and based purely on vicinity by a contact in the department. Participants spanned six different European nationalities, and had different specialisation areas and levels of experience working in the department.

The data was collected through individual interviews, in which participants were presented, one at a time, the same three cases used in experiment I. For every case, once the participant was confident of the instructions, they were left to complete the questionnaire. A maximum of 20 minutes was given for each case.

Regarding the reduction of motivational biases, this was not as straightforward as with the students in experiment I, as to find a true incentive was more complex. It was decided in the end that the best that could be offered in terms of an enticement would

be the promise of a report with the outcomes of the experiment and any interesting results or conclusions, as soon as this was available.

A total of eight participants were interviewed, four produced forecasts using UJ and four using s-SA1.

5.5 Preparing the New Cases

After having conducted experiments I and II with the same three case studies, it was decided to change these to vary the kinds of implementation strategies that the approach was tested with. As was explained in section 5.1, vehicle scrappage schemes were considered ideal as case studies because of their clear policy objectives and well-defined incentive schemes for achieving them. The rationale behind the design of these new case studies and questionnaires will be covered in this section.

Due to the same time constraints affecting the first two experiments, that is the limited availability of participants, each case was designed to last roughly 15-20 minutes. Similarly, the new cases followed the same design as the previous cases: that is with a case description section and then a questionnaire section with an analogy table provided in between for respondents using the s-SA1 method. The same instructions were given in the new cases as for the original ones.

Once again the three descriptions of the strategies were masked in order to avoid them being recognised by respondents. Techniques for such a purpose were similar to the ones previously used and included

- Omitting the names of the nations in question and labelling them merely with an X.

- Not giving specific details of the technology being proposed.

- Omitting dates and using letters to denote a year and numbers to represent elapsed time

- Changing local currency into Pounds Sterling.

Regarding the questionnaire section, this was structured in the same way as the previous cases and included the same questions. As was previously the case, the first

question required respondents to give a point forecast. The effectiveness forecast question for each of three cases is as follows.

Case 1

For the first case, respondents were asked to predict what number of these old, polluting vehicles (of the estimated 4 million vehicles currently on the roads which are eligible for the scheme) will be surrendered in the first 7 months of the strategy.

Case 2

In the second case, a forecast of the percentage of all new vehicles registered in the first 5 months that would have been done so through this governmental incentive was required.

Case 3

Finally, the third case asked for a prediction as to the percentage of the total budget (which was stated to be £1.8 Billion) that will actually be used by people taking up on the offer during the initial 4 month period.

The exact format of the new cases as was presented to participants in experiment III can be found in Appendix 1.

5.6 Experiment III

With the new cases designed, it was decided to run an experiment with non-experts, but trained in forecasting techniques, to test the suitability of the case study design to make sure of its quality, in preparation for experiments IV and V. Hence the aim of this experiment was not so much to further validate the s-SA1 approach when compared to the UJ approach in the hands of non-experts but rather to gain feedback on the design of the experiment to ensure its quality for the two final experiments and consequently the validity of their results.

The third experiment was set up in a similar fashion to the first experiment but with somewhat reduced numbers. The participants were all third year undergraduate

students from the Manchester Business School near to finalising a forecasting course so they were familiar with such techniques.

In a similar way to the first experiment, the class was split into two groups and each assigned one of the two treatments. Reducing motivational biases was, as in experiment I, relatively a straightforward case of offering the incentive of extra credit on the final module grade to the top three forecasters, to ensure the exercise was taken seriously.

Participants were split into two random groups, 1 and 2.

Group 1 size: n = 10

Group 2 size: n = 11

With the participants split into groups 1 and 2, every participant was individually presented with case 1. Participants in group 1 were asked to use UJ and participants in group 2 were asked to use s-SA to produce individual forecasts of PIS effectiveness. Participants were given 20 minutes to complete the task. Such a time span was set in accordance with the time constraint of the double period in which the students were available for the exercise. What's more, this time span was deemed as sufficiently long enough for the participant to understand the case and generate the forecasts.

This process was repeated for case 2 and case 3, as shown in table 5.2. 10 minute breaks were given (for coffee, toilet, etc) between changing cases.

Group	Case 1	Case 2	Case 3
1	UJ	UJ	UJ
2	s-SA	s-SA	s-SA

Table 5.2: Treatments for Experiment III

At the end of the experiment, a 10 minute discussion in which feedback about the exercise could be given was held. The main points brought up by the respondents were

- More detail of the figures had to be given in the description, in order to give the respondent more of a framework of the size of the strategy concerned.

- Re-calling and quantifying the outcomes of analogies was too difficult and further support was required for this.

5.7 Refining the Cases

In light of the feedback from the respondents in experiment III, it was decided that the cases be modified or refined such that more information be given in the description of the strategies for respondents to work with, but also, more support in the handling of the analogies and their outcomes.

The changes and their rationale are summarised in the following points.

Change: For all three cases, standardise what quantitative information is given in the description of each one. Figures to be given: Intended time span of program, monetary value of incentive and overall budget available.

Rationale: In light of feedback from respondents, it was decided that the key parameters for providing such a forecast is the knowledge of the duration of scheme, the money available for the scheme and the economic significance of the incentive. These are regarded as minimum requirements for the task. Extra information about the nature of the PIS will permit the participant with superior expertise in the area to exert this knowledge and improve forecast accuracy.

Change: Improved layout of 'analogies' section.

Rationale: The old layout was not very tidy and somewhat intimidating for participants. Instead of having to fit so much information into a small space which was difficult in the previous vertical format, the new horizontal format was much clearer, as shown in figure 5.4 below.

Figure 5.4: Extract from the New Analogy Table

Change: Addition of categorical outcomes for participants to choose from when making a forecast.

Rationale: It must be remembered that the aim is not to simulate how these forecasts are made in a real life situation because this is next to impossible due to time and other constraints, etc. Rather, the objective is to show that the use of analogies in a structured way improves forecast accuracy when compared to unaided judgment. Hence, providing categorical outcomes, even though these would not occur in real life, is a positive step in this idea of structuring. The new layout of the forecast section is as per figure 5.5.

```
Outcomes
Please select one of the following as your prediction of the percentage of old vehicles surrendered
within the first 7 months because of the strategy:

A                               0% − 20%                              [  ]
B                               21% − 40%                             [  ]
C                               41% − 60%                             [  ]
D                               61% - 80%                             [  ]
E                               81% - 100%                            [  ]


Within your chosen range, please provide a point forecast for the percentage of old vehicles
surrendered within the first 7 months

                                   [      ]
```

Figure 5.5: Extract from the New Forecast Section

An anticipated difficulty in the use of the s-SA1 approach is that respondents have trouble capitalising on the information stemming from their analogies. Hence, participants were now asked, just like Green and Armstrong (2007b), to choose the target outcome (A to E in the figure above) which is most similar to the outcome of each analogy. This would force the participant to link the outcomes of the analogies to the outcome of the target, something which the participant had probably struggled to do in previous experiments. In addition, the participants were asked, after having completed the new analogy table, to choose one of the target outcomes as their final forecast.

This change in analogy support now meant that a purely SA forecast could be obtained, in a similar fashion to the way that was done by the developers of the approach, Green and Armstrong (2007b). That is by using the analogies recalled, the similarity ratings and the outcome intervals to objectively derive a forecast. When participants recalled just one analogy, the SA forecast was fairly trivial. When more than one analogy was recalled, a mechanical rule was decided upon, which would be applied to every participant. As argued by Green and Armstrong, there can be many acceptable mechanical rules, but like the authors, the one adopted here is: the SA forecast will be the outcome of the analogy with the highest similarity rating. In the event of a tie, the analogy for which the respondent has provided the most detail is chosen. The reason

for this being that predictive validity should increase with similarity rating (Green and Armstrong, 2007b).

As explained in sub-section 4.4.3, the usefulness of such interval effectiveness forecasts for decision-making during the formulation stage is questionable and for this reason it was decided that although SA predictions could be derived, it was the point forecasts, and so the s-SA approach, which would continue to be used. Nonetheless, for the sake of completeness, it was decided to investigate the results of the forecasts produced by the SA method. The reasons for this were primarily to see how well the SA approach performs, even if it can only give an interval forecast, but also if there is any consistency between the SA prediction and the outcome interval prediction made by the respondents.

Change: Addition of question asking participant to give a precise numerical forecast, developing on the outcome forecast.

Rationale: For the sake of greater precision in the response.

Change: Addition of two questions, as shaded in figure 5.6 below. One aimed at allowing the respondent to rate his or her suitability for making such predictions and the other aimed at providing insight on forecasting process from that participant's point of view.

7) Roughly, please rate (out of 10)

- your experience with environmental public policy. [____] 0–10

- your experience with cases similar to this one. [____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[____] 0–10

8) If you were contracted to producing such a strategy effectiveness forecast, what approach/process would you adopt? In what sort of time-scale?

Figure 5.6: The Two New Questions

Rationale: The first new question would permit an exploration into whether there is correlation between the respondents' self-belief for such a task and the accuracy of his

or her forecasts. The second new question would allow the collection of new information on the actual forecasting process used in practice as well as some insight on the effect of cultural background on such a task.

The modifications of the new cases as these were presented to participants in experiments IV and V are presented in Appendix 1.

5.8 Experiment IV

With the cases refined, it was decided to conduct two final experiments, with non-experts (experiment IV) and with experts (experiment V). This section is concerned with the design and set up of experiment IV and the next section will deal with that of the fifth experiment.

For this final experiment with non-experts, respondents were found in a postgraduate course of trained forecasters. In other words, whilst being considered non-experts in terms of their knowledge of environmental policy, these participants were familiar with the forecasting techniques they were asked to employ.

Due to the fact that the participants were not students of the authors at the time, unlike previous experiments with students, a different incentive was offered to reduce motivational biases. Rather than offering the incentive of extra credit, a reward of £15 for the best forecaster, £10 for the second best and £5 for the third best was offered. Such a financial reward was deemed an appropriate incentive to ensure sound effort from the respondents.

Again, because the respondents were not students of the author, there was no class time available for the data to be collected like in previous experiments. Rather, permission was requested to speak to the class after a scheduled lecture. The experiment was explained to them and the cases distributed. Half of the class was given s-SA2 and half UJ, and these were selected at random. There was hence no possibility of doing group forecasts on this occasion.

In contrast with previous experiments, respondents were given one week to complete the questionnaires and produce a forecast. The reason for this being is that although

15-20 minutes was deemed enough (as is evidenced from the results of previous experiments), it would be of no harm to let respondents analyse the situation with time and rather, in reality, government officials would spend longer than 20 minutes on such a task.

Hence, the following week, after the same scheduled lecture, the cases were collected. Once the results had been analysed, the three winners were called to collect their rewards.

5.9 Experiment V

In this section, the details on the design and set up of the final experiment, which tested the s-SA2 approach in the hands of experts for the refined cases will be given.

The idea behind this final experiment would be to try and find as many 'true' experts as possible and have them produce forecasts for the three case studies, half using the s-SA2 approach and half using the UJ approach.

Design

The majority of the expert identification approaches suggested by Shanteau et al. (2002) as discussed in sub-section 2.9.1, were not applicable within the time and resource constrains imposed on this research. Personal behavioural characteristics of the sampling frame of the participants such as 'consistency (within) reliability' or 'discrimination ability' just were not available so such techniques had to be discarded. Social characteristics of participants, such as 'social acclamation' or 'consensus' from the field were also unfeasible and so were discarded. There was also no time to conduct 'knowledge tests' prior to accepting that an individual partake in the study.

It must be acknowledged that one has to be very pragmatic when considering the commitment of professionals, as the time investment that they will be prepared to make (if any) will be very limited. Thus, the quickest and most reliable expert identification technique which was considered to best suit the constraints in this final experiment was that of 'certification', in other words, job title.

In order to encourage validity of results, it was decided that as broad a range as possible would be sought. Organisations within the following sectors were considered Academia, Government, Non-Profit Organisations (NPOs), Media, and Consultancies. Within these sectors, a search was carried out to find appropriate professions suitable for such a task and it was found that participants would be contacted in the following organisations, as presented in table 5.3.

	Academia	Government	NPOs	Media	Consultancies
Organisations	Tyndall Centre, IARC, ICEPT, Cambridge University	Department of Energy and Climate Change, Environment Agency, IPCC, DEFRA, DfT, Office for Climate Change, IPPR, Department of Business, Innovation and Skill, BERR	The Climate Group, CarPlus, Low Carbon Vehicles Partnership, Carbon Trust, Greenpeace, Energy Saving Trust, UKERC, Campaign for Better Transport, SMMT	Cash for Clunkers Webpage, The Times Online, The Guardian, The Observer	E4Tech, CE DELFT

Table 5.3: List of Different Organisations Contacted for the Study by Sector.

The rationale for each of the above sectors chosen is as follows.

Academia – These are the 'thinkers' in the environmental policy field and responsible for the majority of the research done on these kinds of policies. The academic research centres contacted all specialise in environmental policy research. They have close ties with government as well as other organisations responsible for the design and implementation of such policies. Hence, individuals within these research centres are likely to have experience with the kinds of implementation strategies being assessed.

Government – These are the 'doers' in the environmental policy field and responsible for actually designing these policies and implementing them. Each department is responsible for the development and implementation of its own policies, although sometimes the latter is outsourced to external agencies on competitive tender. The governmental departments and agencies listed either focus on environmental policy or have groups within them that do (DBIS, BERR). Thus, individuals working within these bodies will certainly have experience with implementation strategies of this kind.

Non-Profit Organisations (NPOs) – These organisations often work between governments, research centres and other interested bodies. Those listed above have as an objective (amongst others) to reduce carbon emissions from road transport, including the average emissions of a national fleet and so will have experience with policies aimed at doing so.

Media – Naturally the media, in particular newspapers with large environmental sectors will have journalists within their ranks who specialise in environmental policy. There will be journalists who will have covered many stories about these kinds of strategies and their effectiveness, gaining a significant level of expertise in the process.

The organisations chosen within each sector were done so based on their track-record as being either highly-regarded within the sector or particularly relevant to the environment policy field. Organisations with a history of knowledge on policy implementation strategies such as vehicle scrappage schemes were targeted. These organisations would then be contacted as a whole or through individual contacts, depending on the situation.

Data Collection

Information for contacting the above organisations was found through researching the web. In some cases, organisations would disclose details about their personnel so direct contact could be made with the individuals deemed most suitable. In other instances, such detailed information was not available and only a general enquiries email or telephone was disclosed. Either way, the data collection process can be summarised with the flowchart in figure 5.7.
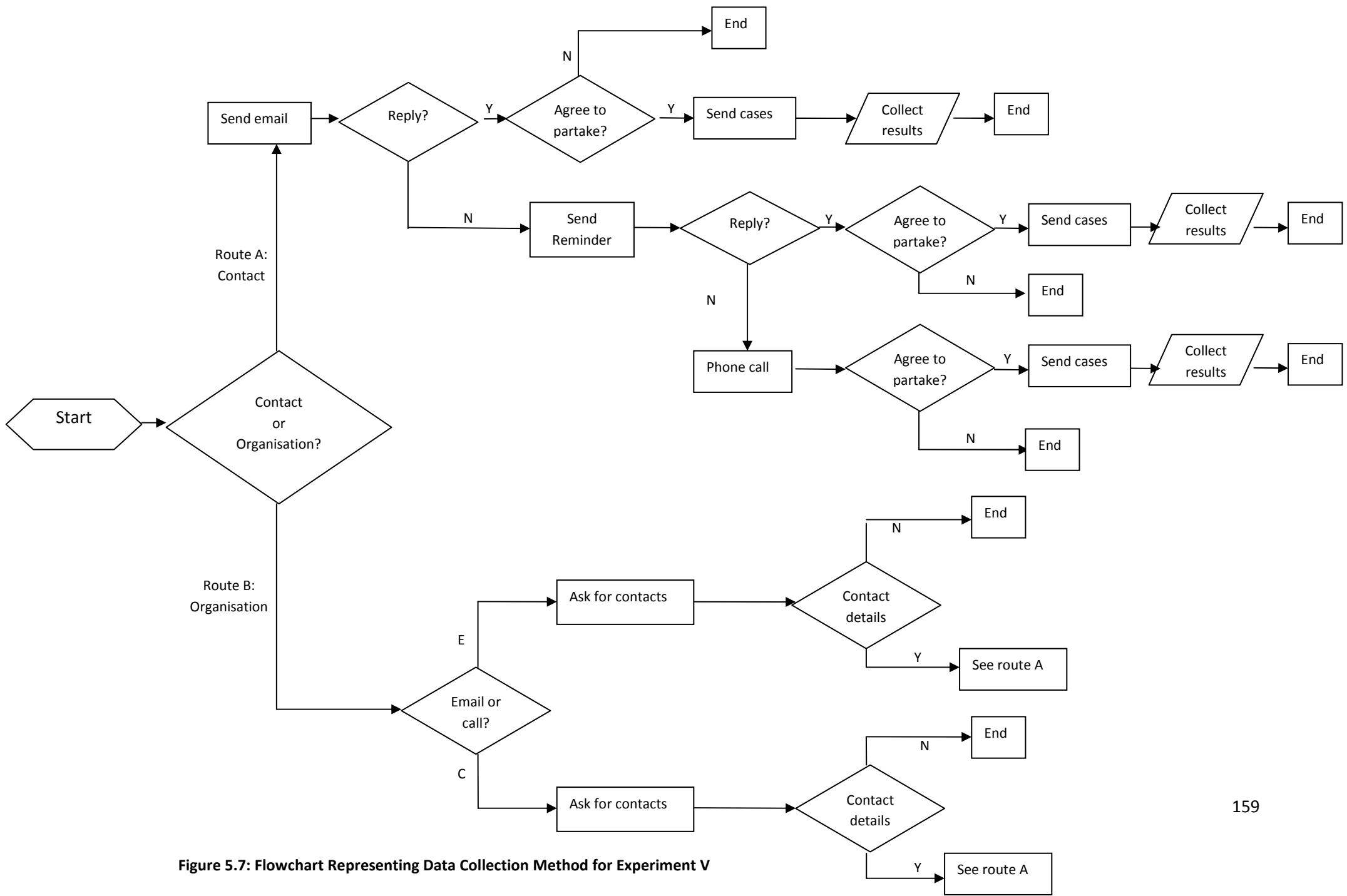
Figure 5.7: Flowchart Representing Data Collection Method for Experiment V

159

Route A

So, if the direct contact details of a suitable individual could be found on the web, route A was taken. An email would then be sent to the individual and given approximately two weeks to respond. If a reply came within two weeks and the individual was willing to participate, the cases were sent out and these were collected within a week's time. If an individual replied negatively, the lead was discarded.

If the individual failed to reply after two weeks, a reminder email was sent and one more week was given for a reply. If a reply was received within this week, either the individual agreed to participate and were sent the cases or they did not and the lead discarded.

If no reply had been received after this extra week after the reminder, a phone call was made to the individual. At this point, the individual either agreed to participate, and was sent the cases, or they did not and the lead was discarded.

Route B

Alternatively, if the personal details of a suitable individual in the organisation could not be retrieved, either an enquiries number or email was always available. The process for either was the same. Whether it was through an email or over the phone, the experiment was explained and details of potentially suitable individuals were requested. If these were provided, the process re-commenced in route A and if not, the lead was discarded.

The results of the literature review presented in chapter 2 suggested that for such a task, usually 5-10 experts are used in the British and EU Governments. For this reason, once 10 sets of results were obtained from 10 experts, the experiment was stopped. Before the 10th was collected, over 90 individuals across all sectors had been contacted.

One of the research questions (RQ2.1) set out by this research was to investigate whether the level of experience of an individual was somehow correlated to forecast accuracy. For the purpose of such an investigation, a rating of experience would be calculated for each participant based on the responses from the questionnaires about

experience and suitability for the task. These ratings for the 10 experts, along with the number of analogies recalled by each (of the ones who were assigned the s-SA treatment) which will be used for investigating RQ2.2 can be found below in tables 5.4 and 5.5.

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Profession	DECC – Policy Official	Tyndall Centre – Researcher in Transport Policy and Behavioural Change	Climate Group – Senior Policy Manager	CE Delft – Senior Researcher/Consultant	EST – Transport Strategy Manager
Analogies Recalled	3	4	3	8	3
Experience Score[1]	9	4	4	8	7

Table 5.4: Expert Ratings for s-SA Participants

	Expert 6	Expert 7	Expert 8	Expert 9	Expert 10
Profession	UKERC/Centre for Transport Research – Senior Lecturer in Transport	ICEPT/Centre for Environmental Policy – Research Associate	E4Tech – Managing Consultant	ICEPT/Centre for Environmental Policy - Researcher	Climate Group – Head of Policy
Experience Score	9.5	6	6	5	5

Table 5.5: Expert Ratings for UJ Participants

As mentioned in section 3.5, it must be recognised that these experience ratings are not, and do not claim to be, objective. They are scores given by to the experts based on their answers to the experience and suitability questions in the questionnaire and will hence be affected by factors such as the respondents' own level of self-belief in his or her ability.

---

[1] This score was based on the average between the ratings given for questions 6 and 7 of the questionnaire, i.e. the 'experience working in an environmental issues setting', 'experience with environmental public policy' and 'experience with cases similar to this one'.

This chapter has summarised the details of each of the five experiments which compose the empirical procedure for exploring the research questions set out in sub-section 3.5.2.

Chapter 6: Results

6.1 Introduction

This chapter will present the results of the five experiments described in chapter 5 along with any other results which are needed in investigating the research questions outlined in sub-section 3.5.2. The chapter is split into four main sections. The first section presents the accuracy measures used throughout the remainder of this results chapter to compare amongst the forecasting approaches and how these were calculated. The second section will present the results of the five experiments in an attempt to see how the structured analogies approach compared against the UJ method. The third section will give greater focus to the results of the fifth experiment and how these results allow for the study of the relationship between the level of expertise of a participant and the accuracy of his or her forecasts. The fourth and final section will explore how the UJ and s-SA2 predictions for the vehicle scrappage scheme effectiveness did in comparison to those anticipated by the local governments.

6.2 Accuracy Measures and Calculations

6.2.1 Accuracy Measures

In order to test the quality of the forecasts produced by either of the two approaches, these were compared against the actual values in each of the cases, as described in section 5.1, and in this way their errors were computed. Different accuracy measures were applied to these errors which enabled comparison between the forecasting approaches. The six different accuracy measures used were Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), Mean Absolute Error (MAE),

Median Absolute Error (MdAE), Geometric Mean of the Relative Absolute Errors (GMRAE) and Geometric Mean of the Relative Mean Absolute Errors (GMRMAE).

MAPE

The Mean Absolute Percentage Error (MAPE) is one of the most common accuracy measures used in the forecasting literature (Armstrong and Collopy, 1992, Makridakis and Hibbon, 2000, Fildes et al., 1998). It is calculated as the mean of the absolute percentage errors, whose formula is

$$APE = 100 * \frac{|F - A|}{A}$$

Where *F* denotes the forecast and *A* denotes the actual (observed) value.

MdAPE

The Median Absolute Percentage Error (MdAPE) is very similar to the MAPE with the difference being in that the median of the APEs is calculated, as opposed to the mean. The MdAPE is particularly useful to overcome extreme values or outliers in the forecasts which would otherwise skew the MAPE.

MAE

The Mean Absolute Error (MAE) is simply the mean of the absolute errors, whose formula is

$$AE = |F - A|$$

Where *F* denotes the forecast and *A* denotes the actual (observed) value. In other words it is the absolute value of the difference between the forecast and the observed value.

MdAE

The Median Absolute Error (MdAE) is very similar to the MAE with the difference being that the median of the AEs is calculated, as opposed to the mean. In similar fashion to

the MdAPE with MAPE, MdAE offers a means of overcoming any biases from extreme values or outliers that might be affecting the MAE.

GMRAE

The Geometric Mean of the Relative Absolute Errors (GMRAE) is a means for comparing an approach directly against a benchmark and as a consequence, indirectly against other approaches. It is calculated as follows. First the RAEs must be computed for each forecast and benchmark. That is,

$$RAE = \frac{|F_\alpha - A|}{|F_B - A|}$$

Where F_α is the forecast of the approach being evaluated and F_B is the forecast of the benchmark approach, and *A* is the actual (observed) value. Once the RAEs have been computed, their geometric mean is taken, that is,

$$GMRAE = (RAE_1 * RAE_2 * RAE_3 * ...... * RAE_n)^{\frac{1}{n}}$$

Where *n* denotes the total number of RAEs. A GMRAE between 0 and 1 suggests the superiority of the approach being evaluated and a GMRAE over 1 suggests the superiority of the benchmark approach.

GMRMAE

The Geometric Mean of the Relative Mean Absolute Errors is very similar to the GMRMAE with the difference that the geometric mean is taken of the relative mean absolute errors and not the relative absolute errors like in the GMRAE metric. Its use was proposed in this research as means for directly comparing the performance of two approaches relative to each other and thereby eliminating the need for a benchmark, which was unavailable for experiments I and II. In other words,

$$GMRMAE = (RMAE_1 * RMAE_2 * RMAE_3 * ...... * RMAE_n)^{\frac{1}{n}}$$

Where *n* denotes the total number of RMAEs.

6.2.2 Calculations

All calculations done when computing the different accuracy measures for both the structured analogies and UJ approach as well as the tables and graphs presented in this chapter were done so in Excel. Details on how this was done follow.

MAPE, MdAPE, MAE, MdAE

For each individual case and for each individual respondent's forecast, the APE was calculated using the actual (observed) value for the effectiveness of the PIS described in that case. Then, for each individual case, the mean of these APEs was calculated to give a MAPE for each of the two approaches for each of the three cases. In order to calculate the MAPE across all cases, the mean of these 'case' MAPEs was computed. Exactly the same procedure was undertaken when calculating the MAE. In the case of the MdAPE and MdAE, the medians of all the APEs and AEs respectively were taken instead of the means.

6.3 RQ1: Structured Analogies vs. Unaided Judgment

6.3.1 Experiment I

The first experiment involved the use of non-experts. Recalling that results were obtained for participants working individually and in groups, the results of the individual experiment according to MAPE and MdAPE are summarised in the tables 6.1 and 6.2 below. For the reasons explained previously, it was not decided to continue with testing the approach when participants work in groups, so the results of these will be used as rationale in the further research section (8.3).

	Case 1		Case 2		Case 3		Overall	
	UJ	s-SA1	UJ	s-SA1	UJ	s-SA1	UJ	s-SA1
MAPE (%)	23.04	28.59	117.12	102.79	18.98	19.14	53.73	49.70
MdAPE (%)	18.59	24.40	77.30	41.84	15.29	16.47	29.24	23.82

Table 6.1: Summary of Experiment I Individual Performance using APE

	Case 1		Case 2		Case 3		Overall	
	UJ	s-SA1	UJ	s-SA1	UJ	s-SA1	UJ	s-SA1
MAE	72710.18	84302.41	16.51	14.49	16.13	16.27	23146.19[1]	29332.61
MdAE	61905.00	78412.50	10.90	5.90	13.00	14.00	26.45	25.00

Table 6.2: Summary of Experiment I Individual Performance using AE

If the 16 different combinations between the case aggregation level, the accuracy measure used (APE or AE) and whether a mean or median measure of location is used, is considered, the s-SA1 approach is the winner on 7 occasions and the UJ on 9 occasions.

Next, it was decided that it would be useful to have an accuracy measure that would enable direct comparison between the two approaches. Something along the lines of the Geometric Mean Relative Absolute Error (GMRAE), which is often used to indirectly compare rivalling approaches by direct comparison of each to a benchmark (in time series analysis, this is often the 'Random Walk'), was considered appropriate. In the absence of any other suitable measure, it was decided that the UJ approach would act like the benchmark and the s-SA1 would be compared to it directly.

Under normal circumstances, the GMRAE is calculated by first finding the RAE of every forecast relative to the benchmark and then computing the geometric mean of these. If the GMRAE is between 0 and 1, then the approach being examined performs better than the benchmark. If the GMRAE is greater than 1, the inverse is true. As will be seen for experiments IV and V, the GMRAE is calculated in this way with the government forecasts acting as a benchmark[2].

In this instance however, due to the fact that different individuals provided the s-SA and UJ predictions respectively, pairings between the two kinds of forecasts could not

---

[1] Overall mean affected by the large errors of case 1

[2] When calculating the GMRAE, the RAEs were computed for each respondent's forecast for each of the cases using the government forecasts as benchmarks. Then, for each individual case, the geometric mean of these RAEs was calculated to give the GMRAE for each of the two approaches for each of the three cases. Then, to calculate the GMRAE across the three cases, the geometric mean of these GMRAEs was taken.

be made and so an RAE was impossible to calculate. Rather, it was decided to calculate the AE of each forecast and calculate a mean AE (MAE) for each of the two approaches in each of the three cases. Then, the relative MAE (RMAE) of each case was taken by dividing the s-SA1 MAE by the UJ MAE. The geometric mean of these MAEs was then used to calculate what will be appropriately called the Geometric Mean Relative Mean Absolute Error (GMRMAE)[1]. Please see Appendix 3 for more details on how the GMRMAE was computed. Table 6.3 presents the results of the GMRMAE.

	Case 1	Case 2	Case 3	GMRMAE
RMAE	1.16	0.88	1.01	1.01

Table 6.3: Performance Comparison Between s-SA1 and UJ with GMRMAE

From these results it can be seen that the GMRMAE is very close to 1.00, suggesting very level performance between the two approaches under consideration, but nevertheless a superiority of UJ.

6.3.2 Experiment II

The second experiment, which involved the participation of 'semi' experts from a large EU Institution, led to the following results. Tables 6.4 and 6.5 present the performance of the s-SA1 and UJ approaches according to MAPE, MdAPE, MAE and MdAE, respectively.

	Case 1		Case 2		Case 3		Overall	
	UJ	s-SA1	UJ	s-SA1	UJ	s-SA1	UJ	s-SA1
MAPE (%)	20.37	25.86	82.98	106.38	22.06	25.00	41.80	52.41
MdAPE (%)	16.50	19.99	59.57	88.65	23.53	20.59	25.28	29.53

Table 6.4: Summary of Experiment II Individual Performance using APE

---

[1] When calculating the GMRMAE, the first thing that was done for each case and for each respondent's forecast was calculate the absolute error with the actual (observed) value. Then, the mean of these was taken for the AEs of each of the two approaches. Then the RMAEs were computed by dividing the s-SA MAE by the UJ MAE. The geometric mean of these was finally calculated to get a final GMRMAE across the three cases.

	Case 1		Case 2		Case 3		Overall	
	UJ	s-SA1	UJ	s-SA1	UJ	s-SA1	UJ	s-SA1
MAE	61875.00	77044.00	11.70	15.00	18.80	21.30	20365.20	25693.30
MdAE	52898.00	61190.00	8.40	12.50	20.00	17.50	4876.50	1471.50

Table 6.5: Summary of Experiment II Individual Performance using AE

As can be seen in tables 6.4 and 6.5, performance between the two and amongst the cases is mixed, with different winners according to the accuracy measure and measure of location used. UJ seems to outperform s-SA1 in 13 of 16 (81.3%) combinations and the vice-versa occurs on 3 of 13 (18.7%) combinations.

As for experiment I, the GMRMAE is calculated for the experiment II results and can be found in table 6.6. The results here show the UJ approach to fairly consistently outperform the s-SA1 in each case and for each accuracy measure used. A GMRAE across all cases and accuracy measures of 1.22 shows an overall noticeable, but not significant, superiority of the UJ in the hands of the 'semi' experts.

	Case 1	Case 2	Case 3	GMRMAE
RMAE	1.25	1.28	1.13	1.22

Table 6.6: Performance Comparison Between s-SA1 and UJ with GMRMAE

6.3.3 Experiment III

The results of this experiment are omitted because its primary objective was not to collect more results but rather to gain feedback about the design of the cases for the purpose of their refinement for experiments IV and V.

6.3.4 Experiment IV

The fourth experiment was the second with non-experts and the first with the refined cases. The results of the s-SA2 and UJ approaches according to MAPE, MdAPE, MAE and MdAE can be summarised as in tables 6.7 and 6.8.

	Case 1		Case 2		Case 3		Overall	
	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2
MAPE (%)	8931.00	9767.00	104.40	65.13	47.14	66.88	3027.60	3299.60
MdAPE (%)	4700.00	11633.00	48.72	38.46	50.54	73.12	78.49	78.49

Table 6.7: Summary of Experiment IV Individual Performance using APE

	Case 1		Case 2		Case 3		Overall	
	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2
MAE	33.49	36.63	20.37	12.70	43.84	62.20	32.57	37.18
MdAE	17.63	43.63	9.50	7.50	47.00	68.00	23.00	43.50

Table 6.8: Summary of Experiment IV Individual Performance using AE

Tables 6.7 and 6.8 show that in the hands of non-experts, UJ performs better than s-SA2 in 11 of 16 (68.8%) combinations and the inverse occurs in 4 of 16 (25.0%) combinations. The remaining combination, the 'Overall MdAPE' is a tie.

Once again, the GMRMAE is used to directly compare the performance of the two approaches relative to each other. These results for experiment IV can be seen in table 6.9 below.

	Case 1	Case 2	Case 3	GMRMAE
RMAE	1.09	0.62	1.42	0.99

Table 6.9: Performance Comparison Between s-SA2 and UJ with GMRMAE

Once again the GMRMAE is very close to 1.00, indicating almost no difference in performance between the two approaches, but this time a slight superiority of the structured analogies method.

With the new cases used in experiment IV and V, government forecasts for the PIS effectiveness are available. As a result, as was explained earlier, it was decided to benchmark the s-SA2 and the UJ approaches against these government forecasts and

use these comparisons to compare performance between the two approaches. In order to do this, the GMRAE was used.

Hence, as any normal GMRAE, the RAEs between the respondents' forecasts and the government forecasts were computed and their geometric means calculated. Finally, the geometric mean of these GMRAEs was calculated to get a GMRAE across all cases for each approach. The results of the use of this accuracy measure are summarised in table 6.10.

	Case 1		Case 2		Case 3		GMRAE	
	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2
GMRAE	61.68	84.03	1.08	0.61	0.63	0.97	3.47	3.67

Table 6.10: Results for GMRAE of s-SA2 and UJ when Benchmarked Against the Government Forecasts

As can be seen from table 6.10, the s-SA2 approach was less accurate than UJ in cases 1, 3 and overall. So the results of experiment IV are mixed, with some accuracy measures indicating the superiority of the s-SA2 and some accuracy measures indicating the superiority of the UJ.

6.3.5 Experiment V

The final experiment of the investigation consisted of testing the s-SA2 approach in the hands of experts and comparing it to forecasts produced via UJ. The results of this fifth experiment according to MAPE, MdAPE, MAE and MdAE are summarised in the tables 6.11 and 6.12 below.

	Case 1		Case 2		Case 3		Overall	
	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2
MAPE (%)	2087.00	886.70	30.26	24.10	21.29	34.84	712.74	315.20
MdAPE (%)	2567.00	1233.00	28.21	7.69	13.98	24.73	43.59	48.72

Table 6.11: Summary of Experiment V Individual Performance using APE

	Case 1		Case 2		Case 3		Overall	
	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2
MAE	7.83	4.87	5.90	4.70	19.80	32.40	11.18	13.99
MdAE	9.63	4.63	5.50	1.50	13.00	23.00	8.50	6.63

Table 6.12: Summary of Experiment V Individual Performance using AE

From the results tables it can be seen that in the hands of capable experts, the s-SA2 approach is more accurate than UJ in 10 of 16 (62.5%) combinations. UJ seems to perform better in 6 of 16 (37.5%) combinations. Incidentally, 4 of these 6 occur in case 3.

As in previous experiments, the GMRMAE is calculated with the results presented in table 6.13.

	Case 1	Case 2	Case 3	GMRMAE
RMAE	0.62	0.80	1.64	0.93

Table 6.13: Performance Comparison Between s-SA2 and UJ with GMRMAE

As can be seen by the figures in table 6.13, the s-SA2 clearly outperforms the UJ in cases 1 and 2 but not in case 3, where it underperforms significantly. This is in agreement with what was previously found. However, the GMRMAE shows that when performance of the two approaches is considered across the three cases, the accuracy measure indicates s-SA as the better of the two by a noticeable distance.

Finally, like for experiment IV, the GMRAE is calculated for each of the two approaches, for each case and then across all cases, when benchmarked against the government forecast. The results of this are as per table 6.14.

	Case 1		Case 2		Case 3		GMRAE	
	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2
GMRAE	12.66	5.56	0.51	0.33	0.21	0.35	1.10	0.86

Table 6.14: Results for GMRAE of s-SA2 and UJ when Benchmarked Against the Government Forecasts

The GMRAE confirms the results found with every previous accuracy measure.

If, like in Makridakis and Hibbon (2000), a performance rank is given to each approach within each metric, for each level of case aggregation, the following table is obtained.

	Case 1		Case 2		Case 3		Overall	
	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2	UJ	s-SA2
MAPE	2	1	2	1	1	2	2	1
MdAPE	2	1	2	1	1	2	1	2
MAE	2	1	2	1	1	2	1	2
MdAE	2	1	2	1	1	2	2	1
RMAE	2	1	2	1	1	2	2	1
GMRAE	2	1	2	1	1	2	2	1
Avg. Rank	2	1	2	1	1	2	1.67	1.33

Table 6.15: Performance Rankings for Each Approach Within Each Metric at Different Aggregation Levels

Table 6.15 is also in agreement with previous accuracy measures as it shows that for cases 1 and 2, s-SA2 performs better than UJ but the inverse is true for case 3. For all three cases aggregated together, this ranking system displays the superiority of the s-SA2 approach.

6.3.6 SA Forecasts for Experiments IV and V

As was explained in sub-section 4.4.3, it was decided to see how the SA performed in the hands of the participants of experiments IV and V when predicting effectiveness intervals, given that these associated their analogies with an outcome interval. The main reasons for this were to see how well the SA approach performs, even if it can only give an interval forecast, as well as to see if there is any consistency between the

SA prediction and the s-SA2 outcome interval prediction made by the respondents. The results are as follows.

Experiment IV

Participant	SA Forecast	Scale-Point Error	s-SA2 Outcome Forecast	Scale-Point Error	Agreement?	SA Forecast Correct?
1	D	3	C	2	No	No
2	A	0	A	0	Yes	Yes
3	D	3	C	2	No	No
4	E	4	C	2	No	No
5	B	1	A	0	No	No

Table 6.16: SA Forecasts for Case 1

Participant	SA Forecast	Scale-Point Error	s-SA2 Outcome Forecast	Scale-Point Error	Agreement?	SA Forecast Correct?
1	E	4	D	3	No	No
2	A	0	B	1	No	Yes
3	D	3	B	1	No	No
4	C	2	A	0	No	No
5	-		B	1	-	-

Table 6.17: SA Forecasts for Case 2

Participant	SA Forecast	Scale-Point Error	s-SA2 Outcome Forecast	Scale-Point Error	Agreement?	SA Forecast Correct?
1	E	0	D	1	No	Yes
2	B	3	B	3	Yes	No
3	D	1	B	3	No	No
4	C	2	A	4	No	No
5	-		A	4	-	-

Table 6.18: SA Forecasts for Case 3

- SA Performance: SA forecast predicts the correct range 23.1% of the time (3 of 13) with a mean scale-point error of 2.00.

- s-SA2 Interval Performance (s-SA2 Outcome Forecast): The predicted interval outcome was correct in 20.0% of the cases (3 of 15) with a mean scale-point error of 1.80.

- SA Consistency: SA forecast is the same as the predicted interval outcome 15.4% of the time (2 of 13).

It could be said that the SA approach is hardly better than chance at predicting the correct interval for PIS effectiveness.

Experiment V

Participant	SA Forecast	Scale-Point Error	s-SA2 Outcome Forecast	Scale-Point Error	Agreement?	SA Forecast Correct?
1	A	0	A	0	Yes	Yes
2	A	0	A	0	Yes	Yes
3	A	0	A	0	Yes	Yes
4	B	1	A	0	No	No
5	A	0	A	0	Yes	Yes

Table 6.19: SA Forecasts for Case 1

Participant	SA Forecast	Scale-Point Error	s-SA2 Outcome Forecast	Scale-Point Error	Agreement?	SA Forecast Correct?
1	A	0	B	1	No	Yes
2	A	0	A	0	Yes	Yes
3	A	0	A	0	Yes	Yes
4	B	1	B	1	Yes	No
5	A	0	B	1	No	Yes

Table 6.20: SA Forecasts for Case 2

Participant	SA Forecast	Scale-Point Error	s-SA2 Outcome Forecast	Scale-Point Error	Agreement?	SA Forecast Correct?
1	A	4	E	0	No	No
2	D	1	D	1	Yes	No
3	A	4	B	3	No	No
4	B	3	B	3	Yes	No
5	A	4	E	0	No	No

Table 6.21: SA Forecasts for Case 3

- SA Performance: SA forecast predicts the correct range 53.3% of the time (8 of 13) with a mean scale-point error of 1.20.

- s-SA2 Interval Performance (s-SA2 Outcome Forecast): The predicted interval outcome was correct in 60.0% of the cases (9 of 15) with a mean scale-point error of 0.93.

- SA Consistency: SA forecast is the same as the predicted interval outcome 60.0% of the time (9 of 13).

6.4 RQ2: Does Level of Expertise Affect Forecast Accuracy?

6.4.1 RQ2.0

The results of experiments IV and V above show that irrespective of the accuracy measure used, the experts of experiment V produced considerably more accurate predictions than the non-experts in experiment IV, for each case and for all cases aggregated.

6.4.2 RQ2.1

In sub-section 3.5.2, it was explained that in order to investigate the second research question, ratings of experience would be given to each of the experts who participated in experiment V (tables 5.4 and 5.5 in section 5.9). Such ratings would be based on the answers to the experience and suitability for the task questions in the questionnaire.

s-SA2 Group

Recalling the table 5.4, the experience ratings along with the accuracy of their s-SA2 forecasts were as follows.

Expert	Experience Rating	Case 1 AE	Case 2 AE	Case 3 AE	MAE
1	9	0.63	1.50	6.00	2.71
2	4	4.63	9.50	23.00	12.38
3	4	12.33	0.50	58.00	23.61
4	8	6.63	1.50	68.00	25.38
5	7	0.13	10.50	7.00	25.38

Table 6.22: Experience Ratings and Forecast AE

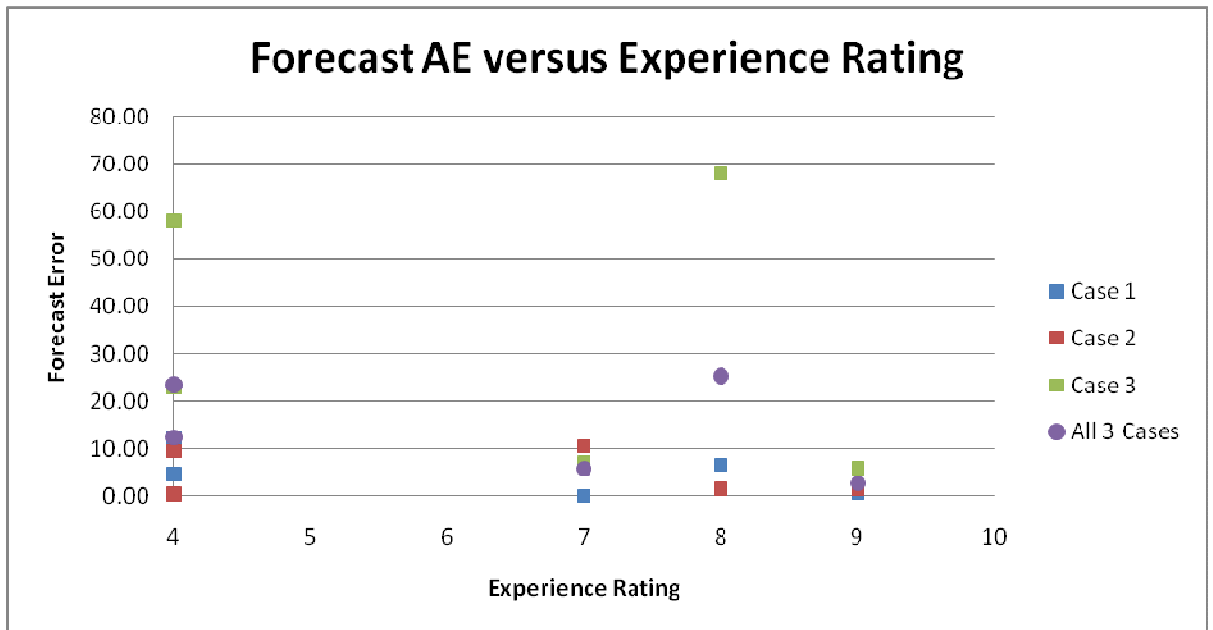If this is plotted, the scatter plot in figure 6.1 is obtained.

Figure 6.1: Graph of AE versus Experience Rating

The scatter plot shows no real pattern in the data. There is no suggestion from any of the individual cases or when the three are combined that forecast error decreases as level of presumed experience increases.

UJ Group

Recalling the table 5.5, the expert ratings along with the accuracy of their UJ forecasts were as follows

Expert	Experience Rating	Case 1 AE	Case 2 AE	Case 3 AE	MAE
6	9.5	12.63	8.50	18.00	13.04
7	6	0.63	5.50	3.00	3.04
8	6	9.63	0.50	13.00	7.71
9	5	1.63	4.50	58.00	21.38
10	5	14.63	10.50	7.00	10.71

Table 6.23: Experience Ratings and Forecast AE

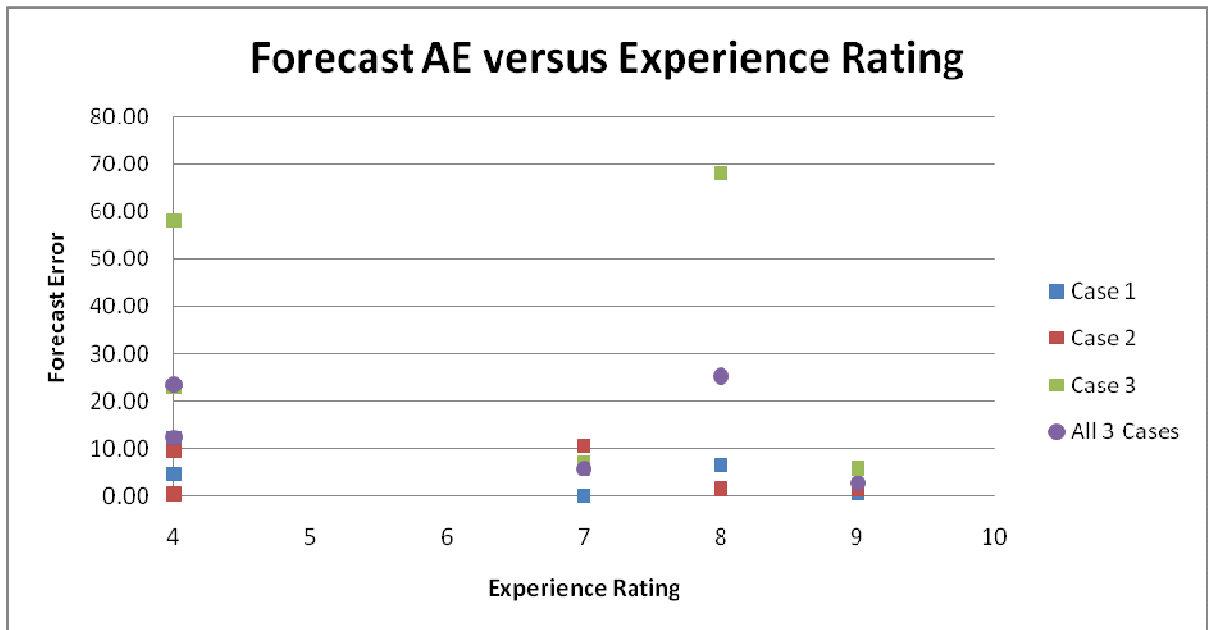If this is plotted, the scatter plot in figure 6.2 is obtained.

Figure 6.2: Graph of Forecast AE versus Experience Rating

If instead of AE, APE is used as an accuracy measure, the results are confirmed.

s-SA2 Group

Expert	Experience Rating	Case 1 APE	Case 2 APE	Case 3 APE	MAPE
1	9	166.67	7.69	6.45	60.27
2	4	1233.33	48.72	24.73	435.59
3	4	1233.33	2.56	62.37	324.57
4	8	1766.67	7.69	73.12	463.53
5	7	33.33	53.85	7.53	31.57

Table 6.24: Experience Ratings and Forecast APE

If this is plotted, the scatter plot in figure 6.3 is obtained.
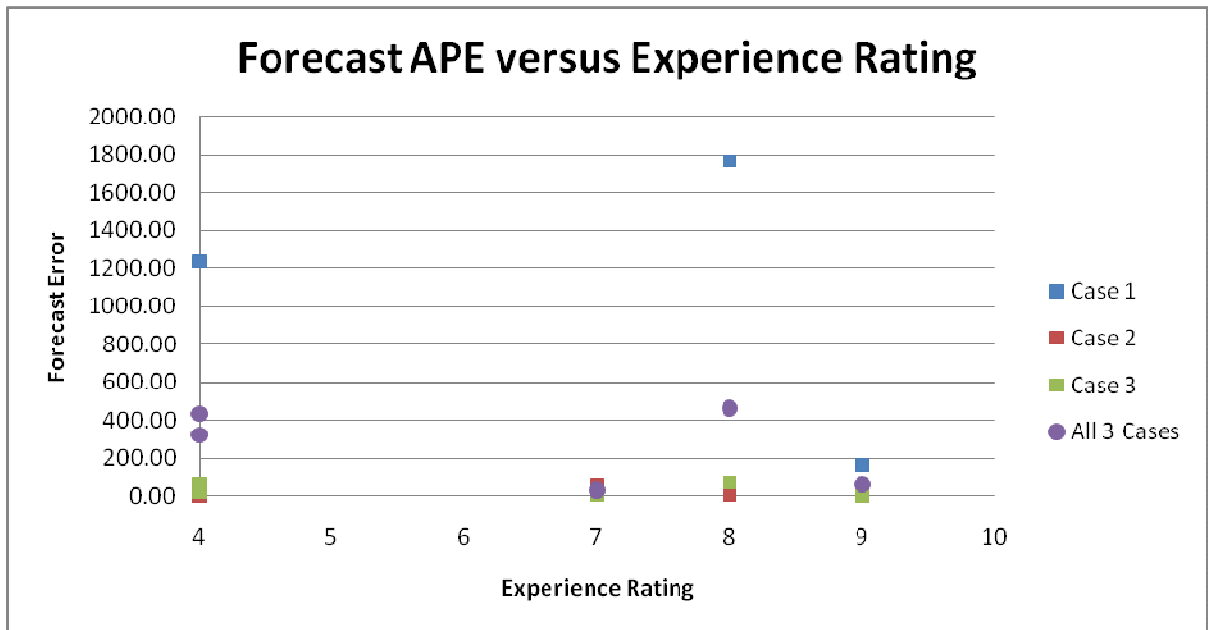
Figure 6.3: Graph of APE versus Experience Rating

Once again, the scatter plot shows no real pattern in the data. This is further evidence that there is no suggestion from any of the individual cases or when the three are combined that forecast error decreases as level of experience increases.

UJ Group

Expert	Experience Rating	Case 1 APE	Case 2 APE	Case 3 APE	MAPE
6	9.5	3366.67	43.59	19.36	1143.20
7	6	166.67	28.21	3.23	66.03
8	6	2566.67	2.56	13.98	861.07
9	5	433.33	23.08	62.37	172.93
10	5	3900.00	53.85	7.53	1320.46

Table 6.25: Experience Ratings and Forecast APE

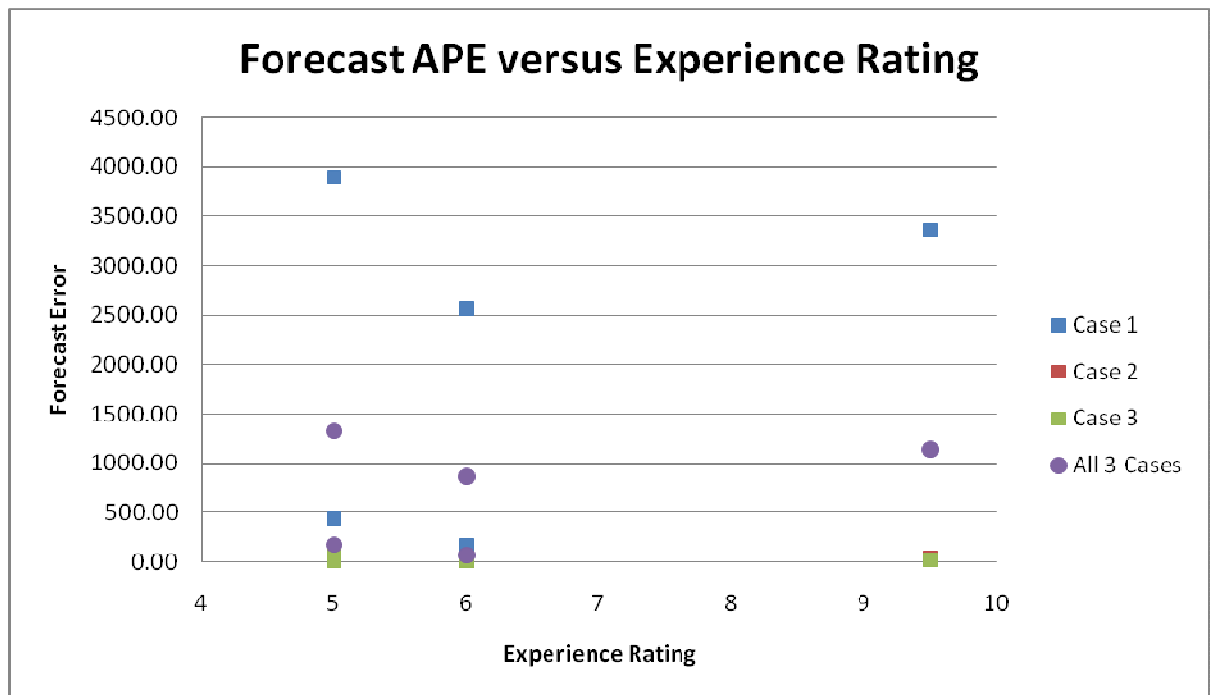If this is plotted, the scatter plot in figure 6.4 is obtained.

Figure 6.4: Graph of Forecast APE versus Experience Rating

Once again, just as for the s-SA2 group, there does not seem to be any pattern in the data. Thus, it would have to be concluded that no real correlation exists between forecast error and experience rating.

6.4.3 RQ2.2

One of the results in Green and Armstrong (2007b) was that experts who could recall more than one analogy produced better forecasts. Hence, for the participants who used s-SA2, it was decided to see if there was any correlation between number of analogies recalled and forecast accuracy. Table 6.26 contains such information with AE being used as the accuracy measure.

Expert	Case 1		Case 2		Case 3		Overall	
	Analogies Recalled	Case 1 AE	Analogies Recalled	Case 2 AE	Analogies Recalled	Case 3 AE	Analogies Recalled	MAE
1	1	0.63	1	1.50	1	6.00	3	2.71
2	2	4.63	1	9.50	1	23.00	4	12.38
3	1	12.33	1	0.50	1	58.00	3	23.61
4	3	6.63	3	1.50	3	68.00	9	25.38
5	1	0.13	1	10.50	1	7.00	3	25.38

Table 6.26: Comparison of Number of Analogies Recalled with s-SA2 Performance

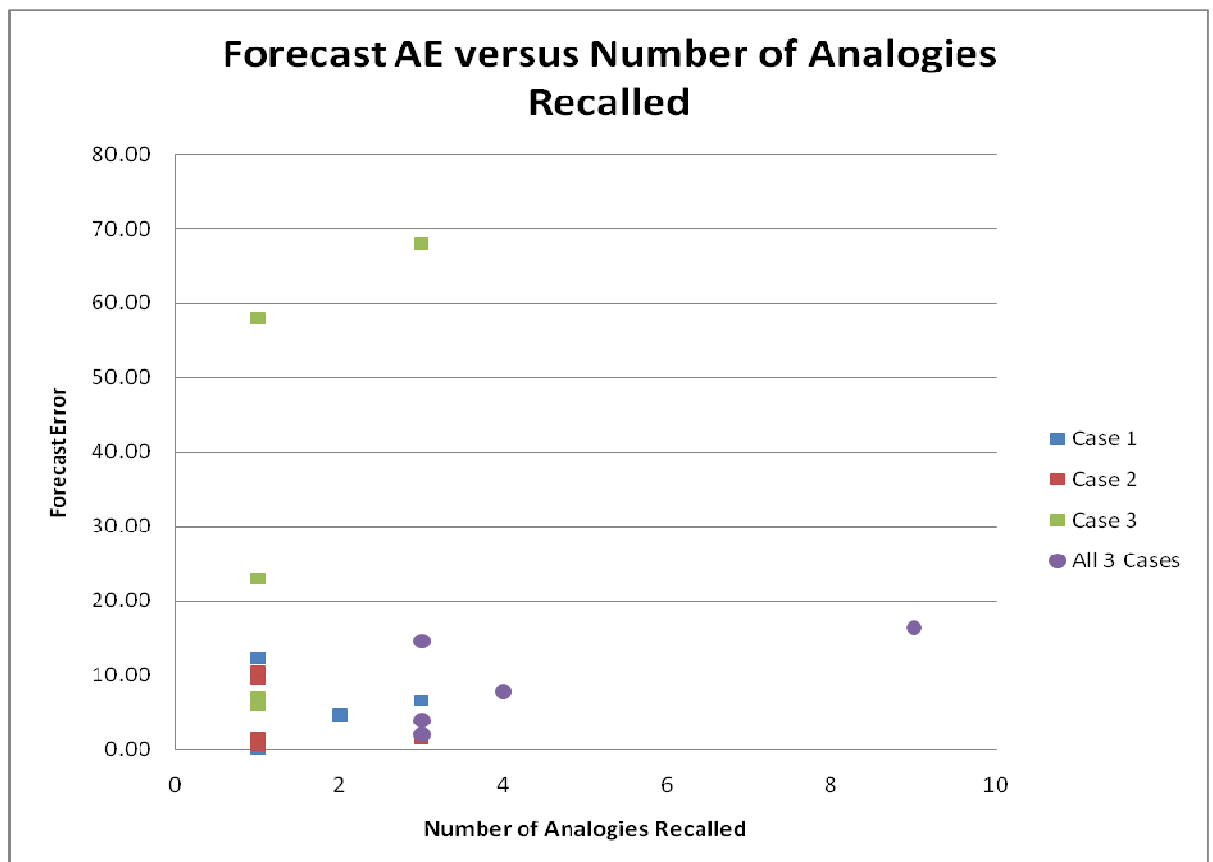These results can be summarised in the scatter plot in figure 6.5.

Figure 6.5: Scatter Plot of s-SA Performance versus Number of Analogies Recalled for Cases 1, 2, 3 and Aggregated for All Three Cases

The results in table 6.26 and figure 6.5 show that there does not seem to be any relationship between number of analogies recalled and forecast accuracy, a result which is in contrast with that of Green and Armstrong (2007b). These results are in agreement with the previous results in showing that there seems to be no correlation between level of expertise and forecast accuracy.

6.5 RQ3: How do the Expert Forecasts Compare with Government Forecasts?

The third research question suggested that the forecasts produced by experts were no worse than those anticipated by the governments who implemented the strategies. In order to explore this issue, the s-SA2 and UJ forecasts are compared in terms of accuracy with those produced by the government for each case and then for the three cases aggregated as seen in tables 6.27 to 6.30 and figures 6.6 to 6.9.

Case 1:

	Case 1		
	UJ	s-SA2	Gov.
MAPE[1]	2.07	0.87	0.09
MdAPE[2]	2.57	1.23	0.09
MAE	7.83	4.87	0.35
MdAE	9.63	4.63	0.35
GMRAE	12.66	5.56	1.00

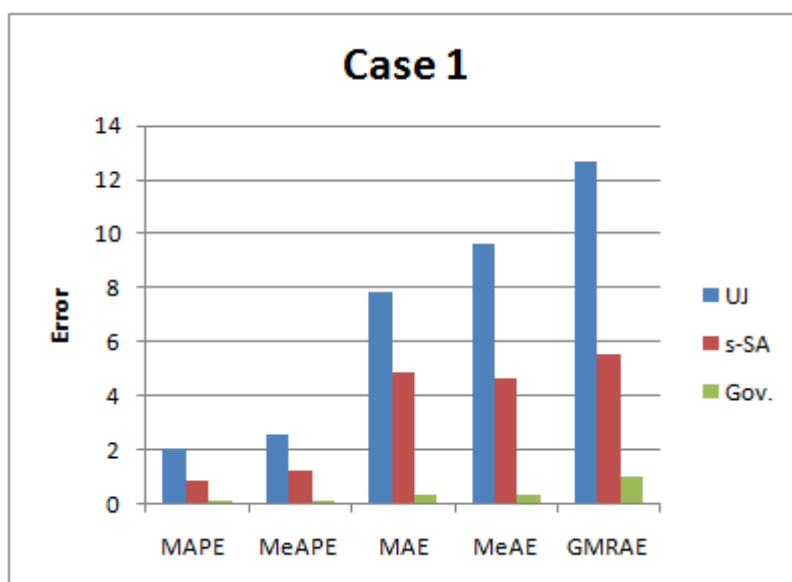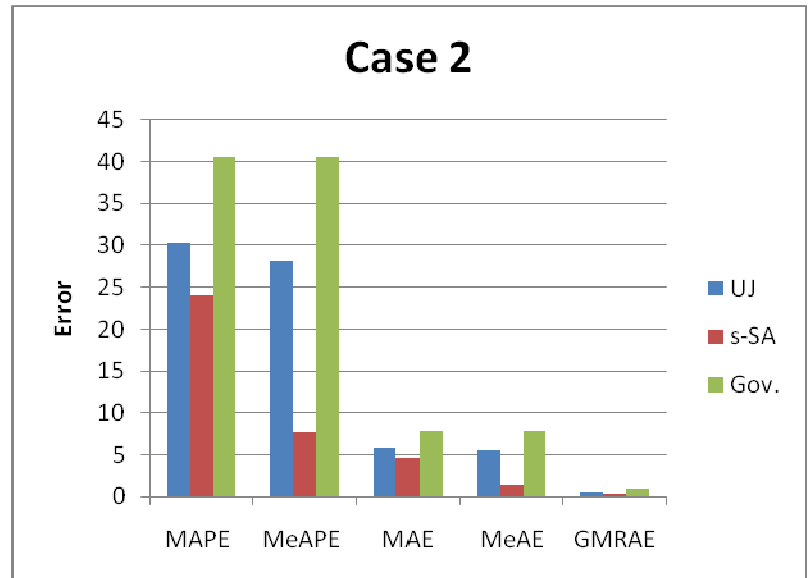**Table 6.27 and Figure 6.6: Expert Forecasts versus Government Forecasts for Case 1**
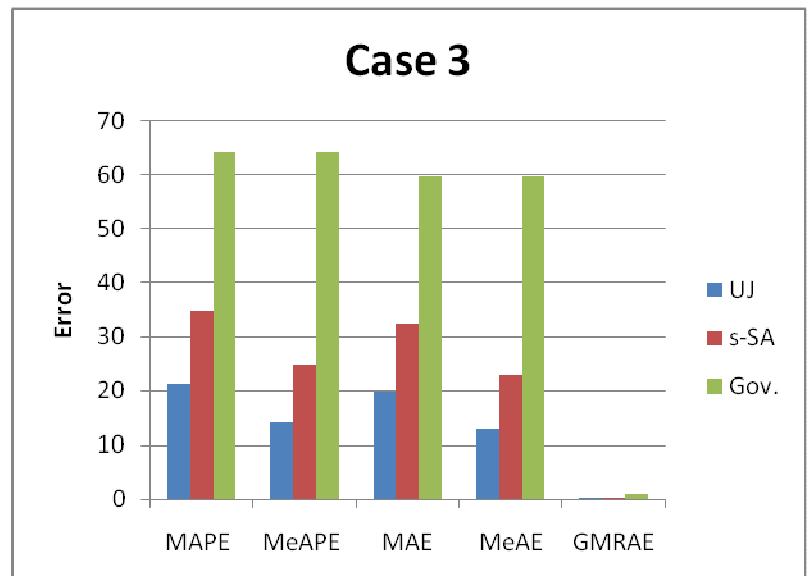
---

[1] X 1000
[2] X 1000

Case 2:

	Case 2		
	UJ	**s-SA2**	**Gov**
MAPE	30.26	24.10	40.51
MdAPE	28.21	7.69	40.51
MAE	5.90	4.70	7.90
MdAE	5.50	1.50	7.90
GMRAE	0.51	0.33	1.00

Table 6.28 and Figure 6.7: Expert Forecasts versus Government Forecasts for Case 2

Case 3:

	Case 3		
	UJ	**s-SA2**	**Gov**
MAPE	21.29	34.84	64.16
MdAPE	13.98	24.73	64.16
MAE	19.80	32.40	59.67
MdAE	13.00	23.00	59.67
GMRAE	0.21	0.35	1.00

Table 6.29 and Figure 6.8: Expert Forecasts versus Government Forecasts for Case 3

Overall:

	Overall		
	UJ	s-SA2	Gov
MAPE[1]	7.13	3.15	0.66
MdAPE	43.59	48.72	64.1
MAE	11.18	13.99	22.6
MdAE	8.50	6.63	7.9
GMRAE	1.10	0.86	1.00
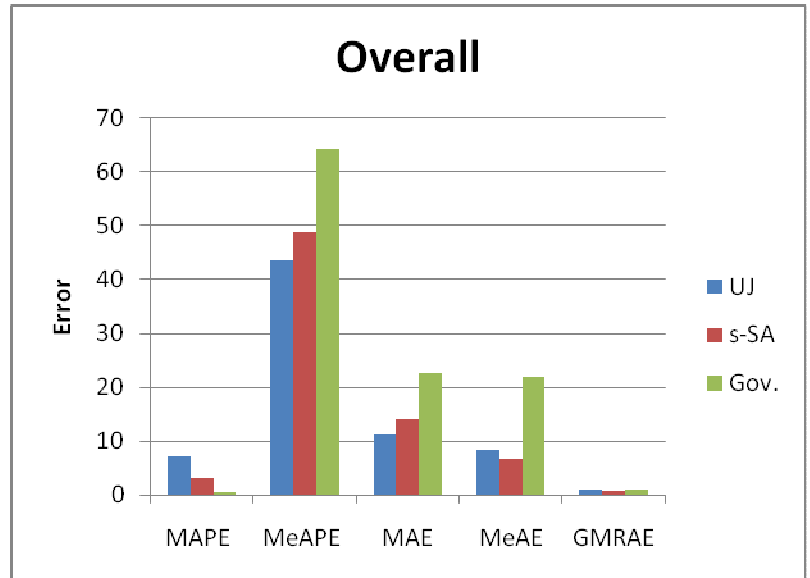
Table 6.30 and Figure 6.9: Expert Forecasts versus Government Forecasts Aggregated Over All Three Cases

Tables 6.27 to 6.30 and figures 6.6 to 6.9 suggest mixed results. In case 1, the Government forecasts is by far the most accurate, but for case 2 and 3 the most accurate predictions are by the s-SA2 and UJ approaches respectively.

These results are confirmed when a table (6.31) of their ranking is constructed.

---

[1] X 100

	Case 1			Case 2			Case 3			Overall		
	UJ	s-SA2	Gov.	UJ	s-SA2	Gov	UJ	s-SA2	Gov	UJ	s-SA2	Gov
MAPE	3	2	1	3	1	2	1	2	3	3	2	1
MdAPE	3	2	1	2	1	3	1	2	3	1	2	3
MAE	3	2	1	3	1	2	1	2	3	1	2	3
MdAE	3	2	1	2	1	3	1	2	3	2	1	3
GMRAE	3	2	1	2	1	3	1	2	3	3	1	2

Table 6.31: Rank Table for Performance of Expert Forecasts Versus Government Forecasts Across All Cases and Accuracy Measures

With an overall mean rank of 1.65, the s-SA2 approach is seen to be the most accurate when all levels of aggregation are considered. In second place is UJ with an overall mean rank of 2.1 and in last place are the Government forecasts with a mean rank of 2.25.

6.6 Conclusions

This chapter has summarised and presented the results of the experiments described in chapter 5 designed to test investigate the research questions set out in sub-section 3.5.2.

Section 6.3 of this chapter showed that UJ is more accurate than structured analogies in experiments I and II. For experiments IV and V, despite performing well, UJ is less accurate than structured analogies. It was also found that if experts obey their analogies, the accuracy of their resulting forecast will improve.

Section 6.4 showed that although it was found that experts produce more accurate forecasts than non-experts, as expected, their level of experience or knowledge of analogies did not have an effect on accuracy.

Finally, section 6.5 showed that the forecasts produced by the experts in experiment V are comparable in accuracy to those produced by the respective governments.

Chapter 7 will focus on the discussion of these results.

Chapter 7: Discussion

7.1 Introduction

Following the results from the literature review, this research put forward three main research questions about forecasting the effectiveness of policy implementation strategies. In order to investigate these questions, a series of experiments, described in chapter 5, were conducted and whose results were presented in chapter 6. This chapter will recall these results and attempt to interpret them by reporting effect size rather than with the use of statistical significance tests for the reasons argued in 4.4.3 and Armstrong (2007).

7.2 RQ1

7.2.1 RQ1.0

7.2.1.1 Experiments I and II

The first research question set out by this research was that a method of structured analogies would be superior to an unaided judgment approach for predicting the effectiveness of a potential policy implementation strategy suggested during the policy formulation stage. In order to go about investigating this, five experiments were conducted. The first three experiments provided the initial results of the structured analogies performance and valuable feedback which would help in the design of the final two experiments, particularly in the redesign of the s-SA1 into the s-SA2.

Experiment I Results

Irrespective of the metric, the results of experiment I, which tested the structured analogies method (in the form of s-SA1) in the hands of non-experts, showed fairly

level performance between the two approaches, but nevertheless, a slight superiority of the UJ approach. The results of experiment II, which used participants from a large EU Institution on the other hand, displayed much more unfavourable results for the s-SA1 with the UJ outperforming it in a large proportion of the time. The GMRMAE confirmed the s-SA1 inferiority in the experiment.

Savio and Nikolopoulos (2009b) reviewed several papers which covered performance in different kinds of studies of experts and non-experts. The review highlighted one particular article, by Bolger and Wright (1994), which argued that experts will only perform better than non-experts if the forecasting situations were 'ecologically valid' (the extent to which the individual is experienced at making such forecasts) and 'learnable' (the degree to which it is possible to master such a forecasting task in a useful way). If either or both are low, the authors argue, there is no reason to believe that experts will perform better than non-experts. In the case of forecasting PIS effectiveness, it is believed that the task is both ecologically valid (the experts have the necessary knowledge for such a task, collected on a daily basis in their professions) and learnable (such a task can be mastered through the study of existing and new strategies along with their levels of effectiveness). In this way, it is believed that the experts will perform better than non-experts when predicting PIS effectiveness.

This same argument however can be used to justify the premise that non-experts will perform equally regardless whether UJ or structured analogies is used. Non-experts will lack the expertise necessary to condition to the forecasting task (despite it itself being learnable) and hence benefit from the proposed error reducing characteristics of the structured analogies method. In this way, the results of this first experiment come as no real surprise as one of the premises for the potential success of the structured analogies approach is the ability to access valuable analogous information that the respondent might posses and aid him or her to structure this information so that a more accurate prediction can be made. If however, no such analogous information exists within the participant, the structured analogies technique will fail to serve its purpose and any such forecast will be little more than an unaided judgment. For this

reason, when participants are non-experts, it is normal that performance of the s-SA1 method be similar to that of unaided judgment.

Experiment II Results

Whilst the results of experiment I were expected, those of experiment II came as a bit of a surprise, considering the participants were taken from the Environment Directorate of an EU Institution. Prior to contact with the participants, their level of suitability for the task was unknown, as they were chosen in a non-random way by a contact held within the department. This was not seen as a problem because although it was a new step in testing the forecasting ability of the s-SA1 approach, the main intention behind this second experiment was to serve as a pilot study for a future experiment involving true experts. Nevertheless, it was still believed that these participants held more relevant knowledge about policy implementation strategies and were more suitable for making such predictions than their predecessors in the first experiment, leading to the belief that their forecasts would be more accurate. This however, as the results are evidence, was not the case. As explained in Savio and Nikolopoulos (2009a, 2009b, 2009c), which documents this experiment and its results, there are two main performance-constraining issues with the participants which could explain such results. These reasons ultimately led to the process of case and questionnaire refinement for experiments IV and V, as was explained in section 5.7.

The first of these was the approach many of the participants took to the exercise. Many of the participants were not at ease with being asked to produce a forecast under experiment conditions. It is still believed that these forecasts were subject to a series of motivational biases, as presented in sub-section 2.9.2.2, which proved detrimental to the accuracy of the forecasts. Motivational biases could have resulted from the following points related to the approach taken towards the exercise.

- In section 5.4 it was explained that the participants were chosen by the contact within the department on a basis of vicinity to their office. This could of course imply that although the respondents may have not been against participating, they were perhaps not overly interested or committed to the task.

- The elicitation method used in experiment II was a one-on-one interview style format which consisted of the administrator presenting each case description and then leaving the participant to make a prediction and complete the questionnaire. Whilst this was done, the administrator was still present in the room. It is hence possible that in combination with the short time span given for each case, the participant felt overly pressured during the exercise and this led to a failure to report an answer to their true capability.

The second reason was the clear lack of analogous information exhibited by the participants. In general, these had difficulty in recalling any analogies and the ones that were recalled had low similarity ratings and their outcomes were not known. As Savio and Nikolopoulos (2009c) shows, the theory behind this difficulty in recalling analogies can be explained by one, if not all, of four constructs, as figure 7.1 below illustrates.
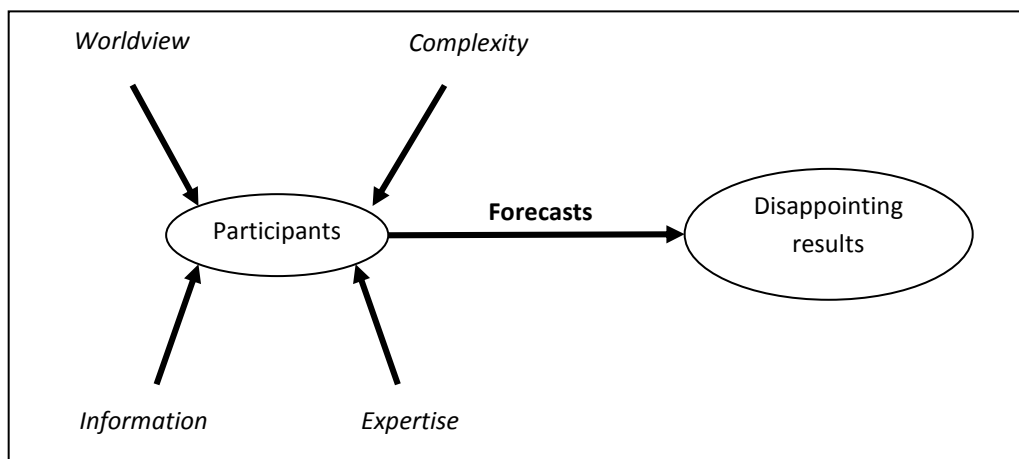
Figure 7.1: Soft systems View of Explanatory Constructs Resulting in Poor Analogy Recall (Savio and Nikolopoulos, 2009c)

Information – there exists a possibility that the case descriptions lacked the sufficient information about the policy and the proposed PIS. Some had complaints about the length and detail of the case descriptions, which were claimed to be too short and vague. The amount and nature of the information given about the real-life PIS given in the case description was heavily constrained by time issues, unlike the experiment in Green and Armstrong (2007b), which was able to provide two-page descriptions for each conflict. It is difficult to compare PIS to conflict situations because the nature of

the task means that figures would have to be included so as to help experts in the framing of the problem, which in turn would establish the correct mindset for the recall of analogies. The danger was that too much information could increase the risk of experts recognising the case and knowing its outcome.

Nonetheless, to avoid future problems of this kind and reduce the chance of any bias stemming from such a source, the nature of the information given in the case description was redesigned for experiment III, as was explained in section 5.7.

Complexity – the issues described in the previous point lead to another possibility as to why there was difficulty in recalling relevant analogies. It could also be possible that the complex nature of polices and implementation strategies make it difficult for humans to recall analogies and their outcomes. Implementation strategies are often more complicated than conflict situations in that there are more parameters associated with them, possibly rendering them more difficult to store and more importantly, to recall. Furthermore, the outcome of a conflict situation is often more straightforward for an individual to digest and then process as it can sometimes be regarded as having a binary outcome. A PIS effectiveness outcome however, although it can also be considered as a 'success' or 'failure', is more usefully measured on a continuous scale, making it more difficult to recall.

In addition, implementation strategies are not only complex in their structure, but also in the cultural context they exist in. As was seen in sub-section 2.4.5, political forces act on the decision as to what shape these strategies will take. So, not only must any potential analogy recalled by an individual be similar in its mechanical structure (similar policy objectives, similar incentives, etc.) but it must also be from a comparable cultural context. Hence, it is suspected that an expert's analogies from one cultural context might not be as applicable for helping predict the effectiveness of a PIS in a different cultural context. With that said, in order for an individual to be in a position to recall useful analogies, they must be made aware of the cultural context of the target PIS, which leaves the complicated task of attempting to convey cultural context in a case description without revealing the country in which it was implemented. Unfortunately, in an effort to eliminate experimental bias, too much

context appears to have been stripped away, hindering experts in their analogy recollection and forecasts, and calling into question the experimental design. In other words, a trade-off had to be made between two sorts of factors that contribute to validity of the experiments.

In attempt to remedy this issue, extra support was given to the participants when the structured analogies method was redesigned from s-SA1 to s-SA2, used for experiments IV and V. The new s-SA2 approach would now allow participants to match the outcomes of their analogies with categorical interval outcomes, in a bid to aid in recalling the outcome of their analogies.

Worldview – another possibility is that the culture factor may have affected analogy recall. Participants came from several different European countries, and interestingly reacted differently depending on their country of origin. Those from more Southern European member states were more comfortable with the exercise than their more Northern counterparts, and this is reflected in the former's more accurate forecasts. It is likely then that nationality and worldview affected the way in which the problem was approached and possibly the way analogies were recalled too.

Soft-systems OR is based on the premise that the way in which problems are structured is a highly subjective exercise. Any problem-structuring method will necessarily be influenced by an individual's 'world-view' or in other words, his or her social and cultural background, experience, education and personal values (Pidd, 2003). Such an argument can also plausibly be applied to forecasting as these notions are the very ones which can lead to biases in judgmental approaches if measures are not taken to minimise them (Armstrong, 1985, chapter 6; Green and Armstrong, 2007b).

Expertise – the final possibility and arguably the most influencing of the factors was that these participants did not possess the sufficient level of expertise required for making such predictions. This was not all surprising considering that despite their positions within the Environment Directorate of this large EU Institution, almost all of them specialised in areas such as 'Nature' and 'Bio-diversity' and not strictly relevant areas to the case studies. So whilst some knowledge of environmental policy relevant

to the case studies was possible, they simply lacked the internal database of analogies to the implementation strategies presented in the cases, needed to fully capitalise on the advantages the s-SA approach can offer.

What's more, the participants were thought to possess an 'intermediate' level of knowledge about such strategies which resulted in a source of cognitive bias, namely due to the use of heuristics. Their level of knowledge is believed to have been high enough to apply rationale and logic to the exercise but not so high as to be able to recall useful information, analogous or other, to aid in the prediction process. Such an occurrence is thought to be the reason as to why a particularly high number of participants resorted to the use of heuristics. Faced with such an exercise but the inability to recall useful analogies, participants had no choice but to call on such ad-hoc techniques to approach the problem with.

Heuristics are a natural way of facing complex problems that humans have (Kahneman et al., 1982; Pidd, 2003). That is, humans will naturally resort to heuristics when faced with structuring a complex problem if no support is given or if no other alternative is available. However, such heuristics will lead to biased results due to their inherently subjective nature (Kahneman et al., 1982; Bolger and Harvey, 1998).

These issues lead back to the discussion, or issue rather, of identifying experts and more importantly their level of expertise. As was discussed in sub-section 2.9.1, identifying an expert for the purpose of a study is very difficult and sometimes it is only ex-post that one realises that the participants were not as knowledgeable as anticipated. The solution to this would then be to locate true experts from sources where past research has shown experts can be found, such as in academia, research centres and specialised governmental agencies (Savio and Nikolopoulos, 2009c), which is precisely what was done for experiment V.

7.2.1.2 Experiments IV and V
Experiments IV and V were the two final experiments of the study which tested the structured analogies approach (in the form of s-SA2) in the hands of non-experts and experts respectively. The cases used were different to the ones in the first two

experiments in that they focused on a particular type of PIS, the vehicle scrappage scheme, and that they had also gone through a process of refining in order to minimise motivational and cognitive biases. The measures for doing this were covered in section 5.7 and included standardising the information given in the case description, a new format to the analogy table to facilitate the analogy recall process, the categorisation of the analogy outcomes to aid in analogy outcome recall, etc.

Experiment IV Results

The results of experiment IV are mixed. The results in tables 6.7 and 6.8 suggest the superiority of the UJ as it performs better than the s-SA2 in more than 2/3 of the combinations. This is supported by the overall GMRAE in table 6.10, which indicates UJ is better over the three cases as it has a lower value. However, the GMRMAE which compares the overall performance of the s-SA2 relative to the UJ gives a value of 0.99, which indicates very similar performance between each approach (a figure of 1.00 being perfect equality).

If the results are considered on an individual case basis, this GMRMAE can be explained fairly easily. The UJ method is more accurate in cases 1 and 3 but considerably less so in case 2. The greater difference in accuracy between s-SA2 and UJ in case 2 than in case 1 and 3, means that when all cases are considered, performance is shown to be relatively on par.

In order to investigate the potential reason(s) for this pattern in performance, attention will be given to tables 6.16, 6.17 and 6.18 which present the details of the SA predictions[1]. From these tables it is clear that in none of the cases did participants recall useful analogies[2] and use these to make a prediction. Furthermore, there was only agreement[3] between the SA forecast and the s-SA2 outcome forecast produced by each individual on 15.4% of occasions. Furthermore, these tables also show that

---

[1] It should be recalled that these SA forecasts were derived using a mechanical rule which selected the analogy with the highest similarity rating as the SA prediction. It should also be recalled that due to the nature of the analogy table used, the SA prediction is always an interval forecast.
[2] Useful analogies are analogies which suggest the correct outcome interval.
[3] By 'agreement' it is meant that the SA outcome interval prediction was the same as the s-SA2 outcome forecast.

the analogies they recalled were no good at making a prediction anyhow as the SA only predicted the correct interval 23.1% of the time. Had these tables shown the analogies to have been used properly in case 2 but not so in cases 1 and 3 then this might have been an explanation for the pattern seen. This was not the case however, as these tables show a poor level of the analogies (poor SA forecasts) and poor following of these (poor agreement percentage).

Hence, it must be concluded that there does not seem to be any rational explanation for such a pattern in performance. That is to say, there is no clear reason why UJ has performed better than s-SA2 in cases 1 and 3 and vice versa in case 2.

What is important is that the parity in performance between the two approaches in the hands of non-experts corresponds and further validates the results from the other experiment with non-experts, experiment I.

Experiment V Results

Experiment V was the final experiment of the study which would compare the performance of the two approaches. In order to avoid the problems encountered in experiment II, the process of finding expert participants was invested considerable time and effort, as documented in section 5.9.

The results of this experiment show a greater consistency than those of experiment IV. Tables 6.11 and 6.12 show that s-SA2 performs better than UJ in almost 2/3 of the combinations. The results of GMRMAE and GMRAE are consistent with this as they too show an overall superiority of the s-SA2 approach. In addition, when all of the accuracy measures are considered and a rank is given to the two approaches in order of performance, as in table 6.15, the s-SA2 is the outright winner. These results then are evidence to suggest support the claim made in the first research question set out by this thesis, that the structured use of analogies improves forecast accuracy when predicting PIS effectiveness.

For experiment V, changes were made based on feedback from previous experiments and considerable effort was put into identifying real experts and this is believed to

have addressed many of the problems presented in sub-section 7.1.1 as reasons for the poor structured analogies performance. In particular, the main reason for the better performance of the s-SA2 in experiment V when compared to experiment II is thought to be attributed to the greater quality of the experts. It is also believed that the greater level of structuring in s-SA2 when compared to s-SA1 played a fundamental part in the improved performance of the structured analogies approach in experiment V (see sub-section 7.2.2).

Nevertheless, every accuracy measure showed that despite s-SA2 performing better overall, the UJ did consistently perform better in case 3, and by some margin. All of the accuracy measures showed s-SA2 superiority by a significant amount in cases 1 and 2 but the inverse was true for case 3. It is due to the results from case 3 that there is not a bigger difference in the overall superiority of the s-SA2 when compared to the UJ.

As was the case in the results of experiment IV, where a similar situation was found, it is difficult to explain why such a phenomenon arises. However, as was done for experiment IV, inspection of the SA performance reveals some interesting results. Tables 6.19, 6.20 and 6.21 show how the SA forecasts compared to the s-SA2 outcome interval forecasts produced by each individual who used the s-SA2 approach, for each of the three cases. There was agreement between the SA forecast and the s-SA2 outcome forecast produced by each individual on 60.0% of occasions and the SA forecast was correct in 53.3% of the time.

- In case 1, the SA forecast for 4 of the 5 participants was correct[1]. In all 4 of these cases, the participants followed[2] their analogies and predicted the outcome correctly. Only one participant made a different prediction to what SA suggested, but got it right.

---

[1] That is to say, the outcome interval predicted by the SA method was correct.
[2] By 'followed their analogies', what is meant is that the individual obeyed what the analogies were suggesting and chose the outcome interval forecast of their most similar analogy. A case in which an individual did not 'follow their analogies' would be one in which the individual chose an outcome forecast which was different to that of their most similar analogy, thereby disobeying what the analogies were suggesting.

- In case 2, the SA forecast for 4 of the 5 participants was correct. However, in only 2 of these cases did the participants follow their analogies and predict the outcome correctly. In the other 2 cases, the participants predicted the outcome just above the one suggested by the SA. One participant followed their analogies but got it wrong.

- In case 3 however, the SA forecast was correct for 0 of the 5 participants. Only 2 of the 5 participants followed their analogies when making an outcome prediction, but none were right. In 2 of the 5 participants, the correct outcome prediction was made whilst disobeying the analogies.

What seems clear from these results is that generally when the SA forecasts are correct, as in cases 1 and 2, the performance of the s-SA2 will be good. When the SA forecasts are bad, so will be the s-SA2 forecasts. To illustrate this, if the MAE of the s-SA2 forecasts given a correct SA forecast is calculated, it is 4.96. If the MAE of the s-SA2 forecasts given an incorrect SA forecast is calculated, it is 24.30. In other words, a correct initial SA prediction will lead to a more accurate s-SA outcome prediction and consequently a more accurate s-SA2 point prediction. This indicates that when an expert has knowledge of, and is in the right frame of mind to recall, useful analogies to the target case, this will help them produce a better s-SA2 forecast.

Although such a result seems fairly trivial, it is not so because as the results in case 2 show, sometimes the participants will think of useful analogies but they will not necessarily obey this information when making a prediction. In case 1 whenever the SA forecast was correct the participant obeyed the information for the s-SA2 prediction, leading to an RMAE of 0.62. In case 2, where the SA forecasts were still suggesting the correct answer but only half of the participants obeyed this information to produce an s-SA2 forecast, the performance declined with an RMAE of 0.80. So not only must useful analogies be recalled but the participant must also obey them in order for the s-SA2 approach to produce better forecasts than UJ. Thus, the more participants follow the analogous information the better the s-SA2 results.

The question must then be raised as to why the analogies recalled in cases 1 and 2 were suggesting the correct answer and those for case 3, were not. Although this is a very difficult question which can probably only really be answered through feedback from the expert participants involved, a few theories arise.

- One possibility could be associated with the way in which the experts framed the cases. The idea of the framing of the case is related to the notions of an individual's worldview and problem-structuring methods, described earlier. The subjective nature of such concepts will mean that an individual's worldview will affect the way in which the problem is framed and then approached (Mingers and Rosenhead, 2004).

  In the email sent to which the cases were attached to participants having agreed to partake, instructions were given to complete the cases in chronological order. Assuming that this instruction was followed, it is likely that the experts, having developed an approach frame for case 1, used the same frame for cases 2 and 3. This is evidenced by the fact that many experts used some of the same analogies for all three cases. This is thought to not have affected performance for cases 1 and 2, considering the strong similarities in their nature. However the substantial differences between cases 1 and 2 and case 3 in two key aspects meant that any failure to adapt the frame could result in the malfunction of the s-SA2 approach, which is believed to have been the case for most experts. These two key differences are as follows.

  (1) Case 3 was different to cases 1 and 2 in the type of PIS effectiveness forecast that was being asked for. In cases 1 and 2, the questionnaire asked for a prediction in terms of the percentage of vehicles (either scrapped or registered), whereas case 3 asked for a prediction in terms of what percentage of the budget would be used.

  Hence, having established their frame in cases 1 and 2, the experts did not adapt their analogies for the change in the measure of PIS effectiveness in case 3 and this is evidenced in their SA predictions.

(2) Case 3 was also different to case 1 and 2 in the scale of the strategy described, in terms of size of incentive and overall budget, which was significantly larger in case 3 than cases 1 and 2. With the participants having developed a frame calibrated to smaller scales for cases 1 and 2, they might have struggled to make the jump in magnitude and adjust their frame, which could have accounted for their erroneous analogies.

- If again it is assumed that the experts completed the cases in chronological order as they were instructed to, another possibility could have been that this very order can be used to explain the poor s-SA2 forecasts in case 3. It is possible that due to it being the third case, participants would have been tired by this point and this lack of motivation could have affected their willingness to recall yet more analogies.

Naturally it can be argued that the factors just described would equally affect the UJ forecasts but it is believed that due to the nature of the s-SA2 approach, which is more demanding on an information and thought-processing level, their negative effects on forecasts would have been magnified.

These arguments can therefore be used to explain why such a pattern of performance was seen across the three cases. A strong s-SA2 performance was seen in cases 1 and 2 because participants were able to recall useful analogies and use this information in their predictions. In case 3, the analogies were not so useful, leading to s-SA2 forecasts which were worse than those of UJ.

In conclusion, the results of experiment V provide evidence for RQ1.0 in that the structured use of analogies does indeed improve the accuracy of PIS effectiveness forecasts when produced by true experts. Experiment IV (and to a lesser extent, experiment I) showed that in the hands of non-experts, there is no significant difference between the performance of UJ and structured analogies, as to be expected. The results of experiment V however, also went on to show that when a true expert is able to frame the problem appropriately and has enough knowledge to recall analogies to the target PIS (thereby generating a correct SA interval prediction), the accuracy of

his or her s-SA2 prediction will improve. In addition, the results indicate that the more participants follow the SA prediction the better their s-SA2 predictions will be.

7.2.2 RQ1.1

In the experiments conducted as part of this research, two forms of the structured analogies method were tested. These were the s-SA1 (used in experiments I and II) and s-SA2 (used in experiments IV and V), where s-SA2 was argued to be more structured than s-SA1.

Having seen in sub-section 7.2.1 that there is evidence to suggest that in the hands of experts, the structured analogies method performs better than UJ for predicting PIS effectiveness, the second research question set out by this thesis suggested that the level of structuring of the structured analogies method in the hands of experts would also affect their forecasting performance. That is, RQ1.1 suggested that the more structured the structured analogies approach, the better it would perform in comparison to UJ.

The results of the experiments show that UJ works better than s-SA1 in experiments I and II and although UJ works well in IV and V, it is not as good as the s-SA2 approach.

This indicates that forecast accuracy of the structured analogies approach improves when participants are given extra support in that they are made to match the outcome of their analogies with a categorical outcome (albeit an interval). This suggestion is supported by the result found in the previous sub-section in that when the SA prediction was correct (i.e. when the experts' most similar analogy matched the correct interval outcome) the s-SA2 was almost always more accurate than UJ.

7.3 RQ2

7.3.1 RQ2.0

The second research question put forward by this thesis suggests that level of expertise is correlated to s-SA forecast accuracy (RQ2.0). More specifically, it is thought that the higher an individual's level of expertise, the better their prediction

will be: comparison of the results of experiments IV and V suggest that this is indeed true. The results in sub-section 6.2.4 and 6.2.5 show that the forecast errors of the participants in experiment V, considered to be the experts, are on average considerably lower than the forecast errors in experiment IV, the non-experts.

Recalling the discussion in sub-section 2.9.1.2, it was concluded that the measurement of an individual's level of expertise is a very difficult task indeed. One attempt to do this was given by Shanteau et al. (2002) and consisted of considering two characteristics which are believed to be present in experts, discrimination and consistency, and establishing a ratio, the CWS, which would indicate a level of expert quality. The efforts of Shanteau et al. (2002) and other studies (Johnson, 1983; Shanteau, 1987) to attempt to define the ideal characteristics which make an individual an expert are contrasted with the efforts of studies such as Bolger and Wright (1994), which attempt to study the factors which make the information extracted from would-be experts, expertise. As was seen earlier, Bolger and Wright (1994) proposed two theoretical constructs which enable the prediction of expert performance in different task domains: if a forecasting task is ecologically valid and learnable, an expert will perform better than a non-expert.

So, on the one side, there is work such as Shanteau et al. (2002) which assume that once experts have been identified, these will be capable of producing expertise and so the research objective is to establish such identification methods, such as their CWS ratio. On the other side, there are studies such as Bolger and Wright (1994), which attempt to study the conditions needed for an expert to perform better, and hence be distinguished from, a non-expert. On either side, the aim is to find a way to demarcate experts and non-experts, either through conditions of performance or common characteristics.

In the case of this research, by the approach documented in section 5.9, the participants selected for experiment V were considered true experts, and as has been seen, their forecasts were more accurate than the non-experts in experiment IV, thereby providing strong evidence of what was put forward by RQ2.0.

What's more, this result provides further evidence for Bolger and Wright (1994)'s hypothesis that experts will perform better than non-expert in a forecasting task which is ecologically valid and learnable, as forecasting PIS effectiveness was reasoned to be at the beginning of this chapter.

7.3.2 RQ2.1

More specifically, RQ2.1 suggests that level of experience within the experts is correlated to s-SA performance. In order to investigate this question the forecast accuracies according to AE and APE of all the participants were analysed in comparison to the experience rating given to each of them, the results of which are presented in section 6.2.

Figures 6.1 to 6.4 indicate that for neither the s-SA participants nor the UJ participants when AE and APE is considered, does such a correlation exist. These results would then tend to suggest that level of experience within the experts does not have an effect on forecast accuracy for either of the methods.

Nonetheless, they must not be considered completely conclusive. There is a strong possibility that the experience rating given to each participant was not a correct representation of reality, despite every effort to make it so. Such experience ratings are very difficult to produce and will necessarily always include a factor of subjectivity. The experience factor was derived from the participants' own answers about their experience and suitability to the task, given in the questionnaire of each case.

Interestingly, these results agree with those found by Green and Armstrong (2007b) when they compared experience level with forecast accuracy. The authors admit that this result came as a surprise and that further investigation is required. Although experience is a common method for identifying experts (see sub-section 2.9.1.1), the results of this study seem to provide further evidence that experience seems to be a weak criterion for the selection of experts for forecasting with structured analogies.

However, both Green and Armstrong (2007b) and this study used very similar criteria for assessing experience, and this could be a reason for the observed results. Both

studies based their experience ratings on questions regarding experience and suitability for the task in the questionnaires. Hence these experience ratings are very subjective and can be ultimately affected by personality factors such as the experts' self-belief. Thus, it will be of interest to explore different methods of assessing experience and seeing how this affects the results.

In sum, although two studies have now provided evidence against it, further research has to be done into measuring experience level and only when an appropriate method has been developed can such a research question be fully and fairly investigated.

7.3.3 RQ2.2

As the third part of this second research question, it was also put forward that, in similar fashion to Green and Armstrong (2007b), that the greater the number of analogies recalled by the experts, the better the forecasts. The results of this investigation can be found in table 6.26 and figure 6.5. The results of RQ2.2 are in line with that of RQ2.1 in that there seems to be no correlation between the number of analogies recalled and forecast accuracy.

Unlike the results for RQ2.1, the results of RQ2.2 are in contrast with the ones found in Green and Armstrong (2007b). They had found that for six of their eight conflict situations, the experts who could think of two or more analogies were more accurate than those who could think of only one. In this study, figure 6.5 shows that this is clearly not the case. Thus, it seems that when using s-SA to predict PIS effectiveness, the ability to recall several analogies seems to make little difference in forecast accuracy.

If the discussion into the measurement of expertise level of an expert from sub-section 2.9.1.2 and above in RQ2.0 are recalled, the results of RQ2.1 and RQ2.2 are interesting. It was explained that such an objective measurement of the quality of the information output from an expert is extremely difficult, if not impossible, due to the complex nature of the notion of expertise. What's more, few methods aside from the CWS ratio (Shanteau et al., 2002) exist for such a purpose in the little literature that has been published on this difficult issue.

The results of RQ2.1 and RQ2.2 become interesting in so much as experience and number of analogies recalled can be considered as two logical constructs which can be considered as defining an expert and level of expertise for the purpose of forecasting PIS effectiveness. This is similarly the case for the study in conflict forecasting of Green and Armstrong (2007b). Hence, if it is assumed that experience level and number of analogies recalled are indicators of the expertise level (which is a logical assumption to make given the situation) of the experts, then the results of this study would suggest that level of expertise within the experts has little effect on forecast accuracy in PIS effectiveness. In the case of Green and Armstrong (2007b) however, such a suggestion is more difficult due to the conflicting evidence between the effect that experience level and the number of analogies recalled has on forecast accuracy. Nonetheless, before any statements are made about the correlation between expertise level and forecast accuracy, a more complete investigation has to be made which is able to measure such a construct in an appropriate way.

7.4 RQ3

The third and final research question of this thesis suggests that forecasts made by experts are no worse than those by made the governments who developed these implementation strategies. Section 6.3 summarises the results in comparing the accuracy of the predictions from the UJ approach, the s-SA approach and the local government using the MAPE, MdAPE, MAE, MdAE and the GMRAE.

For case 1, the forecast produced by the Canadian Government is better than those produced by the UJ and s-SA approach across all accuracy measures. In case 2 however, according to MAPE and MAE, the French Government's prediction was better than UJ but worse than that of s-SA. Finally, in case 3, both the UJ and s-SA forecasts were better than that of the United States Government, across all accuracy measures.

In case 1, the Canadian Government predicted that that by the first twelve months of the scheme, 50,000 vehicles would have been scrapped. Assuming a linear sales model, they expected 29,166 scrapped vehicles in the first seven months. The actual

number of scrapped vehicles by that point was observed to be 15,000 (of 4M). Interestingly, after the cases for experiment V had been designed, the number of scrapped vehicles through the scheme after twelve months was very close to the predicted 50,000 mark. This, in combination with the good forecast results seen in table 8.27, indicates that however the prediction was arrived at, it was done so successfully.

In case 2, the French Government predicted that 28% (250,000 of 900,000) of all new vehicles registered after a year will have been done so through the scheme. Assuming a linear sales model, 104,000 were expected in the first five months. The number of new vehicles registered through the scheme was observed to be 19.5% (175,000 of 900,000) in the first five months. The true figure after a year was indeed very close to the forecast, 26.5% (600,000 of 2,228M) (Lexpansion.com, 2010). So, like case 1, the Government forecast for case 2 was also good.

In case 3, the United States Government predicted that $1000M would be spent in the first four months of the Cash for Clunkers scheme. The actual amount spent after four months was $2,800M.

So, in case 1, the expert forecasts were worse than the Government prediction even in a prediction of PIS effectiveness after seven months. In case 2, although the expert forecasts were better than the Government prediction over mid-strategy horizon, these then improved to be quite accurate over the lifetime of the scheme. In case 3, the expert predictions were seen to be considerably better than those of the United States Government.

When the cases were presented in sub-section 5.1.2, it was stated that little was known about how the forecasts were generated. For cases 1 and 2, the predictions were found through Governmental press releases as no detailed information in the form of reports, proposals or impact assessments were found. Attempts were made to contact the publishers in both cases but no reply was received. For case 3, a policy report from the US Government National Highway Traffic Safety Administration (NHTSA, 2009) was retrieved. Unfortunately, despite giving predictions as to the

anticipated effectiveness of the PIS in terms of vehicle sales, details into how these predictions were made are unknown.

Such a lack of information into how these forecasts were produced makes it difficult to discuss their validity. However, the results of the literature review indicate that these predictions are almost certainly the result of some sort of IA or CBA which was reasoned to be in support of the policy proposal (see section 2.9).

These mixed results provide evidence to suggest that RQ3, expert forecasts are no worse than Governmental forecasts, is true. Although comparisons between the expert forecasts and the Government forecasts are useful as a means for benchmarking one against the other and allow for investigation of RQ3, it must remembered that they are not rivals. The expert forecasts, as proposed in this research, are intended as a quick and inexpensive means for making PIS effectiveness predictions which will allow for a screening of alternative PIS which in turn will provide a selection of potentially effective PIS for further analysis (through IA/CBA ex-ante evaluative techniques).

Nonetheless, it is all the more satisfying and encouraging seeing that these predictions are *comparable* in accuracy to approaches currently used by Governments which undoubtedly have been invested considerably more time, money and effort. This is an important result because as was explained in section 3.5, in order for policy-makers to be confident of such a screening process in selecting potentially effective PIS, the forecasts produced by the structured analogies approach must exhibit a good level of accuracy. Until the exact details of how these kinds of Governmental forecasts are derived, unfortunately, any comparative analysis between the two approaches cannot really go any further.

7.5 Conclusions

The discussion of the results presented in chapter 7 can be summarised with the following three points.

- The results of the experiment V provide evidence that a structured approach to the use of analogies improves accuracy of PIS effectiveness forecasts when produced by experts, in comparison to an unaided judgment approach (RQ1.0). The results showed that a condition for this is that the forecaster be a true expert. Furthermore, if the expert is able to frame the problem appropriately and has enough knowledge to recall analogous strategies to the target PIS, the accuracy of his or her structured analogies prediction will improve. The results also indicate that the more participants follow the SA prediction, the better their structured analogies predictions will be. In addition the results of the experiments show that the more support given to experts when structuring their analogies, the better the structured analogies forecasts will be when compared to UJ (RQ1.1).

- The results of experiments IV and V are evidence that level of expertise does indeed affect forecast accuracy in a positive way (RQ2.0). That is, experts were found to produce considerably more accurate forecasts than non-experts. However, the level of the experts' experience was not found to be correlated with forecast accuracy (RQ2.1). However, this evidence was considered to be inconclusive due to the subjective manner in which experience was rated, which could not be a reflection of reality. Similarly, the number of analogies recalled was not found to be correlated with forecast accuracy either (RQ2.2).

- The results of experiment V suggest that in some cases, the expert forecasts are no worse than those produced by the government (RQ3). Overall, the expert forecasts are comparable to those produced by the government. This is an important result when the role that PIS effectiveness forecasts are proposed to play is considered. That is, policy-makers must have confidence in such an approach for screening alternative strategies in search of potentially effective ones requiring further analysis if they are to be used for such a purpose.

Chapter 8: Conclusions, Limitations and Further Research

The investigation carried out as part of this thesis, which can be defined broadly as policy forecasting research into the decision-making process for policy formulation, revealed some very interesting results. Such results have important implications on two fronts. Firstly, these results will be significant to governments interested in improving their decision-making capabilities at when formulating policies. Secondly, such results provide the next step in the long line of judgmental forecasting approaches research and more specifically that of the structured use of analogies. Moreover, these findings will provide new insight into these approaches as well providing a new rational basis for future research. Thus, this chapter will firstly attempt to summarise the main findings and conclusions of this research. Secondly, the limitations under which the results of this research should be considered will be considered. Finally new possible directions for future research, based on these conclusions will be suggested.

8.1 Summary of Conclusions

First, a bullet-point summary of the main conclusions of this research:

- The literature review in chapter 2 showed that popular forecasting practices in public policy come in the form of ex-ante evaluation techniques which are used for anticipating the impact of a policy or as an aid in the planning and design of the implementation process. In other words, their use has been seen mainly for producing forecasts related to implementation on an operational level.

Furthermore, some evidence that these were also used for defining implementation strategy was also found in the literature.

- The review of judgmental forecasting approaches in chapter 2 found that these can provide much positive insight into a forecasting task where such information is considered to be valuable, and this can lead to an improvement in forecast accuracy. Judgmental approaches are often attractive for this reason but also because they provide a good alternative when no quantitative data is available and also because they are relatively quick and inexpensive to implement. The literature review found no evidence of such approaches currently being used in governmental decision-making.

- The review of the literature in chapter 2 also found that a particular kind of judgmental forecasting approach, based on the use of analogies and referred to as 'forecasting by analogy' (FBA), has been shown to be useful in situations which are difficult to forecast, in which very limited past quantitative data is available, there is a high level of uncertainty and when the use of analogous information is instinctive. However, FBA, like any judgmental approach, was found to suffer from the limitations and biases present in all individuals. Hence, it was argued that a structured approach to the use of analogies is preferred as this was shown to improve forecast accuracy. Furthermore, no evidence of the use of a structured analogies approach for supporting governmental decision-making when defining implementation strategy was found either.

- Chapter 3 proposed that the notion of a Policy Implementation Strategy (PIS) was a more realistic representation of the how implementation strategy is formulated in government. It was argued that technical criteria such as effectiveness can be used to select PIS in the same way it such a criterion has been seen to help decide implementation strategy. Hence it is proposed that PIS effectiveness forecasts be used as a decision-support tool for defining implementation strategy. Such forecasts could be used for a screening of all possible PIS available to identify the most promising, which can then be subject to a more rigorous analysis through ex-ante evaluative approaches such as IA and CBA. Furthermore, it is proposed that a

structured analogies approach be used as the method for producing these PIS effectiveness forecasts because it is quick and inexpensive to implement and is also suited to the nature of such a forecasting task.

- The results of experiment V are evidence that a structured approach to the use of analogies improves accuracy of PIS effectiveness forecasts when compared to an unaided judgment approach (RQ1). In order for this to be true however, the results showed that the forecasters must be a true expert. In addition, if the expert is able to frame the problem in a suitable fashion and has enough knowledge to recall analogous strategies to the target PIS, the accuracy of their s-SA prediction will increase. The results also indicate that the more the forecasters follow their SA prediction, the better their s-SA forecast will be. What's more, the results of the experiments show that the more support given to experts when structuring their analogies, the better the structured analogies forecasts will be when compared to UJ (RQ1.1).

- The results of experiments IV and V provide evidence in support of the proposition that level of expertise has a positive effect on forecast accuracy (RQ2.0). In other words, the experts were found to produce considerably more accurate forecasts than the non-experts. In contrast, the level of the experts' experience was not found to be correlated with the accuracy of their predictions (RQ2.1). However, this evidence was considered to be inconclusive because of the subjective method in which experience of each expert was rated, which it is believed could not be a fair representation of reality. Similarly, the number of analogies recalled by the experts was not found to be correlated with the accuracy of their predictions either (RQ2.2).

- The results of experiment V also suggest that in some cases, the forecasts produced by the experts are no worse than those produced by the government (RQ3). Moreover, overall, an important result is that the experts' forecasts are comparable to those produced by the government. This was considered a very important result because if these PIS effectiveness predictions are to be used for

screening potential PIS, as is proposed, policy-makers must have confidence in their ability to make such assessments.

From these conclusions, four main aspects in which this thesis has provided a contribution to knowledge can be identified:

1. It has built on the evidence in the literature that showed that ex-ante evaluative techniques such as IA and CBA are not only used for designing and planning the policy implementation process, but also for helping define implementation strategy.

2. It has provided evidence to suggest that when used by true experts, a structured analogies approach to making PIS effectiveness predictions is more accurate than an unaided judgment approach, providing further evidence to the results found in Green and Armstrong (2007b).

3. It has provided evidence to suggest that level of expertise is positively correlated to accuracy of PIS effectiveness forecasts. The true expert forecasts were seen to always outperform the non-experts forecasts in a situation which is both ecologically valid and learnable, providing further evidence of the theory put forward by Bolger and Wright (1994).

4. It has presented evidence for the suitability of expert-based PIS effectiveness predictions as a decision-support tool in government for defining implementation strategy because not only are they quick and inexpensive to implement, but they are also comparable in accuracy to the predictions made by government.

8.2 Limitations

In order to provide a better frame for the results found, it is necessary to consider the limitations to which they are bound. The following points can be considered as the primary limitations of the study.

- As suggested in Armstrong (2007), what should be reported together with any forecasting results are effect sizes, rather than statistical significance tests. This effect size can be regarded as one of the limitations of this research. Particularly

for experiment V, time and participant availability constraints meant that only 10 sets of results could be collected. Although this is in line with the number of experts typically used for such exercises in the UK and EU Governments (as reported in chapter 3) and hence is able to provide valid results, it is still too small a sample size to consider any of these results as conclusive. Following the critical realist position taken in this research, they are to be considered as evidence, quite strong in some cases, that needs to be validated further through replication in the aim of providing enough evidence to ultimately develop a generalised theory.

- Like in any study, there will always be limitations in the design of the experiments. In the case of this research, although all decisions were subject to and hence dictated by time and availability constraints, it is possible that results would be sensitive to the following parameters,

  - The time given to participants for providing responses.

  - The selection of participants.

  - The choice of case studies used.

  - The preparation of the cases (length of description, information given in the description, layout, order of questions asked, etc.)

8.3 Further Research

Based on the results and conclusions of this thesis, three suggestions for future work in the line of this research can be made.

8.3.1 s-SA Obey Your Analogies

The results of experiment V not only showed that s-SA2 was more accurate than UJ in predicting PIS effectiveness but also that s-SA2 predictions tend to be more accurate when the expert obeys their analogies. In other words, the results of this study suggest that accuracy is improved if the expert makes a point forecast within the outcome interval suggested by their most similar analogy to the target case. Hence, it

would be interesting to develop on this point and see if there really is a correlation between SA prediction and s-SA2 accuracy.

In the view of such a correlation existing, the s-SA2 approach could be modified to include a step which made experts focus on a point prediction within the interval suggested by their analogies[1]. If found to work, this new semi-structured analogies approach, which will be called s-SA3, could be adapted to include such an anchor in the following way:

Step 1	Description of target situation
Step 2	Analogy recall through the analogy table which asks for descriptions and similarity ratings
Step 3	Match analogy outcomes with target outcomes
Step 4	Forecast PIS effectiveness (interval) *by choosing outcome suggested by most similar analogy*
Step 5	Forecast PIS effectiveness (point) *within the interval from step 4*

Table 8.1: Proposed s-SA3 Approach to Include Step for Obeying Analogies

8.3.2 The New Approach – Hybrid SA-Delphi

A potential next step in this line of research could be to investigate the forecasting ability of the structured analogies approach when experts are allowed to interact with each other. The idea is to have a Delphi style approach, via a web-based application through which the rounds will be performed, in combination with the s-SA2or s-SA3 method.

---

[1] Just as in Green and Armstrong (2007b)'s structured analogies approach, any mechanical rule could be used here to derive a forecast from the analogies. In this case, the outcome of the most similar analogy will be considered as the outcome 'suggested by the analogies'.

The rationale for the proposal of the new approach stems directly from the results of the experiments carried out in this research. The results from experiment I whom in addition to individual forecasts, participants were asked to make predictions in a group, accuracy was seen to improve. Despite being used in the hands of non-experts, such a result leads to the suggestion that perhaps if experts were allowed to interact, and create a situation in which their analogies could be pooled and their discussion encouraged, PIS effectiveness predictions could be improved. The Delphi method, which is a systematic and interactive forecasting approach which uses a group of experts, is considered an appropriate and unbiased system by which such analogy pooling can occur.

Furthermore, the results of this thesis underlined the need for 'true' experts for producing PIS effectiveness forecasts. Hence, as was done for experiment V, an extended investment will be made in order to ensure that the participants used will have the necessary knowledge of analogies so as to capitalise fully on the structured analogies approach.

One lesson that which can be taken from experiment V is that stringent time and availability constraints often limit the level of participation an expert can have. The possibility of assembling enough experts in one place to conduct such a Delphi procedure is very optimistic and realistically very improbable. For this reason, a web-based application is considered an ideal platform for the interaction between the administrator and the participants. In this way the experiment can be carried out without the need for any physical displacements. It also means that the selection of participants is not constrained geographically, thereby dramatically enlarging the sampling frame to a global scale.

Thus, a web-based Delphi style expert interaction system combined with the use of analogies by real experts is considered to be a favourable approach for predicting PIS effectiveness.

The logistics of the new approach are fairly straightforward. The technique, which consists of combining the Delphi method and semi-structured analogies, can be set up as follows.
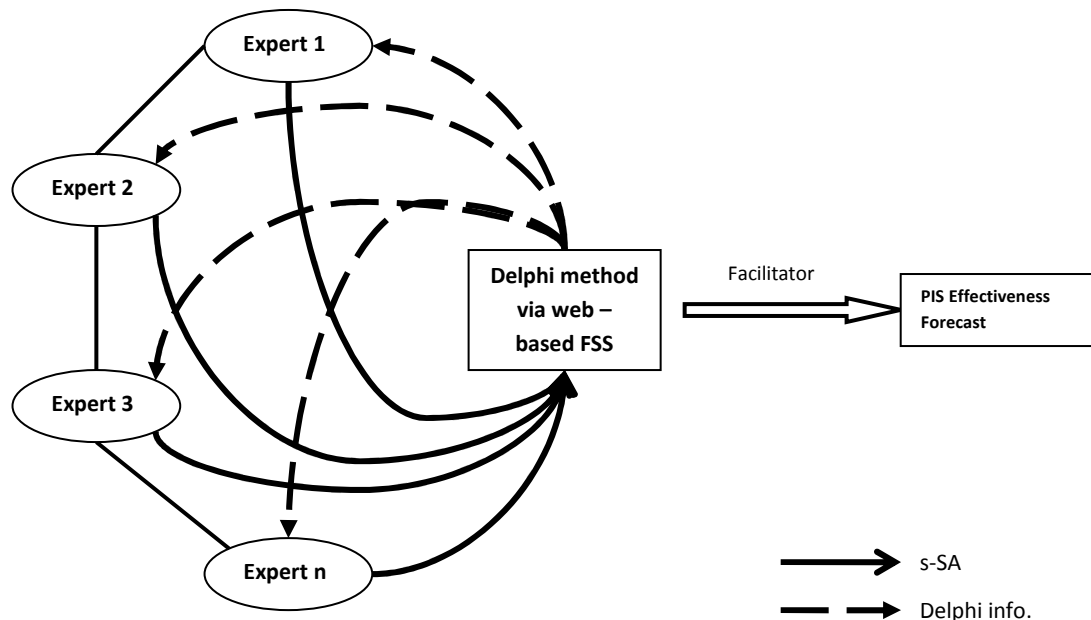

Figure 8.1: A Graphical Representation of the Hybrid SA-Delphi Method

As can be seen in Figure 8.1, **n** experts are presented with the PIS case and asked to apply the principles of the structured analogies approach, i.e. to provide details of analogous PIS, including a brief description of the similarities and differences, a similarity rating and an outcome (where possible). This is done so via a web-based application[1], which is monitored and co-ordinated by the facilitator. After this first round of the Delphi method, the information gathered is fed back to the experts and these are given the opportunity to change their predictions. This process is repeated until the experts no longer wish to make any changes. Finally, the facilitator will use all of the information provided by the experts during the Delphi rounds to derive a final forecast for PIS effectiveness.

---

[1] The format of this software is still to be decided.

8.3.3 Integrating Structured Analogies with Government Forecasts

One of the main results to emerge from forecasting research is the notion that judgmental and quantitative forecasts can be combined to improve accuracy (Bunn and Wright, 1991; Armstrong and Collopy, 1998; Armstrong, 2001b). The idea is to create an approach where the advantages of each are complemented and the limitations of each are minimised (Makridakis et al., 1998). Thus, it would be interesting to see if somehow the structured analogies approach could be combined with any quantitative method used by the government.

References

Abdolmohammadi, M. J. and Shanteau, J. (1992). "Personal Characteristics of Expert Auditors", *Organizational Behaviour and Human Decision Processes* 53 (2), pp. 158-172.

Acutt, M. and Dodgson, J. (1996). "Policy Instruments and Greenhouse Gas Emissions from Transport in the UK", *Fiscal Studies* 17 (2), pp. 65-82.

Allen, G. P. and Fildes, R. (2001). "Econometric Forecasting", in J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers: Norwell, Massachusetts, pp. 303-362.

Angelis, L. and Stamelos, I (2000). "A Simulation Tool for Efficient Analogy Based Cost Estimation", *Empirical Software Engineering* 5 (1), pp. 35-68.

Armstrong, J. S. (1984). "Forecasting by Extrapolation: Conclusions from 25 years of research", *Interfaces* 14 (6), pp. 52-66.

Armstrong, J. S. (1985). *Long-Range Forecasting: From Crystal Ball to Computer*. John Wiley & Sons: New York.

Armstrong, J. S. (1986). "The Ombudsman: Research on Forecasting: A Quarter-Century Review, 1960-1984", *Interfaces* 16 (1), pp. 89-109.

Armstrong, J. S., Ed. (2001a). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers: Norwell, Massachusetts.

Armstrong, J. S. (2001b). "Combining Forecasts", in J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers: Norwell, Massachusetts, pp. 417-439.

Armstrong, J. S. (2001c). "Standards and Practices for Forecasting", in J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers: Norwell, Massachusetts.

Armstrong, J. S. (2007). "Significance Tests Harm Progress in Forecasting", *International Journal of Forecasting*, 23 (2), pp. 321-327.

Armstrong, J. S. (1998). "Integration of Statistical Methods and Judgment for Time Series Forecasting: Principles from Empirical Research", in G. Wright and P. Goodwin, *Forecasting with Judgment*, John Wiley: Chichester, pp. 269-293.

Armstrong, J. S., Green, K. C. and Soon, W. (2008). "Polar Bear Population Forecasts: A Public-Policy Forecasting Audit", *Interfaces* 38 (5), pp.382-405.

Ascher, W. (1978). *Forecasting: An Appraisal for Policymakers and Planners*. John Hopkins University Press: Baltimore, Maryland.

Ayyub, B. M. (2001). *Elicitation of Expert Opinions for Uncertainty and Risks*. CRC Press: New York.

Balch, G. I. (1980). "The Stick, the Carrot, and Other Strategies: A Theoretical Analysis of Governmental Intervention", *Law and Policy Quarterly* 2 (1), pp. 35-60.

Bardach, E. (1977). *The Implementation Game: What Happens After a Bill Becomes a Law.* MIT Press: Cambridge, Massachusetts.

Bardach, E. (1980). "Implementation Studies and the Study of Implements", paper presented at the annual meeting of the American Political Science Association.

Bates, J. M. and Granger, C. W. J. (1969). "The Combination of Forecasts", *OR 20* (4), pp. 451-468.

Bemelmans-Videc, M., Rist, R. C. and Vedung, E. (1998). *Carrots, Sticks and Sermons: Policy Instruments and their Evaluation*. Transaction Publishers: New Jersey.

Better-Regulation-Task-Force (2002). "Local Delivery of Central Policy", *Cabinet-Office*, retrieved on 12/06 2009 from http://archive.cabinet-office.gov.uk/brc/.

Blattberg, R. C. and Hoch, S. J. (1990). "Database Models and Managerial Intuition: 50% Model + 50% Manager", *Management Science* 36 (8), pp. 887-899.

Boardman, A. E., Mallery, W. L. and Vining, A. R. (1994). "Learning From Ex Ante/Ex Post Cost-Benefit Comparisons: The Coquihalla Highway Example", *Socio-Economic Planning Sciences* 28 (2), pp. 69-84.

Booker, J. M. and Meyer, M. A. (1988). "Sources and Effects of Inter-Expert Correlation: An Empirical Study", *IEEE Transactions on Systems, Man, and Cybernetics* 18 (1), pp. 135-142.

Bolger, F. and Harvey, N. (1998). "Heuristics and Biases in Judgmental Forecasting", in G. Wright and P. Goodwin, *Forecasting with Judgment*. John Wiley & Sons: New York, pp.113-138.

Bolger, F., and Wright, G. (1994), "Assessing the Quality of Expert Judgment: Issues and Analysis", *Decision Support Systems* 11 (1), pp.1-24.

Box, G. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day: Oakland, California.

Boyce, J. (2009). *Forecasting the Effectiveness of Policy Implementation Strategies*, A dissertation submitted to The University of Manchester for the degree of Master of Science in the Faculty of Humanities.

Breuning, M. (2003). "The Role of Analogies and Abstract Reasoning in Decision-Making: Evidence from the Debate over Truman's Proposal for Development Assistance", *International Studies Quarterly* 47 (2), pp. 229-245.

Breyer, S. (1979). "Analyzing Regulatory Failure: Mismatches, Less Restrictive Alternatives, and Reform", *Harvard Law Review* 92 (3), pp. 549-609.

Breyer, S. (1982). *Regulation and Its Reform*. Harvard University Press: Cambridge, Massachusetts.

Bunn, D. and Wright, G. (1991). "Interaction of Judgmental and Statistical Forecasting: Issues and Analysis", *Management Science* 37 (5), pp. 501-518.

Cabinet-Office (2003). "The Magenta Book: Guidance Notes for Policy Evaluation and Analysis", *Cabinet-Office*, retrieved on 15/06 2009 from http://www.nationalschool.gov.uk/policyhub/magenta_book/.

Chase, W. G. and Ericsson, K. A. (1981). "Skilled Memory", in J. R. Anderson, *Cognitive Skills and Their Acquisition.* Erlbaum Associates: Hilldale, New Jersey, pp. 73-96.

Chatfield, C. (2003). *The Analysis of Time Series: An Introduction*. Chapman and Hall: New York.

Chen, H. T. and Rossi, P. H. (1980). "The Multi-Goal, Theory Driven Approach to Evaluation: A Model Linking Basic and Applied Social Science", *Social Forces* 59 (1), pp. 106-122.

Dahl, R. A. and Lindblom, C. E. (1953). *Politics, Economics and Welfare: Planning and Politco-Economic Systems Resolved into Basic Social Processes.* Harper and Row: New York.

Davies, P. (1999). "What is Evidence-Based Education?", *British Journal of Educational Studies* 47 (2), pp. 108-121.

Davies, P. (2004). "Policy Evaluation in the United Kingdom", *KDI International Policy Evaluation Forum*: Seoul, Korea.

De Gooijer, J. G. and Hyndman, R. J. (2006). "25 years of Time Series Forecasting", *International Journal of Forecasting* 22 (3), p. 443-473.

DEFRA (2006). "Synthesis of Climate Change Policy Evaluations", *DEFRA*. Retrieved on 24/07 2009 from
http://www.decc.gov.uk/assets/decc/what%20we%20do/global%20climate%20change%20and%20energy/tackling%20climate%20change/programme/synthesisccpolicy-evaluations.pdf.

Denscombe, M. (2003). *The Good Research Guide: For Small-Scale Social Research Projects*. Open University Press: Maidenhead, England.

Derthick, M. (1972). *New Towns in Town: Why a Federal Program Failed.* Urban Institute: Washington.

Doern, G. B. (1981). *The Nature of Scientific and Technological Controversy in Federal Policy Formation*. Science Council of Canada: Ottawa.

Dortmans, P. J. and Eiffe, E. (2004). "An Examination of Future Scenarios Using Historical Analogy", *Futures* 36 (10), pp. 1049-1062.

Dorey, P. (2005). *Policy Making in Britain: An Introduction*. Sage Publications: London.

Dunsire, A. (1978a). *Implementation in a Bureaucracy*. Martin Robertson: Oxford.

Dunsire, A. (1978b). *Control in a Bureaucracy*. Martin Robertson: Oxford.

Dunsire, A. (1990). "Implementation Theory and Bureaucracy", in T. Younis, *Implementation in Public Policy*, Dartmouth, Aldershot, Hants.

Dye, T. R. (1976). *Policy Analysis: What Governments Do, Why they do it, What Difference it Makes*. University of Alabama Press: Alabama.

Easingwood, C. J. (1989). "An Analogical Approach to the Long Term Forecasting of Major New Product Sales", *International Journal of Forecasting* 5 (1), pp. 69-82.

Eder, P. and Leone, F. (1999), "The ECT Angle", *Foresight*, 1 (6), pp. 547 – 561.

Eddy, D. M. (1992). "Cost-Effectiveness Analysis: Is It Up to the Task?", *Journal of the American Medical Association* 267 (24), pp. 3342-3348.

Einhorn, H. J. (1972). "Expert Measurement and Mechanical Combination", *Organizational Behaviour and Human Performance* 7 (1), pp. 86-106.

Einhorn, H. J. (1974). "Expert Judgment: Some Necessary Conditions and an Example", *Journal of Applied Psychology*, 59 (5), pp. 562-571.

Ekins, P., Russel, A. and Hargreaves, C. (2002), "Reducing Carbon Emissions through Improved Household Energy Efficiency in the UK", *Journal of Environmental Policy and Planning* 4 (1), pp. 41-65.

Elmore, R. (1978). "Organizational Models of Social Program Implementation", *Public Policy* 26 (2), pp. 185-228.

Elmore, R. (1979). "Backward Mapping: Implementation Research and Policy Decisions", *Political Science Quarterly* 94 (4), pp. 601-616.

Ericsson, K. A. and Simon, H. A. (1980). "Verbal Reports as Data", *Psychological Review* 87 (3), pp.215-251.

Ericsson, K. A. and Simon, H. A. (1984). *Protocol Analysis: Verbal Reports as Data*. MIT Press: Cambridge, Massachusetts.

Etzioni, A. (1967). "Mixed-Scanning: A 'Third' Approach to Decision-Making", *Public Administration Review* 27 (5), pp. 385-392.

European Commission (2009a). "Impact Assessment Guidelines", retrieved on 21/11 2009 from http://ec.europa.eu/governance/impact/consultation/docs/ia_guidelines_draft_text_final_en.pdf.

European Commission (2009b). "Inter-Institutional Common Approach to Impact Assessment", retrieved on 21/11 2009 from http://ec.europa.eu/governance/impact/key_docs/docs/ii_common_approach_to_ia_en.pdf.

Fildes, R. (1985). "Quantitative Forecasting-The State of the Art: Econometric Models", *The Journal of the Operational Research Society* 36 (7), pp. 549-580.

Fildes, R., Goodwin, P., Lawrence, M. and Nikolopoulos, K. (2009). "Effective Forecasting and Judgmental Adjustments: An Empirical Evaluation and Strategies for Improvement in Supply-Chain Planning", *International Journal of Forecasting* 25 (1), pp.3-23.

Fildes, R. and Nikolopoulos, K. (2006). "Spyros Makridakis: An Interview with the International Journal of Forecasting", *International Journal of Forecasting* 22(3), pp. 625-636.

Forecasters.org. (2009). "About the IIF." Retrieved 19/12, 2009, from http://forecasters.org/about.html.

Forecastingprinciples.com (2009a). "Forecasting Dictionary." Retrieved 16/12, 2009, from

http://www.forecastingprinciples.com/index.php?option=com_content&task=view&id=19&Itemid=18.

Forecastingprinciples.com (2009b). "Selection Tree." Retrieved 16/12, 2009, from http://www.forecastingprinciples.com/index.php?option=com_content&task=view&id=17&Itemid=17.

Forecastingprinciples.com. (2009c). "Methodology Tree." Retrieved 16/12, 2009, from http://www.forecastingprinciples.com/index.php?option=com_content&task=view&id=16&Itemid=16.

Forecastingprinciples.com (2009d). Accessible at www.forecastingprinciples.com.

Geist, M. R. (2010). "Using the Delphi Method to Engage Stakeholders: A Comparison of Two Studies", *Evaluation and Program Planning* 33 (2), pp. 147-154.

Gilbert, N. (2001). "Research, Theory and Method", in N. Gilbert, *Researching Social Life*, Sage: London.

Goldberg, L. R. (1968). "Simple Models or Simple Processes? Some Research on Clinical Judgments", *American Psychologist* 23 (7), pp. 483-496.

Goldfarb, R. S., Stekler, H. O. and David, J. (2005). "Methodological Issues in Forecasting: Insights from the Egregious Business Forecast Errors of Late 1930", *Journal of Economic Methodology* 12 (4), pp. 517 - 542.

Goodwin, P. and Wright, G. (1993). "Improving Judgmental Time Series Forecasting: A Review of the Guidance Provided by Research", *International Journal of Forecasting*, 9 (2), pp. 147-161.

Goodwin, P. (2000). "Improving the Voluntary Integration of Statistical Forecasts and Judgment", *International Journal of Forecasting* 16 (1), pp. 85-99.

Government-Social-Research-Unit (2009). "Policy Hub." from http://www.nationalschool.gov.uk/policyhub/.

Gramblin, E. M. (1990) *A Guide to Benefit-Cost Analysis*. Prentice-Hall: Englewood Cliffs, New Jersey.

Green, K. and Armstrong, J. S. (2007a). "Global Warming: Forecasts by Scientists Versus Scientific Forecasts", *Energy & Environment* 18 (7), pp. 997-1021.

Green, K. and Armstrong, J. S. (2007b). "Structured Analogies for Forecasting", *International Journal of Forecasting* 23, pp. 365 - 376.

Hahn, R. W. and Stavins, R. N. (1992). "Economic Incentives for Environmental Protection: Integrating Theory and Practice", *The American Economic Review* 82 (2), pp. 464-468.

Hall, P. A. (1993). "Policy Paradigms, Social Learning and the State: The Case of Economic Policy Making in Britain", *Comparative Politics,* 25 (3), pp. 275-296.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press: Princeton, New Jersey.

Hammond, K. R. (1996). *Human Judgment and Social Policy*, Oxford University Press: New York.

Hargrove, E. L. (1975). *The Missing Link: The Study of the Implementation of Social-Policy*. Urban Institute: Washington.

Harrington, W., Morgenstern, R. D. and Nelson, P. (2000). "On the Accuracy of Regulatory Cost Estimates", *Journal of Policy Analysis and Management* 19 (2), pp. 297-322.

Harvey, N. (2001). "Improving Judgmental Forecasts", in J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners.* Kluwer Academic Publishers: Norwell, Massachusetts, pp. 59-80.

Heemstra, F. J. (1992). "Software Cost Estimation", *Information and Software Technology* 34 (10), pp. 627-639.

Helmer, O. and Rescher, N. (1958). "On the Epistemology of the Inexact Sciences", *The Rand Corporation* P1513.

Hill, M. (2005). *The Public Policy Process*. Pearson Education Limited: Harlow, England.

Hillman, A. J and Hitt, M. A. (1999). "Corporate Political Strategy Formulation: A Model of Approach Participation, and Strategy Decisions", *The Academy of Management Review*, 24 (4), pp. 825-842.

Hjern, B. (1982). "Implementation Research – The Link Gone Missing", *Journal of Public Policy*, 2 (3), pp. 301-308.

HM-Treasury (2009). "The Green Book: Appraisal and Evaluation in Central Government", *HM-Treasury*. Retrieved from 21/07 2009 from http://www.hm-treasury.gov.uk/d/green_book_complete.pdf.

Hogarth, R. (1980). *Judgment and Choice: The Psychology of Decisions*. Wiley Interscience: Chichester.

Hood, C. (1986). *The Tools of Government*. Macmillan: London.

Hood, C. (1976). *The Limits of Administration*. John Wiley: London.

Howlett, M. and Ramesh, M. (2003). *Studying Public Policy: Policy Cycles and Policy Subsystems*. Oxford University Press: Ontario.

Hughes, O. E. (2003). *Public Management and Administration: An Introduction*. Palgrave-MacMillan: Basingstoke.

IRS. (2010). "Alternative Motor Vehicle Credit." from http://www.irs.gov/newsroom/article/0,,id=157632,00.html.

Janis, I. L. (1972). *Victims of Groupthink*. Houghton-Miffin: Boston, Massachusetts.

John, P. (1998). *Analysing Public Policy*. Continuum: London.

Johnson. P. E. (1983). "What Kind of an Expert Should a System Be?", *Journal of Medical Philosophy* 8 (1), pp. 77-97.

Jordan, A., Wurzel, R. K. W. and Zito, A. R. (2003). "New Instruments of Environmental Governance: Patterns and Pathways of Change", *Environmental Politics* 12 (1), pp. 1-24.

Kagan, R. A. and Axelrad, L. (1997). "Adversarial Legalism: An International Perspective" in P.S Novela, *Comparative Disadvantages? Social Regulations and the Global Economy*. Brookings Institution Press: Washington.

Kahneman, D., Slovic, P. and Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press: Cambridge.

Keeney, R. L. and Von Winterfeldt, D. (1989). "On the Uses of Expert Judgment on Complex Technical Problems", *IEEE Transactions on Engineering Management* 36 (2), pp. 83-86.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*, University of Chicago Press: London.

Knill. C. (1998). "European Policies: The Impact of National Administrative Traditions", *Journal of Public Policy* 18 (1), pp. 1-28.

Kokinov, B. (2003). "Analogy in Decision-Making, Social Interaction, and Emergent Rationality", *Behavioural and Brain Sciences* 26 (2), pp. 167-168.

Krane, D. (2001). "Disorderly Progress on the Frontiers of Policy Evaluation", *International Journal of Public Administration* 24 (1), pp. 95 - 123.

Lasswell, H. D. (1958). *Politics: Who Gets What, When, How*? Meridian: New York.

Lawrence, M., Goodwin, P, O'Connor, M. and Onkal, D. (2006). "Judgmental Forecasting: A Review of Progress Over the Last 25 Years", *International Journal of Forecasting* 22 (3), pp. 493-518.

Lee, W. Y., Goodwin, P., Fildes, R., Nikolopoulos, K and Lawrence, M. (2007). "Providing Support for the Use of Analogies in Demand Forecasting Tasks", *International Journal of Forecasting* 23 (3), pp. 377-390.

Leeuw, F. L. (1998). "The Carrot: Subsidies as a Tool of Government", in Bemelmans-Videc et al. (1998), pp. 77-102.

Lexpansion.com (2009). "De la France Japon, la prime a la casse, est-ce que ca marche?". Retrieved 14/12 2009 from http://www.lexpansion.com/economie/actualite-entreprise/de-la-france-au-japon-la-prime-a-la-casse-est-ce-que-ca-marche_182915.html.

Lexpansion.com (2010). "600.000 bénéficaires de la prime à la casse en 2009". Retrieved on 23/04 2010 from http://www.lexpansion.com/economie/actualite-entreprise/600-000-beneficaires-de-la-prime-a-la-casse-en-2009_225365.html.

Lindblom, C. E. (1959). "The Science of Muddling Through", *Public Administration Review* 19 (2), pp. 78-88.

Linder, S. H. and Peters, G. (1989). "Instruments of Government: Perceptions and Contexts", *Journal of Public Policy* 9 (1), pp. 35-58.

Lipsky, M. (1971). "Street-level Bureaucracy and the Analysis of Urban Reform", *Urban Affairs Review*, 6 (4), pp. 391-409.

Lowi, T. J. (1966). "Distribution, Regulation, Redistribution: The Functions of Government", in R. B. Ripley, ed., *Public Policies and Their Politics: Techniques of Government Control.* W.W. Norton: New York, pp. 27-40.

Lowi, T. J. (1972). "Four Systems of Policy, Politics and Choice", *Public Administration Review* 32 (4), pp. 298-310.

Lowi, T. J. (1985). "The State in Politics: The Relation Between Policy and Administration", in R. G. Noll, *Regulatory Policy and the Social Sciences*, University of California Press: Berkeley, pp. 67-105.

Makridakis, S. (1996). "Forecasting: Its Role and Value for Planning and Strategy", *International Journal of Forecasting* 12 (4), pp. 513-537.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982). "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition", *Journal of Forecasting* 1 (2), pp. 111-153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K. and Simmons, L. F. (1993). "The M2-Competition: A Real-Time Judgmentally Based Forecasting Study", *International Journal of Forecasting*, 9 (1), pp. 5-22.

Makridakis, S. and Gaba, A. (1998). "Judgment: Its Role and Value for Strategy", in G. Wright and P. Goodwin, *Forecasting with Judgment.* John Wiley & Sons: London, pp. 1-38.

Makridakis, S., and Hibon, M. (1979). "Accuracy of Forecasting: An Empirical Investigation", *Journal of the Royal Statistical Society*. Series A (General) 142 (2), pp. 97-145.

Makridakis, S. and Hibon, M. (2000). "The M3-Competition: Results, Conclusions and Implications", *International Journal of Forecasting*, 16 (4), pp. 451-476.

Makridakis, S., Gaba, A., and Hogarth, R. M., (2009). *Dance with Chance: Making Luck Work for You*. Oneworld Publications: Oxford.

Makridakis, S., Wheelwright, S. C., and Hyndman, R. J. (1998). *Forecasting: Methods and Applications*. John Wiley & Sons: London.

Makridakis, S., Assimakopoulos, V., Nikolopoulos, K., Syntetos, A., Thomakos, D. and the Forecasting and Strategy Unit of the National Technical University of Athens (2010), "The M4-Competion", retrieved on 18/02 2010 from http://m4competition.com/.

March, J. G. and Olsen, J. P. (1976). *Ambiguity and Choice in Organisations*. Universitetsforlaget: Oslo.

Markman, A. and Moreau, C. (2001). "Analogy and Analogical Comparison in Choice", in D. Gentner, K. Holyoak and B. Kokinov, *The Analogical Mind*. MIT Press: Cambridge, Massachusetts.

Maslove, A. (1994). *Taxing and Spending: Issues of Process*. University of Toronto Press: Toronto.

Mathiowetz, N. A. (1987). "Response Error: Correlation between Estimation and Episodic Recall Tasks", *Survey Research Methods Proceedings*, National Center for Health Services Research.

Mayntz, R. (1979). "Public Bureaucracies and Policy Implementation", *International Social Science Journal* 31 (4), pp. 633-645.

Mayntz, R. (1983). "The Conditions of Effective Public Policy: A New Challenge for Policy Analysis", *Policy and Politics* 11 (2), pp. 123-143.

McCubbins, A. L. and McCubbins, M. D. (1994). "Learning from Oversight: Fire Alarms and Policy Patrols Reconstructed", *Journal of Law, Economics and Organisation* 10 (1), pp. 96-125.

McCubbins, M. D. and Schwartz, T. (1984). "Congressional Oversight Overlooked: Policy Patrol Versus Fire Alarms", *American Journal of Political Science* 28 (1), pp. 165-179.

McDaniel, P. R. (1989). "Tax Expenditures as Tools of Government Action", in L. M. Salamon, *Beyond Privatisation: The Tools of Government Action*. Urban Institute: Washington.

McDonnell, L. M. and Elmore, R. F. (1987). "Getting the Job Done: Alternative Policy Instruments", *Educational Evaluation and Policy Analysis* 9 (2), pp. 133-152.

McIntyre, S. H., Achabal, D. D. and Miller, C. M. (1993). "Applying Case-Based Reasoning to Forecasting Retail Sales", *Journal of Retailing* 69 (4), pp. 372-398.

Merriam-Webster (2009). "The definition of 'scientific method'". Retrieved on the 12/05 2009 from http://www.merriam-webster.com/dictionary/scientific%20method.

Meyer, M. A. and Booker, J. M. (1987). "Problems with Expert Data in Studies of Inter-Expert Correlation", *Social Statistics Proceedings of the Joint Statistical Meetings*: San Francisco, California, pp. 88-97.

Meyer, M. A. and Booker, J. M. (2001). *Eliciting and Analyzing Expert Judgment: A Practical Guide*, American Statistical Society and the Society for Industrial and Applied Mathematics.

Mills, T. C. and Markellos, R. N. (2008), *The Econometric Modelling of Financial Time Series*. Cambridge University Press: Cambridge.

Mingers, J. and Rosenhead, J. (2004). "Problem Structuring Methods in Action", *European Journal of Operational Research* 152 (3), pp. 530-554.

Mitnick, B. M. (1980), *The Political Economy of Regulation: Creating, Designing and Removing Regulatory Forms.* Columbia University Press: New York.

Mohr, L. B. (1995) *Impact Analysis for Program Evaluation*. Sage: Thousand Oaks, California.

Moran, M., Rein, M., and Goodin, R. E. (2008), *The Oxford Handbook of Public Policy*. Oxford University Press: Oxford.

Nachmias, D. (1979). *Public Policy Evaluation: Approaches and Methods*. St Martin's Press: New York.

Nas, T. F. (1996). *Cost-Benefit Analysis: Theory and Application*. Sage: Thousand Oaks, California.

National-Audit-Office (2010a). "Support to Business during a Recession", Department for Business, Innovation and Skill. Retrieved on 24/04 2010 from http://www.nao.org.uk/idoc.ashx?docId=7e7f7f72-92e9-4e71-8673-fec831691e8f&version=-1.

National-Audit-Office (2010b). "Support to Business during a Recession: Scheme Outlines", Department for Business, Innovation and Skill. Retrieved on 24/04 2010 from http://www.nao.org.uk/idoc.ashx?docId=bc3af0a8-c3d8-4c58-895a-88903087186c&version=-1.

NHTSA (2009). "Consumer Assistance to Recycle and Save Act of 2009", *United States Department of Transport, National Highway Traffic Safety Administration*. Retrieved on 28/03 2010 from http://www.cars.gov/files/official-information/CARS-Report-to-Congress.pdf.

Nielsen, S. B. and Ejler, N. (2008). "Improving Performance: Exploring the Complementarities between Evaluation and Performance Management", *Evaluation* 14 (2), pp. 171-192.

O'Connor, M. and Lawrence, M. (1998). "Judgmental Forecasting and the Use of Available Information" in G. Wright and P. Goodwin, *Forecasting with Judgment*. John Wiley & Sons: London, pp. 65-90.

Palumbo, D. J. (1987). *The Politics of Program Evaluation*. Sage: Beverley Hills, CA.

Parackal, M., Goodwin, P. and O'Connor, M. (2007). "Judgment in Forecasting", *International Journal of Forecasting*, 23 (3) pp. 343-345.

Parsons, W. (1995). *Public Policy: An Introduction to the Theory and Practice of Policy Analysis.* Edward Elgar Publishing: Cheltenham.

Perrels, A. (2001). Efficiency and Effectiveness of Policy Instruments: Concepts and Practice. *Workshop on Good Practices in Policies and Measures*: Copenhagen.

Peters, G. B. and Van Nispen, F. K. M. (1998). *Public Policy Instruments: Evaluating the Tools of Public Administration*. Edward Elgar Publishers: Cheltenham.

Phidd, R. W. and Doern, B. G. (1983). *Canadian Public Policy: Ideas, Structures, Process.* Metheun: Toronto.

Pidd, M. (2003). *Tools for Thinking: Modelling for Management Science*. John Wiley & Sons: Chichester.

Posner, R. A. (1974). "Theories of Economic Regulation", *Bell Journal of Economics and Management Science* 5 (2), pp. 335-358.

Pressman, J. and Wildavsky, A. (1973). *Implementation*. University of California Press: Berkeley, California.

Publicpolicyforecasting.com (2009). Retrieved on 29/09 2009 from http://www.forecastingprinciples.com/index.php?option=com_content&task=view&id =26&Itemid=129.

Radaelli, C. M. (1999). "The Public Policy of the European Union: Whither Politics of Expertise?", Journal *of European Public Policy* 6 (5), pp. 757 - 774.

Rayner, S. (2003). "Democracy in the Age of Assessment: Reflections on the Roles of Expertise and Democracy in Public-Sector Decision-Making", *Science and Public Policy* 30 (3), pp. 163-170.

Reid, D. J. (1968). "Combining Three Estimates of Gross Domestic Product", *Economica* 35 (140), pp. 431-444.

Reid, T. E. (1979). "The Failure of PPBS: Real Incentives for the 1980s", *Optimum* 10 (4), pp. 23-37.

Retireyourride.ca (2009). "Retire Your Ride Clunker Counter". Retrieved on 15/9 2009 from http://www.retireyourride.ca/home.aspx.

Reviere, R., Berkowitz, S., Carter, C. C. and Ferguson, C. G. (1996). *Needs Assessment: A Creative and Practical Guide for Social Scientists*. Taylor and Francis: Washington.

Rich, A. (2004). *Think Tanks, Public Policy and the Politics of Expertise*. Cambridge University Press: Cambridge.

Richards, D. and Smith, M. J. (2002). *Governance and Public Policy in the United Kingdom*, Oxford University Press: Oxford.

Robson, C. (2002), *Real World Research*. Blackwell: Malden.

Rochefort, D. A. and Cobb, R. W. (1994). *The Politics of Problem Definition: Shaping the Policy Agenda*. University Press of Kansas: Wichita.

Rondinelli, D. A. (1983). *Development Projects as Policy Experiments: An Adaptive Approach to Development Administration*. Metheun: London.

Rosenfield, I. (1988). *The Invention of Memory: A New View of the Brain*. Basic Books: New York.

Rossi, P. H., Lipsey, M. W. and Freeman, H. E. (2004). *Evaluation: A Systematic Approach*. SAGE Publications: Thousand Oaks, California.

Sabatier, P. A. (1986). "Top-Down and Bottom-Up Approaches to Implementation Research: A Critical Analysis and Suggested Synthesis", *Journal of Public Policy*, 6, pp. 21-48.

Sabatier, P. A. and Mazmanian, D. A. (1981). *Effective Policy Implementation*. Lexington Books: Lexington, Massachusetts.

Salamon, L. M. (1981). "Rethinking Public Management: Third-Party Government and the Changing Forms of Government Action", *Public Policy* 29 (3), pp. 255-275.

Salamon, L. M. and Lund, M. S. (1989). "The Tools Approach: Basic Analytics", in *Beyond Privatisation: The Tools of Government Action*. Urban Institute: Washington.

Savio, N. and Nikolopoulos, K. (2009a). "Forecasting the economic impact of new policies", *Foresight* 11 (2), pp. 7-18.

Savio, N. D. and Nikolopoulos, K. (2009b). "Forecasting effectiveness of policy implementation strategies: working with semi-experts", *Foresight* 11 (6), pp. 86-93.

Savio, N. D. and Nikolopoulos, K. (2009c). "Forecasting the Effectiveness of Policy Implementation Strategies", *International Journal of Public Administration* 33 (2), pp. 88 - 97.

Schnaars, S. (2009). "Forecasting the future of technology by analogy--An evaluation of two prominent cases from the 20th century", *Technology in Society* 31 (2), pp. 187-195.

Schrodt, P. A. (2002). *Forecasts and Contingencies: From Methodology to Policy*. American Political Science Association: Boston, Massachusetts.

Shadish, W. R., Cook, T. D. and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Houghton-Mifflin: Boston, Massachusetts.

Shanteau, J. (1987). "Psychological Characteristics of Expert Decision-Makers" in J. Mumpower, L. D. Phillips, O. Renn and Y. R. R. Uppuluri, *Expert Judgment and Expert Systems*. Springer-Verlag: Berlin, pp. 289-304.

Shanteau, J., Weiss, D. J., Thomas, R. P., Pounds, J. C., (2002), "Performance-Based Assessment of Expertise: How To Decide if Someone is An Expert or Not", *European Journal of Operational Research*, 136 (2), pp. 253-263.

Shepperd, M. (1997). "Estimating Software Project Effort Using Analogies", *IEEE Transactions on Software Engineering* 23 (11), pp. 736-743.

Simon, H. (1957). *Administrative Behaviour*. Macmillan: New York.

Simpson, D. and Walker, J. (1987). "Extending Cost-Benefit Analysis for Energy Investment Choices", *Energy Policy* 15 (3), pp. 217-227.

Soriano, F. I. (1995). *Conducting Needs Assessments: A Multidisciplinary Approach*. Sage: Thousand Oaks, California.

Stavins, R. N. (1997). "Policy Instruments for Climate Change: How Can National Governments Address a Global Problem?" *Resources for the Future*, Discussion Papers (dp-97-11). Retrieved on 15/03 2009 from http://ageconsearch.umn.edu/bitstream/10757/1/dp970011.pdf.

Suchman, E. (1967). *Evaluative Research*, Russell Sage Foundation: New York.

Summerhillgroup.ca (2009a). "Clean Air Foundation". Retrieved on 12/01 2009 from http://www.summerhillgroup.ca/eng/impact/.

Summerhillgroup.ca (2009b). "Retire-your-ride". Retrieved on 12/01 2009 from http://www.summerhillgroup.ca/eng/impact/programs/retire-your-ride.php.

Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R. and Goodwin, P. (2009). "The effects of integrating management judgement into intermittent demand forecasts", *International Journal of Production Economics* 118 (1), pp. 72-81.

Talbot, C. (2008), "Performance Regimes – The Institutional Context of Performance Policies", *International Journal of Public Administration*, 31 (14), pp. 1569-1591.

Trumbo, D., Adams, C., Milner, M. and Schipper, L. (1962). "Reliability and Accuracy in the Inspection of Hard Red Winter Wheat", *Cereal Science Today* 7.

Tupper, A. and Doern, B. G. (1981). "Public Corporations and Public Policy in Canada", in A. Tupper and B. G. Doern, *Public Corporations and Public Policy in Canada*. Institute for Research on Public Policy: Montreal, pp. 1-50.

Tversky, A. and Kahneman, D. (1974). "Judgments under Uncertainty: Heuristics and Biases", *Science* 185 (4157), pp. 1124-1131.

UK Gov. (2009a). "Impact Assessment Guidance", retrieved on 17/09 2009 from http://www.bis.gov.uk/assets/biscore/better-regulation/docs/10-898-impact-assessment-guidance.pdf.

UK Gov. (2009b). "Impact Assessment Toolkit", retrieved on 17/09 2009 from http://www.bis.gov.uk/assets/biscore/better-regulation/docs/10-901-impact-assessment-toolkit.pdf.

Utton, M. A. (1986). *The Economics of Regulating Industry*. Basil Blackwell: Oxford.

Van Meter, D. and Van Horn, C. (1975). "The Policy Implementation Process: A Conceptual Framework", *Administration and Society* 6 (4), pp. 445-488.

Weber, M., Ed. (1946). *From Max Weber: Essays in Sociology*. Translated, edited and with an introduction by H. H. Gerth and C. Wright Mills. Oxford University Press: New York.

Weiss, C. H. (1972). *Evaluation Research: Methods for Assessing Program Effectiveness*. Prentice-Hall: New Jersey.

Weiss, C. H. (1993). "Where Politics and Evaluation Research Meet", *Evaluation Practice* 14 (1), pp. 93-106.

Wetherley, R. and Lipsky, M. (1977). "Street-level Bureaucrats and Institutional Innovation: Implementing Special Education Reform", *Harvard Educational Review*, 47 (2), pp. 171-197.

Wholey, J. S. (1979). *Evaluation: Promise and Performance*. Urban Institute: Washington.

Wikipedia.org (2009). "Cash for Clunkers". Retrieved on 18/10 2009 from http://en.wikipedia.org/wiki/Car_Allowance_Rebate_System.

Wolf, C. (1987). "Markets and Non-Market Failures: Comparison and Assessment", *Journal of Public Policy* 7 (1), pp. 43-70.

Zerbe, R. O. (1998). "Is Cost-Benefit Analysis Legal? Three Rules", *Journal of Policy Analysis and Management* 17 (3), pp. 419-456.

Appendices

Appendix 1

This appendix contains the exact cases given to the participants in the five experiments conducted for investigating the research questions posed by this thesis.

The Original Three

First, the original three cases used in experiments I and II.

Case 1

s-SA

Manchester Business School

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description and

2/ try to think of several analogous situations and

3/ about how similar your analogies are to the case.

4/ Fill-in the questionnaire

Case 1 – New Eco-friendly Technology Adoption

Description

In light of the battle against climate change, a new eco-friendly technology has been developed that could significantly reduce the total CO2 emissions of nation X. This new technology, seen as a viable "green" alternative, is rather more expensive to purchase than the more polluting one currently in use. Nation X, worried about its total CO2 emissions, wishes to promote the adoption of the new technology through a nationalised tax incentive scheme. Such a scheme would reward the purchase of the new technology with a considerable tax return on the price paid (starting at a max of ± 15% of original price). After the sale of 60,000 units, the % of the price returned phases out over the next 3 quarters (i.e. once 60,000 are sold with a ± 15% return, the % returned gradually decreases over time to 0%)

(A) In the table below, please briefly describe

      (i) your analogies,

      (ii) their source (e.g. your own experience, media reports, history, literature, etc.), and

      (iii) the main similarities and differences between your analogies and this situation.

(B) Rate analogies out of 10 (0 = no similarity… 5 = similar… 10 = high similarity).

(C) Give the outcome of your analogy (i.e. what result did your analogy have)

(A)	(i) Description	(ii) Source	(iii) Similarities and Differences	(B) Similarity Rating	(C) Outcome
a.					
b.					
c.					
d.					
e.					
e.g.	New technology promoted by Gov. to combat climate change	Media	Similarities: new technology introduced for improvement purposes Differences: Different context, not necessarily for directly reducing CO2	8	Reduced waste (any kind) by 24%

Questionnaire

1)

The following table shows the sales for the new eco-friendly technology. Y refers to the year in which the tax incentive scheme was initialised. Please give your sales forecasts for years Y+1, Y+2 and Y+3 (in light of the introduction of the incentive scheme).

Year	Y-5	Y-4	Y-3	Y-2	Y-1	Y
Sales (units)	14,453	20,145	36,249	50,554	84,931	212,112

Forecast year Y+1: [   ]

Forecast year Y+2: [   ]

Forecast year Y+3: [   ]

2)   Roughly, how long did you spend on this task?

         {include the time you spent reading the description and instructions}   [____] mins

3)   How likely is it that taking more time would change your forecast?

         {0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have as a environmental issues management specialist?

                                                                                      [   ] years

7) Please rate your experience (out of 10) with cases similar to this one   [____] 0–10

UJ

Manchester Business School

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description and

2/ Fill-in the questionnaire

Case 1 – New Eco-friendly Technology Adoption

Description

In light of the battle against climate change, a new eco-friendly technology has been developed that could significantly reduce the total CO_2 emissions of nation X. This new technology, seen as a viable "green" alternative, is rather more expensive to purchase than the more polluting one currently in use. Nation X, worried about its total CO_2 emissions, wishes to promote the adoption of the new technology through a nationalised tax incentive scheme. Such a scheme would reward the purchase of the new technology with a considerable tax return on the price paid (starting at a max of ± 15% of original price). After the sale of 60,000 units, the % of the price returned phases out over the next 3 quarters (i.e. once 60,000 are sold with a ± 15% return, the % returned gradually decreases over time to 0%)

Questionnaire

1)

The following table shows the sales for the new eco-friendly technology. Y refers to the year in which the tax incentive scheme was initialised. Please give your sales forecasts for years Y+1, Y+2 and Y+3 (in light of the introduction of the incentive scheme).

Year	Y-5	Y-4	Y-3	Y-2	Y-1	Y
Sales (units)	14,453	20,145	36,249	50,554	84,931	212,112

Forecast year Y+1: [   ]

Forecast year Y+2: [   ]

Forecast year Y+3: [   ]

2)   Roughly, how long did you spend on this task?

{include the time you spent reading the description and instructions}   [____] mins

3)   How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have as a environmental issues management specialist?

[   ] years

7) Please rate your experience (out of 10) with cases similar to this one   [_____] 0–10

Case 2

s-SA

<div align="center">

Manchester Business School

</div>

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description and

2/ try to think of several analogous situations and

3/ about how similar your analogies are to the case.

4/ Fill-in the questionnaire

Case 2 – New CO2 Differentiated Tax Scheme

Description

Reducing CO2 emissions from 'a popular mode of transportation' is a key step in combating the negative effects of climate change. In order to meet CO2 emission targets, Gov. X introduced a strategy based on 3 pillars (voluntary commitments from manufacturers, improvement in consumer information and promotion of 'greener' vehicle types via fiscal measures) in year Y to reduce such emissions to the required level by year Y + 13. The first two pillars are expected to reduce CO2 emissions the majority of the way. The last pillar, which consists of a tax scheme which differentiates between vehicle types based on energy efficiency and favours the "greenest" ones, has the aim of reducing CO2 emissions the rest of the way.

(A) In the table below, please briefly describe

(i) your analogies,

(ii) their source (e.g. your own experience, media reports, history, literature, etc.), and

(iii) the main similarities and differences between your analogies and this situation.

(B) Rate analogies out of 10 (0 = no similarity… 5 = similar… 10 = high similarity).

(C) Give the outcome of your analogy (i.e. what result did your analogy have)

(A)	(i) Description	(ii) Source	(iii) Similarities and Differences	(B) Similarity Rating	(C) Outcome
a.					

b.				
c.				
d.				
e.				
e.g.	New technology promoted by Gov. to combat climate change	Media	Similarities: new technology introduced for improvement purposes Differences: Different context, not necessarily for directly reducing CO2	6	Reduced waste (any kind) by 24%

Questionnaire

1)

Assuming the average emissions for a vehicle type at year Y was α, please give a forecast of the **% decrease** of this average (in light of the introduction of the 3 pillar strategy described above) over the 13 year period (i.e. at Y+13).

Forecast for % decrease at Y+13: [   ]

2)   Roughly, how long did you spend on this task?

{include the time you spent reading the description and instructions}   [___] mins

3)   How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have as a environmental issues management specialist?

[   ] years

7) Please rate your experience (out of 10) with cases similar to this one   [_____] 0–10

UJ

Manchester Business School

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description and

2/ Fill-in the questionnaire

Case 2 – New CO2 Differentiated Tax Scheme

Description

Reducing CO2 emissions from 'a popular mode of transportation' is a key step in combating the negative effects of climate change. In order to meet CO2 emission targets, Gov. X introduced a strategy based on 3 pillars (voluntary commitments from manufacturers, improvement in consumer information and promotion of 'greener' vehicle types via fiscal measures) in year Y to reduce such emissions to the required level by year Y + 13. The first two pillars are expected to reduce CO2 emissions the majority of the way. The last pillar, which consists of a tax scheme which differentiates between vehicle types based on energy efficiency and favours the "greenest" ones, has the aim of reducing CO2 emissions the rest of the way.

Questionnaire

1)

Assuming the average emissions for a vehicle type at year Y was α, please give a forecast of the **% decrease** of this average (in light of the introduction of the 3 pillar strategy described above) over the 13 year period (i.e. at Y+13).

Forecast for % decrease at Y+13: [   ]

2)   Roughly, how long did you spend on this task?

{include the time you spent reading the description and instructions}   [____] mins

3)   How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have as a environmental issues management specialist?

[   ] years

7) Please rate your experience (out of 10) with cases similar to this one   [_____] 0–10

Case 3

s-SA

<div align="center">

Manchester Business School

</div>

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description and

2/ try to think of several analogous situations and

3/ about how similar your analogies are to the case.

4/ Fill-in the questionnaire

Case 3 – Get Digital

Description

The following policy was initiated by the Greek Government and is aimed at supporting the use of computers by students. More specifically, the policy entails governmentally financing laptops for University 'freshers' in 2009. The policy says that each student will be refunded 80% of the value of the laptop of his/her choice with a limit of 400 €. It should be noted that the Undergraduate Office plays an active role in the policy as they are responsible for contacting the would-be students in order to inform them and remind them of the offer.

(A) In the table below, please briefly describe

       (i) your analogies,

       (ii) their source (e.g. your own experience, media reports, history, literature, etc.), and

       (iii) the main similarities and differences between your analogies and this situation.

(B) Rate analogies out of 10 (0 = no similarity… 5 = similar… 10 = high similarity).

(C) Give the outcome of your analogy (i.e. what result did your analogy have)

(A)	(i) Description	(ii) Source	(iii) Similarities and Differences	(B) Similarity Rating	(C) Outcome
a.					

b.				
c.				
d.				
e.				
e.g.	Similar policy in another country for high speed internet	Media	Similarities: funding to students Differences: different country, funding for internet instead of a laptop	8	Response rate of policy was about 80%

Questionnaire

1)

What percentage of students will take up the offer from the Government?

Take up percentage: [   ]

2)   Roughly, how long did you spend on this task?

{include the time you spent reading the description and instructions}   [____] mins

3)   How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have as a environmental issues management specialist?

[ ] years

7) Please rate your experience (out of 10) with cases similar to this one   [____] 0–10

UJ

Manchester Business School

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description and

2/ Fill-in the questionnaire

Case 3 – Get Digital

Description

The following policy was initiated by the Greek Government and is aimed at supporting the use of computers by students. More specifically, the policy entails governmentally financing laptops for University 'freshers' in 2009. The policy says that each student will be refunded 80% of the value of the laptop of his/her choice with a limit of 400 €. It should be noted that the Undergraduate Office plays an active role in the policy as they are responsible for contacting the would-be students in order to inform them and remind them of the offer.

Questionnaire

1)

What percentage of students will take up the offer from the Government?

Take up percentage: [   ]

2)   Roughly, how long did you spend on this task?

{include the time you spent reading the description and instructions}   [___] mins

3)   How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have as a environmental issues management specialist?

[ ] years

8) Please rate your experience (out of 10) with cases similar to this one   [____] 0–10

The New Cases

Second, the new cases based on the vehicle scrappage scheme used in experiment III.

Case 1

s-SA

Manchester Business School

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description

2/ Try to think of several analogous situations

3/ Try to think about how similar your analogies are to the case.

4/ Fill-in the questionnaire

Case 1 – Operation Green

Description

The Government of Nation X, worried about the number of old, polluting vehicles on its roads, are planning on implementing Operation Green, intended to encourage owners of these high polluting vehicles to get them off the road and offer a reward in exchange for doing so. The program is committed to improving air quality by responsibly recycling vehicles and aims to retire vehicles over a period of 4 years.

To be eligible for a reward, vehicles must be at least 14 years old, in running condition, registered and insured for the last 6 months prior to application. Rewards for such a surrender include a public transit pass or a membership to a car-pooling scheme, £200 (approximately), or a discount on the purchase of a vehicle made in the last 5 years.

(A) In the table below, please briefly describe

       (i) your analogies,

       (ii) their source (e.g. your own experience, media reports, history, literature, etc.), and

       (iii) the main similarities and differences between your analogies and this situation.

(B) Rate analogies out of 10 (0 = no similarity… 5 = similar… 10 = high similarity).

(C) Give the outcome of your analogy (i.e. what result did your analogy have)

(A)	(i) Description	(ii) Source	(iii) Similarities and Differences	(B) Similarity Rating	(C) Outcome
e.g.	Scrappage scheme with similar aim to replace old vehicles with more efficient new ones	Media	Similarities: same policy, same objective Differences: different incentives offered, different budget available	8	Hugely successful – programme resulted in the registration of an extra X% of new vehicles in that period
a					
b					
c					
d.					
e					

Questionnaire

1) Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. For this, they request that you predict how many of these old, polluting vehicles (of the estimated 4 million vehicles currently on the roads which are eligible for the scheme) will be surrendered in the first 7 months of the strategy.

Forecast   [   ]

2)   Roughly, how long did you spend on this task?

{*include the time you spent reading the description and instructions*}   [____] mins

3)   How likely is it that taking more time would change your forecast?

        {0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

[   ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy.   [____] 0–10

- your experience with cases similar to this one.   [____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[____] 0–10

UJ

Manchester Business School

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description and

2/ Fill-in the questionnaire

Case 1 – Operation Green

Description

The Government of Nation X, worried about the number of old, polluting vehicles on its roads, are planning on implementing Operation Green, intended to encourage owners of these high polluting vehicles to get them off the road and offer a reward in exchange for doing so. The program is committed to improving air quality by responsibly recycling vehicles and aims to retire vehicles over a period of 4 years.

To be eligible for a reward, vehicles must be at least 14 years old, in running condition, registered and insured for the last 6 months prior to application. Rewards for such a surrender include a public transit pass or a membership to a car-pooling scheme, £200 (approximately), or a discount on the purchase of a vehicle made in the last 5 years.

Questionnaire

1) Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. For this, they request that you predict how many of these old, polluting vehicles (of the estimated 4 million vehicles currently on the roads which are eligible for the scheme) will be surrendered in the first 7 months of the strategy.

Forecast [   ]

2)   Roughly, how long did you spend on this task?

{include the time you spent reading the description and instructions}   [____] mins

3)   How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

[ ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy.   [____] 0–10

- your experience with cases similar to this one.   [____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[____] 0–10

Case 2

s-SA

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description

2/ Try to think of several analogous situations

3/ Try to think about how similar your analogies are to the case.

4/ Fill-in the questionnaire

Case 2 – Reducing Average Vehicle CO_2 Emissions

Description

Reducing CO_2 emissions from everyday road vehicles is a key step in combating the negative effects of climate change. In order to meet average CO_2 emission targets, Gov. X wishes to introduce a strategy which they hope will significantly reduce their average CO_2 emissions from these vehicles. The proposed strategy is designed to encourage owners of dated, inefficient vehicles to purchase newer, more efficient vehicles via a rebate incentive. The strategy promises a reward of £950 when a vehicle emitting less than 160g of CO_2/km is purchased and a (functioning) vehicle of more than 10 years is handed in for destruction. This strategy is planned to stay in place for 12 months.

(A) In the table below, please briefly describe

> (i) your analogies,

> (ii) their source (e.g. your own experience, media reports, history, literature, etc.), and

> (iii) the main similarities and differences between your analogies and this situation.

(B) Rate analogies out of 10 (0 = no similarity… 5 = similar… 10 = high similarity).

(C) Give the outcome of your analogy (i.e. what result did your analogy have)

(A)	(i) Description	(ii) Source	(iii) Similarities and Differences	(B) Similarity Rating	(C) Outcome
e.g.	Scrappage scheme with similar aim to	Media	Similarities: same policy, same objective Differences: different	8	Hugely successful – programme

	replace old vehicles with more efficient new ones		incentives offered, different budget available		resulted in the registration of an extra X% of new vehicles in that period
a					
b					
c					
d.					
e					

Questionnaire

1) Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. For this, they require you to predict what percentage of all new vehicles registered in the first 5 months will have been done so through this governmental incentive?

Forecast   [   ]

2)   Roughly, how long did you spend on this task?

{include the time you spent reading the description and instructions}   [____] mins

3)   How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

[   ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy.   [_____] 0–10

- your experience with cases similar to this one.   [_____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[_____] 0–10

UJ

Manchester Business School

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description and

2/ Fill-in the questionnaire

Case 2 - Reducing Average Vehicle CO$_2$ Emissions

Description

Reducing CO2 emissions from everyday road vehicles is a key step in combating the negative effects of climate change. In order to meet average CO2 emission targets, Gov. X wishes to introduce a strategy which they hope will significantly reduce their average CO2 emissions from these vehicles. The proposed strategy is designed to encourage owners of dated, inefficient vehicles to purchase newer, more efficient vehicles via a rebate incentive. The strategy promises a reward of £950 when a vehicle emitting less than 160g of CO2/km is purchased and a (functioning) vehicle of more than 10 years is handed in for destruction. This strategy is planned to stay in place for 12 months.

Questionnaire

1) Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. For this, they require you to predict what percentage of all new vehicles registered in the first 5 months will have been done so through this governmental incentive?

Forecast   [   ]

2)   Roughly, how long did you spend on this task?

{include the time you spent reading the description and instructions}   [____] mins

3)   How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

[   ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy.   [_____] 0–10

- your experience with cases similar to this one.   [_____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[_____] 0–10

Case 3

s-SA

<div align="center">

Manchester Business School

</div>

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description

2/ Try to think of several analogous situations

3/ Try to think about how similar your analogies are to the case.

4/ Fill-in the questionnaire

Case 3 – Killing Two Birds with One Stone

Description

In light of the battle against climate change, national governments are always looking for ways to improve their overall efficiency and thereby reducing their total carbon emissions. As one of many nations suffering from the negative effects of the recent global recession on the automotive industry, Nation X has devised a strategy which will simultaneously inject a much needed boost into this sector as well as putting greener and more efficient vehicles on the road.

This strategy rewards the purchase of a new vehicle (with a retail price of less than £27,000) when an older vehicle (of less than 25 years) is traded in. The size of this reward, which will come in the form of credit towards the purchase of the new vehicle, ranges from £2,100 to £2,600 depending on the difference in efficiency between the vehicle traded in and the one purchased. The overall available budget for such a strategy is £1.8 billion. The incentive offered by the strategy is due to be in place for 5 months or until the budget is exhausted.

(A) In the table below, please briefly describe

　　　　(i) your analogies,

　　　　(ii) their source (e.g. your own experience, media reports, history, literature, etc.), and

　　　　(iii) the main similarities and differences between your analogies and this situation.

(B) Rate analogies out of 10 (0 = no similarity… 5 = similar… 10 = high similarity).

(C) Give the outcome of your analogy (i.e. what result did your analogy have)

(A)	(i) Description	(ii) Source	(iii) Similarities and	(B) Similarity	(C) Outcome

257

			Differences	Rating	
e.g.	Scrappage scheme with similar aim to replace old vehicles with more efficient new ones	Media	Similarities: same policy, same objective Differences: different incentives offered, different budget available	8	Hugely successful – programme resulted in the registration of an extra X% of new vehicles in that period
a					
b					
c					
d.					
e					

Questionnaire

1) Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. Given that the overall budget for the proposed strategy is £1.8 billion, they require you to give a prediction as to the percentage of this budget that will actually be used by people taking up on the offer during this 5 month period.

Forecast [   ]

2)   Roughly, how long did you spend on this task?

{include the time you spent reading the description and instructions}   [____] mins

3)   How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

[ ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy.   [____] 0–10

- your experience with cases similar to this one.   [____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[____] 0–10

UJ

Manchester Business School

Name:

Each attached file contains a case description and a short questionnaire. Please follow these steps for each case:

1/ Read the description and

2/ Fill-in the questionnaire

Case 4 – Killing Two Birds with One Stone

Description

In light of the battle against climate change, national governments are always looking for ways to improve their overall efficiency and thereby reducing their total carbon emissions. As one of many nations suffering from the negative effects of the recent global recession on the automotive industry, Nation X has devised a strategy which will simultaneously inject a much needed boost into this sector as well as putting greener and more efficient vehicles on the road.

This strategy rewards the purchase of a new vehicle (with a retail price of less than £27,000) when an older vehicle (of less than 25 years) is traded in. The size of this reward, which will come in the form of credit towards the purchase of the new vehicle, ranges from £2,100 to £2,600 depending on the difference in efficiency between the vehicle traded in and the one purchased. The overall available budget for such a strategy is £1.8 billion. The incentive offered by the strategy is due to be in place for 5 months or until the budget is exhausted.

Questionnaire

1) Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. Given that the overall budget for the proposed strategy is £1.8 billion, they require you to give a prediction as to the percentage of this budget that will actually be used by people taking up on the offer during this 5 month period.

Forecast [   ]

2)   Roughly, how long did you spend on this task?

        {include the time you spent reading the description and instructions}   [____] mins

3)   How likely is it that taking more time would change your forecast?

        {0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4)   Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

                                                                                [ ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy.   [____] 0–10

- your experience with cases similar to this one.   [____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

                                                                                [____] 0–10

The Refined Cases

Third, the new cases used in experiment III but with the modifications made to the PIS description, the analogy table and the questionnaire, which were used in experiments IV and V.

Case 1

s-SA

Manchester Business School

Name:

Each of these files contains a short description of a policy and the proposed strategy for implementation. Please read the description and follow the instructions.

Case 1 – Operation Green

Description

The Government of Nation X, worried about the number of old, polluting vehicles on its roads, is planning on implementing 'Operation Green', intended to encourage owners of these high polluting vehicles to get them off the road and offer a reward in exchange for doing so. The program is committed to improving air quality by responsibly recycling vehicles and aims to retire vehicles over a period of 4 years.

To be eligible for a reward, vehicles must be at least 14 years old, in running condition, registered and insured for the last 6 months prior to application. Financial rewards for such a surrender include either £200 cash, or a public transit pass or a membership to a car-pooling scheme or a discount on the purchase of a vehicle made in the last 5 years, with the latter three having an approximate value of £200. The overall budget available for the program is £55 million.

Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. They would like to know what percentage of these old, polluting vehicles (of the estimated 4 million vehicles currently on the roads which are eligible for the scheme) will be surrendered in the first 7 months of the strategy.

Instructions

In the tables provided below (supplementary tables are found at the very end), please describe any analogous strategies to the one described in the case that you may think of. Please include details on

- the similarities and differences between your analogous strategy and the target strategy

- their source (e.g. your own experience, media reports, history, literature, etc.)

- a similarity rating between your analogous strategy and the target strategy (0 = no similarity… 5 = similar… 10 = high similarity)

- the outcome of your analogous case (which of the outcomes A-E found at the bottom, is most similar, in terms of effectiveness, to the outcome of your analogy?)

1. Analogies

Example Analogy

Description	Scrappage scheme with similar aim to replace old vehicles with more efficient new ones
Similarities and Differences	Similarities: same policy, same objective Differences: different incentives offered, different budget available

Source ____Media__   Similarity Rating _8___   Closest Outcome__E___

Analogy 1

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Analogy 2

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Analogy 3

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome _____

Outcomes

Please select one of the following as your prediction of the percentage of old vehicles surrendered within the first 7 months because of the strategy:

A	0% – 20%	[ ]
B	21% – 40%	[ ]
C	41% – 60%	[ ]
D	61% - 80%	[ ]
E	81% - 100%	[ ]

Within your chosen range, please provide a point forecast for the percentage of old vehicles surrendered within the first 7 months

[   ]

2. Questionnaire

1) Roughly, how long did you spend on this task?

   {include the time you spent reading the description and instructions}   [____] mins

2) How likely is it that taking more time would change your forecast?

   {0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

3) If you knew that this case was from the UK, how likely would you be to change your forecast?

   {0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4) Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

[   ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy.   [_____] 0–10

- your experience with cases similar to this one.   [_____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[_____] 0–10

8) If you were contracted to producing such a strategy effectiveness forecast, what approach/process would you adopt? In what sort of time-scale?

Supplementary Tables

Analogy 4

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Analogy 5

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Manchester Business School

Name:

Each of these files contains a short description of a policy and the proposed strategy for implementation. Please read the description and follow the instructions.

Case 1 – Operation Green

Description

The Government of Nation X, worried about the number of old, polluting vehicles on its roads, is planning on implementing 'Operation Green', intended to encourage owners of these high polluting vehicles to get them off the road and offer a reward in exchange for doing so. The program is committed to improving air quality by responsibly recycling vehicles and aims to retire vehicles over a period of 4 years.

To be eligible for a reward, vehicles must be at least 14 years old, in running condition, registered and insured for the last 6 months prior to application. Financial rewards for such a surrender include either £200 cash, or a public transit pass or a membership to a car-pooling scheme or a discount on the purchase of a vehicle made in the last 5 years, with the latter three having an approximate value of £200. The overall budget available for the program is £55 million.

Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. They would like to know what percentage of these old, polluting vehicles (of the estimated 4 million vehicles currently on the roads which are eligible for the scheme) will be surrendered in the first 7 months of the strategy.

Instructions

Outcomes

Please select one of the following as your prediction of the percentage of old vehicles surrendered within the first 7 months because of the strategy:

A	0% – 20%	[ ]
B	21% – 40%	[ ]
C	41% – 60%	[ ]
D	61% - 80%	[ ]
E	81% - 100%	[ ]

Within your chosen range, please provide a point forecast for the percentage of old vehicles surrendered within the first 7 months

[   ]

2. Questionnaire

1) Roughly, how long did you spend on this task?

{include the time you spent reading the description and instructions}   [____] mins

2) How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

3) If you knew that this case was from the UK, how likely would you be to change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4) Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

[   ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy.   [____] 0–10

- your experience with cases similar to this one.   [____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[____] 0–10

8) If you were contracted to producing such a strategy effectiveness forecast, what approach/process would you adopt? In what sort of time-scale?

Case 2

s-SA

<div align="center">**Manchester Business School**</div>

Name:

Each of these files contains a short description of a policy and the proposed strategy for implementation. Please read the description and follow the instructions.

Case 2 – Reducing Average Vehicle CO_2 Emissions

Description

Reducing CO_2 emissions from everyday road vehicles is a key step in combating the negative effects of climate change. In order to meet average CO_2 emission targets, Gov. X wishes to introduce a strategy which they hope will significantly reduce their average CO_2 emissions from these vehicles. The proposed strategy is designed to encourage owners of dated, inefficient vehicles to purchase newer, more efficient vehicles via a rebate incentive. The strategy promises a reward of £950 when a vehicle emitting less than 160g of CO_2/km is purchased and a (functioning) vehicle of more than 10 years is handed in for destruction. This strategy is to stay in place for 12 months or until the funds are exhausted. The overall budget for the scheme is £210 million.

Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. For this, they require you to predict what percentage of all new vehicles registered in the first 5 months (the total being 900,000) will have been done so through this governmental incentive?

Instructions

In the tables provided below (supplementary tables are found at the very end), please describe any analogous strategies to the one described in the case that you may think of. Please include details on

- the similarities and differences between your analogous strategy and the target strategy

- their source (e.g. your own experience, media reports, history, literature, etc.)

- a similarity rating between your analogous strategy and the target strategy (0 = no similarity… 5 = similar… 10 = high similarity)

- the outcome of your analogous case (which of the outcomes A-E found at the bottom, is most similar, in terms of effectiveness, to the outcome of your analogy?)

1. Analogies

Example Analogy

Description	Scrappage scheme with similar aim to replace old vehicles with more efficient new ones
Similarities and Differences	Similarities: same policy, same objective Differences: different incentives offered, different budget available

Source ____Media__   Similarity Rating _8___   Closest Outcome__E___

Analogy 1

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Analogy 2

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Analogy 3

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Outcomes

Please select one of the following as your prediction of the percentage of all new vehicles registered in the first 5 months will have been done so through this governmental incentive scheme:

A	0% – 20%	[ ]
B	21% – 40%	[ ]
C	41% – 60%	[ ]
D	61% - 80%	[ ]
E	81% - 100%	[ ]

Within your chosen range, please provide a point forecast for the percentage of all new vehicles registered in the first 5 months

[   ]

2. Questionnaire

1) Roughly, how long did you spend on this task?

> *{include the time you spent reading the description and instructions}* [____] mins

2) How likely is it that taking more time would change your forecast?

> *{0=almost no chance (1/100) … 10=practically certain (99/100)}* [__] 0–10

3) If you knew that this case was from the UK, how likely would you be to change your forecast?

> *{0=almost no chance (1/100) … 10=practically certain (99/100)}* [__] 0–10

4) Do you recognise the actual case described in this file? Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

[ ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy. [____] 0–10

- your experience with cases similar to this one. [____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[____] 0–10

8) If you were contracted to producing such a strategy effectiveness forecast, what approach/process would you adopt? In what sort of time-scale?

Supplementary Tables

Analogy 4

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Analogy 5

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

UJ

Manchester Business School

Name:

Each of these files contains a short description of a policy and the proposed strategy for implementation. Please read the description and follow the instructions.

Case 2 – Reducing Average Vehicle CO₂ Emissions

Description

Reducing CO_2 emissions from everyday road vehicles is a key step in combating the negative effects of climate change. In order to meet average CO_2 emission targets, Gov. X wishes to introduce a strategy which they hope will significantly reduce their average CO_2 emissions from these vehicles. The proposed strategy is designed to encourage owners of dated, inefficient vehicles to purchase newer, more efficient vehicles via a rebate incentive. The strategy promises a reward of £950 when a vehicle emitting less than 160g of CO_2/km is purchased and a (functioning) vehicle of more than 10 years is handed in for destruction. This strategy is to stay in place for 12 months or until the funds are exhausted. The overall budget for the scheme is £210 million.

Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. For this, they require you to predict what percentage of all new vehicles registered in the first 5 months (the total being 900,000) will have been done so through this governmental incentive?

Instructions

Outcomes

Please select one of the following as your prediction of the percentage of all new vehicles registered in the first 5 months will have been done so through this governmental incentive scheme:

A	0% – 20%	[ ]
B	21% – 40%	[ ]
C	41% – 60%	[ ]
D	61% - 80%	[ ]
E	81% - 100%	[ ]

Within your chosen range, please provide a point forecast for the percentage of all new vehicles registered in the first 5 months

[   ]

2. Questionnaire

1) Roughly, how long did you spend on this task?

     {include the time you spent reading the description and instructions}   [___] mins

2) How likely is it that taking more time would change your forecast?

     {0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

3) If you knew that this case was from the UK, how likely would you be to change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4) Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

[ ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy.   [_____] 0–10

- your experience with cases similar to this one.   [_____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[_____] 0–10

8) If you were contracted to producing such a strategy effectiveness forecast, what approach/process would you adopt? In what sort of time-scale?

Case 3

s-SA

Name:

Each of these files contains a short description of a policy and the proposed strategy for implementation. Please read the description and follow the instructions.

Case 3 – Killing Two Birds with One Stone

Description

In light of the battle against climate change, national governments are always looking for ways to improve their overall efficiency and thereby reducing their total carbon emissions. As one of many nations suffering from the negative effects of the recent global recession on the automotive industry, Nation X has devised a strategy which will simultaneously inject a much needed boost into this sector as well as putting greener and more efficient vehicles on the road.

This strategy rewards the purchase of a new vehicle (with a retail price of less than £27,000) when an older vehicle (of less than 25 years) is traded in. The size of this reward, which will come in the form of credit towards the purchase of the new vehicle, ranges from £2,100 to £2,600 depending on the difference in efficiency between the vehicle traded in and the one purchased. The overall available budget for such a strategy is £1.8 billion. The incentive offered by the strategy is due to be in place for 5 months or until the budget is exhausted.

Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. Given that the overall budget for the proposed strategy is £1.8 billion, they require you to give a prediction as to the percentage of this budget that will actually be used by people taking up on the offer during this 5 month period.

Instructions

In the tables provided below (supplementary tables are found at the very end), please describe any analogous strategies to the one described in the case that you may think of. Please include details on

- the similarities and differences between your analogous strategy and the target strategy

- their source (e.g. your own experience, media reports, history, literature, etc.)

- a similarity rating between your analogous strategy and the target strategy (0 = no similarity… 5 = similar… 10 = high similarity)

- the outcome of your analogous case (which of the outcomes A-E found at the bottom, is most similar, in terms of effectiveness, to the outcome of your analogy?)

1. Analogies

Example Analogy

Description	Scrappage scheme with similar aim to replace old vehicles with more efficient new ones
Similarities and Differences	Similarities: same policy, same objective Differences: different incentives offered, different budget available

Source ____Media__   Similarity Rating _8___   Closest Outcome__E___

Analogy 1

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Analogy 2

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Analogy 3

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Outcomes

Please select one of the following as your prediction of the percentage of this budget that will actually be used by people taking up on the offer during this 5 month period:

A	0% – 20%	[ ]
B	21% – 40%	[ ]
C	41% – 60%	[ ]
D	61% - 80%	[ ]
E	81% - 100%	[ ]

Within your chosen range, please provide a point forecast for the percentage of this budget that will actually be used by people taking up on the offer during this 5 month period

[　　　　]

2. Questionnaire

1) Roughly, how long did you spend on this task?

       {include the time you spent reading the description and instructions}   [____] mins

2) How likely is it that taking more time would change your forecast?

       {0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

3) If you knew that this case was from the UK, how likely would you be to change your forecast?

       {0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4) Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

       [ ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy.   [_____] 0–10

- your experience with cases similar to this one.   [_____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[____] 0–10

8) If you were contracted to producing such a strategy effectiveness forecast, what approach/process would you adopt? In what sort of time-scale?

Supplementary Tables

Analogy 4

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

Analogy 5

Description	
Similarities and Differences	

Source _____   Similarity Rating _____   Closest Outcome_____

UJ

Manchester Business School

Name:

Each of these files contains a short description of a policy and the proposed strategy for implementation. Please read the description and follow the instructions.

Case 3 – Killing Two Birds with One Stone

Description

In light of the battle against climate change, national governments are always looking for ways to improve their overall efficiency and thereby reducing their total carbon emissions. As one of many nations suffering from the negative effects of the recent global recession on the automotive industry, Nation X has devised a strategy which will simultaneously inject a much needed boost into this sector as well as putting greener and more efficient vehicles on the road.

This strategy rewards the purchase of a new vehicle (with a retail price of less than £27,000) when an older vehicle (of less than 25 years) is traded in. The size of this reward, which will come in the form of credit towards the purchase of the new vehicle, ranges from £2,100 to £2,600 depending on the difference in efficiency between the vehicle traded in and the one purchased. The overall available budget for such a strategy is £1.8 billion. The incentive offered by the strategy is due to be in place for 5 months or until the budget is exhausted.

Before implementing this strategy, the Government of Nation X is trying to evaluate whether this policy implementation strategy will be effective and to what extent. Given that the overall budget for the proposed strategy is £1.8 billion, they require you to give a prediction as to the percentage of this budget that will actually be used by people taking up on the offer during this 5 month period.

Instructions

Outcomes

Please select one of the following as your prediction of the percentage of this budget that will actually be used by people taking up on the offer during this 5 month period:

A	0% – 20%	[ ]
B	21% – 40%	[ ]
C	41% – 60%	[ ]
D	61% - 80%	[ ]
E	81% - 100%	[ ]

Within your chosen range, please provide a point forecast for the percentage of this budget that will actually be used by people taking up on the offer during this 5 month period

[   ]

2. Questionnaire

1) Roughly, how long did you spend on this task?

277

{include the time you spent reading the description and instructions}   [____] mins

2) How likely is it that taking more time would change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

3) If you knew that this case was from the UK, how likely would you be to change your forecast?

{0=almost no chance (1/100) … 10=practically certain (99/100)}   [__] 0–10

4) Do you recognise the actual case described in this file?   Yes [__] No [__]

If so, please identify it: [_____]

5) How many people did you discuss this forecasting problem with? [_____] people

6) Roughly, how many years experience do you have working in an environmental issues setting?

[ ] years

7) Roughly, please rate (out of 10)

- your experience with environmental public policy.   [_____] 0–10

- your experience with cases similar to this one.   [_____] 0–10

- your suitability for predicting the effectiveness of policy implementation strategies such as these.

[_____] 0–10

8) If you were contracted to producing such a strategy effectiveness forecast, what approach/process would you adopt? In what sort of time-scale?

Appendix 2

Acronyms

ARCH – Autoregressive Conditional Heteroskedasticity

ARIMA – Autoregressive Integrated Moving Average

CBA – Cost Benefit Analysis

CEA – Cost Effectiveness Analysis

CWS Ratio – Cochran-Weiss-Shanteau Ratio

DECC – Department of Energy and Climate Change

DEFRA – Department for Environment, Food and Rural Affairs

EC – European Commission

ECT – Environmental Clean Technology

EST – Energy Saving Trust

EU – European Union

FBA – Forecasting by Analogy

FSS – Forecasting Support System

GARCH – Generalised Autoregressive Conditional Heteroskedasticity

GHG – Green House Gas

GMRAE – Geometric Mean of Relative Absolute Error

GMRMAE – Geometric Mean of Relative Mean Absolute Error

IA – Impact Assessment

ICEPT – Imperial (College) Centre for Energy, Policy and Technology

IPCC – Intergovernmental Panel on Climate Change

LTM – Long Term Memory

MAE – Mean Absolute Error

MAPE – Mean Absolute Percentage Error

MBI – Market Based Instrument

MdAE – Median Absolute Error

MdAPE – Median Absolute Percentage Error

NEPI – New Environmental Policy Instrument

NGO – Non-Governmental Organisation

NPM – New Public Management

NPO – Non-Profit Organisation

PM – Performance Measurement

PIS – Policy Implementation Strategy

SA – Structured Analogies (as proposed by Green and Armstrong, 2007b)

s-SA1 – Semi-Structured Analogies1 (used in experiments I, II and III)

s-SA2 – Semi-Structured Analogies2 (used in experiments IV and V)

s-SA3 – Semi-Structured Analogies3 (proposed for future research)

SIG – Special Interest Group

STM – Short Term Memory

UJ – Unaided Judgment

UKERC – United Kingdom Energy Research Centre

VA – Voluntary Agreement

VED – Vehicle Excise Duty

Appendix 3

Details of Attached Papers

Copies of papers that have been published as a result of the research carried out for this thesis are attached and their details are as follows:

Savio, N. and Nikolopoulos, K. (2009a). "Forecasting the economic impact of new policies", *Foresight* 11 (2), pp. 7-18.

Savio, N. D. and Nikolopoulos, K. (2009b). "Forecasting effectiveness of policy implementation strategies: working with semi-experts", *Foresight* 11 (6), pp. 86-93.

Savio, N. D. and Nikolopoulos, K. (2009c). "Forecasting the Effectiveness of Policy Implementation Strategies", *International Journal of Public Administration* 33 (2), pp. 88 - 97.

Papers

Forecasting the economic impact of new policies

Nicolas Savio and Konstantinos Nikolopoulos

Abstract

Purpose – Once a policy proposed by the European Commission is approved by European Parliament or Council, its implementation strategy is the responsibility of the member states. Often, there will be several parallel strategies shaped by a series of incentives financed by the government and naturally, the aim is to choose the most cost effective one. For strategy and planning as well as budgeting purposes, forecasts of the adoption rate of these policy implementation strategies will be an indicator as to their effectiveness. A new hybrid approach combining structured analogies and econometric modelling is proposed for producing such forecasts.

Design/methodology/approach – With every different policy, there will be different qualitative and quantitative data available for producing such implementation strategy adoption rate forecasts. Hence, the proposed hybrid approach, which combines the strengths and reduces the weaknesses of each of its constituents, can be adjusted to match the quantity and nature of the available data.

Findings – This paper reveals a lack of emphasis on such a forecasting application in the existing literature, while stressing its importance to governmental decision makers. What is more, the paper reveals a lack of documentation of this forecasting process in large governmental structures.

Practical implications – If shown to improve the ability to produce such forecasts, the proposed approach could be very beneficial to decision makers when faced with several possible implementation strategies.

Originality/value – The use of expertise is quite common in forecasting policy impact but in an unstructured way. The advanced model proposes structuring the use of analogies in an objective manner. Furthermore, combining with econometric modelling, the incorporation of valuable quantitative information is made possible.

Keywords Financial forecasting, Research methods, Social policy, European Union

Paper type Conceptual paper

1. Introduction

Relative to other areas of forecasting research, policy impact forecasting has seen considerably less interest from the field in the past. Possible reasons could include the high levels of complexity and low success rate associated with forecasting the impact of a new policy. Nonetheless, it must be said that such a forecasting application is key for any government when assessing whether to implement a policy and hence very important for long range planning and strategy in any governance structure.

The notions of sustainable development, climate change and nature conservation are undoubtedly high priorities on the European Union's (EU's) political agenda. It is the EU's responsibility to set the European standard and lead the way for other countries and organisations around the world in combating the negative effects of humanity's development. This stresses the importance of environmental policies and their successful implementation. As part of the latter, the ability to forecast the economic impact of these new policies and hence improve decision making for the relevant governance structures is vital, assuming of course that better forecasts improve decision making.

Nicolas Savio and Konstantinos Nikolopoulos are both based at the Manchester Business School, Manchester, UK.

The following paper focuses on how to assess the economic impact of new environmental policies and recognises the importance of such a sound prediction and the potential benefits it can offer. More specifically, it focuses on policies that are implemented through initiatives (also known as policy implementation strategies) promoting a new type of eco-friendly technology via incentives. The rate of adoption of the new technology is seen as an indication of the effectiveness of the initiative. With various possible policy implementation strategies, of different shape and specifications, available for attaining the same objectives, any government will strive to opt for the most cost effective. Having fixed budgets for such projects, ex-ante forecasts of such adoption rates are strong indicators of the financial investment required by each alternative implementation strategy and hence very important for deciding which is to be chosen. Clearly, this will be significant for

budgeting purposes. As a result, the paper proposes a method based on combining econometric modelling and structured analogies for producing such ex-ante adoption rate forecasts.

The paper starts with a discussion on policy implementation strategies and the need for such forecasts. The proposed approach for producing these forecasts and the requirements for developing such an approach are then described. Finally, the paper concludes and makes suggestions for further research.

2. Policy implementation strategies

As the main executive branch of the EU, one of the European Commission's (EC's) principle roles is to co-ordinate the policies between the member states to be adopted by the EU. They do so by proposing legislation and consequently overseeing their implementation. The principle of subsidiarity says, however, that it is the duty of the member states to implement this legislation and hence the strategy for doing so is entirely their prerogative.

Many policy implementation strategies, or initiatives, stem from and are made possible by the advent of new environment and clean technologies (ECT), used to confront existing problematic situations. A significant challenge faced by governance structures is to reduce the delay between the identification of such problems and the development and implementation of the policy strategies to cope with them (Eder and Leone, 1999). Appropriate foresight exercises and forecasting approaches are needed to help policy makers design policies capable of ''pre-empting rather than remedying environmental damage'' (Eder and Leone, 1999, p. 548). In order to do so, we need to improve our ability to assess the consequences of ECT progress. Such a need is particularly relevant in view of the EU's over-arching goal of sustainable development where Johnston (2001) rightly highlights a need for both ''intergenerational'' and ''transversal'' solidarity.

Tavares (2002) emphasises the importance of technology foresight in development policies and similarly, ECT foresight is important in environmental policies. The task of foreseeing the impact of new ECTs is a rather complicated one because their adoption is dependant on so many conditions. For example, oil price has a direct influence on the economic rate of adoption of alternative energy technologies (Dearing, 1999). This difficulty is then ''transmitted'' to the ability to forecast the impact and adoption of the policy implementation strategies promoting the new ECTs.

A particularly high-profile environmental policy that has seen an increase in political pressure in recent times is that of climate change. Research and development institutes in both the public and private sector, on both a European and member state scale, are continuously trying to develop new eco-friendly technologies to improve efficiency, quality, reliability, etc. with this policy in mind. The government then attempts to promote the change over from the old wasteful technologies to the newer, more efficient ones through government-funded policy implementation strategies, or initiatives.

It is often the case that the EU policy will have objectives or targets that, through legislation, will be compulsory for the member states and it is these new technologies that will provide the means for hitting these targets. To illustrate these ideas, the example of road transport emission reduction can be used as shown in Table I.

As May (2004) rightly explains, the automotive industry is undergoing a transition period in which the actors in the industry must work together with the governance structures in reducing the environmental impact of road vehicles. Certainly, there must be co-ordination between the developers of the ECT and the developers of the policies that aim to promote their adoption in order to maximise the chance of success.

Taking the above example, several strategies will exist for attaining such a target set by the legislation and it will be in the interest of the government of that member state to evaluate which initiative strategy will be the most effective (through a cost/benefit analysis for example). This is illustrated in Figure 1.

As can be seen in Figure 1, when implementing such legislation, the governments of the member states will have different, rivalling, strategies available for attaining the same objectives. Each of these strategies will be shaped differently with different characteristics in terms of incentives, which will have a direct bearing on the rate at which the new technology is adopted, an indication of the effectiveness of the strategy. This then will imply how much money is to be invested in the implementation.

Based on the characteristics of the strategy, ex-ante forecasts for such an initiative adoption rate can be made. These ex-ante forecasts will provide a means for rating the effectiveness of the policy implementation strategy, or initiative, before it is introduced, thereby making these a key tool in the governmental decision making. Taking the example from Figure 1, we obtain Table II.

EU policy Combat the negative effects of climate change

284

EU legislation/directive to member states for
such a policy
Reduce average vehicle CO_2 emissions by 2012
to 120 g/km
Member state strategy to comply with directive Initiative launching new type of eco-friendly
engines (e.g. hybrid), with promotion of change
over via incentives

In such a situation, a typical decision-making process for choosing which of the two policy implementation strategies to adopt, would have to consider aspects such as:

ʙ Will both A and B attain the required target?

ʙ How long will it take for each strategy to hit the target?

ʙ Which strategy then is the most cost effective? Etc.

Any governance structure will have a fixed budget available for funding such initiatives. The way the initiative is perceived by the public will depend on various criteria, such as the perceived attractiveness of the incentive(s), the image of the new technology being produced (perceived ease of use, perceived usefulness), etc. Such criteria will then determine how the new policy is perceived by the public and will be an indication as to how quickly it will be adopted. This notion of limited funding for these initiatives stresses the importance of being able to produce ex ante forecasts of how each strategy, defined by its different characteristics, will perform in achieving the desired targets and objectives set out by the EU policy.

The reason for this is because such predictions of policy take up rates will be an indicator as to the economic investment needed for its implementation and so will be decisive in helping the concerned decision maker select the most cost effective strategy to adopt. Decision making in governance structures is just as important, if not more so, as in profit organisations, because such decisions will affect millions of people as well as the multitude of environmental, social and economic systems in place.

Forecasting the economic impact of new policies in this way is often challenging because each situation is unique. The high level of uncertainty surrounding the forecasts is generally consistent but the extent and availability of quantitative as well as qualitative data can vary considerably from one situation to the next. Like for any forecasting task, it is important to use all available information, of both nature, that will help improve forecast accuracy and to use analysis methods that capitalise fully on their availability.

Currently, during any policy proposal process, in order to assess the impact the policy will have on an environmental, social and economic scale, the EC conducts impact assessments (IA). Different models[1] are available for such assessments and the availability of data for each particular case as well as the requirements of the IA will determine how quantitative or qualitative the analysis is and which model is to be used (see Table III). In many cases a mixture between the two is used and it is common to make use of external expertise if necessary.

Similarly, the UK Government also carries out an impact assessment during the proposal process of its own policies. But generally, when proposing an initiative, aimed at complying with a EU directive, the different strategies are examined with cost/benefit analyses with the help of relevant economists/experts in area, as well as with different stakeholders.

EU legislation target – reduce average vehicle CO_2 emissions by 2012 to 120 g/km
Possible policy implementation strategies Cost (£)
A
10 per cent subsidisation of hybrid cars purchase price L
100 per cent discount on road tax (road tax exemption) M
100 per cent discount on congestion charge (congestion charge exemption) N
Total L þ M þ N
B
40 per cent subsidisation of hybrid cars purchase price X
85 per cent discount on road tax Y
85 per cent discount on congestion charge Z
Total X þ Y þ Z

In all of these cases of policy forecasting, whenever quantitative data are available it is common for governmental analysts to use some sort of econometric model. Such models are useful for establishing the causal relationships between economic variables and using them to make predictions. In the absence of quantitative data, popular judgmental methods

include unaided judgment, the Delphi technique, panel groups as well as analogies, the choice of which depending on the availability of the experts. Gordon (2007) presents a tweaked (roundless) version of the Delphi method (which he calls the RT Delphi) for forecasting energy scenarios, which he shows to produce the same results as Delphi but in a shorter time span. It remains to be seen if the RT Delphi provides as much information and has a wide a range of applications as the traditional version.

Unfortunately, however, in most cases, when expertise is used, it is elicited in an unstructured way, and this, as will be seen later on, is subject to many limitations. A structured approach to the use of expertise could make full use of this domain knowledge whilst minimising its negative effects. Moreover, the procedure used by the EC for eliciting and analysing expert judgment is poorly documented. If a formal procedure exists for doing so, literature relevant to it is nonexistent or inaccessible.

Of particular importance to this paper is the use of expertise for forecasting the governmental initiative adoption rate, as described earlier. The use of analogies for such a

CGE
models
Sectoral
models
Macro-econometric
models
Environmental impact
assessment models
Micro-simulation
models
Range of coverage of measure
Single market analysis without economy-wide
impacts £
Single market analysis with economy-wide
impacts £ £
Multi-market effects in secondary markets £ £
Ecosystem £
Purpose of the model analysis
Simulation (long-term) £ £ £ £
Forecasting (short/medium term) £
Effects to be analysed
Economic effects (within given model framework) £ £ £
Ecological effects of economic activities £ £ £ £
Ecological effects £
Distributional effects
Between countries £ £ £ (£)
Between sectors £ £
Between households £ £ £
Degree of disaggregation
Between sectors or households
Potentially high £ £
Potentially low £
Within a sector
Potentially high £
Potentially low £ £
Effects on:
GDP £ £
Ecological damages £
Unemployment £ £
Public budget £ £
International trade £ £
Emissions £ £ £ £
Emission/deposition £
Household income £ £ £
Source: EC (2006, p. 22)

task provides an attractive prospect as valuable information can be taken from analogous initiatives (from the past or other circumstances) and applied to the target situation.

3. Structured analogies

The use of analogies for the purpose of forecasting has been the focus of notable research with fairly positive outcomes. One such study by Duncan et al. (2001) explored the use of analogies in the context of time series forecasting, while Nikolopoulos et al. (2007) used analogies to forecast television audience ratings. McIntyre et al. (1993) used analogies by buyers to build an expert system to forecast the effects of sales promotions. Although demonstrating the utility of analogies for forecasting, these studies failed to consider the

issues surrounding the use of analogies for the forecaster and the potential difficulties in using them. In other words, they overlooked the need to offer the forecaster support when using analogies.

However, as two recent studies (Lee et al., 2007; Green and Armstrong, 2007) have recognised, the unstructured use of analogies can be constrained by the various cognitive limitations of the human mind. Lee et al. (2007) and Green and Armstrong (2007) claim that a forecaster needs support when using analogies and consequently propose approaches to overcome these problems.

Lee et al. (2007) examine the process of using analogies for forecasting and identify three main stages involved in their use:

1. Recall. The need to actually retrieve the past cases.

2. Similarity judgments. The need to determine the suitability of the retrieved cases through assessing their similarity.

3. Adaptation judgments. The need to make adjustments to these past cases to suit the target situation in order to produce the forecasts.

The paper recognises the limitations of the human mind and the effect such limitations can have on the use of analogies. It provides evidence of the difficulties faced by the forecaster in each of these three stages when deprived of any support and how providing such aid could be beneficial.

The authors argue that the recall of past cases can be hindered by human memory limitations (only a small number of cases may be recalled), incorrect recollection of case details and an erroneous recollection method. Therefore, to this they propose the implementation of a database of past cases that will act as memory support and an aid for the recall process.

In the same way, the paper argues, similarity judgments can be difficult for the forecaster to make due to cognitive limitations and so the authors suggest that this too should be supported. They hypothesise that "providing similarity support in addition to memory support will lead to more accurate forecasts than providing memory support alone" (Lee et al., 2007, p. 4).

Finally, the authors argue the need for support for the adaptation judgments and hence further hypothesise that providing support for all three stages will provide more accurate forecasts than had just similarity and memory been supported.

These ideas are the base for a forecasting support system (FSS) developed by the authors to help assess the effects of sales promotions on demand forecasting tasks. The FSS is intended to aid users with similarity and adaptation judgments when drawing up information from the past. Their study showed that such an FSS could substantially improve forecast accuracy under certain conditions (see Lee et al., 2007, pp. 10-12).

Although the ideas in Lee et al. (2007) for supporting the use of analogies for forecasting are valid and the results found positive in the context of accuracy, the development and use of such an FSS will be quite costly and will not be suitable to all situations. Furthermore, such an

FSS assumes that past cases will be available for the building of a memory support database, which will not always be the case.

A similar study by Green and Armstrong (2007) proposes a formal method for using analogies for forecasting to overcome the potential biases associated with the human interface when performing such a task. The authors acknowledge the benefits of using analogies in many forecasting situations, but believe that experts will tend to choose easy to recall analogies, which in turn could be the similar cases that confirm their beliefs. In other words, they believe that if analogies are used in an unstructured manner, people are prone to using inferior analogies that will be undoubtedly subject to different biases.

Stemming from evidence of the success of structured use of information as opposed to unstructured use in judgmental forecasting (Armstrong, 1985, Ch. 6), Green and Armstrong (2007, p. 366) propose that a structured approach to forecasting by analogies (FBA) would encourage experts "to consider more information from the analogies, and to process it in a more effective way" which leads them to hypothesise that the use of analogies would improve forecast accuracy provided they were used objectively and in a structured manner. In order to do so, the paper suggests that the use of analogies should be structured following a five-step procedure[2]:

1. Description of the target situation. The administrator (the person responsible for the collection and weighting of the experts' advice) prepares a comprehensive but brief description of the situation by seeking advice either from expected unbiased experts (in the case of EU policy, these can be found in academia for example) or from experts with expected differing biases (again in the case of EU policy, these can be found in the different stakeholder organisations, lobby groups, etc.).

2. Selection of experts. The administrator recruits a set of experts who are likely to know

287

about situations analogous to the target situation. These experts should be chosen on the basis of how much knowledge they have on these analogous situations, the variability in their responses and the importance of obtaining accurate forecasts.

3. The experts each identify and describe analogies. The experts are asked to describe as many analogous cases as possible (without considering the extent of similarity to the target situation) and to match their analogies' outcomes with the target outcomes.

4. The experts each rate the similarity of the situations. The experts are asked to list similarities and differences between the analogies and the target situation and then rate the level of this similarity.

5. Derivation of the forecasts. To promote logical consistency and replicability, the administrator should decide on the rules for deriving a forecast from experts' analogies. The forecaster then transcribes this information and uses it to forecast. In this way, the knowledge and expertise available from analogous cases can be extracted more efficiently and centred on providing relevant information for the forecasting of the target variable.

Their study, which compares the structured analogies approach with unaided judgment and chance, looks at eight different conflict situations, whose outcome must be predicted. The study reports that when predicting decisions made in eight conflict situations, 46 per cent of structured analogies forecasts were accurate, compared to the 32 per cent accuracy of unaided experts' forecasts, which were little better than chance. They also report that when the experts showed an increased expertise in the domain, the accuracy of the structured analogies forecasts increased further.

Although self proclaimed as ''objective'' by its developers, the structured analogies approach exhibits the usual problems associated with the use of experts for forecasting. That is to say, that in a few steps in the process, subjectivity cannot be avoided. The fact that the administrator, responsible for selecting, eliciting and analysing the expertise, is different for each task means that the overall process remains essentially subjective in nature. Nevertheless, Green and Armstrong's (2007) method makes good progress in making a

highly subjective process as objective as possible and can hence be seen as considerable advance in the evolutionary line of forecasting by analogies.

Both studies have proposed a way of formalising the use of analogies through structuring the FBA procedure and have found that this leads to improved forecast accuracy, and even more so when the knowledge of the experts increases[3].

Having seen how analogies can be useful in situations that are quite difficult to forecast, such as conflict forecasting or technology forecasting, one would expect FBA to be useful in forecasting tasks associated with policy impact, a similarly difficult situation. Forecasting the impact of a new policy can be treated as a non-periodic special event. More often than not in such cases, data are not available to the forecaster and so recourse to analogies is an attractive prospect. Furthermore, having examined the evidence supporting the structuring of FBA to increase forecast accuracy, forecasting the impact of new policies via structured analogies would be a rational step forward.

4. Strength of combining

Moreover, such a structured use of analogies could provide an ideal compliment to the forecasts produced by the econometric model. There is no evidence to suggest that a combination between quantitative and qualitative methods is currently used by governmental analysts when forecasting policy impact. Several studies have shown the potential benefits of combining quantitative forecasts with expertise in the domain. In a study on the benefits of combining forecasts, Armstrong (2001) uses previous research on the subject to make a valid contribution on the benefits of combining and when such an approach is suitable. Briefly, he concludes that combining offers a means of minimising the weaknesses of each constituent while consolidating their strengths. In order to achieve best results, and of particular relevance to this research, he advises to:

ʙ Combine forecasts derived from substantially differing methods and drawing from difference sources of information.

ʙ Adjust the weighting on each forecast depending on the level of confidence in the accuracy of that forecast.

He reasons that combining is a recommendable strategy in forecasting when the situation involves a high degree of uncertainty, when the identification of the most accurate method is not obvious and to avoid large errors (where the three are not necessarily exclusive).

This therefore seems to be particularly relevant since all of the conditions described above seem to match the characteristics of policy impact forecasting. The combination between the econometric forecast and the structured analogies prediction will undoubtedly draw information from different sources as well as being largely contrasting methods. Hence, a combination between these two methods will provide an approach that consolidates the strengths of both whilst minimising their weaknesses.

288

The hybrid model

The weighting of the forecast from each of the methods will depend on the availability of information/data in each of the cases. A simple illustration showing the nature of the proposed hybrid model for forecasting adoption rate (as a percentage of the target population that has converted to the new eco-friendly technology) would be:

$$\%_{Adoption} = W_1 F_{SA} + W_2 F_{Eco}$$

where W_1 and W_2 are weights given to the structured analogies forecast, F_{SA}, and the econometric forecast, F_{Eco}, respectively. These two weights can be adjusted depending on the amount of data of that type available. Clearly, if no quantitative data were available for a given policy, the full weight would be on the structured analogies forecast and the approach would be purely judgmental.

A soft systems view of the whole situation of policy development and implementation is depicted in Figure 2. The diagram presents the main actors in the situation:

ʙ The EC. This organisation is responsible for proposing legislation that is in line with EU policies, such as sustainable development. Furthermore, they are responsible for overseeing the implementation of such policies.

ʙ The governments of the member states (the UK Government). The principle of subsidiarity says that it is the member states that are responsible for the implementation of the EU policies and how they do it is up to them. If a new eco-friendly technology can play a part in such an implementation, these governments will develop policy implementation strategies, or initiatives, aimed at promoting the use of such technologies (in the private sector and the public) through incentives (subsidised prices, compensation schemes, etc.) to promote the change over. There will be a fixed budget for investment into these initiatives so such governments will need to forecast the adoption rate of each short listed strategy so as to have an idea of the economic impact of each alternative. This will provide an indication as to the effectiveness of each and will hence allow identifying the most cost effective option.

ʙ The private sector. This will include the companies responsible for developing the new technologies to be promoted by the governance structures. They will have business targets/objectives so they will want to forecast the rate at which their technology is being adopted (which can be viewed as a measure of performance). They will also include private companies that offer their employees benefits for adopting the new technologies.

ʙ The public. This will be the main target group of the policies, at which the new technology is aimed.

In the centre of the diagram sits the approach for forecasting adoption rate, and thereby the economic impact of the initiative promoting it. Similarly, the proposed approach can be used for producing the adoption rate forecasts required by the companies.

5. Evaluating the new methodology

In order to develop and test the proposed approach against current methods, data is required. In an attempt to attain such data, two prominent sources stood out as sensible starting points: the EC and the UK Government. The aim was to obtain as many cases where three conditions held:

1. the policy proposed led to the development and implementation of a governmental initiative aimed at hitting a policy target;

2. a forecast of the adoption rate of this initiative was produced during the ex-ante cost/benefit analysis (or other); and

3. a follow-up study was conducted to determine the actual adoption rate.

As illustrated by Figure 3.

With this research proposing a tool that could potentially benefit and facilitate regulation, it was natural to approach the Better Regulation directorates of both governments in search of the cases described above.

As a starting point, a meeting was held with officials of the Secretariat General (which has better regulation as one of its key responsibilities) of the EC in Brussels, Belgium (4 December 2007). Having presented the main ideas of the research, the main conclusion drawn was that the EC was not sure what data were actually available. Being a transparent organisation, the EC publishes all past impact assessment (IA) reports along with their policy proposal documents. However, the EC emphasised that follow-up studies (as in condition 3 above) are only carried out for ''big money'' projects. Hence, we are left in the situation of conducting an exhaustive search of the EC's database to find possible cases with which to test our approach.

Organised in a different structure to the EC, the task of better regulation falls to the

Department for Business, Enterprise and Regulatory Reform (BERR). This time, a conference call was held with an official responsible for the IA template used in UK policy development. Very much like the EC meeting, the main conclusion was that the UK was not exactly sure what was available in terms of data and that it was simply the case of conducting an exhaustive search through their database (also published online), which although larger (dating back to 1998), was far less organised (several broken links). Other possible sources of data yet to be contacted are the Department for Environment, Food and Rural Affairs (DEFRA) of the UK Government and the European Environmental Agency (EEA).

6. Conclusions and further research

Hence, there seems to be no formal procedure in place at EU or UK policy level for producing implementation strategy adoption rate forecasts. If there is indeed one, it has been poorly documented and access to it remains complicated.

Nonetheless, the need for such forecasts is great; not only for budgeting reasons but also for long range planning and strategy. Such forecasts will be an indication of the expected effectiveness of the strategy and will consequently serve as an important criterion for the decision of which strategy to finally opt for and execute.

Similarly, there exists little or no literature documenting any empirical work done on forecasting policy implementation strategy adoption rates. With no past research on this specific application of policy impact forecasting, it will be difficult to compare the quality of the proposed model's forecasts with those produced by any other.

What is more, there is no system in place at these governance structures to control the quality of the forecasts being produced. It is often the case that follow up studies are not carried out to verify the quality of any of the ex-ante forecasts produced. As a consequence, the absence of such a system makes it difficult to truly progress along the right line to ensure the improvement of such policy impact forecasts. Such a documentation would put healthy pressure on decision makers to analyse the situation properly and would propel the need for better forecasting approaches.

Moreover, access to the required cases for testing the forecasting ability of the proposed model, as documented in section 5, is proving difficult. Such cases are not readily available and it will be an important challenge in this research to obtain and use them to test the proposed approach.

Policy implementation strategies exist not only in the public sector, but also in the private, in the shape of corporate policies. Many firms around the world are in the process of becoming "greener" and as a result, are implementing corporate policies in line with this (e.g. encouraging employees to change to hybrid cars by offering them several incentives).

Similarly, the proposed approach could be used to forecast technology adoption rate (without the use of incentives) as this is also an area that suffers from the use of expertise in an unstructured way. The nature of the model means that it is flexible enough to be applied to areas like the ones just mentioned.

Finally, it must be said that all of the ideas presented in this paper could just as easily be explored in a non-European context. This research could be carried out in an American or Asian context very easily, using local data.

Notes

1. Computable general equilibrium (CGE), sectoral models, macro-econometric models, environmental impact assessment models, and micro-simulation models (EC, 2006).
2. This information could be obtained through questionnaires, interviews and panel groups with key decision makers.
3. Forecast accuracy is shown to improve when experts can recall more than one analogy (Green and Armstrong, 2007; Lee et al., 2007; Hoch and Schkade, 1996).

References

Armstrong, J.S. (1985), Long-range Forecasting: From Crystal Ball to Computer, John Wiley & Sons, Chichester.

Armstrong, J.S. (2001), "Combining forecasts", Principles of Forecasting: A Handbook for Researchers and Practioners, Kluwer Academic Publishers, London, pp. 417-39.

Dearing, A. (1999), "Have we the foresight for sustainable development?", Foresight, Vol. 1 No. 2, pp. 131-43.

Duncan, G.T., Gorr, W.L. and Szczypula, J. (2001), "Forecasting analogous time series", Principles of Forecasting: A Handbook for Researchers and Practitioners, Kluwer Academic Publishers, London, pp. 195-213.

Eder, P. and Leone, F. (1999), "The ECT angle", Foresight, Vol. 1 No. 6, pp. 547-61.

European Commission (EC) (2006), "Annexes to European Commission impact assessment guidelines" available at: http://ec.europa.eu/governance/impact/docs/key_docs/sec_2005_0791_anx_en.pdf (accessed 25 September 2007).

Gordon, T.J. (2007), "Energy forecasts using a "roundless" approach to running a Delphi study",

290

Foresight, Vol. 9 No. 2, pp. 27-35.

Green, K. and Armstrong, J.S. (2007), ''Structured analogies for forecasting'', International Journal of Forecasting, Vol. 23 No. 3, pp. 365-76.

Hoch, S.J. and Schkade, D.A. (1996), ''A psychological approach to decision support systems'', Management Science, Vol. 42 No. 1, pp. 51-64.

Johnston, P. (2001), ''Environmental citizenship: Europe's new policy challenge'', Foresight, Vol. 3 No. 2, pp. 95-101.

Lee, W.Y., Goodwin, P., Fildes, R., Nikolopoulos, K. and Lawrence, M. (2007), ''Providing support for the use of analogies in demand forecasting tasks'', International Journal of Forecasting, Vol. 23 No. 3, pp. 377-90.

McIntyre, S.H., Achabal, D.D. and Miller, C.M. (1993), ''Applying case-based reasoning to forecasting retail sales'', Journal of Retailing, Vol. 69 No. 4, pp. 372-98.

May, G. (2004), ''Europe's automotive sector at the crossroads'', Foresight, Vol. 6 No. 5, pp. 302-12.

Nikolopoulos, K., Goodwin, P., Patelis, A. and Assimakopoulos, V. (2007), ''Forecasting with cue information: a comparison of multiple regression with alternative forecasting approaches'', European Journal of Operational Research, Vol. 180 No. 1, pp. 354-68.

Tavares, L.V. (2002), ''Development policies in the EU and technology foresight: an experiment in Portugal'', Foresight, Vol. 4 No. 2, pp. 33-45.

Corresponding author

Konstantinos Nikolopoulos can be contacted at: kostas.nikolopoulos@mbs.ac.uk

Forecasting the Effectiveness of Policy Implementation Strategies

Nicolas Savio[a]; Konstantinos Nikolopoulos[a]
[a] Manchester Business School, University of Manchester, Manchester, United Kingdom

Forecasting the Effectiveness of Policy Implementation Strategies

Nicolas Savio and Konstantinos Nikolopoulos

Manchester Business School, University of Manchester, Manchester, United Kingdom

Primarily, policies are intended to address economic, social and environmental problems. When implementing a policy, any government will be faced with the decision as to what strategy to adopt in order to meet the objectives set out by the policy in the most cost effective way. Several such Policy Implementation Strategies (PIS) may be available, making such a decision not so straightforward. With limited funds available, such a decision has particular importance for budgeting. This paper proposes forecasting PIS effectiveness as a decision support tool. The nature of Structured Analogies (SA) is considered suitable for generating such forecasts. A simpler version of SA, semi-structured analogies (S-SA), where experts do not need to recollect the exact outcome of analogies, is tested. Empirical findings suggest that in the hands of non-experts, the S-SA approach improves forecast accuracy when compared to unaided judgment. Accuracy improves further when forecasts are produced in groups.

Keywords: policy implementation, strategy, forecasting, structured analogies, judgement

Predominantly, policies are intended to provide solutions to problems. For whatever their nature, policy makers at the central authority strive to design them such that the policy objectives are met at the lowest possible environmental, social, and economic sacrifice. A great deal of time and effort is invested in policy design and implementation to make this so. Naturally, such a task will inevitably involve

foresight exercises for making a series of predictions on aspects such as policy impact and implementation effectiveness; a common tool for the former is to conduct an Impact Assessment, for example. Often, policy implementation can take many forms and can be accomplished through various, alternative, competing strategies. Such strategies, hereafter *Policy Implementation Strategies (PIS)*, are of particular interest to this research.

The main focus of this article is on predicting the effectiveness of PISs, as such forecasts are recognized to be important and beneficial to decision making (Wright & Goodwin (2009), Makridakis (1996)). What's more, such performance measurement can be used a means for the long-term improvement of the public policy process. This is all the more important in a world where public management is becoming leaner and results orientated.

More specifically, the research centers on policies which are implemented through *incentive schemes*[1] promoting new types of eco-friendly technology. With various possible policy implementation strategies available, of different shape and specifications, for attaining the same objectives, any government is faced with a decision as to which to adopt. Such schemes are an interesting choice because there is usually limited capital available for their implementation, putting an even greater importance on getting the decision right. Having fixed budgets for such projects, they will naturally move towards the most cost-effective. Ex-ante forecasts of effectiveness are strong indicators of the financial investment required by each alternative implementation strategy and hence very important for such a decision. In view of this, such PIS effectiveness forecasts can be viewed as indicators of the economic impact these new policies will have on the governmental budget. For these reasons, such forecasts are considered very valuable in their role as a decision support tool in governmental decision making. In light of this, the article proposes its own approach for generating ex-ante PIS effectiveness forecasts, based on the use of structured analogies (Green & Armstrong (2007)).

This article presents the results of a pilot experiment conducted to test a new semi-structured analogies (S-SA)

Correspondence should be addressed to Konstantinos Nikolopoulos, Manchester Business School, Booth Street East, Manchester M15 6PB, UK. E-mail: kostas.nikolopoulos@mbs.ac.uk

[1]Usually in the form of financial remuneration from the Government for adopting the new technology

approach, when compared to unaided judgment (UJ), in the hands of non-experts. The aim of this experiment is to provide preliminary insight into the forecasting accuracy of the SSA approach and is the first stage of a larger research project which ultimately aims to test the approach in the hands of experts. The article starts with a discussion on policy implementation strategies and decision making. The aim here is to introduce some of the key concepts associated with this research and to make clear what exactly it is we're trying to forecast. The following section is concerned with performance measurement and how PIS effectiveness forecasts tie in with existing public policy doctrines on this concept. This section also provides an exploration as to how this is

currently being tackled on an EU and UK level as well as presenting the new approach advanced by this research. The section after that offers a review and discussion on the role of experts when producing such forecasts, as in several contexts, it has been shown that the forecasts they produce are no better than that of non-experts. The next section describes the set up, design and results of the pilot experiment aimed at evaluating the new method in the hands of nonexperts. The final section of this article provides a discussion of these results and draws general conclusions as well as proposing a continuation path for this research.

POLICY IMPLEMENTATION

At an EU level, the principle of subsidiarity says that the implementation of any policy or attainment of any policy objectives proposed by the European Commission (EC) and passed by the Council, is the responsibility of each Member State. Failure to do so can result in penalties imposed upon by the EC. Effectiveness of policy implementation has been the topic of considerable research with several papers offering reviews and critiques (Slack (1981), Sinclair (2001), Pitts (2007)).

A popular discussion topic concerns policies aimed at combating the negative effects of climate change. As a superpower, the EU has the responsibility to lead the way in such a battle as outcomes there will undoubtedly influence and affect global trends (Levi-Faur & Vigoda-Gadot (2006)). In both the private and public sectors alike, there is an emphasis on the introduction of environment and clean technologies ((ECT), Eder & Leone (1999)) aimed at improving efficiency, quality, and reliability of operations and processes.

By promoting such a green movement, national governments can use these new ECTs as tools for meeting any policy objectives imposed by global, European, or local political and/or legislative forces. Such a promotion is usually done through state funded policy implementation strategies (Savio & Nikolopoulos (2009)). Policy implementation strategies are initiatives, schemes, programs, or projects offering incentives for the take up of an ECT.

Policy Implementation Strategies and Decision Making

A policy implementation strategy (PIS), coined by Savio and Nikolopoulos (2009), will provide the means through which a national government can attain any local or EU policy objectives or targets imposed on it. This idea is illustrated with Table 1:

Here, EU legislation sets a target of reducing household carbon emissions by 33 percent by 2020, stemming from a policy to combat the negative effects of climate change. In this example, the Member State could launch a PIS promoting the adoption of several measures to improve household efficiency through several incentive schemes offering financial rewards for their implementation. However, this is just one of many alternative strategies possible for attaining such a target as seen in Figure 1.

K alternative PISs for complying with EU policy/legislation are shown. With such a selection of alternatives, the Member State government is faced with a tough decision as to which to opt for. Each alternative PIS will differ in characteristics and incentives offered, which will have a bearing

on their effectiveness. Table 2 illustrates.

Such a decision is of significant economic importance as any governance structure will have a fixed budget available for funding such PISs. This stresses the importance of opting for the most cost-effective PIS and hence places a substantial amount of significance on the decision.

PERFORMANCE MEASUREMENT AND FORECASTING PIS EFFECTIVENESS

Performance measurement (PM) has seen considerable interest within policy administration research in recent times. This emphasis on quality control comes as no surprise considering the popularity of the new public management (NPM) doctrine, in which PM is a key idea.

Such was the popularity of PM that research and applications have been seen across a wide range of jurisdictions over the last 20 years. As a means for comparing, categorizing and evaluating effectiveness of such PM regimes across the

TABLE 1

Example of EU Policy Implementation

EU Policy
Combat the Negative Effects
of Climate Change
EU legislation/directive to Member
States for such a policy
Reduce household carbon emissions
by 33% by 2020
Member State strategy to comply
with directive (PIS)
Initiative involving considerable
financial returns on different
measures to make households more
efficient (e.g. better insulation, more
efficient boilers, differentiated
council tax scheme, etc.)

90 SAVIO AND NIKOLOPOULOS

board, Talbot (2008) proposed a theoretical framework for doing precisely this. Such a framework would allow comparative studies in total performance regimes irrespective of what sector and what point in time. In other words, it was now possible to start amassing a body of knowledge about performance regimes, i.e., when and in what context were they were successful or not and why.

Although studies show that PM is indeed a "way forward" (Haas (2008)) and provides helpful results in several aspects (Andrews & Hill (2003), Bovaird et al. (2003), Mausolff & Spence (2008), Compin (2008)), there are still concerns when the implementation stage is considered (Frank & D'Souza (2004)).

Considered the new paradigm, public value management (PVM), is where many believe the future of the field lays (Talbot (2009)). PVM is seen as a fusion between ideas surrounding efficiency, achievement, and performance with broader, softer notions about the policy role of public managers. Recalling Moore's (1995) strategic triangle, creating public value requires operational feasibility, an aim of creating something of value as well as legitimacy and political sustainability. In other words, PVM can be seen to be NPM with an added importance on feasibility and value creation with one's actions.

Specific to this research, the idea is to use PIS effectiveness forecasts a decision support tool when selecting which PIS to adopt. Basing such a decision on effectiveness ties in

with the need to improve efficiency of the public policy
process (as advocated by the NPM doctrine), whilst ensuring
the objectives of the policy are met. Such a result is then
seen as to have added value in the public domain, provided
the process is operationally feasible as well as politically
sustainable and legitimate, as PVM suggests.
For financial reasons, it will be in the interest of the government
of that Member State to evaluate which of the PISs
will be the most effective and at what cost. PIS effectiveness
is defined as the extent by which the strategy moves the current
situation towards the desired target set by the policy.

FIGURE 1 Graphical illustration of policy development and implementation using policy implementation strategies.

Member States

PIS 1
- Policy objective(s)
- ECT
- *Incentives 1*
Attainment of targets set
out by policy through
effective implementation
Government
Policy
Policy Targets
Principle of
Subsidiarity

PIS 2
- Policy objective(s)
- ECT
- *Incentives 2*

PIS K
- Policy objective(s)
- ECT
- *Incentives K*
K Policy Implementation Strategies
(PIS) available for selection under a
fixed budget allocated for funding
such initiatives!
Cost/benefit analyses!

TABLE 2

Alternative Policy Implementation Strategies Example

EU Legislation Target Reduce Household Carbon Emissions by 33% by 2020

Possible policy implementation strategies Cost (£)

A:
- 45% return on insulation materials L
- 25% purchase price return on new generation efficient boilers M
- Differentiated council tax scheme with greater returns for greener households N
Total L+M+N

B:
- 80% return on insulation materials X
- 35% purchase price return on new generation efficient boilers Y
Total X+Y

As can be seen by the examples in Table 3, such a definition
allows for a quantitative measurement of effectiveness.
Ekins et al. (2002) used a similar criterion for measuring
effectiveness in a number of energy efficiency schemes of
UK households. Such a quantification leads us to think that
a priori information about the effectiveness of a PIS could
serve as an important criterion for the decision described in
the previous section. Hence, forecasts of PIS effectiveness
could prove an important decision support tool.
At this point, it's important to mark the distinction
between what is regarded as "policy effectiveness" and
"PIS effectiveness." These must not be confused. Policy
effectiveness works more on the policy level and would
cover aspects such as "to what extent is reducing average
CO_2 emission in cars actually combating climate change."
PIS effectiveness works more on the strategy level and
would cover "to what extent is subsidising the purchase of

hybrid cars reducing average CO_2 emissions." Several studies have dealt with the former (Scott, 2007, & Glachant, 2001) but the latter has received little identifiable attention from academia or governmental publications.

Forecasting PIS effectiveness isn't all that straightforward since each situation can substantially differ in various aspects such as the perceived attractiveness of the incentives(s) and the public image of the new ECT being introduced (its perceived ease of use and perceived usefulness). This will give an indication as to how the new PIS is perceived by the public and will ultimately provide an indication as to how successful it will be at attaining the policy targets.

The high level of uncertainty surrounding the forecasts is generally consistent but the extent and availability of quantitative as well as qualitative data can vary considerably from one situation to the next. Like for any forecasting task, it's important to use all available information, of both nature, that will help improve forecast accuracy and to use analysis methods that capitalise fully on their availability.

Current Practice at an EU and UK level

Savio and Nikolopoulos (2009) write that governmental analysts commonly use proportionate analysis (EC Impact Assessment Guidelines (2009)) when faced with such a forecasting task. That is to say, the potential impact of the PIS will determine the level of effort in quantification and cost invested into the process. Generally, when quantitative data are available and when the situation is of a complexity worthwhile modeling, an econometric model is used. In the absence of such data or in a situation too complex to model and so such an effort isn't warranted, judgmental approaches (including unaided judgment, the Delphi technique, panel groups, and forecasting by analogies (FBA)) are used.

The exact procedure used by the EC and the UK Government for eliciting and assessing expertise is poorly documented, if at all. No literature describing any sort of formal method for doing so has been found. All that is certain is that when such expertise is called upon, it's done so in an unstructured manner (Savio & Nikolopoulos (2009)), leading to many limitations (Lee et al. (2007), Green and Armstrong (2007)). A structured approach to the use of expert judgment could capitalize fully on the judge's experience while minimizing their biases (Savio & Nikolopoulos (2009)).

It is argued that an attractive prospect for forecasting PIS effectiveness is through the use of analogies as relevant information could be taken from PISs from the past under similar circumstances and used for the benefit of the target situation (Savio & Nikolopoulos (2009)). Experts in the area, who deal with (or are exposed to and are hence familiar with) PISs on a regular basis will inevitably (subconsciously) construct a "mental database" in which they store different strategies together with their characteristics (targets, incentives, costs, etc). For many of these stored PISs, the database will also contain information on whether the strategy was successful or not, and to what degree (i.e., PIS effectiveness). Such a database would put them in a position to, given a new PISs with certain characteristics,

generate forecasts on aspects such as effectiveness. In other words, their previous knowledge has served to "train" their minds and theoretically, by associating these analogous cases to a target case, this will enable them to produce better predictions (in comparison to the "un-trained" mind of a non-expert).

TABLE 3
Examples of PISs and Their Effectiveness Measurement

EU Legislation/Directive to Member States for Such a Policy	Member State Strategy to Comply with Directive (PIS)	Measurement of PIS Effectiveness
Reduce average vehicle CO2 emissions by 2012 to 120g/km	PIS launching new type of eco-friendly engines (e.g., hybrid), with promotion of change over via incentives	Impact on hybrid sales (in number of vehicles)
Reduce overall household energy consumption	PIS promoting the change-over/adoption of new energy efficient lightbulbs though subsidized pricing	Change in percentage of new energy efficient lightbulbs present in households
Reduce average vehicle CO2 emissions by 2012 to 120g/km	PIS offering a £2000 discount on the purchase price of a new car if a car of 10+ years is traded in for scrapping	Impact on sales of new cars (in number of vehicles)

The Semi-Structured Analogies (S-SA) Approach

Recalling the discussion in Savio and Nikolopoulos (2009), analogies can be useful in situations which are quite difficult to forecast (Lee et al. (2007), Green & Armstrong (2007)), such as conflict forecasting or technology forecasting, and so we'd expect FBA to be useful in policy forecasting tasks, a situation of comparable difficulty. With studies (Lee et al. (2007), Green & Armstrong (2007)) showing that FBA accuracy is improved when done so in a structured manner, this research proposes the use of semi-structured analogies (S-SA) for producing PIS effectiveness forecasts. The S-SA approach is slightly different to the one used in Green and Armstrong (2007) in that the forecasts are generated by the experts themselves and not the administrator.

The reason for this is that many times, the experts will recall analogies but will struggle to remember their quantitative outcome or even give a similarity rating, data which is necessary for the administrator to derive the final forecasts objectively (Green & Armstrong, 2007). Hence, the way S-SA works is that experts are aided in structuring their knowledge by recalling analogies and listing them with their descriptions in the Table provided (see Appendix 2). It is hypothesized that this structuring will improve forecast accuracy (when compared to a non-structured approach, referred to as "unaided judgment").

EXPERTS VERSUS NON-EXPERTS IN FORECASTING

It is clear that the hope is for the forecasts produced by the experts using S-SA will be more accurate than unaided judgment. For completeness sake however, this research also wishes to test the ability of the proposed approach in the hands of non-experts. Various studies have covered the performance of experts vs. non-experts in various different

contexts (Bolger & Wright (1994), Wilkie-Thomson et al. (1997), Faulkner & Corkindale (2009), Andersson et al. (2005), Tetlock (2006), Sjoberg (2009))).

What Does the Literature Say?

Sjoberg (2009) explored the ability of different groups (with varying level of expertise) in predicting the outcome of the Swedish elections. He found that the group comprised of random members of the public out-performed the panel of experts. He attributes this victory to information from opinion polls and the wishful thinking effect along with bias on the part of the experts. Staying within the political context, Tetlock (2006) offers a large study in which it puts expert judgment up against actual outcomes, simple statistical methods, uninformed non-experts and well-informed non-experts. The study found that the expert predictions were hardly, if at all, more accurate than those of informed non-experts.

Similar studies include Andersson et al. (2005), who explored the performance of experts when predicting the 2002 FIFA World Cup and Faulkner and Corkindale (2008), who looked at experts' ability at predicting the success of new product. Andersson et al. (2005) found that experts and non-experts were just as good as each other in such a context and attributed this result to inefficient use of expert information as well as too much reliance on a "quick and easy" prediction heuristic.

It must be said that another likely explanation for this result is that a sporting event such as a football world cup is inherently unpredictable because any team's success goes beyond measurable indicators such as quality, form, world ranking, etc., and includes extraneous, abstract factors (mainly psychological), which are difficult to account for. Faulkner and Corkindale (2008) results showed that experts are no better than the public in forecasting the success of a new product. The authors evaluated the knowledge of the experts and they were not actually more predisposed to the task than the public.

Faulkner and Corkindale (2008) seems to be an example where not enough care was taken in defining and selecting the group of experts. In fact, it is seldom that one sees the word "expert" defined in the literature. Often, the term is used very broadly and such a lack of definition (albeit unintentional) can lead to disappointment when performance is considered simply because what was thought to be expert judgment, simply isn't. Perhaps this lack of definition is due to the fact that this is such a difficult task in any context. How do we know someone is an expert in politics or technology or environmental policy? In no area is expertise in the subject scrupulously delineated. Expertise is shaped by individual training and experience and so can never really be objectively measured (making comparison difficult), and so one is never entirely sure the level of expertise one is getting when conducting a study involving experts (Shanteau et al. (2002)).

A success story for expert judgment is a study carried out by Wilkie-Thomson et al. (1997), which compared expert and novice performance in the context of financial forecasting. The authors argue that the significant success of experts in this context is due to their specific knowledge of the time series used (currencies) in addition to their general knowledge of financial forecasting.

The above studies show that the performance of experts compared to non-experts really depends on the difficulty of the task and the quality of the expertise available. A significant paper on the topic is Bolger and Wright (1994), because it reveals, after examining the literature, that, of 20 studies conducted in different areas, 6 showed superior expert performance, 9 showed inferior expert performance, and 5 showed equality.

Having reasoned that expertise should be measurable in improved accuracy with respect to non-expert predictions in order to be useful, such poor expert performance is explained in terms of two key dimensions. According to the authors, the

"ecological validity[2]" and the "learnability[3]" of the forecasting tasks will determine the level of expert performance. Scenarios in which both these dimensions are high will suggest good expert performance. When one or both are low, non-experts could perform just as well (Bolger & Wright (1994)).

Expert versus. Non-Expert in a PIS Context

Within the framework of this research, which will be the first time the expert vs. non-expert debate is applied to a PIS context, the forecasting task (involving experts) is thought to be of high ecological validity and learnability so experts are expected to perform better than non-experts (university students). The mental database of analogies described in previous sections will undoubtedly provide the knowledge and experience required to generate accurate expert forecasts (more so than non-experts for that matter) for the target case. The real question will be if a structural approach to the use of this database will improve upon unaided expert judgmental forecasts. This then leads us on the pilot experiment of this research, which had as aim to explore how nonexperts, using unaided judgment and S-SA, performed when forecasting the effectiveness of three real life policy implementation strategies.

TESTING S-SA IN PRACTICE

In order to test the proposed approach and get a first feel for its forecasting ability, a pilot study was conducted with a group of non-experts.

Methodology

A total of 31 students from the Manchester Business School were used as participants in the pilot experiment. All the participants were undergraduate students well into a "forecasting and applications" module and so were already familiar with judgmental forecasting approaches such as unaided judgment and structured analogies. They were told that this was a chance to apply the methods they had learned and extra credit on their final grade for the course was offered for the best forecasters as an incentive.

The participants were asked to forecast the effectiveness of a PIS using either an unaided judgment (UJ) approach or a purely semi-structured analogies (S-SA) approach (depending on their assigned group number), individually and then in groups.

The experiment consisted of presenting three different disguised cases of real policy implementation strategies in which three conditions held:

1. The policy proposed led to the development and implementation of a governmental PIS aimed at

hitting a policy target.

2. A *forecast* of the effectiveness of this PIS was produced during the ex-ante cost/benefit analysis (or other).

3. A measure of the *actual* PIS effectiveness is available and accessible.

This is illustrated by the diagram in Figure 2, below. The three cases were taken from different global contexts and were written in a way that would (hopefully) render them difficult to recognize. The three descriptions included details of the policy objectives as well as details of the PIS (incentives offered, etc.). See Appendix 1 for the descriptions of the three cases. Depending on which approach, unaided judgment (UJ) or semi-structured analogies (S-SA), was to be used, participants were given instructions accordingly.

If S-SA was to be used, participants were asked to (1) read the description; (2) try to think of several analogous situations; (3) rate the similarity of the analogous case to the target situation and (4) fill in the questionnaire. A table (see Appendix 2) was provided for steps 2 and 3. When UJ was to be used, steps 2 and 3 were omitted and participants were asked to simply fill in the questionnaire after reading the description. The questionnaires for both treatments for all three cases comprised of four parts; (1) to fill in the required forecasts; (2) a question on the time taken to complete the task; (3) a question on the likelihood of a change in forecast if given more time and (4) a question on whether or not the case had

2The extent to which the results of the task can be applied to every day working contexts.

3The degree to which expertise can be conditioned in the task domain in a useful manner (no feedback of accuracy will essentially waste the experience of a judge).

FIGURE 2 Evaluating PIS

X Years

Time

Policy
Implementation
Forecasted PIS
effectiveness
PIS
implementation
Actual effectiveness of
PIS

94 SAVIO AND NIKOLOPOULOS

been recognized. The questionnaires for the three cases can be found in Appendix 3.

Set Up

Participants were split into two random groups, 1 and 2.

Group 1 size: n = 15

Group 2 size: n = 16

Each of these two groups was then split into a further three groups (1a, 1b, 1c and 2a, 2b, 2c) of 3/4 participants in each for the group exercises.

With the participants split into groups 1 and 2, every participant was individually presented with case 1. Participants in group 1 were asked to use UJ and participants in group 2 were asked to use S-SA to produce individual forecasts of PIS effectiveness. Participants were given 15 minutes to complete the task. Such a time span was set in accordance with the time constraint of the double period in which the students were available for the exercise. What's more, this time span was deemed sufficiently long enough for the participant to understand the case and generate the

forecasts.

Participants were then asked to get into their subgroups and each was presented with case 1 again. Subgroups 1a, 1b, and 1c were asked to produce group forecasts using UJ and subgroups 2a, 2b, and 2c were asked to produce group forecasts using S-SA. Again, 15 minutes was given for this task. This process was repeated for case 2 and case 3, as shown in Table 4. Ten-minute breaks were given (for coffee, toilet, etc.) between changing cases.

Results

The results of the experiment are summarized in Table 5. The results show that:

When performance across all three cases, both individually and in groups, is considered, the S-SA approach is more accurate with a MAPE of 48.0% and a MeAPE (median absolute percentage error) of 24.7% compared to UJ with a MAPE of 52.7% and a MeAPE of 27.6%

The S-SA method also outperforms UJ when forecasts are produced individually and quite significantly when forecasts are produced in small groups. This could be an indication that the S-SA approach can be improved when individuals work in groups and hence more analogies can be recalled.

Having run ANOVA for the four different treatments (UJ individual, UJ grouped, S-SA individual, and S-SA grouped), statistically, there is no evidence to suggest that the MAPE of the four treatments differ. This is true for performance across all three cases as well as for each individual case.

TABLE 4
Treatments

Group Case 1 Case 2 Case 3
1 UJ UJ UJ
2 S-SA S-SA S-SA
1a, 1b, 1c UJ UJ UJ
2a, 2b, 2c S-SA S-SA S-SA

TABLE 5
Summary of Results

Individual Peformance Grouped Performance
Treatment 1. Unaided Judgment MAPE MeAPE MAPE MeAPE
Across all cases (n = 44) 53.73 29.24 Across all cases (n = 9) 48.06 23.56
Case 1 23.04 18.59 Case 1 29.52 24.99
Case 2 117.12 77.30 Case 2 100.95 91.49
Unaided Judgment across
all cases (n = 53)
MAPE MeAPE Case 3 18.98 15.29 Case 3 13.73 11.76
52.76 27.66 MAPE MAPE
Winner Case 1 7.43 Winner Case 1 17.51
Winner Case 2 6.38 Winner Case 2 6.38
Winner Case 3 0.00 Winner Case 3 5.88
Average of Winners 4.60 Average of Winners 9.92
Treatment 2. Structured Analogies
MAPE MeAPE MAPE MeAPE
Across all cases (n = 46) 49.70 23.82 Across all cases (n = 9) 39.33 41.18
Case 1 28.59 24.40 Case 1 32.89 24.73
Case 2 102.79 41.84 Case 2 65.48 56.03
Structured Analogies across
all cases (n = 55)
MAPE MeAPE Case 3 19.14 16.47 Case 3 13.73 11.76
48.00 24.73 MAPE MAPE
Winner Case 1 1.45 Winner Case 1 23.68
Winner Case 2 6.38 Winner Case 2 41.84
Winner Case 3 0.00 Winner Case 3 5.88
Average of Winners 2.61 Average of Winners 23.80

DISCUSSION AND CONCLUSIONS

We consider the findings of this preliminary empirical study quite relevant to public administration and effective decision making in a policy context in that they provide the first steps into the evaluation of a new approach for producing PIS effectiveness forecasts.

In more detail, the important conclusions to be drawn from this research are:

There exists little or no literature documenting any empirical work done on forecasting PIS effectiveness. With no past research on this specific application of policy impact forecasting, it will be difficult to compare the quality of the proposed model's forecasts with those produced by any other.

No literature was found in relation to the use of expertise in governmental forecasting. Such an absence means it is difficult to know exactly what level the governmental analysts are working at. Consequently, this then makes it difficult for academia to identify a suitable line of progress in the area so as to improve on such forecasts.

Policy implementation strategies are needed when assessing different possibilities of how to meet policy objectives stemming from political or legislative forces. The need for accurate PIS effectiveness forecasts is, hence, great as these serve as an important decision support tool for choosing which PIS to adopt. This then, has important implications for governmental planning and strategy.

The results of the pilot study are encouraging. They've shown that even in the hands of non-experts, the proposed S-SA approach (despite being purely semi-structured analogies in this experiment) provides more accurate forecasts than unaided judgment. This is evidence that allowing the participant to structure his thoughts (expertise in the case of experts) before providing forecasts improves accuracy, as would be expected. Accuracy is improved further when forecasts are produced in groups where more analogies can be recalled.

The Next Step

The discussion on experts and non-experts in forecasting indicates that experts are expected to perform better than non-experts when it comes to forecasting PIS effectiveness. Hence, having conducted a pilot study with encouraging results, the next step is to carry out another experiment but this time with the use of experts. This second experiment will be similar in nature to the pilot one but, realistically, on a different scale (depending on the candidate response rate). This will allow for further model validation.

REFERENCES

Andersson, P., Edman, J, & Ekman, M., (2005). Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts, *International Journal of Forecasting*, *21*, 565–576.

Andrews, M., & Hill, H. (2003). The impact of traditional budgeting systems on the effectiveness of performance based budgeting: A different viewpoint on recent findings, *International Journal of Public Administration*, *26*(*2*). 135–155.

Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: Issues and analysis, *Decision Support Systems*, *11*, 1–24.

Bovaird, T., Loeffler, E., & Martin, J. (2003). From corporate governance to local governance: Stakeholder-Driven community score-cards for UK

local agencies? *International Journal of Public Administration*, *26* (*8 & 9*), 1037–1060.

Compin, N. S. (2008). State DOT performance programs: From program development to strategic planning, *International Journal of Public Administration*, *31*, 616–638.

Eder, P., & Leone, F. (1999). The ECT angle. *Foresight*, *1*(*6*), 547–561.

Ekins, P., Russel, A., & Hargreaves, C. (2002). Reducing carbon emissions through improved household energy efficiency in the UK, *Journal of Environmental Policy and Planning*, *4*, 41–65.

European Commission Impact Assessment Guidelines. (2009). Retrieved Feb. 27, 2009 from http://ec.europa.eu/governance/impact/docs/ key_docs/iag_2009_en.pdf

Faulkner, M., & Corkindale, D. R. (2009). Are experts better than potential users in predicting the uptake of an innovation? Extending the use of the Juster Scale. *Technological Forecasting and Social Change*, *76*(7), 910–916.

Frank, H. A., & D'Souza, J. (2004). Twelve years into the performance measurement revolution: Where we need to go in implementation research, *International Journal of Public Administration*, *27* (*8 & 9*), 701–718.

Glachant, M. (2001). The need for adaptability in EU environmental policy design and implementation, *European Environment*, *11*, 239–249.

Green, K., & Armstrong, J. S. (2007). Structured analogies for forecasting, *International Journal of Forecasting*, *23* (*3*), 365–376.

Haas, P. J. (2008). Introduction to the Symposium Performance Measurement: Can it improve performance? *International Journal of Public Administration*, *31*, 593–594.

Lee, W. Y., Goodwin, P., Fildes, R., Nikolopoulos, K., & Lawrence, M. (2007). Providing support for the use of analogies in demand forecasting tasks, *International Journal of Forecasting*, *23* (*3*), 377–390.

Levi-Faur, D, & Vigoda-Gadot, E. (2006). New public policy, new policy transfers: Some characteristics of a new order in the making, *International Journal of Public Administration*, *29*, 247–262.

Makridakis, S. (1996). Forecasting: Its role and value for planning and strategy, *International Journal of Forecasting*, *12*, 513–537.

Mausolff, C., & Spence, J. (2008). Performance measurement and program effectiveness: A structural equation modeling approach, *International Journal of Public Administration*, *31*, 595–615.

Moore, M. (1995). *Creating public value: Strategic management in government*. Cambridge, MA: Harvard University Press.

Pitts, D. W. (2007). Implementation of diversity management programs in public organisations: Lessons from policy implementation research, *International Journal of Public Administration*, *30*, 1573–1590.

Savio, N., & Nikolopoulos, K. (2009). Forecasting the economic impact of new policies, *Foresight*, *11*,(*2*), 7–18.

Scott, E. M. (2007). Setting, and evaluating the effectiveness of environmental policy, *Environmetrics*, *18*, 333–343.

Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not, *European Journal of Operational Research*, *136*, 253–263.

Sinclair, T. A. P. (2001). Implementation theory and practice: Uncovering policy and administration linkages in the 1990s, *International Journal and Public Administration*, *24* (*1*), 77–94.

Sjoberg, L. (2009). Are all crowds equally wise? A comparison of political election forecasts by experts and the public, *Journal of Forecasting*, *28*, 1–18.

Slack, J. D. (1981). Policy implementation effectiveness in cross-national analysis, *International Journal of Public Administration*, 3 (2), pp. 219–242.

Talbot, C. (2008). Performance regimes—The institutional context of performance policies, *International Journal of Public Administration*, *31*, 1569–1591.

Tetlock, P. E. (2006). *Expert political judgment: How good is it? How can we know?* Princeton, NJ Princeton University Press.

Wilkie-Thomson, M. E., Onkal-Atay, D., and Pollock, A. C. (1997). Currency forecasting: an investigation of extrapolative judgment. *International Journal of Forecasting*, *13*(4), 509–526.

Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: Enhancing the scenario method, *International Journal of Forecasting*, *25* (4), 813–825.

APPENDIX 1

PIS descriptions for each case presented to participants

Case 1—New Eco-Friendly Technology Adoption

Description

In light of the battle against climate change, a new ecofriendly technology has been developed that could significantly reduce the total CO_2 emissions of nation X. This new technology, seen as a viable "green" alternative, is rather more expensive to purchase than the more polluting one currently in use. Nation X, worried about its total CO_2 emissions, wishes to promote the adoption of the new technology through a nationalized tax incentive scheme. Such a scheme would reward the purchase of the new technology with a considerable tax return on the price paid (starting at a max of $\pm 15\%$ of original price). After the sale of 60,000 units, the % of the price returned phases out over the next 3 quarters (i.e., once 60,000 are sold with a $\pm 15\%$ return, the % returned gradually decreases over time to 0%)

Case 2—New CO2 Differentiated Tax Scheme

Description

Reducing CO_2 emissions from "a popular mode of transportation" is a key step in combating the negative effects of climate change. In order to meet CO_2 emission targets, Gov. X introduced a strategy based on 3 pillars (voluntary commitments from manufacturers, improvement in consumer information and promotion of "greener" vehicle types via fiscal measures) in year Y to reduce such emissions to the required level by year Y+13. The first two pillars are expected to reduce CO_2 emissions the majority of the way. The last pillar, which consists of a tax scheme which differentiates between vehicle types based on energy efficiency and favours the "greenest" ones, has the aim of reducing CO_2 emissions the rest of the way.

Case 3—Get Digital

Description

The following policy was initiated by the Greek Government and is aimed at supporting the use of computers by students. More specifically, the policy entails governmentally financing laptops for University "freshers" in 2009. The policy says that each student will be refunded 80% of the value of the laptop of his/her choice with a limit of 400 €. The Undergraduate Office plays an active role in the policy because they are responsible for contacting the would-be students in order to inform them and remind them of the offer.

APPENDIX 2

Analogies table offered to participants using S-SA

(i) Description (ii) Source (iii) Similarities and Differences (B) Similarity Rating (C) Outcome

a.
b.
c.
d.
e.
e.g. New technology promoted by
gov. to combat climate change
Media Similarities: new technology
introduced for improvement
purposes
Differences: Different context,
not necessarily for directly
reducing CO2
8 Reduced waste (any
kind) by 24%

APPENDIX 3

Questionnaires for each case

Case 1—New Eco-Friendly Technology Adoption
Questionnaire
1. The following table shows the sales for the new ecofriendly technology. Y refers to the year in which the tax incentive scheme was initialised. Please give your sales forecasts for years Y+1, Y+2 and Y+3 (in light of the introduction of the incentive scheme) as well as a prediction interval (interval that the forecast can range with 95% confidence).
Forecast year Y+1: [ ] Interval: from [ ] to [ ]
Forecast year Y+2: [ ] Interval: from [ ] to [ ]
Forecast year Y+3: [ ] Interval: from [ ] to [ ]
2. Roughly, how long did you spend on this task?~~ *{include the time you spent reading the description and instructions}* [__] mins
3. How likely is it that taking more time would change your forecast? *{0 = almost no chance (1/100) . . . 10 = practically certain (99/100)}* [_] 0–10
4. Do you recognize the actual case described in this file? Yes [_] No [_]
Case 2—New CO2 Differentiated Tax Scheme
Questionnaire
1. Assuming the average emissions for a vehicle type at year Y was *a*, please give a forecast of the **% decrease** of this average (in light of the introduction of the 3 pillar strategy described above) over the 13-year period (i.e., at Y+13) as well as a prediction interval (interval that the forecast can range with 95% confidence).
Forecast for % decrease at Y+13: [ ]Interval: from [ ] to [ ]
2. Roughly, how long did you spend on this task? *{include the time you spent reading the description and instructions}* [__] mins
3. How likely is it that taking more time would change your forecast? *{0=almost no chance (1/100) ... 10=practically certain (99/100)}* [_] 0–10
4. Do you recognize the actual case described in this file? Yes [_] No [_]
Case 3—Get Digital
Questionnaire
1. What percentage of students will take up the offer from the Government? Could you also provide a prediction interval for this forecast (interval that the forecast can range with 95% confidence)?
Take up percentage: [ ]Interval: from [ ] to [ ]
2. Roughly, how long did you spend on this task? *{include the time you spent reading the description and instructions}* [__] mins
3. How likely is it that taking more time would change your forecast? *{0=almost no chance (1/100) . . . 10=practically certain (99/100)}* [_] 0–10
4. Do you recognize the actual case described in this file? Yes [_] No [_]
If so, please identify it:
[_____]

Year Y-5 Y-4 Y-3 Y-2 Y-1 Y
Sales (units) 14,453 20,145 36,249 50,554 84,931 212,112

Forecasting effectiveness of policy implementation strategies: working with semi-experts

Nicolas D. Savio and Konstantinos Nikolopoulos

Abstract

Purpose – Policy implementation strategies (PIS) are schemes designed by a government with an aim of hitting targets or attaining objectives set out by a policy. Forecasting by analogies (FBA) has been shown to be successful in situations of high uncertainty and low quantitative data as is that of PIS effectiveness forecasts. What is more, a structured approach to FBA helps the expert structure his thoughts in an organized manner before making a prediction, which is hypothesized to improve accuracy. This paper aims to discuss these issues.

Design/methodology/approach – This research suggests a semi-structured analogies (S-SA) approach for such a task. A pilot experiment was carried to test the performance of the S-SA approach in the hands of semi-experts when compared with unaided judgment (UJ).

Findings – The results of the experiment showed that for this level of expertise, there is no statistical evidence to suggest any approach is better than the other. Possible explanations of this result is that analogy recall of experts was hindered by four constructs: information, complexity, worldview, and expertise. It was concluded that the structured analogies approach for forecasting PIS effectiveness must be investigated further by means of a study involving "true experts".

Research limitations/implications – The sample sizes were small.

Practical implications – Forecasting PIS effectiveness is seen as an important tool for deciding upon which PIS to ultimately adopt (as there may be many available!) and this then has important implications for governmental budgeting.

Originality/value – The paper offers further insight into the performance of a structured analogies approach to forecasting PIS effectiveness in the hands of individuals with a mid-level of expertise.

Keywords Forecasting, Skills, Government

Paper type Research paper

Introduction

Policy implementation strategies (PIS) are schemes designed by a government with an aim to hitting targets or attaining objectives set out by a policy. This becomes clearer if we look at the different components of the term more closely:

в Policy. Because the policy implementation strategy is ultimately designed for hitting a policy target or attaining a policy objective.

в Implementation. Because the policy implementation strategy is the final stage in the policy implementation process.

в Strategy. Because the policy implementation strategy is "the strategic way" which the government chooses to hit the required target.

Environmental policies tend to have as objectives to protect the environment, e.g. save energy, etc.; particularly for such policies these policy implementation strategies will involve an incentive scheme offering financial recompense for changing over to a more efficient alternative (often, a "greener" technology to reduce carbon emissions for example).

Nicolas D. Savio is based at the Business Systems Division, Manchester Business School, Manchester, UK. Konstantinos Nikolopoulos is Fellow of the Herbert Simon Institute, Manchester Business School, Manchester, UK.

Hence, forecasting policy implementation strategy effectiveness is seen as an important tool for deciding upon which policy implementation strategy to ultimately adopt, as there may be many available; and this then has important implications on governmental budgeting (Savio and Nikolopoulos, 2009a). A central theme within policy administration research over the last decades has been that of performance measurement (PM), a key idea of the new public management (NPM) regime (Haas, 2008; Talbot, 2008). The idea of a lean, results-based management style in public organizations is in line with the ideas of effectiveness, cost based decision making put forward in this paper. More recently, the emergence of the public

307

value management (PVM) regime, considered to be the new paradigm in this line of research (Talbot, 2009), has seen the ideas of NPM fused with softer notions about value creation in the policy process by public organizations. It is argued that the availability of accurate predictions of policy implementation strategy effectiveness will enable government officials to make informed decisions based on an a priori analysis of the situation and a well founded rationale, thereby adding value to the policy process.

To date, the way in which these forecasts are prepared in governmental bodies is poorly documented and unclear (Savio and Nikolopoulos, 2009a). Usually, when faced with tasks of high complexity for which the level of knowledge required is beyond that available within the governmental organization, outsourcing to ''experts'' in the private sector or academia is often done (Brown and Potoski, 2006; Peters and Savoie, 2000). Such a cooperation between the two actors will allow for a decision to be made with the support of a knowledgeable source (Heath and Gonzalez, 1995). What is more, studies have shown that most of the time, decision makers do indeed follow this expert advice (Harvey and Fischer, 1997; Sniezek et al., 2004). When forecasting policy implementation strategy effectiveness there is often little or no quantitative data available and so recourse to expert judgment is a popular and sometimes only viable approach (European Commission Impact Assessment Guidelines, 2009). So the question becomes:

What method should be used to predict such strategy effectiveness and how well does this method perform?

It is thought that when experts currently produce such forecasts in real life, they do so in an unstructured manner, often using unaided judgment, which is subject to many biases which can in turn reduce forecast accuracy (Savio and Nikolopoulos, 2009a). Research suggests (Green and Armstrong, 2007) that a structured approach to such a forecasting task would reduce bias and improve accuracy.

Forecasting by analogies (FBA) (Lee et al., 2007) has been shown to be successful in situations of high uncertainty and low quantitative data as is that of policy implementation strategy effectiveness forecasts. What is more, a structured approach to FBA helps the expert structure his thoughts in an organized manner before making a prediction, which is hypothesized to improve accuracy. In this paper, we document an experiment carried out recently to test a new, structured approach to FBA. This approach, termed semi-structured analogies (S-SA), is tested in the hands of policy developers of a large European governing body and are considered to be semi-experts: that is experts with mid-level expertise[1].

The problem of identifying experts and their level of expertise

The problem when wanting to use expert knowledge for a task is being able to identify an expert and then elicit the necessary information from him or her (Shanteau et al., 2002). The latter will be covered later in this paper but it is the former that is more problematic in this research.

The word ''expert'' is seldom defined in the literature and this is no wonder as it is such a difficult and dangerous task. It is very difficult, not to say virtually impossible, to measure the level of expertise of an individual in a certain area, so it is very tough to anticipate what you are going to get when working with experts. Very often you only realize the level of expertise after the forecasting exercise is complete[2]; this was the case in this study as well. For policy implementation strategy forecasts, it is generally considered that the ''true experts'' can identified and targeted in academia and industry. Governing bodies generally outsource such forecasting tasks to specialized institutes or academic centers, known in the

field to be experienced and trained for such a task. A good discussion on the use of experts in forecasting tasks is offered in Savio and Nikolopoulos (2009b).

Working with semi-experts

An experiment was carried to test the performance of the S-SA approach in the hands of semi-experts when compared with unaided judgment. S-SA was first used in Savio and Nikolopoulos (2009b) when it was tested in the hands of non-experts. The S-SA approach, which is based on the structured analogies (SA) approach in Green and Armstrong (2007)[3], is one promising alternative when participants fail to give quantitative outcomes to their analogies. These exact outcomes are essential for the administrator in order to produce forecasts with the SA method. Nevertheless, S-SA helps the participant to structure his/her knowledge of analogous policy implementation strategy situations by providing a table in which the he/she can describe and give similarity ratings of analogies before producing a forecast.

Participants were taken from the environment directorate of a large European governmental organization. Most participants were policy developers in this department. Participants spanned six different European nationalities, with different backgrounds and levels of experience working in the area.

Responses were obtained through individual interviews in which participants were

308

presented cases along with a description of the real life policy and policy implementation strategy and then asked to produce policy implementation strategy effectiveness forecasts as well as completing a questionnaire (see Appendix 1). For participants using S-SA, extra 30 minutes were given prior to producing forecasts so as to recall, describe, rate similarity and state outcome of any analogies (even though these last two would be difficult) with the use of a table (see Appendix 2). For participants using unaided judgment (UJ), this step was omitted and participants simply produced a forecast after reading the description.

A total of eight participants were interviewed, four produced forecasts using UJ and four using S-SA. Participants were told that the exercise would take no longer than 30 minutes (ten minutes for each case) so time shortage was not an issue. For participants using UJ, completion of each case ranged from three to six minutes and for participants using S-SA, completion of each case ranged from three to ten minutes.

Perhaps the main weakness of this study, the sample size, was determined by the difficulty in locating "true experts", and until a thorough experiment with at least 50 experts is conducted, we can not present statistically significant results. However the following results give a first insight into the potential of the forecasting approaches and will have useful implications when the upcoming, larger study is considered.

Results

The performance of the participants over the two approaches can be summarized in Table I. As can be seen in Table I, UJ seems to outperform S-SA across all three cases, but by very little. Statistically, we cannot say there is any difference between the two approaches. It is

	Case 1 (MAPE*) (%)	Case 2 (MAPE) (%)	Case 3 (MAPE) (%)	Accuracy over 3 cases (MAPE) (%)
UJ	20.4	82.9	22.1	41.8
S-SA	25.9	106.4	25	52.4

Note: Mean Absolute Percentage Error (MAPE) is a popular accuracy measure used in forecasting which represents forecast error as a percentage of the actual outcome

also important to note that with both treatments the observed errors are very high – on average more than 40 percent and in the second case almost 100 percent; this is indicative of the difficulty of the nature of the forecasting task that the semi-experts were facing. Interestingly, similar to the results found in Green and Armstrong (2007), for the participants using S-SA, those who were able to recall more analogies and with greater detail (albeit without always giving an outcome) were more accurate.

So what went wrong?

The results indicated that S-SA forecasts are no better than UJ in the hands of semi-experts. This contrasts the results found in Savio and Nikolopoulos (2009b) that showed S-SA to be better than UJ in the hands of non-experts (although still not statistically significant). Now, in light of such results, an explanation must be found.

Firstly, it was noted that during the interview process, many of the participants were uneasy producing forecasts under the experiment conditions. Several complained at the brevity of the description and asked to be given additional information. In response to this, it was explained to them that the descriptions were written in a way as to disguise the real-life policy they stemmed from but to give enough information on the policy implementation strategy characteristics to make an effectiveness prediction. This seemed to calm nerves but failed to satisfy the participants completely. It is believed this problem arose because of the level of knowledge the participants had on cases like the ones presented. This level was high enough that most viewed the exercise as a game which they had to solve to produce the "correct" forecast in a rational process but not as high for them to be able to recall multiple analogies with outcomes. This then may have led to an over-analysis of the situation that ended in confusion and this could have affected forecasts in a negative way.

Secondly, the participants, in general, struggled to recall analogies and the majority of the ones recalled were not considered too relevant nor did the participant know their outcome (quantitatively). Incidentally, for the participants using S-SA, it was those same that recalled

309

more analogies who were the more relaxed when performing the task.

We can attempt to elucidate these results by means of four explanatory constructs as illustrated in Figure 1.

1. Information. It is possible that not enough information regarding the policy and the policy implementation strategy were provided in the case description. This is unlikely however, because as already mentioned, the description was written, it is argued, with enough detail and information for a participant to use for producing a forecast. What is more, any

more information regarding the policy and policy implementation strategy would have increased the risk of the case being identified.

2. Complexity. It could be the case that due to the complex nature of policies and policy implementation strategies, it may be too difficult for participants to recall analogies and their outcomes. Unlike a conflict situation (Green and Armstrong, 2007), whose outcome is generally binary, policy implementation strategy effectiveness (which can be influenced by a series of inter-related factors) is measured on a continuous scale, so outcome is more difficult to recall.

3. Worldview. Very interestingly, the participants of different nationalities tended to react differently to the exercise. Those from Latin European nations were more at ease with the exercise than those from Nordic nations. This is reflected in the results with the Latin countries performing better than the Nordic ones in terms of forecast accuracy. So it could be that a participants worldview and cultural background affect the approach as well as attitude taken towards the exercise and this can consequently affect recall of analogies.

4. Expertise. Finally, it could well be that these participants were simply not "expert enough". Although they worked as policy developers in an Environment directorate of a large European governing body, perhaps they lacked the necessary "analogy database" needed for the S-SA approach to work as intended.

The solution

The next step is to investigate further and to test the structured analogies (SA) approach with participants deemed to be real experts; these should be primarily sourced form from academia and industry. Intuition suggests that with real experts, a greater database of analogies will be available for recall and the standard SA approach can be used. In order to do this, we need similarity ratings and quantitative outcomes; this was the essential element of the success of SA in the study of Green and Armstrong (2007). If it results that experts give good analogies but qualitative outcomes then the S-SA would have to be used again. But hopefully this time, the greater expert knowledge "database" will come through and be reflected in the forecasts.

Furthermore, this greater knowledge database is expected to overcome some of the problems experienced in the semi-expert experiment; such as the case description provided will contain enough information for the expert to work with, but not so much that he/she will recognize the case and consequently produce a biased forecast. As a result experts will be more comfortable with the task and the risk of confusion in by the task lower; thus the risk of over-analysis will be significantly reduced.

Conclusions

It must be recognized that the sample of participants used was very small, resulting in statistically insignificant results. Nonetheless, the results provide a useful insight into the performance of the methods and allow for some important conclusions to be drawn, which can be summed up in the following points.

There is no statistical evidence to suggest that unaided judgment or structured analogies perform differently in the hands of "semi-experts" when forecasting the effectiveness of a policy implementation strategy.

Analogies of the target policy implementation strategy and their outcomes are not as easy to recall as in conflict forecasting (but this could be due to inadequate expertise of participants used).

It is difficult to know or define what or who an expert is and quantify his/her level of expertise. In some cases the true level of expertise is only realized after one has finished the forecasting exercise and observed experts' ability to recall information.

Further research should focus on conducting a similar study with more and "true experts" using more cases and alternative forecasting approaches. The level of knowledge of the participants in a future study could well be enough to allow for the recall of several analogies, along with a similarity rating and a quantitative outcome of effectiveness. This would allow the for the full use of the SA method. Otherwise, it could be interesting to test the approach

giving participants categorical outcomes for the target situation and asking them to relate the outcomes of these analogies to these, in the same way Green and Armstrong (2007) did. Another possibility could be to simply provide participants with a scale against which they could rate the net benefit of their analogous policy implementation strategy (e.g. High net cost ($2$10) to High net benefit (b10)). A final alternative could be to ask participants to give enough detail about their analogy so as to allow the administrator to obtain enough objective information regarding its outcome in order to produce a forecast.

Notes

1. For the sake of distinguishing between absolute ''non-experts'' as would be students and ''true experts'' as would be academics or top-managers working in the area.
2. Hence the name ''semi expert'' used throughout, to mark a difference between non-experts and ''true experts''.
3. But with forecasts being produced by the experts and not by the administrator.

References

Brown, T. and Potoski, M. (2006), ''Contracting for management: assessing management capacity under alternative service delivery arrangements'', Journal of Policy Analysis and Management, Vol. 25 No. 2, pp. 323-46.

European Commission Impact Assessment Guidelines (2009), available at: http://ec.europa.eu/ governance/impact/docs/key_docs/iag_2009_en.pdf (accessed 27 February 2009).

Green, K. and Armstrong, J.S. (2007), ''Structured analogies for forecasting'', International Journal of Forecasting, Vol. 23 No. 3, pp. 365-76.

Haas, P.J. (2008), ''Introduction to the symposium performance measurement: can it improve performance?'', International Journal of Public Administration, Vol. 31, pp. 593-4.

Harvey, N. and Fischer, I. (1997), ''Taking advice: accepting help, improving judgment and sharing responsibility'', Organizational Behaviour and Human Decision Processes, Vol. 70, pp. 117-33.

Heath, C. and Gonzalez, R. (1995), ''Interaction with other increases decision confidence but not decision quality: evidence against information collection views of interactive decision making'', Organizational Behaviour and Human Decision Processes, Vol. 61, pp. 305-26.

Lee, W.Y., Goodwin, P., Fildes, R., Nikolopoulos, K. and Lawrence, M. (2007), ''Providing support for the use of analogies in demand forecasting tasks'', International Journal of Forecasting, Vol. 23 No. 3, pp. 377-90.

Peters, B.G. and Savoie, D.J. (2000), Governance in the 21st Century: Revitalizing the Public Service, McGill-Queen's Press, Montreal.

Savio, N. and Nikolopoulos, K. (2009a), ''Forecasting the economic impact of new policies'', Foresight, Vol. 11 No. 2, pp. 7-18.

Savio, N. and Nikolopoulos, K. (2009b), ''Forecasting the effectiveness of policy implementation strategies'', International Journal of Public Administration (in press).

Shanteau, J., Weiss, D.J., Thomas, R.P. and Pounds, J.C. (2002), ''Performance-based assessment of expertise: how to decide if someone is an expert or not'', European Journal of Operational Research, Vol. 136, pp. 253-63.

Sniezek, J.A., Schrah, G.E. and Dalal, R.S. (2004), ''Improving judgment with prepaid expert advice'', Journal of Behavioural Decision Making, Vol. 17, pp. 173-90.

Talbot, C. (2008), ''Performance regimes: the institutional context of performance policies'', International Journal of Public Administration, Vol. 31, pp. 1569-91.

Talbot, C. (2009), ''Public value: the next 'big thing' in public management?'', International Journal of Public Administration, Vol. 32, pp. 167-70.

Appendix 1

Appendix 2

Corresponding author

Nicolas D. Savio can be contacted at: Nicolas.Savio@postgrad.mbs.ac.uk

(i) Description (ii) Source
(iii) Similarities and
differences (B) Similarity rating (C) Outcome
a. Gov. campaign in Hungary to
promote home use of
computers
Experience Possibility to buy a computer
at a great discount
9 Successful 80 percent
b. Gov. campaign in Hungary to
subsidize improvement of
household efficiency
Experience – 9 Successful 90 percent

311

c. Gov. campaign in Hungary
aimed at helping graduates
find employment
Experience Employer receives fringe
benefits from the state for
employing graduates
7.5 –
d.
e.
e.g. New technology promoted by
Gov. to combat climate
change
Media Similarities: new technology
introduced for improvement
purposes Differences:
Different context, not
necessarily for directly
reducing CO_2
8 Reduced waste (any
kind) by 24 percent