Making Essential Choices with Scant Information

Front-End Decision Making in Major Projects

Edited by Terry Williams, Kjell Sunnevag and Knut Samset

Making Essential Choices with Scant Information

This page intentionally left blank

Making Essential Choices with Scant Information

Front-End Decision Making in Major Projects

Edited by

Terry M. Williams
Professor, University of Southampton, UK

Knut Samset
Professor, Norwegian University of Science, Norway

and

Kjell J. Sunnevåg
Head of Research, Norwegian Competition Authority, Norway

palgrave
macmillan

Softcover reprint of the hardcover 1st edition 2009 978-0-230-20586-4

First published 2009 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN 978-1-349-30224-6 ISBN 978-0-230-23683-7 (eBook)
DOI 10.1057/9780230236837

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

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

A catalog record for this book is available from the Library of Congress.

10 9 8 7 6 5 4 3 2 1
18 17 16 15 14 13 12 11 10 09

Contents

List of Figures

List of Tables

Acknowledgements

This book is a result of research funded by the Concept Research Programme on front-end management of public investment projects. More information on the Programme is given on www.concept.ntnu.no.

Contributors

Bjørn Andersen is Professor at the Norwegian University of Science and Technology and Research Director of a research department at SINTEF Technology & Society: Productivity and Project Management. His research and teaching experience is within the area of project management, risk and quality management, project organization, benchmarking of projects, etc. Bjørn Andersen was part of the international research team in the IMEC research program that benchmarked 60 projects worldwide. He has authored or co-authored several papers and books on these topics.

Tim Bedford is Professor of Risk and Decision Analysis at the University of Strathclyde, Glasgow. He previously worked at King's College Cambridge and Delft University of Technology. His current research interests are in probabilistic risk analysis, uncertainty modelling and decision analysis. He is currently a member of the Board of Directors of ESReDA, a Fellow of the UK Safety and Reliability Society and a member of the EPSRC Mathematics Programme Strategic Advisory Panel. He was Chair of Mathematical Methods in Reliability Conference, 2007. He co-wrote "Probabilistic Risk Analysis: Foundations and Methods". His industry collaboration includes work for Aveco, ESTEC, MoD, DSTL, Railway Safety, Magnox, Netherlands Ministry for Water Management, Philips and NASA.

Roger M. Cooke is a Chauncey Starr Senior Fellow at the Resources for the Future (RFF). He retains a part-time position as Professor of Applied Decision Theory at the Department of Mathematics at Delft University of Technology in the Netherlands, where he set up the master's programme in Risk and Environmental Modelling. His research has widely influenced risk assessment methodology, particularly in the areas of expert judgement and uncertainty analysis. He is recognized as one of the world's leading authorities on mathematical modelling of risk and uncertainty.

Terry Cooke-Davies is Executive Chairman of Human Systems International Limited, a Visiting Professor at Bond University, Queensland and a Visiting Fellow at Cranfield University, UK. He holds a PhD in project management, and his research interests include the project management practices of large national and multinational organizations, organizational project management maturity and the relevance of complexity theory to project management. He is co-author with Paul Dinsmore of "The Right Projects, Done Right" and has authored many journal papers, book chapters and magazine articles.

Bent Flyvbjerg is Professor of Planning at Aalborg University, Denmark and part-time Chair of Infrastructure Policy and Planning at Delft University of Technology, The Netherlands. He is founder and director of Aalborg University's research programme on large-scale infrastructure planning. His books include *Decision-Making On Mega-Projects* (with Hugo Priemus and Bert van Wee, eds.), *Megaprojects and Risk* (with Nils Bruzelius and Werner Rothengatter), *Making Social Science Matter* and *Rationality and Power*.

Kåre P. Hagen is Professor in Economics at the Norwegian School of Economics and Business Administration (NHH) since 1979. In the period 1991–2000, he was serving as research director at the Center for Economic Policy Studies at the Institute for Research in Economics and Business Administration (SNF), Bergen. His main research areas are public economics, cost-benefit analysis and network economics. He has had several stays as visiting researcher at universities abroad including the Centre for Operation Research and Econometrics (CORE) at The Catolique University, Louvain, Cowles foundation for research in economics at Yale University, Nuffield College at Oxford University.

Kees van der Heijden Professor Cornelius (Kees) A.J.M. van der Heijden is an associate Fellow at Said Business School, University of Oxford, and Emeritus Professor at the University of Strathclyde, Glasgow, where he taught General and Strategic Management until 2004. Prior to 1990, he headed the business environment division of Royal Dutch/Shell, London where he oversaw scenario planning, including monitoring and analyzing the business environment and advising top management on strategic implications. Previously, he headed Shell's internal strategic consultancy group. He is a Global Business Network cofounder. He is the author of *Scenarios: The Art of Strategic Conversation* and many other publications.

Bettina von Helversen is a research scientist at the Center for Adaptive Behavior and Cognition, Max Planck institute for Human development. The Center investigates reasoning and decision making under uncertainty at the levels of both individuals and social groups. Von Helversens research interests has been inter al categorization and estimation strategies, ontogenetic development of heuristic strategies, adaptiveness of basic cognitive capacities to environmental structures, as well as legal decision making.

Brian Hobbs holds the Project Management Research Chair at the University of Quebec at Montreal (UQAM). He has been a professor in the Master's Program in Project Management for more than twenty years. This program, of which he is a past director, is accredited by PMI's Global Accreditation Center. He served a three-year term on PMI's Standards Members Advisory Group (MAG) ending in 2002 and joined the Research MAG in July 2006. He is a reviewer for both the Project Management Journal and the International

Journal of Project Management. He has presented many papers at both research and professional conferences worldwide.

Geir Kirkebøen is Professor in the Psychology Department at the University of Oslo. He has an interdisciplinary background, educated both as a computer scientist and a psychologist. He holds a doctorate in Artificial Intelligence (Expert Systems) and had a postdoctoral fellowship on a history of science topic. Currently he is responsible for a program where the students combine computer science and psychology of judgment and decision making. His research areas are judgement and decision making and history of science (Descartes' psychology). He has had several stays as visiting scholar at universities abroad, including Sydney University (Australia) and Washington University (St. Louis, USA).

Roger Miller is the Jarislowsky Professor of Innovation and Project Management at École Polytechnique in Montréal and a founding partner of SECOR, a strategy consulting firm with offices in Montréal, Toronto and Paris. He has been a fellow at the Center for International Affairs of Harvard University and in MIT's International Motor Vehicle Program. His work has focused on (1) strategy and industry dynamics (2) public policies in science and technology and (3) large engineering project management. Roger Miller is a Fellow of the Canadian Academy of Engineering.

Peter W.G. Morris is Professor of Construction and Project Management at The Bartlett University College, London. He is also Executive Director in INDECO, a consultancy in project management and has conducted research with firms such as AstraZeneca, BG, BP, Camelot, FKI, Fujistu, GSK, Land Securities, Pfizer, Shell, Skanska, Spirent, the UK Government and Unilever on a wide range of issues, including strategy, design management, knowledge management, and procurement. Professor Morris is a Fellow of the Institution of Civil Engineers, a Fellow of the Chartered Institute of Building and a Fellow of the Association for Project Management. He was previously Chairman of the Association for Project Management.

Petter Næss is Professor of Urban Planning at Aalborg University, Denmark. Næss has for many years carried out research on sustainable urban development with a prime focus on the influence of urban structure on travel behaviour. He has published about 50 research reports and more than 20 refereed scientific articles in leading international journals within these fields, and is the author of books published in Norway (1997), Denmark (2005) and the United Kingdom (2006) on the influence of urban structural conditions on travel behaviour and energy use.

Greg Parnell is Professor of Systems Engineering at the United States Military Academy at West Point. He is also senior Principal with Innovative Decisions Inc. He has extensive decision analysis, operations research and systems engineering teaching experience at United States Military Academy, Virginia Commonwealth University and Air Force Institute of Technology (AFIT). He recently chaired a National Research Council Study on Biterrorism, is a member of several advisory panels of the National Security Agency Advisory Board and was a former member of Department of Energy's Environmental Management's National Prioritization Team. He has broad leadership and management experience as a Research and Development (R&D) manager and systems engineer, as well as being an experienced senior decision analyst. His research has focused on decision analysis, risk analysis, resource allocation and systems engineering for defence, intelligence, homeland security, R&D and environmental applications.

Philip Pugh practices as an independent consultant, in which role he has been the originator of cost forecasting models covering a wide range of defence equipment and now established in use within government and industry in the UK, Europe and Australia. Prior to this, Philip Pugh was employed within the (UK) Ministry of Defence where he was successively responsible for the collation of cost forecasts for central committees, funding of engine technology demonstrator work and forecasting costs of major weapons and electronic projects. His most recent publication is "The Source Book of Defence Equipment Costs" (2007).

Knut Samset is Professor of Project Management at the Norwegian University of Science and Technology in Trondheim (Norway). He is founder and partner of the consultancy company Scanteam and director of the Concept Research Program on Front-end Management of Major Investment Projects. His current research is on project governance, appraisal and quality assurance of major investments. He has published a large number of research reports and papers and is the author of books on technology assessment, project design, evaluation and front-end management of projects.

Benjamin Scheibehenne is a researcher at the Cognitive Science Program, Indiana University, Bloomington. He received his PhD at the Center for Adaptive Behavior and Cognition, Max Planck Insitute for Human Development. His current interests are cognitive models of choice and the prediction of individual preferences with a special focus on food choice and consumer choice.

Kjell J. Sunnevåg now at the Norwegian Competition Authority, was previously a research manager at the Institute for Research in Economics and Business Administration (SNF) in Bergen, Norway. His research interests have been public economics, environmental economics, economics of scarce resources, as well as investment analysis and the use of real options for economic evaluation in practical decision making. In 2006, he edited a book on various aspects of up-front decision making based on scant information, commissioned by the Concept Research Program.

Terry M. Williams is Professor of Management Science, Director of the Management School at the University of Southampton, UK. He previously worked in Operational Research (OR) at Engineering Consultants YARD developing Project Risk Management and acting as Risk Manager for major projects, then at Strathclyde University, UK as Professor of OR and Head of Department. He continues research and consultancy modelling the behaviour of major projects, both post-project review and pre-project risk, including work in major claims in Europe and North America. Terry is a speaker on modelling in project management and has written 60 journal articles and books on modelling complex projects and post-project reviews. He is a member of a number of research networks worldwide and sits on PMI's Research Members Advisory Group. He edits a learned Journal and is a PMP.

Mark Winter is a member of the Centre for Research in the Management of Projects at Manchester Business School. A strong exponent of the need to bring academia and industry closer together in project management; he recently led the UK government-funded research programme, *Rethinking Project Management*. He is also a major contributor to industry-based programmes including the Project Management Professional Development Programme run by the University of Manchester for a consortium of companies led by Rolls-Royce. He has worked with many organizations across different sectors and is particularly interested in the front-end of projects and the application of soft systems methodology.

George Wright is Professor of Management at Durham Business School, UK.

Fergus Bolger is an associate Professor of Psychology at Bilkent University, Turkey.

Gene Rowe is a senior research scientist at the Institute for Food Research, UK. Together, the authors have published in journals such as Organizational Behavior and Human Decision Processes, Risk Analysis, and Technological Forecasting and Social Change.

Part I

Introduction

1
Decisions Made on Scant Information: Overview

Terry Williams

This chapter brings an overview to an edited book that looks at how decisions can be made at the front-end of major projects, in circumstances where information is usually scant. The book examines how projects can be successfully aligned with the desired direction; how sufficient, appropriate and valid information can be gathered at the front-end; how information can be analysed; and finally how decisions can be made. Each chapter of the book is written by an expert in the field, and each chapter speaks for itself. However, some key themes run throughout the book. These include the need for alignment between organisational strategy and the project concept; dealing with complexity, in particular the systemicity and interrelatedness within project decisions; consideration of the ambiguity implicit in all major projects; taking into account psychological and political biases within estimation of benefits and costs; consideration of the social geography and politics within decision-making groups; and preparation for the turbulence within the project environment, including the maintenance of strategic alignment.

Projects

Business is becoming increasingly projectised or project oriented, and global spend on projects is now many billions of dollars annually. Developing new technological products, building new capital assets, or carrying out unique large-scale enterprises all require major projects to be undertaken. The place of projects has become increasingly important in the life of corporations and, indeed, of nations.

Projects have always been important in human development. This is true both for projects with a tangible output (e.g. the Pyramids or the Great Wall of China) and for projects which bring about a change in the organisation of society (e.g. the explorations of Columbus and his "discovery" of America).

While society continuously tries to incrementally improve the way it operates, throughout history projects have formed the major stepping stones for step-changes. This has been even more the case in recent decades, and, indeed, projects are becoming more important to industrial life. Even 15 years ago, the preface to Turner (1993) extrapolates from statements by British Telecom to suggest that the annual spend on projects in the UK would be around £250 million.

In recent decades, projects have become more complex, in parallel with the growth in complexity of technology. They have also become more time-constrained (Williams, 1999). Moreover, there has been an increase in extremely large projects. Kharbanda and Pinto (1996), for example, list over 40 projects in process in the mid-1990s in India, China and South-East Asia alone, each one forecast to cost over $1 billion. These are mainly construction projects, but engineering projects are also becoming larger in some industries, as the investment needed to develop new products increases (the break-even point of an aircraft development programme is generally held to be at least 300 units, e.g.).

So clearly, projects are becoming important, one well-known definition of "project" being "a temporary endeavour undertaken to create a unique product or service" (Project Management Institute, 2000). During the past 50 years, tools and techniques have been developed to manage projects better. The field generally known as "Project Management" became well defined and developed in the 1950s, particularly in the US Navy Atlas and Polaris programmes. As methods were formulated and codified, there arose a Project Management "profession", represented by the US-based Project Management Institute, or PMI, with about a quarter of a million members (as well as a number of much smaller national organisations); there are university degrees in the subject, an ANSI standard, and professional qualifications such as the PMI's PMP qualification. There are over 500 books on project management topics, and in the past 40 years, over 3500 research articles have been published in the English language within the project management field (Pinto et al., 2003).

"Projects", then, *ought* to be recognised, and considered an important and successful part of corporate and public life. But ask the archetypical "man in the street" about projects, and it is clear that the reputation of projects and project management is that they are generally unsuccessful. A key word often associated with them in the public's mind is the English colloquialism "white elephant" (something whose cost and subsequent upkeep is much greater to the owner than its value, deriving from the reputed practice of monarchs giving sacred white elephants as gifts). A newspaper article from 2005, for example, begins, "The Millennium Dome, the great white elephant languishing in east London. . . ." (Wray, 2005). Even if the result is seen to be useful, there are often reports in the media of large public construction projects that have suffered huge cost or time overruns, such as Denver's $5 billion airport 200% overspend (Szyliowicz and Goetz, 1995), or the UK's Scottish Parliament coming in "10 times over budget and more than three years late" (Tempest, 2004). Pinto (2007) quotes from an *Infoworld* article describing, "a US Army study of IT projects [that] found that 47% were delivered to the customer but not used; 29% were paid for but not delivered; 19% were abandoned or reworked; 3% were used with minor changes; and only 2% were used as delivered" (p. 7).

Private industry's problems and overspends are not usually publicised as much, although the famous database of public and private projects by Morris and Hough (1987) concluded that "the track record of projects is fundamentally poor, particularly for the larger and more difficult ones. . . . Projects are often completed late or over budget, do not perform in the way expected, involve severe strain on participating institutions or are cancelled prior to their completion after the expenditure of considerable sums of money." (p. 7). And while all this may be in the past, our current record seems no better. One of the biggest current projects in the UK is the "National Programme for IT in the National Health Service" (expected expenditure to be over £12 billion during 2004–2014). A recent Parliamentary report concluded, "Four years after the start of the Programme, there is still much uncertainty about the costs of the Programme for the local NHS and the value of the benefits it should achieve . . . The Department is unlikely to complete the Programme anywhere near its original schedule" (House of Commons, 2007).

This is not to say that all projects are managed badly – indeed, the management and governance of many projects has made considerable improvements in recent years. Although the value of project management is notoriously difficult to quantify (Thomas and Mullaly, 2007), one example is given in a UK Government report on the performance of its Office of Government Commerce (OGC) in 2007. This states that "the OGC has achieved some notable successes since it was set up in 2000. . . . The OGC has also established Gateway reviews as a means to help departments improve their record in project delivery. Over 1,500 Gateway reviews have been completed since their introduction in 2001 on more than 700 separate projects and programmes in central government, resulting in over £2.5 billion value for money savings. . . ." (HM Treasury, 2007).

However, there is no doubt that "projects" have generated the reputation in the public perception for overruning and overspending. So what explains this strange dissonance? Why have the study and improvements in Project Management resulted in such a dire reputation for projects? It is the contention of this book that the concentration on project management has been much too narrow. There is more expertise now in delivering efficiently and successfully a well-defined pre-specified project within a well-defined constant environment. This has proved very valuable in certain circumstances. But a much wider view needs to be taken. The initial choice of project concept is of critical importance. This represents the one key decision of many made during the lifetime of a project, which is likely to have the largest impact on long-term success or failure. Here, as discussed by Miller and Hobbs' Chapter 18, the project concept is much more than just the technical solution – it includes the entire business case, all the various organisations involved, and the various mechanisms and arrangements involved in the inter-organisational relationships.

This has been known for a long time – in the UK public arena, it was brought to general attention by the influential "Downey" report 40 years ago, which laid down the policy that early "project definition" should take up 15% of the cost and 25% of the time of the project (Ministry of Technology, 1969). But development here has been very much slower than the development of tactics for the execution phase. There have been some developments – for example, in the public arena in Norway, this shift in emphasis has led to the "Concept" programme (Concept, 2007), and the governance framework for public projects established in the UK has an emphasis on the "Gateway 0" analysis (Office of Government Commerce, 2007). These developments have recognised a key issue for major projects: that they have to be established within a turbulent environment, where the idea of specifying a well-defined project goal, which remains constant, is often not applicable. Front-end planning needs to recognise and plan for this turbulence (e.g. by ensuring that Gateway 0s continue intermittently throughout the project). However, there remain considerable challenges during this period.

There is one particular problem that characterises decisions at this point, and that shapes this book – indeed, has prompted recognition of the need for this book. The major decisions are made at this early stage, when epistemic uncertainty is at its highest, but when available information is most restricted. So how can good decisions be made? What information is needed? How valid is that information? These are some of the areas explored in the book. This chapter refers to sections which discuss whether less detailed information can actually be an advantage to decision-makers, by providing focus and flexibility, and by removing distracting and unhelpful details.

This book, then, is about that key stage – making decisions at the start of the project. It looks at both the public and private sectors, and many types of projects. This chapter introduces the book and outlines some of the arguments that will appear in the succeeding chapters.

Strategy drives the project

Industry is starting to realise the importance of projects in executing its strategy.

> Let's face it – we all function in a highly competitive global economy where yesterday's laggard is tomorrow's market leader. Globally, boards of directors and senior business managers are looking for ways to compete

> more effectively in this highly volatile environment. Some organisations strategise well, but their execution leaves something to be desired. Others execute well, but haven't developed a strategic plan to drive the execution. I'm here to argue that project management is the very tool that bridges strategic planning and execution, resulting in better bottom line results on a far more predictable basis.
>
> (Fraser, 2005)

A project has to start with the corporate – or public sector – strategy. Projects and programmes are mechanisms for bringing about changes, in particular large one-off capital expenditures, so clearly there is a need to look at how strategy drives the definition of projects. It is important to look beyond the simple success criteria on which project management has traditionally concentrated (delivering the planned output within cost and schedule) to the value that a project can give. Morris (Chapter 3) quotes the famous IMEC study by Miller and Lessard (2001), who distinguish between efficiency and effectiveness of project success, the latter pointing to the value generated by the project. He points out that the projects in this study were much more efficient than they were effective. Samset takes this distinction further in Chapter 2, quoting a fivefold success criterion widely used in international development projects – efficiency, effectiveness, relevance, impact and sustainability. Although assessing the social profitability of a given project in the public sector is not simple, Hagen (Chapter 19) looks at an economic comparison of what the project is likely to generate in terms of services, compared to the costs of inputs, including market prices and social opportunity costs.

It is therefore essential to identify explicitly the strategy of the organisation and ensure that the goals or objectives of any project will "further the sponsoring organization's chosen corporate strategy and contribute to its overall goals". This is the recommendation of Cooke-Davies in Chapter 6, which looks at the front-end alignment of projects. This chapter, together with Chapter 3, by Morris, considers the strategy of the organisation, and the importance of developing projects to pursue that strategy, with emphasis on the value the project produces for the organisation, rather than simple efficiency of execution. Of course, to do this, the organisation's needs must be made explicit. Chapter 5, by Naess, examines the relationship between needs analysis, goal formulation and impact assessment, and includes some methods for needs analysis. This chapter also highlights some problems with current practices which will occur as themes within the book, both qualitative, particularly a lack of a view of the systemicity within the analysis, and biases in the quantitative analysis, which will be discussed below, when estimating costs and benefits.

Clearly, getting this alignment right is critical to the value of a project. In Chapter 2, Samset points out the seriousness of "when a project fails in strategic terms, even if it successfully produces intended outputs. Strategic failure means that the choice of concept turns out to be the wrong one." Project management has been developing in this area for some time. Turner (1993), for example, described the importance of the alignment of business strategy and portfolio, or programme, objectives. Morris points out in Chapter 3 that the UK Association of Project Managers' "Body of Knowledge" now "emphasises context, strategic imperative, commercial drivers, technology, the traditional control functions, and, not least, people. Hence, the whole of Section 1 is concerned with how projects fit within their business/sponsor's context." Indeed, the Gower Handbook of Project Management (3rd edition) has "Implementing Strategy through programmes of projects" as its second chapter (Partington, 2000) (the forthcoming 4th edition has a similar chapter).

Note that we have used various words – "projects", "programmes" and "portfolios", and we will not debate the meanings of these words here – Morris's Chapter 3 gives some of this debate, including the key distinction by the OGC, following their view that a key benefit of programme management is the alignment of projects to organisational strategy (Office of Government Commerce, 1999). The track record of projects discussed above shows that there is still some way to go, but at least these chapters lead us to concentrate on the critical issues.

The relationship between strategy and project management is not one-way. Morris describes how strategy implementation is accomplished with project management, but project management can also contribute to strategic management. He points out that project management's contribution "can add value to the emerging strategy and ensure that benefits are reaped from its realisation." The importance of project management in producing value for organisations is demonstrated below, where governance within a turbulent environment is discussed.

Goals

Having said that there is a need to identify corporate goals and objectives, and to align projects with these, is not to gainsay the difficulty of this in practice. Roth and Senge (1996) describe management decision-making in the real, complex world, classifying problems by firstly the underlying complexity of the problem situation itself, i.e. "dynamic complexity", and secondly the complexity of the group effect, i.e. "behavioural complexity". The dynamic complexity in the underlying problem is overlaid – and sometimes dwarfed – by issues of different stakeholders having different perceptions of "reality", different understandings of "the problem", different assumptions, different values, different objectives, etc. Problems that are complex in both dimensions, they call "wicked messes".

Projects in a typical management environment, public or private, can often be said to be in such a "wicked mess". "Projects are complex, ambiguous, confusing phenomena wherein the idea of a single, clear goal is at odds with the reality" (Linehan and Kavanagh, 2004). Engwall (2002), indeed, describes the establishment of the perfectly correct goal as a "futile dream".

Winter, in Chapter 7, quotes Morris, stating, "at the front end . . . we often have quite messy, poorly structured situations where objectives are not clear, where different constituencies have conflicting aims". His chapter goes on to provide a well-established methodology, known as "Soft Systems Methodology", which was developed by Peter Checkland, for gaining understanding about such situations and using this to help orient the front-end of projects. This methodology recognises the subjectivity implicit in all human situations, including projects, and is therefore able to develop learning and understanding at the start of a project. This subjectivity is also key to the ongoing execution of projects, as participants "make sense" (Weick, 1995) of the project and work towards project delivery.

Considerations of goals are affected by the social geography of an organisation, and so the behavioural-complexity aspect of the "wicked mess" must be considered. Groups, by their very existence, condition decision-making, and where strong structures or strong power gradients exist between members of the organisation, decision-making might seem to become less "rational". Perhaps the best known effect is "Groupthink" (Janis, 1973), where the individuals within a group conform in their thinking with what they think is the group consensus. Equally well known is Habermas's (1984) theory of communicative rationality: where communication is dominated by discourse unfettered by the coercive use of power, there will be good exchange of rationality, but where power is being used to limit free communication, there might be failures in decision-making. In assessing how judgements are made about the future, consideration must also be given to the many aspects of the group of decision-makers within an organisation: the different levels of power, interest, credibility, difference between expressions and perceptions, the various aspects of social geography, etc. Miller and Hobbs (Chapter 18) expand on this, stating that the project concept should meet the needs of many stakeholders, both those within the project organisation and those in the wider environment. Furthermore, the development of the concept is related to the social process of building the project consortium.

Judgements about the future

Deciding the organisational strategy is intimately related to making judgements about the medium and long-term future. Firstly, the organisation has to consider how its environment might change in the future. Kharbanda and Pinto (1996) give a well-known example of a Mitsui project to build a plant in Iran, for converting natural gas into petrochemical products. The contracts were negotiated with the government (then under the Shah); the Revolution occurred during project development; then war broke out between Iran and Iraq; and so the story continues for nearly a decade after this point, before finally Mitsui abandoned the project, with a loss of £2–3 billion. A fairly extreme case maybe, but in planning out any project, the organisation has to consider possible future events and how it should react and interact with these possible events. Van der Heijden (Chapter 4) discusses the need for making such judgements, and the issues involved, together with practical and well-tried methods for their development.

A number of issues are involved in making such judgements, some of which occur as themes throughout this book.

Firstly, the people involved are not supremely rational decision-makers. Real managers are human beings, which means that, at best, they display "bounded rationality" (Simon, 1972). They are limited in the extent to which they can make a fully rational decision. Not only are they lacking complete information about the present, and have uncertainty about the future, but they are also limited in the extent to which they can solve complex problems. Indeed, as Miller and Hobbs point out in Chapter 18, the assumptions underlying rational decision-making frameworks are often simply not valid in the circumstances of a real project. Thus such decision-makers adopt choices that are merely "good enough" or "satisficing" (Isenberg, 1991). Moreover, rather than evaluating projects at a one-off point at the beginning, with full information about costs and benefits, many project sponsors look at projects from "evolutionary perspectives" (Miller and Hobbs, 2005). Here sponsors are seen to act as champions, "shaping projects in response to" changes in the environment.

A further reason for bounded rationality, which particularly affects projects, is the existence of cognitive biases natural to humans. How these biases can be involved in estimating the costs and plans of projects is discussed in the next section; they also affect views of the future and the benefits likely to be gained from a project. Flyvbjerg quotes some of his extensive evidence of project benefit overestimation and cost underestimation in Chapter 8. He describes the biases involved, dividing them into technical (e.g. due to inadequate forecasting techniques or honest mistakes), psychological (e.g. "optimism bias") and political-economic explanations (i.e. reasons to deliberately claim an optimistic view of the future). This work is also referred to by Naess in Chapter 5. Much academic evidence for the middle category comes from the famous work of Kahneman and Tversky. Kirkebøen (Chapter 9) also looks at these "planning fallacy" biases, quoting the Sydney Opera House example. Pugh, in Chapter 16, gives more supporting evidence for such effects. Figure 1 of Chapter 16 provides an additional explanation of why forecasts may be "excessively optimistic". Large projects can be of very long duration, involving judgements far into the future. For the public sector, this raises issues about the role of the discount rate and the required social rate of return, which is covered by Hagen in Chapter 19.

A third aspect of the difficulty faced by boundedly rational decision-makers in reaching judgements about the future is the systemicity and interconnectedness involved in the various aspects (the dynamic-complexity aspect of the "wicked mess"). Parnell discusses the complexity of project planning in Chapter 12 and looks at multiply-related uncertainties about the future. Van der Heijden's methods (Chapter 4) explicitly address this, "actively search[ing] for predetermined elements in the causal systemic network in which the project is embedded". Naess (Chapter 5) also cites too narrow a needs assessment as a significant problem.

Finally, judgements about the future are, again, made within the social geography of the group or coalition and so are subject to the same effects of behavioural complexity discussed above, under "Goals". Again, the methods outlined by van der Heijden in Chapter 4 take full account of the multiple perspectives and worldviews of the decision-makers. The methodology covered by Winter in Chapter 7 concurs with this.

Of course, judgements about the future are not one-off events at the start of the project but continue as the project makes its way through the actualisation of that future. This will be discussed in the section on the turbulent environment of the project.

Estimating the project

One key element of judgement about the project is the estimation of its cost and schedule – a fundamental part of project management, but one which seems to pose unique difficulties for major projects. Flyvbjerg (Chapter 8) examines the various reasons behind the "pervasive misinformation" which persistently trouble project estimation – not only technical explanations, such as inadequate data or lack of experience, but the main headings of optimism bias and strategic underestimation of costs.

This is not to say that human estimators are not to be used. While the evidence of cognitive and political biases is well established, there is also much evidence of the skill of experts in using innate tacit knowledge to estimate projects. Chapter 10, by Scheibehenne and von Helversen, outlines some well-known "decision heuristics" and concludes that, under the conditions of uncertainty about the future and systemicity encountered during front-end project decision-making, "heuristics provide a feasible way to make decisions. Contrary to the common view of heuristics as second-best solutions . . . the research program . . . has provided substantial evidence that heuristics often achieve an astonishingly high performance using just a fraction of the time and the amount of information required by standard decision strategies." Indeed, it will be demonstrated that the use of only relevant information can make the heuristics more robust.

One key problem, ingrained in the lore of project managers, is the uniqueness of their projects. Project managers are used to a definition of projects such as that of Buchanan and Boddy (1992): "A project is a unique venture with a beginning and an end, conducted by people to meet established goals with parameters of cost, schedule and quality." They tend to think of each project as unique – which it may well be in many aspects. However, this view of the project can nearly always be balanced with evidence from past projects. Kirkebøens takes up Kahneman's idea, in Chapter 9, that this means taking an "outside" instead of an "inside" view on the project. Flyvbjerg (Chapter 8), when considering what planners can do about the biases he has identified, concentrates on this idea of the "outsider" view (relating the same well-known anecdote about Kahneman's work) to produce the idea of Reference Class Forecasting. This technique is now well established in a number of countries. Parametric forecasting is discussed by Pugh in Chapter 16 and is based on his experience of applying this in the public sector. This is another method for trying to avoid the over-reliance on project uniqueness, as it tries to bring the evidence of past projects to bear on the project-estimation problem. It, too, is well established in a number of countries, particularly in the defence sectors.

Estimation of uncertainties and the likelihood of risks are crucial to the estimation in any project. Wright (Chapter 11) (and, briefly, Bedford in Chapter 14) discusses the difficulties that humans – including so-called experts – encounter in making such assessment and how estimation can be improved. This difficulty is particularly acute in projects, where much uncertainty is epistemic (due to a lack of knowledge), rather than aleatoric (due solely to probabilistic uncertainty). Bedford (Chapter 14) divides the areas of uncertainty further into lack of understanding about: the major uncertainties and their interactions; the degree of project uniqueness; and the way in which future decisions will affect outcomes. He describes some probabilistic models for exploring the first two and aiding the third. The first of these – looking at interactions of risks, or risks under conditions of systemicity – is a significant problem in risk analysis. Parnell (Chapter 12) explores further the assessment of multiple-related risks.

In practice, estimates of uncertainty are made by groups, with all the same issues of groupthink, consensus, politics, etc., as outlined above. Cooke takes the discussion further in Chapter 13, by looking into these issues and derives methods for performance-based expert judgement models. Finally, estimates, once made, need to be updated as more evidence becomes available. While this may lie beyond the strict boundaries of this book, an element of updating is still required, even in the front-end stage of a project. Bedford, Pugh and Sunneveg all refer to Bayesian methods of updating (in Chapters 14, 16 and 17, respectively).

Governance in a turbulent environment

The need to align projects with the strategy of the organisation has already been discussed. It is the role of project governance to ensure that "effective governance of project management ensures that an organisation's project portfolio is aligned to the organisation's objectives, is delivered efficiently and is sustainable" (Association of Project Management, 2002). Morris emphasises the need for project governance to ensure that projects deliver strategic value (Chapter 3).

However, the environment in which most projects operate is complex and turbulent, and conventional project management is not well suited to such conditions. The irony of this unsuitability is pointed out by Malgrati and Damiani (2002), who contrast "one of the main reasons for the spread of project management in companies, namely environmental complexity and uncertainty ... and exposure to external change", with the philosophical underpinnings of traditional project management. They conclude that "The Cartesian clarity of inner structures clashes with the increasing porosity of projects to complex contexts that they seek to deny (referring to a paper by Ulri and Ulri from 2000). ... The risk, in short, is that the idealistic 'island of order' may suddenly turn into a more realistic, very classic, 'iron cage'."

For projects to be aligned with organisational strategy – *and stay aligned* – it is important to recognise the turbulence of the environment and build in the capability to cope with this turbulence at the start of the project. As Miller and Hobbs discuss in Chapter 18, this is equally important when the project is being undertaken by a heterogenous consortium or group of organisations, where processes and structures need to be developed to deal with turbulence.

Firstly, then, flexibility needs to be built into the project strategy, both in the front-end concept stage and in later stages. Olsson (2006) shows the need for tactical flexibility within a defined strategy; and Samset (Chapter 2) points out the danger in seeking predictability. He warns that "prediction [can] become a prescription ... it shifts the decision-maker's focus from finding the best solution to ... [making] his own idea or prescription come true". Premature lock-in to an inappropriate concept can be a major danger to project success.

It has already been suggested that projects are not a simple execution of well-developed plans but are often Weickian sense-making activities (Weick, 1995), as the project management team cope with ambiguity, uncertainty and complexity. This sense-making within ambiguity takes place within the turbulent environment, making the project management task that much more complex. Cicmil et al. (2006) contrast "traditional approaches based on rational, objective, and universal representations of 'the project' with a phronetic analysis of the ambiguous, fragmented and political reality of project situations". Front-end decision-making has to develop a project strategy, while recognizing these ambiguities.

The governance framework thus has to recognise these realities of project life and be sufficiently versatile to enable projects to adapt, be flexible and avoid premature lock-in. When there is restrictive "straitjacket" governance, there is a danger of projects gradually becoming unaligned with organisational goals. Miller and Hobbs, again based on the IMEC work, conclude that "Governance regimes for megaprojects are time-dependant and self-organizing. They involve a network of actors in a process through which the project concept, the sponsoring coalition, and the institutional framework co-evolve."

This is not to say that governance framework should be unstructured. There is a clear need for staging. The OGC's "Gateway 0" analysis has already been mentioned as one example. This is designed to be repeated periodically, to ensure that, as an ongoing strategic assessment, during which the need for the programme is confirmed, it is likely to achieve the desired outcomes. Morris discusses the role of staging and gatekeepers in the ongoing governance of a project (Chapter 3).

Scant information

Chapter 10, by Scheibehenne and von Helversen, concludes that "less can be more", and that having less information can actually help the decision-makers. This is also emphasised by Samset in Chapter 2. A restricted, but carefully selected, sample of relevant facts and judgemental information may be an advantage in the effort to establish a broad overall perspective and identify and test alternative strategies. Omitting details and less relevant information helps avoid "analysis paralysis", when decision-makers are presented with large amounts of detailed information too early in the decision-making process. Furthermore, Samset points out that accurate quantitative information tends to quickly become out of date. This is a problem, since the front-end phase in major projects may last for years, even decades. Samset refers to the "half-life of information", and carefully extracted qualitative information about a well thought-out project concept can provide reliable and valid input to the decision for the whole of the front-end phase.

Similarly, the exposition of parametric analysis by Pugh in Chapter 16 enables the forecast to "concentrate upon total costs and avoid being drawn into excessive detail. To descend prematurely into detail is to base forecasts upon what is not yet known and can only be conjectured." Scheibehenne and von Helversen (Chapter 10) also point to the danger that, in circumstances of uncertainty, risk and unforeseen consequences, decision-makers will give spurious credence to a decision made on the basis of detailed information.

There is clearly a need to make the most of the information that is available, and some techniques for this are discussed by Anderson in Chapter 15. It is also clear that bad project decisions have been made due to lack of information. Kharbanda and Pinto (1996) refer to the decision-making involved in the Sydney Opera House, without having "a basic design [or] a realistic estimate of time and cost involved. Feasibility analysis was almost nonexistent". However, it is an important theme of this book that when decisions are made at the very front-end of a project – when uncertainty is at its highest, and available information most restricted – the lack of detailed information can actually be a benefit rather than a hindrance, in providing focus and flexibility to the decision-maker.

This book

This book considers how decisions can be made at the front-end of major projects, in circumstances where information is usually scant. After two chapters introducing the area of front-end management of projects and decision-making at this stage, Part 2 examines how projects can be successfully aligned with the desired direction; Part 3 discusses how information – sufficient, appropriate and valid – can be gathered at the front-end; Part 4 concentrates on how information can be analysed; and finally in Part 5 the focus is on how decisions can be made. This sequence is not a recommendation for a decision-making process but simply a logical way to organise the chapters. Indeed, the book consists of independent contributions from leading experts in their fields, so it does not provide a single approach, far less a single "recipe": rather, it explores the issues involved in such decision-making.

It is not, therefore, the place of this chapter to conclude for the whole book, as each chapter is designed to speak for itself. However, some key ideas run throughout the book and are worth mentioning. There is firstly a need for alignment between organisational strategy and the project concept. But even when that is achieved, there is a need to deal with complexity, particularly the systemicity and interrelatedness within project decisions, as well as to consider the ambiguity implicit in all major projects. When calculating benefits and costs, estimation is affected by certain recurring issues, particularly psychological and political biases, and decision-making needs to take account of these. Major decisions are generally not taken by individuals in isolation, so there is a need to consider the social geography and politics within decision-making groups and organisational consortia. Finally, projects, once launched, do not travel a simple straight line, but move into an environment where the circumstances are constantly changing. This raises issues of governance and, particularly, the maintenance of strategic alignment.

This book seeks to lay the foundations for effective decision-making at the front-end of projects, where information upon which to base decisions is limited. Each chapter is founded on a solid base of sound theory but also demonstrates how the theories can be of practical value to real decision-making within major projects. The editors hope that the book will be of practical use in developing projects more successfully.

References

Association of Project Management. (2002). *Directing Change; A Guide to Governance of Project Management*. Association of Project Managers: London. Available at www.apm.org.uk

Buchanan, D. and Boddy, D. (1992). *The Expertise of the Change Agent: Public Performance and Backstage Activity*. Prentice-Hall: London.

Cicmil, S., Williams, T., Thomas, J., and Hodgson, D. (2006). Rethinking project management: researching the actuality of projects. *International Journal of Project Management* **24**: 675–686.

Concept. (2007). "The Concept Research Programme". http://www.concept.ntnu.no/index_engelsk.htm

Engwall, M. (2002). The futile dream of the perfect goal. In: Sahil-Andersson, K. and Soderholm, A. (eds). *Beyond Project Management: New Perspectives on the Temporary–Permanent Dilemma*. Libe Ekonomi, Copenhagen Business School Press: Malmo, Sweden, pp. 261–277.

Fraser, I. (2005). "Indispensable for Business Results": Why Directors should be advocates for Project Management. *Boardroom*. October 2005: 3.

Habermas, J. (1984). *The Theory of Communicative Action Vols I and II*. Polity Press: Cambridge, UK.

House of Commons. (2007). *House of Commons Committee of Public Accounts. Department of Health: The National Programme for IT in the NHS. Twentieth Report of Session 2006–2007*. HC 390, 11th April 2007. The Stationery Office Limited: London.

Isenberg, D. (1991). How senior managers think. In: Henry, J. (ed.) *Creative Management*. Open University Press: Milton Keynes, UK.

Janis, I. (1973). *Victims of Groupthink*. Houghton Mifflin: Boston.

Kharbanda, O. and Pinto, J. (1996). *What Made Gertie Gallop? Learning from Project Failures*. Van Nostrand Reinhold: New York.

Linehan, C. and Kavanagh, D. (2004). From project ontologies to communities of virtue. Paper presented at the *2nd International Workshop, "Making Projects Critical"*. 13–14th December 2004: University of Western England, UK.

Malgrati, A. and Damiani, M. (2002). *Rethinking the New Project Management Framework: New Epistemology, New Insights*. Proceedings of the PMI Research Conference, Seattle 2002, pp. 371–380. Project Management Institute: Newtown Square, PA, US.

Miller, R. and Hobbs, B. (2005). Governance regimes for large complex projects. *Project Management Journal* **36** (2): 42–50.

Miller, R. and Lessard, D. (2001). *The Strategic Management of Large Engineering Projects: Shaping Institutions, Risks and Governance*. MIT Press: Cambridge, MA.

Ministry of Technology. (1969). *Report of the Steering Group on Development Cost Estimating (The Downey Report)*. HMSO: London.

Morris, P. and Hough, G. (1987). *The Anatomy of Major Projects: A Study of the Reality of Project Management*. Wiley: Chichester, UK.

Office of Government Commerce. (1999). *Managing Successful Programmes. Office of Government Commerce*. The Stationery Office: Norwich, UK.

Office of Government Commerce. (2007). *The OGC Gateway™ Process. Gateway to Success*. http://www.ogc.gov.uk/documents/cp0002-Gateway_Gateway_to_Success.pdf

Olsson, N. (2006). Management of flexibility in projects. *International Journal of Project Management* **24**: 66–74.

Partington, D. (2000). Implementing strategy through programmes of projects. In: *Gower Handbook of Project Management* (3rd edition). Gower: Aldershot, UK.

Pinto, J. (2007). *Project Management: Achieving Competitive Advantage*. Pearson Education Inc.: Upper Saddle River, NJ.

Pinto, J. Cleland, D. and Slevin, D. (eds). (2003). *The Frontiers of Project Management Research*. Project Management Institute: Newtown Square, PA, US.

Project Management Institute. (2000). *A Guide to the Project Management Body of Knowledge (PMBOK)*. Project Management Institute: Newtown Square, PA, US.

Roth, G. and Senge, P. (1996). From theory to practice: research territory, processes and structure at an organizational learning centre. *Journal of Organizational Change Management* **9**: 92–106.

Simon, H. (1972). Theories of bounded rationality. In: McGuire, C. and Radner, R. (eds). *Decision and Organization*. North-Holland: Amsterdam.

Szyliowicz, J. and Goetz, A. (1995). Getting realistic about megaproject planning: the case of the new Denver International Airport. *Policy Sciences* **28** (4): 347–367.

Tempest. (2004). Scottish parliament opens for business. *Matthew Tempest*, guardian.co.uk, Tuesday 7th September 2004. From http://www.guardian.co.uk/politics/2004/sep/07/scotland.devolution

Thomas, J. and Mullaly, M. (2007). Understanding the value of project management: first steps on an international investigation in search of value. *Project Management Journal* **38** (3): 74–89.

Treasury, HM. (2007). Transforming government procurement. London. January 2007. Available at http://www.hm-treasury.gov.uk

Turner, J.R. (1993). *The Handbook of Project-Based Management*. McGraw-Hill: Maidenhead, UK.

Weick, K. (1995). *Sensemaking in Organizations*. SAGE Publishers: London.

Williams, T. (1999). The need for new paradigms for complex projects. *International Journal of Project Management* **17** (5): 269–273.

Wray, R. (2005). No place like Dome for major new venue. *The Guardian*, Thursday 26th May 2005.

2

Projects, Their Quality at Entry and Challenges in the Front-end Phase

Knut Samset

When news about a project hits the headlines, the public and media seem to be more concerned about its immediate outputs than with the long-term outcome of the investment. The bulk of planning resources are used to help do things right – not to do the right thing. The initial concept tends to remain largely unchallenged during the front-end phase and ends up as the chosen one – regardless of how inept it might turn out to be. This introductory chapter looks at some features common to major projects and the above paradoxes and suggests that more resources and efforts need to be used in the earliest stage of the front-end phase in order to improve a project's strategic performance. It argues that absence of information should not be an excuse but a challenge. The basis for choosing the concept can be improved considerably by systematically drawing on lessons from the past, intuition, surveys, public hearings, the use of expert judgement, stochastic estimation, scenario techniques, etc., combined with available factual information.

Features of a project

A project is a restricted undertaking assigned to produce a specified output within an agreed time limit and budget. It is undertaken in order to solve a problem or respond to certain needs in society. Projects are considered unique in the sense that even when the objectives and outputs are the same they may differ in varying degrees, since the context in which they operate and the uncertainty to which they are exposed differ from project to project. Projects are temporary in the sense that they have a definite beginning and end. They may involve a single person or many thousands of people. They may require less than a hundred hours to complete or several million hours. Projects may involve a single unit of one organisation or may cross organisational boundaries.

The tasks that projects are assigned for solution are defined in terms of more or less precise and realistic objectives. Being a temporary arrangement, and because the undertaking is more or less unique, uncertainty is often greater than normally expected within permanent organisations. Because of the uncertainty associated with planning and implementation, the extent to which a project will attain its goal is also uncertain.

There are several reasons for the increasing use of projects today. One is that many tasks in society are so huge and complex that individual organisations lack the competence or capacity to carry them out alone. Another is that a project focuses on and visualises the task and, therefore, has a motivating effect on everyone involved; responsibilities are clarified and the different parties are made accountable. Moreover, the project is an expedient way to transfer risk from the financing to the implementing party. It is also a convenient way to organise, which allows participants to pool resources and co-operate towards a common goal.

Nevertheless, there are numerous examples of projects that have involved high additional costs for society, both during and after they have been implemented. Cost overrun and delays are common. There seems to be a contradiction between the increasing use of projects and the fundamental problem that many of them appear to be failures in the public view. (Morris and Hough, 1991; also Flyvbjerg et al., 2003). Occasionally news about a project hits the media headlines. This happens when costs exceed budget or a project is significantly delayed. However, these are highly restricted and premature measures of a project's success. Judged in a broader perspective, a successful project is one that also significantly contributes to the fulfilment of its agreed objectives. Moreover, it should have only minor negative unintended effects, its objectives should be consistent with needs and priorities in society, and it should be viable in the sense that the intended long-term benefits resulting from the project are achieved. These requirements were first formulated for US-funded international development projects by the USAID in the 1960s. They were subsequently endorsed by the UN, OECD and the European Commission. In essence, they demand that five requirements or success factors are fulfilled – *efficiency, effectiveness, relevance, impact* and *sustainability*. These are tough requirements that go far beyond the issues usually covered by the media, but they do not seem to be the main concern of many planners and decision makers (Samset, 2003). What is termed *efficiency* represents only the immediate indications of a project's success in delivering its outputs. Clearly, there are many examples of projects that score highly on efficiency but subsequently prove to be disastrous in terms of their effect and utility. There are also numerous projects which fail to pass the efficiency test but still prove to be tremendously successful both in the short and the long term.

Tactical and strategic performance

The distinction when applying the success criteria above is the project's tactical and strategic performance. Success in tactical terms typically would be to meet short-term performance targets, such as producing agreed outputs within budget and on time. These are essentially project management issues. Strategic performance, however, includes the broader and longer-term considerations as to whether the project would have sustainable impact and remain relevant and effective over its lifespan. This is essentially a question of getting the business case right by choosing the most viable project concept.

One example of a project that is viable in strategic terms but inefficient tactically is the University Hospital in Oslo, Norway, which was completed in 2000, a year behind schedule and with considerable cost overrun. Newspapers printed comprehensive coverage of developments during the construction phase, and a public inquiry was subsequently commissioned to establish the reasons for the problems. Clearly, cost overrun was considerable in absolute terms, but in relative terms it was equivalent to only a few months' operational costs for the entire hospital and was therefore insignificant from a lifetime perspective. The overall conclusion after a few years of operation was that the University Hospital was a highly successful project.

More serious by far is when a project fails in strategic terms, even if it successfully produces intended outputs. Strategic failure means that the choice of concept turns out to be the wrong one. It could be the wrong solution to the problem in hand or a partial solution only. In some cases, the project might create more new problems than it solves so that these outweigh the benefits. In some cases, the initial problem no longer exists once the project is completed. One such example is an on-shore torpedo battery built inside the rocks on the northern coast of Norway in 2004. The facility is huge and complex, designed to accommodate as many as 150 military personnel for up to 3 months at a time. It was officially opened as planned and without cost overrun. However, only a week later it was closed down by Parliamentary decision. Clearly, a potential enemy would not expose its ships to such an obvious risk. The concept of permanent torpedo batteries was reminiscent of the Second World War, and had long since been overtaken by political, technological and military development, when the decision to build was taken in 1997, years after the Cold War was called off. What was quite remarkable was that this problem, which can only be characterised as a strategic failure, got little attention in the media, possibly because the project was a success in tactical terms.

Uncertainty and time

Projects are exposed to uncertainty in varying degrees, and this is often used to explain their failures. Uncertainty characterises situations where the actual outcome of a particular event or activity is likely to deviate from the estimate or forecast value. Uncertainty may have many and various causes, related to the situation itself – the design of the project, the time perspective, available information, the implementation of the project, etc. (Ritchie and Marshall, 1993). Obviously, decision making becomes difficult when uncertainty is high. Availability of relevant information reduces uncertainty from the decision-maker's point of view. It is widely believed that uncertainty is highest at the initial stage, when the project concept is conceived, and that it tends to reduce rapidly as information accumulates over time. This is illustrated in Figure 2.1. It follows that the utility of adding information is at its highest in the earliest stage. The decision-maker's flexibility and cost of making amendments is visualised with a mirroring graph. Decision makers can juggle with different ideas and strategic solutions to a problem in the initial stages, but once decisions have been made, essential choices become locked, and it is more difficult and expensive to change the overall design. Therefore, major issues such as agreeing on the most effective solution to a problem and the choice of concept need to be dealt with as early as possible – later on is too late. Less essential issues such as avoiding major cost overrun can be handled later, for example, when the final budget is agreed.

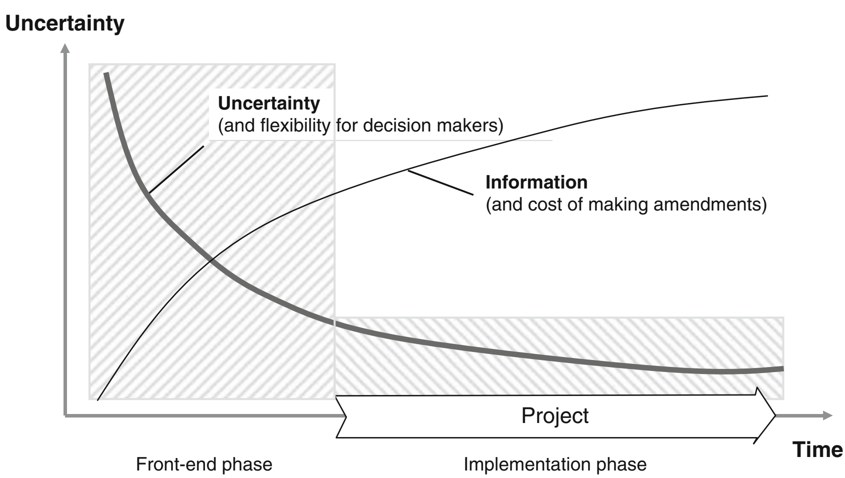


Figure 2.1 Uncertainty versus available information in a project. In general, uncertainty is highest at the outset and reduces as time passes and relevant information is generated

In Figure 2.1, the distinction is made between the front-end and the implementation phase. The graph suggests that the potential to reduce uncertainty and risk is largest up-front and decreases substantially when the project is implemented. It is a paradox, therefore, that most of a project's planning resources may be spent on detailed planning and engineering, while too little is usually spent on getting the idea right from the start.

Where projects fail strategically, it is likely that the problem can be traced back to decisions in the earliest phases, when the initial idea was conceived and developed. What happens during the front-end phase is therefore essential for a project's success. A study by the World Bank based on a review of some 1125 projects concluded that 80% of the projects with a satisfactory "quality-at-entry"[1] were successful, while only 35% of those with an unsatisfactory quality-at-entry achieved success (World Bank, 1996). Improved front-end management is therefore likely to pay off in a wider life cycle perspective, as evinced by the IMEC study (Miller and Lessard, 2000). One way of improving quality-at-entry is by challenging initial ideas and applying simple analyses, extracting and making use of previous experience from similar undertakings, and consulting with stakeholders.

In most cases, the key issue at the earliest stage is to shed sufficient light on the underlying problem that provides the justification for the project, and the needs that the project is meant to satisfy. Detailed information about possible alternative solutions is less relevant. This illustrates what seems to be a major dilemma, since most projects originate as one specific solution to a problem, while the problem itself may not be analysed sufficiently, and alternative solutions may not have been considered at all. Typically, the preferred concept originates in the mind of one individual, based on intuition and experience, rather than systematic analysis of problems, needs, requirements, etc. Most of the information generated is associated only with the initially identified solution. A second dilemma is that this information, which may be very detailed and specific, tends to lock decisions into the initially preferred concept – to the extent that this will inevitably be the one that is finally chosen. It is all too rare that alternative concepts are identified and analysed to the extent that they get a fair trial in the subsequent decision process.

This is a book about decisions and their implementation in the earliest phase of a project. This is when fundamental choices are made, when uncertainty is at its highest, freedom to choose is at its optimum and available information is most restricted. Adding information, therefore, makes sense – but only to a certain degree. The crucial issue is not the volume but what type of information is needed. Some available information might not be relevant in the decision-making process, and some information that is needed will not be available until later.

This seems to indicate a paradox. Contrary to the idea depicted in Figure 2.1, the amount of available information upfront might not be the issue. In the initial phase of a project, the priority is to establish an overall perspective, and to analyse the problem in its context, considering the needs and priorities of stakeholders, users and affected parties, in order to come up with a sensible strategy. Opportunities and risks should also be considered. Experience suggests that creativity, imagination and intuition can be more valuable at this stage than large amounts of data. Therefore, lack of information in the earliest phase may not necessarily be a problem: it can even be a strength. Many planners have learnt that in the early phase of a project it can be an advantage to operate primarily with qualitative expressions and only to a very limited degree with quantitative data. Only relevant, or "scant" information should be used to help identify the most viable concept for the project.

Information, validity and time

Investment projects are initiated for a purpose. The output of a project could be a road, an airport or an IT-system. It should have an effect, e.g. improved traffic flow or productivity. It should also produce long-term benefits in, say, economic terms. Figure 2.2 illustrates three separate processes and their corresponding objectives and suggests that there are three major planning perspectives in all projects. The project itself is a focused and restricted intervention that represents the first link in an intended chain of cause and effect. *Outputs* should deliver against, and be measured against, the project's defined *goal*. There should eventually be long-term effects as expressed in the project *purpose*. The project idea needs to be considered carefully upfront in the light of these three perspectives. This does not always happen. Too many projects are designed within a perspective essentially restricted to the delivery of the outputs, with few considerations regarding its immediate effects. In those

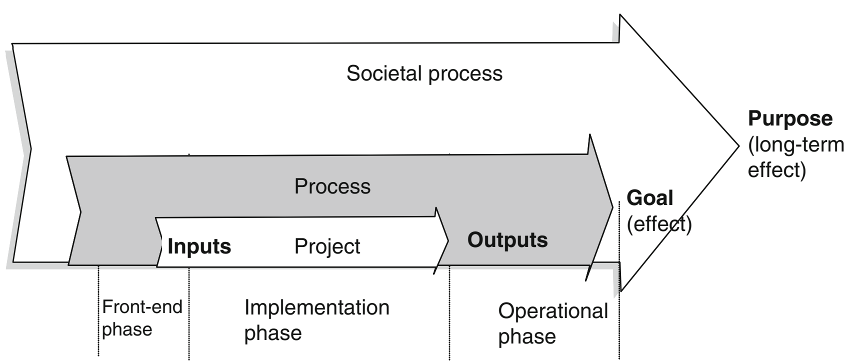


Figure 2.2 Three planning perspectives on a project

cases, where the long-term strategic effects are being considered, it is often the case that what is expressed as the project's purpose is overly ambitious and beyond its reach. This creates a problem during implementation, since the agreed objective provides unclear guidance. The same problem occurs upfront. Overly ambitious objectives give no guidance for selecting a viable concept. The long-term objective will need to be restricted to the effects that can be attributed to the project after it has been implemented.

In order to succeed in strategic terms, planners need to have a broad and long-term perspective. This is necessary to allow different concepts to be considered. For instance, the need to improve traffic flow on a specific road will require a road project. But the answer to the economic problem caused by traffic congestion could involve entirely different concepts, e.g. improving parking facilities in the suburbs for commuters, increasing tariffs for rush-hour traffic, improving the commuter train service, re-routing the traffic permanently, etc. While a broader perspective allows more concepts to be considered, it also makes deliberation more difficult. Planners will have to try and look deeper into the future, and the absence of valid information will be more noticeable. Uncertainty will be higher and conclusions more hypothetical and tentative. The problem is amplified by the fact that the front-end phase in large investment projects, particularly public schemes such as roads, airports, hospitals and defence projects extends over several years. As a result, the entire decision process becomes a stepwise sequence of analyses and decisions, and the outcome uncertain.

Project life cycle and stakeholders

The *front-end phase* commences when the initial idea is conceived and proceeds to generate information, consolidate stakeholders' views and positions and arrive at the final decision of whether or not to finance the project. In other words, the initial idea is transformed into the choice of concept. This may take years, even decades, in some large public investment projects. The key stakeholder during the front-end phase is the commissioning party, who is supposedly attempting to arrive at a rational choice[2] of concept, in dialogue, and sometimes in opposition, with other stakeholders. Such decisions will clearly have implications for the planning and implementation of the project, but more so for its effect and utility. The project management perspective should be secondary upfront, and the focus should be on the justification and potential benefits of the anticipated project, as seen in the operational perspective. Once this is agreed, subsequent decisions during the front-end phase are likely to have a more restricted effect on the choice of concept, focusing more on issues to do with budgeting, planning and implementation, and thereby entering the realm of project management issues.

The more fundamental challenges would typically be to deal with problems such as tactical budgeting, whereby responsible agencies at various levels tend to underestimate costs in order to increase the chance of obtaining government funding for a project. Another challenge is to increase the likelihood that the most relevant project concept is chosen, and yet another is to ensure a transparent and democratic process, avoiding adverse effects of stakeholder involvement and political bargaining. Challenges are abundant and complex.

The *implementation phase* commences once the decision to finance is made and includes detailed planning, mobilisation of resources and implementation, resulting in the delivery of the project's outputs. The main stakeholders are the contractors, while the commissioning party's involvement depends largely on the contractual arrangement. The contractors have a restricted view of the project, and their motivation is to deliver the agreed outputs according to specifications, while at the same time making a profit. For the contractor, the initial choice is of limited significance: his responsibility is to implement whatever he is commissioned to do.

The *operational phase* commences once the outputs have been delivered, set in operation or are being used. The main stakeholders are termed the "users". Decision makers are responsible for the operation at this stage and will have to make do with what has been produced, with limited possibilities to make strategic changes. The users are just the passengers on the ride and largely detached from the foregoing decision processes with no possibility to influence.

In a typical case the three groups of decision makers have different interests and perspectives on the project. They operate in separate sequences without much interaction. There is, of course, a need for some sort of alignment of interests, and in many projects the contractors and users may be able to influence decisions during the front-end phase to a limited extent.

Decisions and performance

A technocratic model for decision making prescribes that decision and analysis should follow in a logical and chronological sequence that will eventually lead to the selection and go-ahead of the preferred project, with no unforeseen interventions or conflicts, as illustrated in Figure 2.3. In reality, the process may seem somewhat anarchic. It may be complex, unstructured and affected by chance. Analysis may be biased or inadequate. Decisions may be affected more by political priorities than by rational analysis. Political priorities may change over time. Alliances and pressures from individuals or groups of stakeholders may change. The amount of information is large and may be interpreted and used differently by different parties. The possibility for disinformation is considerable.

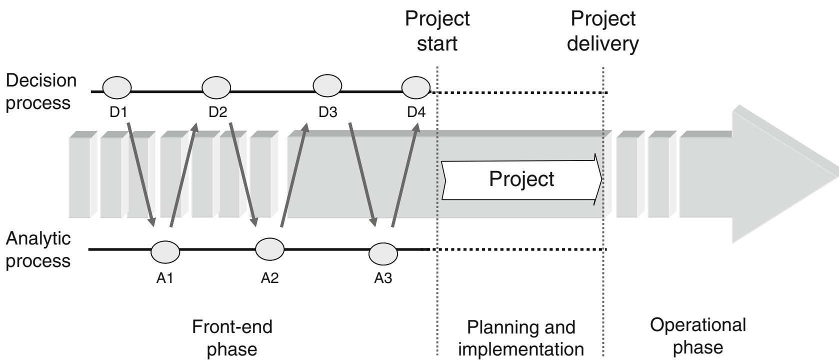


Figure 2.3 A model of technocratic decision-making upfront in projects

This more complex type of decision making is more in agreement with democratic principles than the technocratic expert-driven rationalistic process. However, the early selection of a concept tends to survive decision making, regardless of which process is adopted. This makes a strong case for proper research to identify the most viable concept upfront. However, time factor, complexity and lack of predictability also imply that the outcome of rationalistic planning upfront tends to alter over time.

Exact quantitative information tends to be more affected by time than by the choice of concept. On the one hand, it is obvious that the higher the precision, the more rapidly outdated is the information.[3] It is tempting to speak of the "half-life of information". For instance, exact information about the demand in a fast-developing market will have limited value after months or even weeks. On the other hand, there are many examples to suggest that qualitative assessments tend to remain valid for much longer. Consider the assessment of users' fundamental preferences within a market segment. While it might not be possible to make a valid prediction of the actual demand 3 years into the future, it may be judged that demand will continue for a long time and can therefore be relied upon in strategic planning upfront.

Restricted quality of information upfront may not be a major problem, since the need for precise information is low. It increases as the time for detailed planning approaches. In other words, the utility of exact information tends to reduce with the time-span. The opposite seems to be more of a problem: when decision makers are confronted with an abundance of detailed information at an early point in time, it may result in what is referred to as "analysis paralysis".

In major investment projects the front-end phase may last for many years and include several parliamentary election periods with shifting governments. In such cases, the outcome of the decision process is difficult to foresee. This could be a considerable problem. On the other hand, it could give space for ideas and decisions to mature, objectives to be aligned with policy and political preferences, stakeholders to be involved, the public to be informed, etc. Democratic decisions need to take time. Starting with a well thought-out strategy may be an advantage but is no guarantee for the best choice when the final decision is made. In some cases, the result may be entirely different from the initial choice. In other cases, the lengthy and unpredictable decision process may result in an optimal decision – even though the initial choice was entirely wrong.

Decisions based on scant information

There is a general understanding that good decisions depend on the quality of available information. In reality the situation may be more complex. What is considered quality information may vary. Experience suggests that many decision makers tend to be indulgent about what information is required, and they are willing to make major decisions even when available information is insufficient. Even in situations where relevant information is available, many decisions are influenced more by intuition and personal preferences than by facts and analysis (Mintzberg, 1994; also Henden, 2004).

The practical implications of this may not necessarily be adverse. Studies show that intuition may result in good decisions and even be preferable to rational analysis in some cases where there is an urgent need for decisions – even in complex cases. What seems to be necessary for intuition to work is that decision makers have thorough knowledge and insight into the field, which would typically require years of relevant experience – together with at least a minimum of analysis.

A crucial issue, in addition to the quality of information, is the extent to which it can be used. One common problem is that the complexity and amount of detail available may be beyond the comprehension of the decision makers and therefore inhibit it being used as intended. This type of information may be required as the decision process moves into the final stages. Upfront, however, the main challenge is to develop a realistic overall understanding of the situation in order to identify an appropriate strategy. The major requirements that will have to be fulfilled in order to solve the initiating problem should be identified, and this should eventually guide the selection of project concept. The challenge upfront will be to identify what is essential and restrict the amount of information that needs to be communicated. This will improve communication and the likelihood that information will be used, whilst retaining the flexibility that is necessary in an ongoing democratic decision-making process. The Pareto principle, also termed the 80/20 rule, can be applied to illustrate the need for information upfront. The idea is that in a chain of cause and effect, only a fraction of the causes will be the root of most of the consequences. To identify and visualise these few will be the main challenge.

With the limitations mentioned above regarding decision-makers' ability and willingness to make so-called rational decisions, it is essential to provide the best possible information as a basis for the key decisions. There is no single formula for what is the best decision basis. The UN/OECD model mentioned earlier is one example of a simple generic model that can be applied. This model requires that available information provides qualified answers upfront as to whether the anticipated project will be cost-effective, if its objective is pointing in the right direction and can realistically be achieved, whether the overall impact will be acceptable and if its long-term utility can justify the investment. It can be applied to the two examples mentioned earlier:

(1) The torpedo battery project evolved out of a front-end phase that lasted for more than a decade. After it was decided to go ahead, the project was implemented and completed as planned, within budget and with a cost comparable with similar projects elsewhere. The project was therefore considered successful if judged on its *efficiency* alone. The effect of such a facility would be measured in terms of its defence capability or ability to deter. However, the torpedo battery will have no such effect since the facility will never be in operation. It terms of its *effectiveness*, therefore, the project failed completely. The reason for this is that the project is no longer *relevant*, since the political scenario and warfare technology have changed over recent decades, and the anticipated military threat no longer exists. The project had some short-term *impact* in terms of local enterprise and employment, but the long-term effects will be negligible. The project is not *sustainable*, since the government is not willing to pay for operations. It is unlikely that the facilities can be realised at a price that covers investment costs or utilised in a way that generates sufficient income. Taken together, the five success criteria above provide a solid basis to conclude that the project is most unsuccessful.

(2) In the case of the national University Hospital, there was considerable cost overrun during construction and also 1 year's delay. *Efficiency* was therefore less than expected. However, there is no doubt that the project is *relevant*. It is the main national hospital, which provides some highly specialised and state-of-the-art expertise not found elsewhere in the country, and it also serves as a key educational institution. It has proved to operate *effectively*. One of its secondary effects is that of urban development in the city centre on the vacated premises after it was moved to the periphery. The project therefore scores highly on *effectiveness, relevance* and *impact*. In terms of its *sustainability*, there is no doubt about the commitment of government to secure operational

> funding in the future. Additional costs inflicted during construction were marginal in a wider time perspective and have little effect on its future economic viability. On this basis, the conclusion is that the project is highly successful.

The above decision model is widely used to evaluate projects mid-term or ex-post. It can also be applied in the earliest phase when ideas and concepts are being considered. However, not all success criteria can be operationalised and substantiated with reliable estimates or information upfront. Much will have to be based on judgement and weak or uncertain information. Some of the information needed does not exist. The possibility to predict is limited or non-existent. This applies especially to assessments regarding the efficiency and effectiveness of the anticipated project in the future. However, the key success criteria in any project are those that express the justification and long-term viability, in other words, the project's relevance and sustainability. Ensuring that a project is relevant is essentially a question of aligning its objectives with needs and priorities that justify the undertaking, on the one hand, and what can realistically be expected as its effects on the other. These are issues associated with needs, strategy, market demands, utilisation, etc. Securing sustainability and avoiding adverse impacts is essentially a question of understanding the complexity of the contextual situation in which the project is implemented and operated. This includes its institutional setting, market demands and restrictions, stakeholders' needs and priorities, and technological and environmental opportunities and challenges. Experience suggests that in many cases it may be possible to make at least reasonable qualitative assessments of a project's relevance and sustainability at an early stage. Such conclusions have often proved to remain valid during the front-end phase and beyond. Since they are fundamental success factors, the assessments should be updated and reviewed regularly throughout the front-end phase.

The conclusions regarding the two projects above were made in retrospect. Taking a second look at these projects from the ex-ante position suggests that the success criteria can also be applied upfront. The main problem with the torpedo battery was that it was not relevant. That is why it was closed down immediately and therefore will not fulfil its objectives. In this case, the assessment of relevance is essentially one of establishing a realistic military scenario. This has changed over many years since the Second World War. However, there is no doubt that there was sufficient information to foresee the relevance of the project a decade before the final decision to go ahead was made. The Cold War had been called off. The use of missile technology renders permanent installations like the torpedo battery obsolete. Even if these facts were not entered into the equation upfront, it should have been done later and reviews of the project's relevance been made at regular intervals – especially in view of political developments such as the downfall of the Soviet Union.

Based on this insight in retrospect, and projecting backwards to see to what extent the problems could have been foreseen at the time when the concept was decided, the conclusion is obvious: In this case there are no excuses. Current military technology and the political scenario should have excluded the concept altogether.

Regarding the University Hospital, it was evident at an early stage that the project was *relevant* and would be *sustainable*, and that this would be the case throughout the front-end phase and beyond. This provided a solid foundation for further planning and decision making. With this type of certainty upfront, most projects would probably succeed in the long term.

Projecting backwards in this case gives a different answer. Here the strategic decisions were sound. What went wrong had to do with emerging new technologies and added responsibilities, which were captured during the engineering phase, after the budget was decided, and subsequently required additional funding and time. These are minor issues, and it would not be feasible to suggest that initial decisions should be able to capture problems at this precision level.

Qualitative versus quantitative information

What is of interest in these examples is not the conclusions but the principle that decisions need to be based on a foundation of assessment and intuition. The solidity of assessment depends on the selection of decision criteria and the underlying information used to substantiate these. Each decision criterion needs to be substantiated with a number of parameters or indicators, producing an information hierarchy. In the example above, the decision model comprised five decision criteria (or two in the ex-ante situation).

This principle applies, regardless of the type of information used. It can be factual or judgemental, quantitative or qualitative. For the assessment at the highest level of the hierarchy to be useful and trustworthy, the selected decision criteria need both to capture the essential aspects that ought to be considered and be sufficiently comprehensive. Underlying supporting information needs to be valid and reliable. Reliability is a question of whether you can trust the information, this being determined by the quality of sources and the way it is collected. Validity is a term used to express the extent to which an indicator provides information that corresponds to what is to be measured. The type of indicators chosen will determine the validity of the assessment. Using several indicators at the disaggregate level helps improve validity at the aggregate level, provided each indicator is valid.

Take the national University Hospital as an example: Measuring the effectiveness of the project will require indicators that are relevant to the production of health services, e.g. the capacity of the hospital and the degree to which the capacity is utilised, the comprehensiveness of health services, the quality of these services, etc. Further down in the information hierarchy, these indicators can be disaggregated into several more detailed indicators so that the aggregate will provide a more precise assessment of the effectiveness of the hospital. Utilisation of capacity, for instance, can be expressed in terms of the utilisation of hospital beds, of polyclinic services, of different types of medical equipment, of personnel, etc.

What characterises such an information hierarchy is that it contains a mixture of qualitative and quantitative information. The gradient is towards qualitative information at the higher levels and more towards quantitative information at the bottom. When considering such a model along a timeline, in this case the different stages of a project's front-end phase, it is obvious that the information hierarchy will not be deeply rooted at the earliest stage of the process, for the simple reason that information does not exist. Assessment will have to rely heavily on judgement. Reliability will be severely restricted. The challenge will therefore be to ensure validity, by selecting the right type of indicators. As the decision process progresses, the information hierarchy will extend according to what information is available, which to an increasing degree will be factual and quantitative.

The absence of factual information is often used as an excuse for not seeking information upfront, rather than as a challenge to try and improve the basis for major decisions. The potential for improvements is substantial. A study of some 250 projects concluded that the main problems encountered were caused by uncertainties that were largely predictable (Samset, 1998). One explanation might be that successes and failures tend to repeat in projects with similar goals and characteristics. Taking the accumulated lessons of success and failure from such projects as the lead when looking for information relevant to the design and appraisal of an anticipated project may therefore be highly worthwhile. Surprisingly, this is rarely done even when such information is easily accessible.

Analysts and decision makers tend to underestimate the value of judgemental information. Expert judgement, whether based on intuition, experience, collective focus group sessions or a combination of these has proved to produce surprisingly accurate forecasts (Bazerman, 1994). Experiments from cognitive psychology (Goodwin and Wright, 1996) suggest that people are usually better at guessing the outcome of future events than answering general knowledge questions; that groups can generate better assessments of probabilities than individuals working alone by sharing knowledge and experience; and that assessment improves when complex events are broken down into more simple events. Learning from the past and from experience is useful in the attempt to understand future developments. For professional use in designing projects, a systematic approach and an open-mindedness allows alternative concepts to be considered. This is utilised in methods such as Delphi techniques, successive stochastic estimation, Monte Carlo simulation, etc., which have produced estimates that in many cases outperform what is achieved by deterministic methods and at a much lower cost.

The ability to predict is nevertheless limited. In general, future events are predictable only to a limited extent. Some events can be predicted in quantitative terms on the basis of assessments of past trends. In a situation which is predictable to some extent, uncertainties may cause only minor deviations along the general direction indicated by trends in the past. Other events occur more spontaneously and represent qualitative changes or discontinuities. Sudden changes in the economy, a new technological breakthrough, an act of terrorism or a change of government or management may suddenly change the general picture altogether. Often, neither the timing nor the effect of such events can be predicted or understood before they occur. Uncertainty therefore depends not only on the amount of information which is actually available to the decision maker but also on the nature of the process itself. In many cases, vital information is by definition not available beforehand to shed light on development before it actually happens.

These are some of the reasons why certainty in decision making is a rare occurrence that can be expected only when the time perspective is short, and the outcome is highly predictable. In the transition from uncertainty to strategy, the decision is likely to be affected by the weaknesses inherent in the prediction itself and the decision-maker's personal bias. In the transition from decision to outcome, prediction is likely to be affected by a number of unforeseen, confounding events – some of which may be unpredictable, critical events that may alter the direction altogether – and, ultimately, the involved parties' effort to make the prediction itself come true.

Predictability in itself is neither a necessity nor an objective in decision making. Too often the prediction becomes a prescription. It is no longer explorative, but normative, in the sense that it shifts the decision-maker's focus from finding the best solution to employing his resources and skills to make his own idea or prescription come true. This is one reason why many projects fail in the long run.

The cost of information

Information collected to shed light on future events will always be incomplete. A common definition of the term "uncertainty" that links this phenomenon to decision making is that uncertainty is an expression of the lack of information to make the right type of decision (Galbraight, 1979). Information, then, becomes the main agent to reduce uncertainty. This is a narrow and inconclusive definition, but the assumption that uncertainty is highest at the outset when information is most limited is a tremendous motivation for investing in information at an early stage, on the assumption that this will reduce uncertainty, as illustrated in Figure 2.1. The result could be wasted resources, where broad and often expensive pre-studies may provide masses of information beyond what is useful for the appraisal of a perceived project. Again, it is the quality and not the quantity of information that counts.

The cost of collecting information on a specific topic usually increases progressively with the amount of information collected, as indicated in Figure 2.4. This is because more information requires more in-depth studies or more wide-ranging information searches. On the other hand, the gain in utility of additional information tends to decrease. This is because there is usually a critical amount of information that is needed to get the necessary insight into a situation. Additional information will be of limited use. Maximising the utility/cost-ratio will set a limit to the amount of information that is useful.

This emphasises the need to invest in relevant information at the earliest stage of a project, while at the same time limit the search to what is useful for decision making at this stage. A targeted search for information regarding the main uncertainties likely to affect the project is more cost-effective than an unguided search, since it makes it possible to increase the share of relevant information and reduce the total amount. Assuming that success and failure are largely repetitive in the same type of projects, and that the uncertainties causing major problems in projects are largely predictable, the chance is that

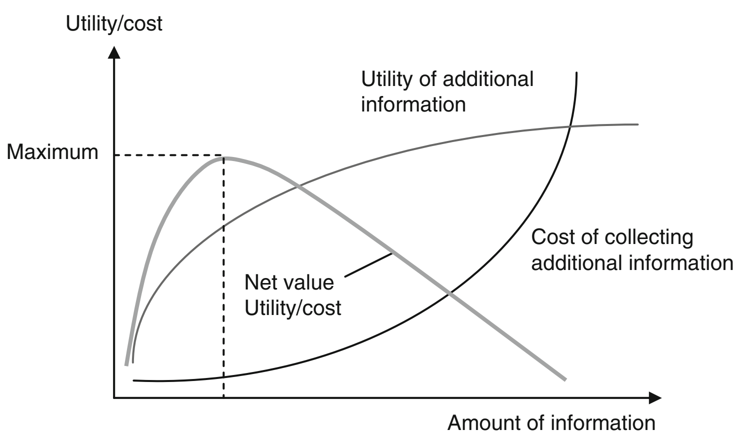


Figure 2.4 Cost and utility of information when additional information is collected to improve management of a project. As a general rule, the gain in utility tends to decrease and cost tends to increase. Consequently, in terms of cost efficiency, there is a limit to the amount of information that should be collected

a pre-study guided by accumulated experience from similar projects could produce cost-effective, relevant information.

The second aspect that illustrates the utility of investing in information becomes evident when the project is seen in a lifetime perspective, as illustrated in Figure 2.5. The diagram depicts the size of cash-flow during the main phases of a project. In most projects, the bulk of planning resources are invested in detailed planning after the decision to go ahead. The purpose is to improve tactical performance during implementation, in terms of reduced costs and shortening of the implementation period. These are vital measures to ensure success both in the short and the long term. However, as discussed earlier, cost overrun and reduced progress are marginal problems in many projects when seen in a lifetime perspective. They represent the project manager's focus on a project, which is the most restricted one. The utility of investing in planning can be considerable in absolute terms, but marginal in relative terms, i.e. compared with the accumulated economic result during the operational phase.

Seen from a strategic perspective, it makes much sense to invest in improving the basis for decision making during the front-end phase. This is likely to positively affect the long-term economic result and utility of the project. Compared with current practice in most projects, a small investment upfront can be highly cost-efficient if it can improve long-term utility – the more so,

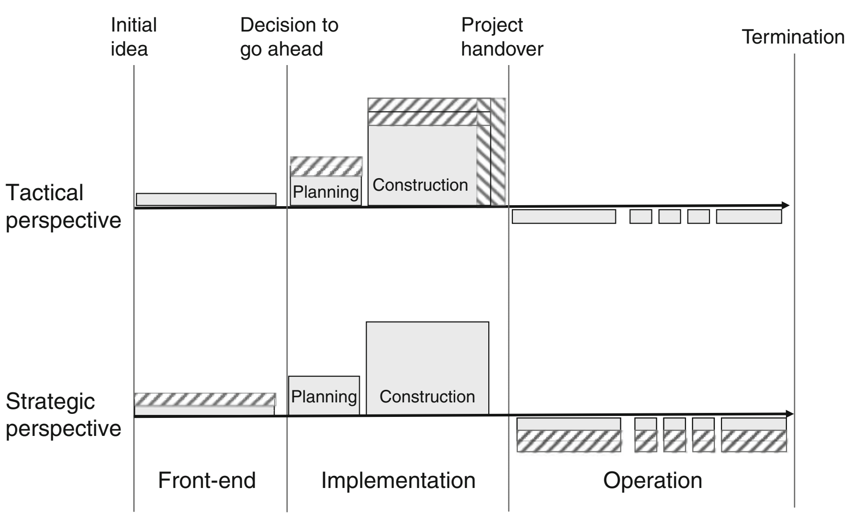


Figure 2.5 Investments in information during the front-end and planning phase aimed to improve the tactical versus strategic performance of a project, respectively

because the investment needed is likely to be minor compared with what is invested in detailed planning.

This is of course not a question of either/or, but of doing both. The challenge is to ensure that the project succeeds both tactically and strategically. Resources invested upfront aim to improve strategic performance and long-term utility, i.e. "quality-at-entry".

The difference between the two projects discussed in this chapter is that the University Hospital was successful in strategic terms but less so tactically. The torpedo battery was a complete failure in strategic terms. When that happens it does not make much difference if tactical performance is excellent.

Notes

1. Quality-at-entry was used as an indicator to characterise the identification, preparation and appraisal process that the projects had been subjected to upfront.
2. Rational choice in the sense that it is based on reason and that maximising utility is a main concern.
3. We need of course to make a distinction between lasting information, for example, physical data on the one hand, and less durable information such as economic estimates on the other hand.

References

Bazerman M. (1994). *Judgment in Managerial Decision Making*. John Wiley & Sons, Inc.: New York.

Flyvbjerg B, Bruzelius N, and Rothengatter W. (2003). *Megaprojects and Risk. An Anatomy of Ambition*. Cambridge University Press: Cambridge, UK.

Galbraight JR. (1979). *Designing Complex Organisations*. Addison-Wesley: Reading, MA.

Goodwin P and Wright G. (1996). *Decision Analysis for Management Judgment*. John Wiley & Sons, Inc.: London, UK.

Henden G. (2004). *Intuition and Its Role in Strategic Planning*. PhD Thesis, BI Norwegian School of Management: Oslo.

Miller R and Lessard DR. (2000). *The Strategic Management of Large Engineering Projects: Shaping Institutions, Risk and Governance*. MIT Press: Cambridge Mass.

Mintzberg H. (1994). *The Rise and Fall of Strategic Planning*. Prentice Hall: UK.

Morris PWG and Hough GH. (1991). *The Anatomy of Major Projects. A Study of the Reality of Project Management*. John Wiley & Sons, Ltd.: Chichester, UK.

Ritchie B. and Marshall D. (1993). *Business Risk Management*. Chapman & Hall: London, ISBN 0–412–43100–9.

Samset K. (1998). *Project Management in a High-Uncertainty Situation. Uncertainty, Risk and Project Management in International Development Projects*. PhD dissertation, Norwegian University of Science and Technology: Trondheim, May 1998.

Samset K. (2003). *Project Evaluation. Making Investments Succeed*. Tapir Academic Press: Trondheim.

World Bank. (1996). *Evaluation Results 1994*. The International Bank of Reconstruction and Development: Washington, DC.

Part II

Aligning Projects

3

Implementing Strategy Through Project Management: The Importance of Managing the Project Front-end

Peter W.G. Morris

Introduction

This chapter, if it succeeds, will open wider two doors in management theory and practice: one, to show concretely how corporate and business strategy implementation can be effected via project management, something rarely addressed in the academic literature; two, to show how project management, as a discipline, can contribute to strategic management. In doing, so it should clarify the confusion which has crept into perceptions of what programme and project management are. It will illustrate the importance of work done in the project "front-end". It will stress why value and benefits, and other measures of effectiveness may be more significant than the traditional project management efficiency measures of budget, schedule and scope attainment. And finally, it will show the importance of leadership in achieving this.

The underlying paradigm adopted is that of managing projects as organisational entities: the chapter does not take project management as an execution-only oriented discipline, with some other discipline being concerned with project definition (a view which prevails in the field and which can be inadequate and misleading). It addresses instead the discipline (or disciplines) required to manage projects successfully. This includes what needs to be managed in developing and defining projects, as well as in building, completing and operating them.

Strategy implementation

Business, or corporate strategy, occupies a central place in management teaching, research and practice, and yet when it comes to the implementation of strategy there is almost nothing in the scholarly literature on the role of projects or project management. This seems strange, since projects are the natural vehicles for implementing strategy.[1]

For many strategy theorists, having cast one's strategy, the next task concerns the decisions affecting project selection and resourcing – the analytical world of managing the project portfolio (Archer and Ghasemsadeh, 1999; Cooper et al., 1998; Rajegopal et al., 2007). A much more popular and populist field, however, is the management actions of the individual executive in shaping and implementing strategy and, in particular, the role of the CEO (Montgomery, 2008). The term *strategy,* indeed, comes from the Greek *strategos* meaning "the General", or more literally "the leader of the army". Thus Mintzberg, for example, has stressed the need to craft strategy to fit the organisation's context, and the conditions unfolding around its realisation (1987a), emphasising the importance of reacting effectively to "emergent" events, rather than simply following "deliberate" processes and plans (1987b). Ploy, perspective and pattern, all requiring insight and judgement, are as important, Mintzberg contends, as position and plan.

Other authors approach issues in strategy implementation from a more process-oriented perspective (Brown and Eisenhardt, 1997; Bryson and Delbecq, 1979; Eden and Ackermann, 1998; Eisenhardt, 1989; Loch, 2000; Thompson and Strickland, 1995), a perspective taken to almost mechanistic steps by Kaplan and Norton (2004, 2008). Others, such as Artto and Dietrich (2004) and Burgelman and Doz (2001), take a more synoptic view, stressing process, but also highlighting the importance of the CEO's strategic judgement. Similarly, Kotter (1996) emphasises the role of leadership, while proffering an eight-step process to implementing strategy: a sense of urgency, organisation, vision and strategy, communicating, empowering, creating short-term wins, consolidating, and institutionalising.

EDS, the consultancy company, offers a very similar sequence for managing organisational change (Pendlebury et al.,1998) with project management – as execution/delivery management – making an unusual appearance, as their fourth step – "Delivering". Change management, in fact, has an extensive literature on strategy implementation (Bossidy et al., 2002; Bridges, 1996; Cameron and Green, 2005; Cohen, 2005; Gilbert et al., 2007; Luecke, 2003), though generally with little, if any, reference to projects or project management.

No one in fact, from the business literature angle, is really connecting the implementation of business strategy with the way projects and programmes are shaped and executed. Yet, these are the normal vehicles for managing capital expenditure and, some would argue, for managing a wider variety of change.[2] Even the Association for Project Management (APM), the UK's professional project management body, in its otherwise excellent guidance on project governance, only refers to the need for "a coherent relationship [to] be demonstrated between the business strategy and the project/program portfolio" (APM, 2004). It says nothing about how project management can help shape the emerging strategy in such a way that its realisation aligns with intention, let alone anything about how project management can improve that strategy.

Perhaps the problem is that the projects community typically approaches strategy implementation from too linear a perspective, focusing almost exclusively on moving through the development life cycle towards completion (Anderson and Merna, 2003; Artto et al., 2008a, 2008b; Englund and Graham,1999; Milosevic, 1989, 2002; Milosevic and Srivannaboon, 2006; Morris and Jamieson, 2004). This has strengths, in that it emphasises delivery, but is weaker with respect to "loop-closing" feedback activities, such as lessons learned (Sense, 2008), and post-completion appraisal (Figure 3.1). Kaplan and Norton's iterative cycle is preferable: develop the strategy, translate it, plan operations, monitor and learn, test and adapt the strategy, and repeat. Programme management, however, has strong elements of learning between projects (OGC, 2007; Thiry, 2004a).

But possibly the neglect of project management's potential role in strategy implementation is better explained as being due to the failure of senior management – and non-project academics – to engage intellectually with the mechanics of implementation, and, by the same token, with the world of project management. Maybe this is because so few have an engineering or projects background. Maybe it is a personality trait. Perhaps though it is partly the fault of the project management community for not presenting both itself and its contribution to strategic management better. Thus while two recent studies (Crawford, 2005; Thomas et al., 2002) found that senior

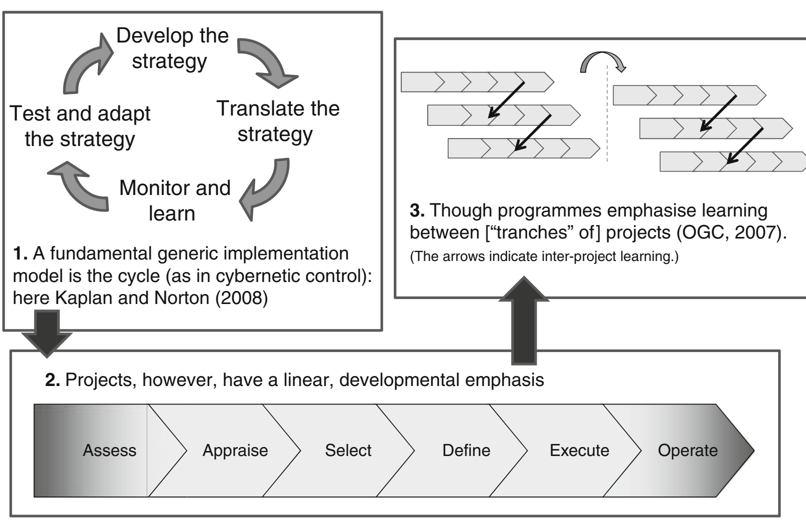


Figure 3.1 The linear nature of projects

management believes project managers should not involve themselves in strategic issues, this chapter, in contrast, proposes that project management has a real role to play in strategy implementation and, by extension, with strategy formulation.

It is the contention of this chapter that those shaping and executing the projects which flow from an enterprise's strategy can, and should, challenge and contribute to the unfolding of that strategy; that project management has more of a contribution to make to strategy implementation than simply ensuring alignment and being efficient in execution; that it can add value to the emerging strategy and ensure that benefits are reaped from its realisation.

What leads me to suggest that the findings of the Crawford and Thomas studies might be mistaken? The reason, I contend, is the conceptual framework currently associated with project management. It is not that the research has factual or methodological mistakes but is one rather of perspective – of paradigm: the paradigm that project management is only about execution management, and that project managers are not involved in managing the front-end definition phases of project development. If this is the prevailing view, then we should not be surprised if people see little role for project managers in shaping and implementing strategy. But this paradigm would change, if research findings were taken on board which emphasise the importance of managing the front-end, definitional stages of project development in order to achieve successful project outcomes.[3] This research suggests that we should indeed broaden the role of project management, and we shall be looking in this chapter at the consequences of doing so, but first we have some confusion regarding the definition of project management to deal with in order to clear our way forward.

Project management: paradigm lost

There is still substantial divergence of views and terminology about what is required to manage projects successfully, and what is the role of project management. There is a widespread, dangerous move to box project management into an execution or delivery paradigm, with the linkage to the sponsor's strategy having to pass through a higher programme management activity. It begins with the Project Management Institute (PMI) in the USA but is reinforced by the UK's Office of Government Commerce (OGC). Both differentiate programme management from project management, and allocate responsibility for corporate and business strategy to the former, with virtually no direct input from the latter – a view which this chapter will argue is illogical and unhelpful.

In its *Guide to the Project Management Body of Knowledge*, PMI gives its view of project management, or rather its "standards" for Project Management (PMBOK®), Program Management, and Portfolio Management (PMI, 2004, 2006a, 2006b).

Project Management is defined in the PMBOK® as follows: "the application of knowledge, skills, tools and techniques to project activities to meet project requirements" (PMI, 2004). (Nothing is mentioned about people or behaviour.) The key phrase is, "to meet project requirements". The requirements are assumed to exist; they are not developed within the purview of project management. No one is identified as managing the requirements' definition process. In fact, PMBOK® is a little confused, for a line or two later it states that, "managing projects includes identifying requirements and establishing clear and achievable objectives". PMBOK® works off an essentially execution view of the discipline: the definitional stages of the project – in which goals and targets are defined, requirements elicited, concept shaped, options explored and strategy developed – are completely ignored; they are done by someone else. (Do not get me wrong: execution efficiency is core to project management, but it is not sufficient: as we shall see. It neither addresses where many of project delivery problems originate from nor where value can be best built in. As we shall see, both of these occur in the "front-end" prior to the project targets being set.)

This view of the discipline is reinforced in the next chapter of the PMBOK® where, in Figures 3.2 and 3.3, the earliest output from the project is the project charter, developed in response to the "idea". From the charter flows the preliminary scope statement, and from this flows the Project Management Plan. There is no discussion of the project sponsor, of his aims or of his strategy. Nor is there any mention of a project strategy.

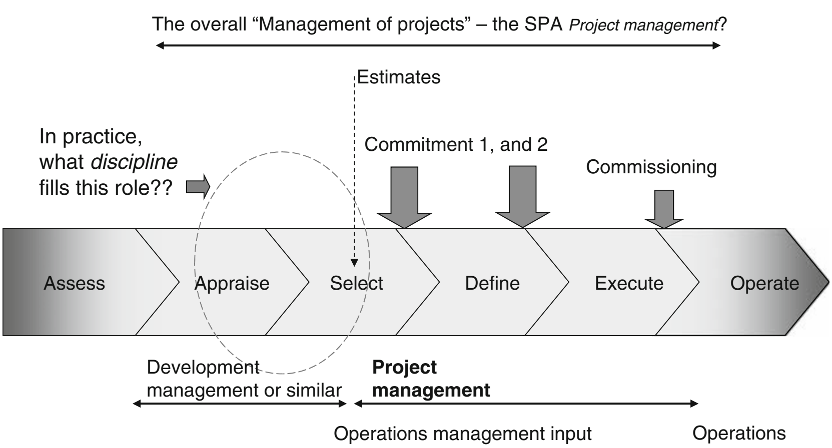


Figure 3.2 Where does project management sit? Is project management highly execution-oriented, or is it the discipline of managing projects so that they are developed and delivered to give best business value?

	Population %	
1. A process was used for optimising the value of proposed project/ programme strategy	55	
Of which:		
Value was expressed as benefit over resources used.	80	
The process was formalised as value management	55	
of which value management workshops were held at strategic stages in the life of the project.	40	
Those not using a process for optimising the value of project/programme strategy believed they should.	55	
2. Value engineering was practised on programmes and projects		25
Of which:		
Value engineering (optimising the value of the technical configuration) was distinguished from value management.	80	
Those not practising value engineering on programmes and projects thought they should.	56	
3. The value optimisation process was integrated with risk management		75
Those that did not thought it should be done.		40

Figure 3.3 Survey findings – value management

Many of the factors which cause projects to be delivered late, over budget, not to work properly or not to deliver intended benefits, were *not*, in fact, those typically covered by PMBOK® but were those which either arose earlier in the project definition stage, or lay with the sponsor or were issues "external" to the immediate project team (Morris, 1994; Morris and Hough, 1987). This finding was subsequently reinforced in major studies by Miller and Lessard (2001), Flyvbjerg et al. (2003) and the CIA (Meier, 2008) and supported by analyses such as those by Standish (1994), Williams (2005) and many others.[4] Strongly informed by this research, the UK's Association for Project Management developed a much broader, more catholic view of the domain.

The APM Body of Knowledge (APM, 2006) emphasises context, strategic imperative, commercial drivers, technology, the traditional control functions and, last but not least, people. Hence, Section 1 of the APM BOK is wholly given over to how projects fit within their business/sponsor's context. Section 2 covers "Planning the Strategy" (although the term project strategy then disappears to be replaced by "Project Management Plan").

Project and programme management strategy

Both APM and PMI treat the Project Management Plan interchangeably with the project strategy. PMI sees the Project Management Plan as "defining, integrating and coordinating all subsidiary plans" (PMI, 2006a). APM describes it as "the high level plan that will enable the project to reach a successful conclusion" (APM, 2006). What is missing in these definitions is any sense of direction – of movement or development. Similarly, OGC avoids the active verb "plan" and falls back onto nouns: "Corporate strategy defines the organisation's approach to achieving its corporate objectives ... [and] gives rise to the formulation of... policy. Strategies are aimed at ... particular goals... delivered by a series of plans" (OGC, 2007).

Strategy is more than planning. It is about vision, about Mintzberg's five "p"s. Artto et al. (2008a), after an exhaustive review of the term, offer the following definition: "Project Strategy is *a direction* in a project that contributes to success of the project in its environment." Strategy shapes and gives momentum to the project's course. The project and its strategy are dynamic (a view which accords with the perspectives of Montgomery, Kotter and others noted earlier). Further, Artto et al. contend that projects and their strategies do not have to adopt "an obedient servant's role" with respect to their sponsors or other stakeholders. Instead there is opportunity for the project to exercise a more pro-active role in strategy formulation and implementation, which is precisely the argument of this chapter.

In most project-based organisations, project strategy refers *de minimus* to an execution strategy – often called a Project Execution Plan (PEP). It is less common to find projects with either an earlier "front-end" development or overall project plan (or both). Systems Engineering best practice advocates that projects and programmes should not start without a [Systems Engineering] Management Plan defining how the project will be organised, engineered and conducted and that this should be used as a live management tool throughout development (INCOSE, 2004). OGC in *PRINCE2* includes the Project Initiation Document, which is less of a strategy, and more of a statement of intent, and the Brief, which is similar (objectives, direction, resourcing, etc.). (Though leading thinking in architecture talks now about strategic briefing (Blyth and Worthington, 2001).)

OGC's *Managing Successful Programmes* (MSP), which aims to provide good practice guidance to implementing corporate strategy through change programmes, talks about the Programme Vision, Programme Brief and the Programme [Preparation] Plan but not specifically about a programme strategy, although reference is made to strategy subsets, such as Stakeholder Strategy and Benefits Management Strategy.

Managing the front-end

As Aristotle said, defining the question is half the answer.[5] Too often in projects there is a tendency to spend insufficient time in the front-end, developing a robust project definition – the old "rush to code" danger.[6] How much time should be spent on the front-end? Miller and Lessard (2001) quote 3–35%, but this is so broad as to be of little use. IPA, the benchmarking firm working largely in the oil and gas sector, has a sophisticated metric of

"Front-End Loading".[7] IPA's metric works largely because they have a huge amount of homogeneous data. The problem with seeking a generic metric is that projects cover such a range of types and contexts. This is true even within a single sector, e.g. software.

Probably the best known – infamous – set of metrics on software development is that by Standish (1994).[8] Among the factors identified by Standish as contributing to the disturbingly high failure rate of ICT projects, are the front-end factors of not having requirements adequately defined, inadequate involvement of the user and inappropriate choice of technology. More recent findings published by Standish (2006) have shown an improving situation, one reason for which, they claim, is a move towards shorter projects. In fact the suspicion is that many of these projects may be little more than tasks. Certainly in many software projects, 2 years is considered lengthy. This philosophy is taken to the extreme in Agile development, where 90-day cycles are considered the maximum project duration which can be reliably planned.

Agile works on the premise that in software development, it is impossible to estimate accurately much beyond 90 days, that schedule and functionality have primacy over budget, that waiting for the whole thing to be finished before requirements are verified is too long and that the best way to elicit requirements is through close pairing between the user/customer (Leffingwell, 2007). In Agile, project management becomes almost a subset of execution. This extreme focus on the immediate short-term begs the question: where is the system architecture and the overall project strategy developed? In Scrum and Extreme Programming (XP) it is missing, but Rational Unified Planning (RUP) combines the Agile principles of iteration and team work, while acknowledging the front-end activities of building the business case and stakeholder management in the initial Inception stage, and project planning and system architecture in the subsequent Elaboration stage (Kruchten and Kroll, 2003).[9]

Many of the skills required at the front-end are special and specific to the functional expertise needed to create the development offering, i.e. financial, economic, legal, technical, commercial, etc. Some, however, will be project management related: resource requirements, procurement strategies, risk management, cost and schedules estimates, scheduling and phasing options, etc. The discipline of project management thus has a clear input into the front-end development work required in shaping the project definition.[10] But what is missing in the traditional paradigm of project management is any treatment of how the front-end definitional stages of a project should be managed – by whom, and by which discipline? For, over and above this "execution" input, this whole set of front-end activities needs managing – not just managing as in administering but managing in such a way that the project as a whole offers the best value possible (see Figure 3.2).

Development is a key area in projects. The craft of the developer is to build in and optimise the value of the proposed project offering. This is pre-eminently a skill of the owner, or in Miller and Lessard's words, "the sponsor" (Miller and Lessard, 2001). For Miller and Lessard, the key factor with regard to achieving satisfactory project performance is the competence of the sponsor, in particular the ability to deal with socio-economic and political "exogenous" and supply-chain "endogenous" factors. Since best practice these days calls for aligned supply chains (Cox et al., 2003), this surely holds true not just for the sponsor's Single Point of Accountability (SPA), but for everyone working on (or influencing, whether as a supplier or as a more general stakeholder) the project definition – whether an execution-focused project manager, a front-end project manager or the overall project SPA, as we shall see later in the example of Rolls-Royce.

Creating value and reaping benefits: the purpose of projects

Miller and Lessard (2001) make the distinction between efficiency and effectiveness measures of project accomplishment. The former are the "on time, in budget" traditional project management set of metrics; the latter are essentially about value-for-money and how well benefits are realised. Interestingly, the 60 large engineering projects Miller and Lessard studied did a lot worse on effectiveness (45% met their objectives, 18% were alright, without crises, 17% had to be restructured after crises, and 20% were abandoned or taken over) than on efficiency (82% met cost targets, 72% were on schedule).[11]

Value can be defined as the ratio of benefits to investment, or quality over cost, function over cost, or performance over resources (Thiry, 2004b). Prahalad (1994) was an early promoter of the importance of value creation as part of the strategy function but did so from a quasi Balanced Scorecard approach rather than from an implementation concern. He argued that strategy should be seeking to simultaneously create value in four distinct "markets": capital (investors), product (consumers), labour (specialised talent) and technology (suppliers). There has recently been research into improving projects' value propositions from the perspective of the capital budgeting process (Akalu, 2001; Yeo and Qiu, 2003) and on how project performance can influence shareholder value (Turner, 2000). Practically, one of the best ways project management can improve value is simply by shortening the project's duration, although there are a number of other project management levers. A more generic approach, however, is the use of Value Management (VM).

Value Management is the term used in the USA as an umbrella for Value Analysis (VA) and Value Engineering (VE) (Society of American Value Engineers, 1997). In other countries, it is seen as "the combined application of value methodologies and other methodologies at the organizational level ... to improve organizational effectiveness" (BSI, 2000). VM is based essentially on three key principles:

- a multi-disciplinary approach to analyzing value in terms of stakeholders' needs and objectives;
- a structured decision process underlying creative thinking;
- a focus on analyzing functions rather than just accepting pre-defined solutions.

In many ways, value management can be considered as a state of mind – a disposition to seek out value. Many project people believe that the early Optioneering phase of a project's development (surrounding "Appraise" in Figure 3.2) is essentially a form of value management. Nevertheless, formal VM involves conducting a series of workshops in which the project's opportunities for improving its value propositions are rigorously and objectively analyzed. These workshops are typically (Simister, 2007):

- VM1: review mission and strategic fit
- VM2: project scope – define performance of elements of the project
- VM3: test design options – VE
- VM4: wrap-up.

VM and VE are very common in the engineering, manufacturing and construction-based industries, but less so in ICT/software development. Yet, amazingly there is no mention whatsoever of Value in *PRINCE2* (OGC, 2002a), and it has only the briefest of mentions in *Managing Successful Programmes* (OGC, 2007). It is almost completely missing from PMBOK®. (In fact, Value Engineering is referred to twice as a technique but with no explanation; there is neither a discussion of the concept nor of the much bigger concept of Value Management.) These omissions may seem astonishing at first. Reflection on the reigning paradigm of project management, however, explains why they are absent: project management is about executing orders, not about shaping instructions. Value is not the concern of this paradigm: it is "someone else's baby"! The reality, even amongst PMI's membership, is different. In research conducted at UCL for PMI in 2005/2006, 55% of those surveyed[12] had a process for optimising the value of the project (Figure 3.3).

Creating value usually entails taking on board risk. Advising on and managing risk is another major contribution that the project manager should be making to senior management as the strategy develops, both in terms of the evaluation of options and the calculation of estimates and their uncertainty spreads. Curiously, little work has been done linking value management and risk management.[13] In fact, considerable overlap and confusion exists between several important interrelated topics in this area – principally, value, risk, uncertainty, opportunity and benefit. Latterly, risk, uncertainty and opportunity have received a lot of attention and agreement on terminology, with the project management and insurance communities bending the English language to define risk, not as the possibility of a negative event[14] but as an uncertain event (ICE, 1998; OGC, 2002b; Simon et al., 1997). There has been little work, however, on the relation between value and the recently emerging practice of managing benefits.

Benefits, strategy and confusion over project and programme management

The focus on benefits in project and programme management is squarely on their management. This involves clearly identifying what the project or programme should be delivering, deriving measures for these benefits and (very importantly) identifying owners for them. It must be ensured they are "harvested" effectively, and lessons learned from so doing are fed back so that future projects are changed and shaped (strategy, configuration, plans, etc.), and future benefits are improved in a cost-effective way – i.e. with value enhanced.

The project management community is split about who has responsibility for managing benefits. The prevailing view from the major institutions is that benefits management is the provenance of programme rather than project management. *PRINCE2* is the honourable exception, having a whole section (13) on the business case, one part of which is benefits definition. (The business case is shown as potentially affecting the project planning, not the other way around.) Development of the business case is the responsibility of the sponsor (the Executive) but "this may be delegated to the Project Manager". The business case is developed during Initiation and approved at each stage gate. Benefits are assessed at the Post-Project Review.

Benefits are not mentioned at all in PMBOK®, PMI's project management "standard". They are, however, everywhere in PMI's "*Standard for Program Management*" (PMI, 2006a). For PMI, programme management is about the delivery of benefits – project management is not. As PMI's standard makes crystal clear, "some organisations refer to large projects as programs. The management of large individual projects or a large project that is broken into more easily managed subprojects remains with this discipline of project management... If a large project is split into multiple related projects with explicit management of the benefits, then the effort becomes a program." What about projects that are not split into multiple projects: should not all projects, and their sub-projects, pay explicit attention to managing the benefits they are supposed to be delivering?[15] After all, projects are done for a purpose – often a business benefit. Would you not want the manager of your project to be looking at how he or she could improve its business benefit?

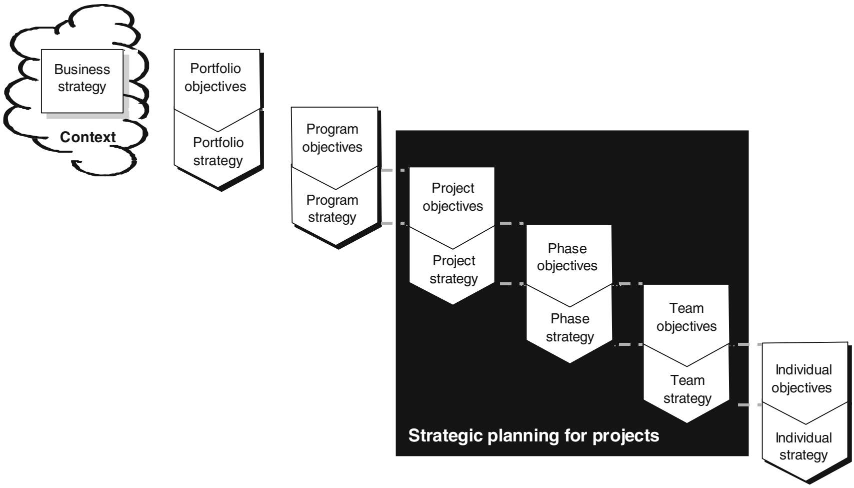
OGC's *Managing Successful Programmes* (OGC, 2007) shares the PMI view: "Programmes deal with outcomes, projects deal with outputs" (OGC, 2007).[16] This is an unnecessary and dangerous maxim which far exceeds the more sensible advice of *PRINCE2*: "Projects within a programme environment may not in themselves deliver business benefits. They may be required to deliver products that are prerequisites for other projects" (OGC, 2002a). Similarly, *The Gower Handbook of Programme Management* states that "projects do not deliver benefits, but create deliverables" (Reiss et al., 2006).

It could be argued that it is only semantics, but words are powerful. They create concepts and shape practices, and here the clear message from these two highly influential institutions is that working in projects need not involve concerns about business benefit. All my research and practical experience suggests it should be otherwise.

What is happening is that the need to relate front-end, value-adding development activities and back-end benefits realisation activities (necessary on almost all projects) is getting confused with the distinction between projects and programmes. Programmes, states PMI, are "group[s] of related projects managed in a coordinated way to obtain benefits and control not obtainable from managing them individually" (PMI, 2006a). Or, similarly, programmes are "temporary, flexible organisations created to coordinate, direct and oversee the implementation of a set of related projects and activities in order to deliver outcomes and benefits related to the organisation's strategic benefits" (OGC, 2007). Both these definitions are acceptable. But, if working on a single shot project, in order to recognise that its business value and benefits can be managed, is it necessary to insist that these are programmes and not projects and that it is a case of programme management and not project management?

One could go on. *The Gower Handbook of Programme Management,* for example, states that "It is the purpose of project management to take a project from the definition stage to delivery of previously agreed defined products. It is the purpose of programme management to define projects…." (Reiss et al., 2006). Is all project definition work, therefore, done by programme management? Few organisations would agree that this is the case.

The Japanese project management Body of Knowledge, *P2M*, seems more sensible. Project management is elevated to "the total framework of practical professional capability [required] to deliver …. a given mission." Value creation is seen as the sum of this endeavour (ENAA, 2002). But as regards PMI and OGC, Reiss and others are of like mind: programmes and programme management are the means through which strategy is implemented, not projects or project management.



This is a classic p.m. diagram (Turner, 1999; Youker, 1993) suggesting a hierarchical cascade from portfolios to programmes to projects – and omitting any feedback loops! In fact the hierarchy need not apply at all: portfolios can exist below programmes; projects do not necessarily have to be subservient to programmes

Figure 3.4 The portfolio-programme-project hierarchical cascade

This insistence on seeing the front-end definition activities, and the linkage through strategy, as being necessarily through programmes, and then on to projects, may be the consequence of thinking in hierarchic terms rather than life cycle developmental terms. It may, on the other hand, be due to not thinking through what really are the differences between programmes and projects (particularly major ones).[17]

As regards the former, all illustrations of the relationship between projects and programmes show projects sitting beneath programmes, which sit below strategy (see Figure 3.4). In reality, however, not all projects, even all small ones, sit beneath a programme. Where they do not, but are stand-alone, their front-end definition is perforce part of the project.

As regards the latter, ultimately the difference between projects and programmes can become very hard to distinguish. Is a new airport terminal a project or a programme? It is a programme probably, when one considers all the IT, baggage handling, airline-side activities, security, etc., in addition to the building itself and the fact that all these need to be working in order for benefits to be realised. Was the Channel Tunnel a programme? Is a large Private Finance Initiative (PFI) hospital a programme? [Yes, to all, for the same reason.] Is a railway or refinery shutdown, or a new building a project? For a work-package manager, the work-package might clearly be a project (but the package should normally be able to have its benefits identified). For the owner's project manager, however, there will typically be several major work-packages-cum-projects, and for him/her, it would appear to be technically a programme.

Thus the principles of programme management would seem to apply to (a) the larger or more complex the project/undertaking, or (b) the higher up the WBS one is working, and the more interaction of disparate work-packages [projects] one has to manage. Does this mean that talk about projects and project management should cease in such cases? This would surely be impractical, and it would be absurd to propose such a thing. Common usage would resist it. Hence there remains the sensible situation where projects and project management should acknowledge and apply many of the features of programme management, including the need to (1) improve value and derive benefits from the project, (2) manage the project's front-end effectively and to minimise unwanted risks, (3) develop a strategy for the project to cover both of the above and to contribute to the enterprise's business and corporate strategy – which is the argument put forward in this chapter.

Rather than compartmentalising responsibilities and diminishing the potential relevance of project management (Morris, 2003), we should surely encourage more inclusive a discipline (one which I have termed "the management of projects" (Morris, 1994; Morris and Pinto, 2004)). As I wrote with my colleagues Lynn Crawford, Damian Hodgson, Miles Shepherd and Janice Thomas in 2006, following a 2-year project management research visioning programme involving over 40 leading project management academics and practitioners (Winter et al., 2006): "Program management is more strategic than project management, but to deny the discipline of project management a strategic, holistic role is surely dangerous, ill-grounded, unnecessary, and limits the membership of [the professional] project management associations and project manager roles to the more junior practitioner level, leaving senior practitioners without a professional home. Project Managers, Program Managers, Project Sponsors, Portfolio Managers, Project Directors, Project Services Managers, etc. are in the [research programme's] view, all members of a project management 'job family' sharing a responsibility for the 'management of projects' " (Morris et al., 2006).

Different types of strategy

Recent research into the linkage between business strategy and project management has positioned business strategy as being pre-eminently about competitiveness (Srivannaboon, 2006; Srivannaboon and Milosevic, 2004). Certainly competitiveness is a central pre-occupation of business strategy (Porter, 1980, 1985), but there may well be specific topics that demand their own strategy, not least as they bear on implementation. Technology strategy, for example, is of dominant concern in many organisations (Christensen, 1999; Christensen and Raynor, 2003; Thomke, 2003). Lack of alignment between corporate and project/programme technology strategies was recently found by the US General Accounting Office to be a major source of cost growth and schedule slippage (GAO, 2006).

Technology was a major cause of disruption and failure on many projects pre-mid-90s (Morris and Hough, 1987), though it generally seems that our understanding of how to manage and achieve technological innovation and uncertainty has improved in recent years (Crawford, 2005), e.g. via prototyping, rapid applications development (RAD) and improved requirements management (Davis et al., 2004; McKinlay, 2007; Stevens et al.,1998). Even fast-tracking and concurrent engineering can now be done without the hazards of old-fashioned concurrency (Morris, 1994). Concurrent engineering is no longer simply beginning to build before design is complete. It now involves joint design and production teams, information modelling of the product that is to be built and integrated scheduling of production and front-end activities (Prasad, 1996). Companies that work in environments where these issues are common generally have well-developed management practices for dealing with them (although design management and briefing still cause difficulties in building construction). The engineering construction industries, for example, have a list of largely technology-based "Value Improving Practices" that, following research in the UK by ACTIVE[18] and the European Construction Institute, and benchmarking by IPA, have been seen should be addressed in the front-end (see Figures 3.5 and 3.6).

However, projects and programmes where there is deliberate technology push, not surprisingly, still experience problems. The GAO found in its 2005/2006 study of 54 major programmes that only 10% had mature technologies at programme start, only 43% had "critical technology maturity" by Design Readiness Review and only 67% by Production initiation (GAO, 2006). Also, "programs with mature technology at key milestones had average cost overruns of 4.8% while programs with immature technology had average cost overruns of 34.9%" (Meier, 2008).

Resource-based strategy is another major strategy stream directly influencing project implementation (Barney and Clark, 2007; Collis and Montgomery, 1995). This has a very wide remit, from capacity planning, through processes (dynamic capabilities) (Eisenhardt and Martin, 2003), to [core] competences (Boyatzis, 1982, 1993; Hamel and Prahalad, 1994). Conceptually, even Supply Chain Management (Cox et al., 2003) could be seen as being within this grouping. All can be major issues in the setting up of and carrying out projects.

Similarly, companies and projects (or programmes) will often need commercial and financial strategies; or sustainability, Health, Safety, Security and Environment (HSSE) strategies, stakeholder management strategies, benefits management strategies, etc.

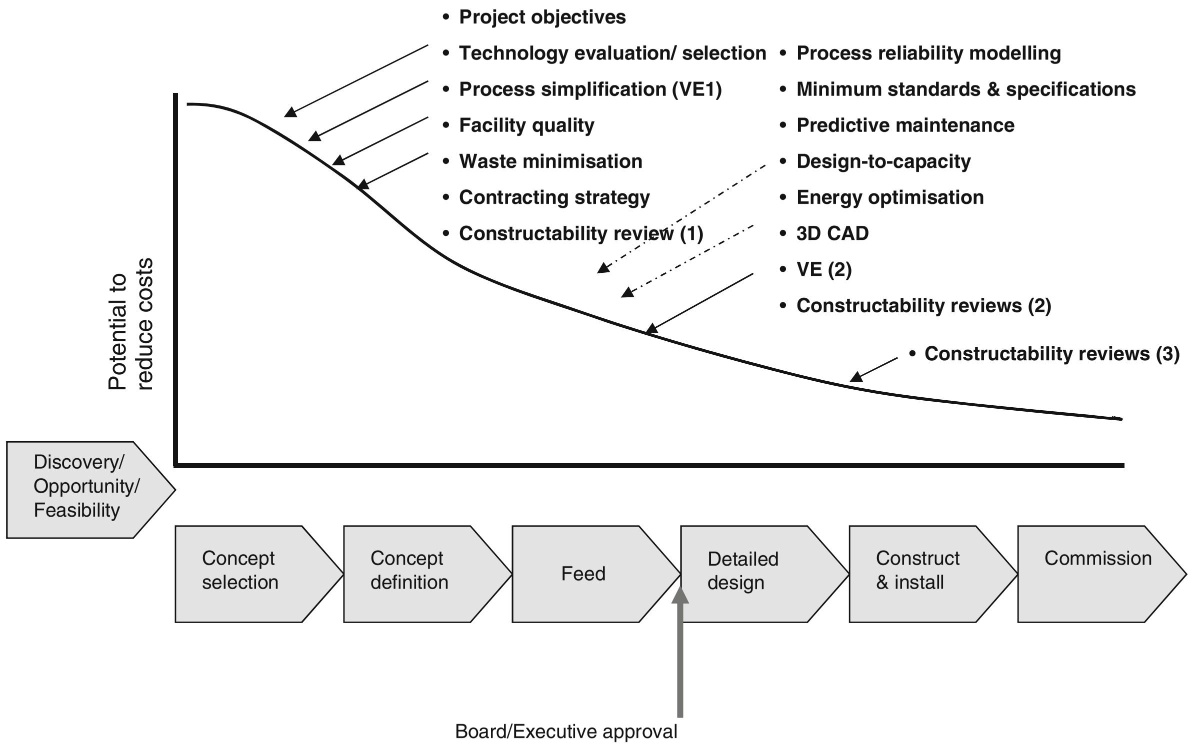


Figure 3.5 Front-end loading activities

Class of facility	Quality: definition of the best value facility classifications to meet the Business Plan requirements
Technology selection	Identification and evaluation of the technology most appropriate to meet the defined business need
Process simplification	Reduction of unnecessary investment and operating process costs
Design to capacity	Avoidance of over-sizing components and systems to meet the defined business need
Design to cost	Controls cost throughout the design process by defining not-to-exceed cost targets for each system or sub-system of the project or product
Waste minimisation	Reduction of waste at source and re-use of waste for cost-effectiveness
Customised standards, specs, practices	Customising of standards, specifications and practices that are appropriate to the application and not excessive to the defined needs of the specific facility
Energy optimisation	Maximisation of total return on investment by judicious selection and use of plant utilities and equipment
Facility optimisation	Reduction in overall return in investment and operating costs by combining or making unnecessary one or more chemical or physical processing steps
Constructability	Analysis of a design by experienced construction personnel to reduce costs and save time during construction
Value engineering	Identification of alternatives for meeting functionality and quality requirements at the least life cycle cost
Strategic master planning	Needs identification and alignment of organisation and long range infrastructure development
Life cycle value impact assessment	In conjunction with external stakeholders, a triple bottom line (social, environmental and financial) approach to determining most appropriate solution for the total life cycle
Systems optimisation	Efficiency reviews to optimise an in-service process or facility
Reliability improvement	Provide an effective way to cost justify maintenance activities, decrease equipment downtime and identify solutions with a high return
Risk assessment and management	Risk-based decision making and management of risks
Supply chain optimisation	Integration of the entire supply chain (e.g. suppliers and installers) for full potential in terms of optimising cost, schedule and quality
Partnering	Trust-based process that focuses owner and service provider(s) on creative cooperation and avoidance of confrontation for mutual financial benefit
Performance criteria and measures	Performance Criteria and Measures allow stakeholders to explicitly model required performance (e.g. functionality, schedule, various impacts, etc.) for a specific situation and use the model to evaluate alternate solutions, independent of cost

Figure 3.6 Process engineering industry VIPs (value improving practices)

Project governance: setting the conditions for delivering strategic value

What has become clear in recent years is that since strategy flows from the business into projects and programmes (with some flow returning from projects to the business), the way that this is organised, controlled and led will have a strong influence on the conduct and character of the project.

The British Airports Authority (BAA), for example, employed, at least until it was acquired by Ferrovial in 2005, the Objectives, Goals, Strategies, Measures (OGSM) method developed by Procter & Gamble, where objectives are translated into goals, and for each set of goals, strategies are developed against which metrics are placed. This process repeats at progressively descending levels of the organisation, goals and metrics at one level giving rise to objectives and strategies at the level below (Figure 3.7) (Morris and Jamieson, 2004).

BAA's project development process is typical of most capital expenditure projects. The development manager develops a Statement-of-Need for the project; the development team develops options for meeting that need, one of which is developed in greater detail. A Project Board is responsible for

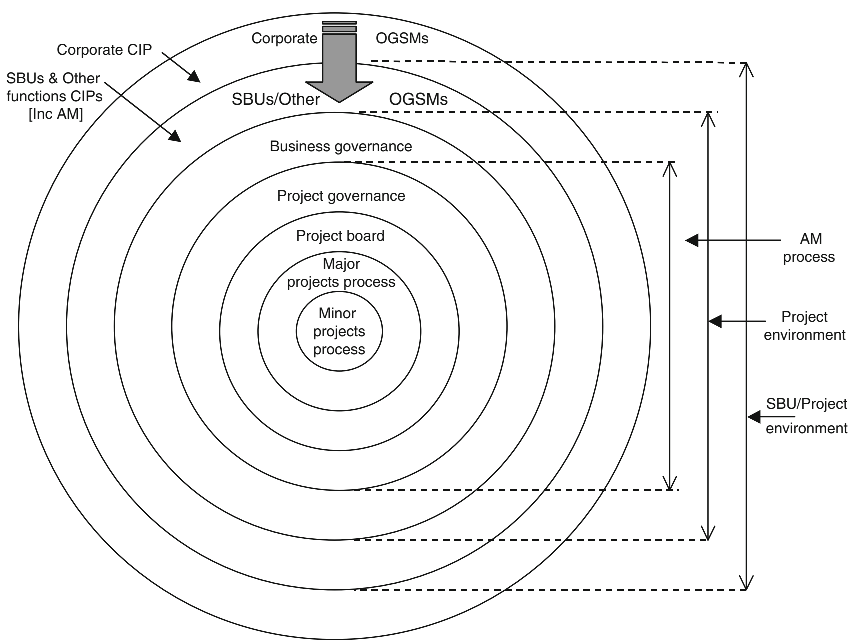


Figure 3.7 Corporate, business unit and project environments

ensuring that the project meets the needs of the business and maximises shareholder value, and that the strategic contribution, value-for-money and uncertainty assessments are undertaken appropriately. The Board Chairman is accountable for successful project delivery. Governance decides at the end of the "Define Need" stage whether the project can proceed to the "Define Solution" stage. A project execution team is responsible for "Delivering the best value solutions". This team used to be headed by a Project Manager, but this title was changed to Project Leader, to emphasise the importance of the execution team contributing to the overall asset development process by acting creatively and adding value. In other words, this is a process similar to that advocated above, although interestingly the language is all about value rather than benefit.

This project development process is almost identical to that followed in most oil and gas companies, although here there is usually more emphasis on the role of the gatekeeper (Cooper et al., 1998). The gatekeeper is the person best qualified to assess the project at the stage gate review, both in terms of its ongoing business contribution and its project assurance (conformity to expected practice). The gatekeeper may be a different person from the sponsor/the sponsor's Single Point of Accountability. Ongoing studies by a UCL research team (so far unpublished) show that the rigour with which the Gatekeeper, in particular, conducts the gate review exerts a very substantial impact on the effectiveness with which project management practices are applied. The interesting point is that many gatekeepers have no formal training in project management! That said, a professional ought to be able to conduct himself/herself properly, independently of their client's behaviour. Yet, because the prevailing paradigm is that there is little need, or expectation, for project managers to participate in, or contribute to, the sponsor's business decisions, project managers often fail to advise the sponsor or gatekeeper adequately, not least in the area of strategy. (*PRINCE2* valuably contests this, having a whole section devoted to the business case, but even then nothing on strategy.)

The same principles apply for suppliers as for owners. It makes sense for Tier 1 suppliers to get as close as possible and to deliver to their clients' strategies. But even Tier 2 and Tier 3 suppliers need to align their project execution with their clients' and their own internal strategies and to ensure management does this in both development and execution. One of the most common complaints in supplier organisations is that commitments made by Sales & Marketing to win the tender are not effectively integrated with project delivery. Reviewing strategy at the Tender Submission gate is one way of addressing this shortcoming. Rolls-Royce, for example, a Tier 2 supplier, has a rigorous project stage-gate process, in which strategy is one of several project elements which are systematically reviewed at each gate (Morris and Jamieson, 2004).

An example of project strategy

So, what does a project strategy look like? Big pharma provides an interesting, surprisingly generic, example.

All "big pharma" companies follow approximately the same drug development cycle process (if only because they are obliged to, for regulatory reasons). What makes them so interesting from the strategy implementation perspective is (a) the very high failure rate of candidate (new) chemical entities and consequent need for the companies to juggle resources across their development portfolio, for expected efficacy and commercial attractiveness (value and benefit); (b) the fact, perhaps unique amongst project industries, that the front-end does not really consist of developing and designing the product but is instead devoted to uncovering its safety and efficaciousness (from its reactions in animals and man in pre-clinical and clinical tests) as quickly as possible. Most compounds fail these tests and are removed from the pipeline ("attrited"). Such is the rate of attrition, particularly in the early stages, that only relatively short term, flexible project strategies are prepared (Figure 3.8).

Strategy is truly emergent and interacts constantly with project management. Nevertheless, three specific project strategy plans are prepared for the project in quite a "deliberate" manner. The first is an operating plan that is prepared prior to moving to pre-clinical testing. Several months prior to the commencement of the compound being approved for early development, a Stage 2 strategic planning document is prepared. A Stage 3 strategy is produced at the end of Stage 2, defining the implementation plan for full development activities. The project leadership team develops these strategy documents, ensuring that they are kept updated and reflect any changes in strategy that may be needed.

The scope and sequencing of these strategic plans is not dissimilar to that of many companies in, for example, ICT, oil and gas, power and other utilities, new product development and many other capital development industries.

Applying value management practices to key elements of project strategy is a distinctive element in the way that drug development programmes and projects are managed. Crucially, however, the value optimisation process takes place around the strategic character of the project in relation to the portfolio, including the rest of the programme (the family of candidate products based on a similar chemical entity or part of the same brand), where this is appropriate.

Leadership and project effectiveness

Most pharmaceutical project management organisations distinguish between the project leader/director role and the project manager. Typically, the former has a strong feeling for the science of the development (and

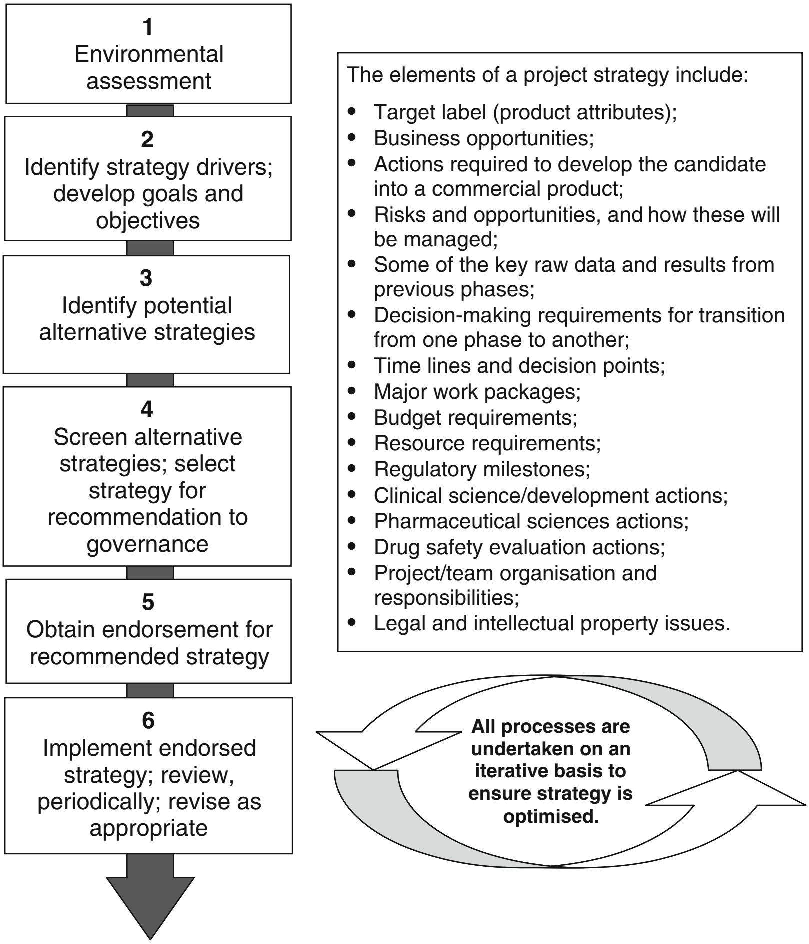


Figure 3.8 Drug development project strategy

ought to have a good grasp of the commercial possibilities – but combining the two is often a challenge); the latter is more concerned with the operational management of the project. The reasons for this split are largely historic, the differences being similar to those between a movie director and producer, as discussed in Foulkes and Morris (2004). In practice, the precise sharing of responsibilities often follows the characteristics and wishes of the two individuals filling these roles on any one project. The important thing is that the overall "management of the project" space gets properly filled on the project. Significantly, the project leader/director typically assumes a

	Management	**Leadership**
Creating an agenda	Planning and budgeting	Establishing Direction
Developing a human network for achieving the agenda	Organizing and Staffing	Aligning people
Execution	Controlling and Problem Solving	Motivating and Inspiring
Outcomes	Predictability and order; consistent delivery	Produces change

Figure 3.9 Kotter's distinctions between management and leadership

much more prominent role in shaping project strategy, although this is not always the case. The same point was made in the BAA case. The problem, once again, is that of project management being trapped in an execution-only mode. As a result, it is seen as management (subservient tactician) to strategic leadership.

The leader/manager split is reminiscent of Kotter's distinction (see Figure 3.9). Kotter's "management" dimension is exactly the "efficiency" model that underlies PMBOK®: planning, organising, controlling – i.e. "consistency" (Kotter, 2000). Yet shaping and delivering projects requires that directions be established, value optimised and opportunities created. Projects need to produce business value as well as deliver predictable outcomes. Both are needed. But whereas most project managers are happy to see themselves as efficient execution tacticians, the prize is for project managers to begin thinking and acting as though effectiveness mattered. They should be thinking about how the project, as it is developed, can enhance the value of the sponsor's strategic proposition. This is not a role just to be abdicated to programme managers. Project management competency frameworks should reflect this broader view. Learning and development, and project-based learning and knowledge management should be picking up on this too. As project management moves in this direction, so will its contribution at the senior management level become increasingly recognised.

Conclusions and summary

The effective management of projects is more than just execution-oriented "project management". Projects are undertaken to create value and deliver benefits. Shaping the interaction between the sponsor's goals and the way the project (or programme) is to be developed, in the best way possible, is absolutely crucial – probably one of the most important aspects of managing a project. This, essentially, is the project strategy challenge. Accomplishing it involves the interaction of both the sponsor's/stakeholders' and the project's strategies. Unfortunately this linkage, indeed the subject of strategy generally, is largely unrecognised in most project management literature or professional standards; it has been claimed, illogically and against the tenor of leading research, as the prerogative of programme management. It needs instead to be claimed for the whole "family" of those involved in the management of projects.

Notes

1. There are many forms and definitions of corporate and business strategy. For the purposes of this chapter, we are assuming here that the two are largely synonymous. As we shall see, there are various types of strategy within overall business and corporate strategy – technology, resourcing, Health and Safety, etc.
2. The UK's Office of Government Commerce positions programme management as being about the management of organisational change, whether specification-led technology/engineering infrastructure, business transformation, or political and societal (OGC, 2007).
3. Kuhn used the term "paradigm shift" to reflect the situation when the interest of a scientific community changes its perspective beyond what the classic model – the "exemplar piece of research" – has heretofore suggested the field to be. Most scientists work in narrowly defined aspects of the overall field. Revolutions in the way that field is seen occur when new evidence accumulates to the point that new normative models and ways of perceiving the domain are generated (Kuhn, 1962). This may arguably be happening now in project management. Thus while PMI's "standards" remain trapped in PMBOK's processes and execution culture, research journals such as the *International Journal of Project Management* and PMI's *Project Management Journal* reflect a much broader set of perspectives and paradigms (see for instance the editorials by Turner (2006) and Bredillet (2007, 2008)), as do the PMI Research Conferences and the International Project Management Association (IPMA) Congresses. In fact there has recently been an entire journal, *International Journal of Managing Projects in Business*, launched with this as its subject.
4. For example, Cooke-Davies (2004); Dvir et al. (2003); Laufer and Hoffman (2000); Lechler (1998); Munns and Bjeirmi (1996); Shenhar et al. (1997); Wateridge (1995).
5. *Ethics* (Book 1.C.4): literally: "For the beginning is thought to be more than half the whole".
6. The CIA study on cost and schedule growth on major US defence and intelligence programs is classic in its emphasis on front-end factors: "most unsuccessful programs fail at the beginning. The principal causes of growth on these large-scale programs can be traced to several causes related to overzealous advocacy, immature technology, lack of corporate technology roadmaps, requirements instability, ineffective acquisition strategy, unrealistic program baselines, inadequate systems engineering, and workforce issues" (Meier, 2008, p. 59).
7. Quote: "Every project is carried out to meet a particular business need. The question is: What assets should be built in order to meet that need? Business Front-End Loading (Business FEL) is a tool for determining which is the 'right' project to meet the needs of business. IPA's Business FEL tool assesses the level of definition of a number of critical items that are used to determine what, if any, asset should be built to meet a particular business need. The 'right' project is one that best meets the business needs and meets, or exceeds expected financial performance. The 'right' project provides an asset that fits the business strategy. We use our

Business FEL Index to help both project and business teams understand the risks associated with a capital project. The index is made up of three components: the business case, the business/engineering interface, and the conceptual engineering and facilities planning factor." (www.ipaglobal.com).

8. It showed 31% of projects were cancelled before completion; 52.7% cost 189% of original estimates; only 16.2% finished on time, in budget; completed projects had only 42% of original features; and so on (78.4% deployed with only 74.2% of original features) – figures which have to be challenged on methodological grounds, as, for example, by Jørgensen and Moløkken-Østvold (2006).
9. Efforts are currently underway to show how Agile (specifically DSDM) can and should integrate with *PRINCE2* (Richards, 2007).
10. This idea of an input from "production" into "design" as part of an integrated front-end team is the logic behind Concurrent Engineering, but should not be seen as being limited to Concurrent Engineering: the logic works for all projects. The challenge is how to organise it.
11. The higher achievement efficiency rates compared with the effectiveness rates superficially may not seem surprising. The case could be made, however, that one should be seeking higher effectiveness rates than efficiency rates, since it is more important to achieve business benefit and value for money than mere completion to an estimated future cost or schedule target. In any case, a stretch target modestly exceeded may make more business sense than an easy target comfortably achieved.
12. We obtained data from 75 members of PMI Chapters in Europe – people at various levels of seniority, in small, medium and large enterprises in a diverse range of business sectors, such as aerospace, automotive, IT, telecommunications, pharmaceuticals, retail, transportation and publishing, and academia and consultancy.
13. A recent exception is Dallas (2007) but the treatment is more from a processes view than a conceptual one.
14. "Risk: hazard, chance…, or possibility of bad consequence, loss, etc., exposure to mischance" Oxford English Dictionary.
15. Literally, the definition is referring only to "a large project split in multiple related projects" and not necessarily to all projects. But the language invites the inference that this may indeed be true for all projects. No mention is made of what happens for large projects which are not so split. And in the end the definitions become very soft, as we shall see shortly. What is a large project? What is the difference between a not-so-large project split into multiple related sub-projects? Next to none I would contend.
16. Some parts of *Managing Successful Programmes* (17.2, 17.4, Figure 17.2) however do stress the need for projects to understand how their outputs will lead to benefits, and that projects need to work with operational staff (e.g. in quality reviews) to assess whether project outputs are likely to be fit-for-purpose and lead to improvements: see Table 9.1, 12.17, Table 12.1, p. 237, 9.3.3, 18.2.2.
17. Artto et al. (2008c) have produced an interesting analysis of the usage of the terms programme and projects. Based on an analysis of 517 articles on programmes and 1164 on projects, they concluded that "when compared to projects, programs result in broader, fuzzier and more far-reaching effects with long-term implications in the future. Outcomes for projects are concrete business results which contribute in a foreseeable manner to business success" [sic]. They go on to note a number of gaps in the conceptual foundation of programmes and

their management, namely: ignorance of original theoretical roots of [particularly] large project studies [sic again]; neglect of inter-project coordination; and neglect or inter-organisational issues.

18. The on-shore "Achieving Cost reduction through Technology, Innovation and Value Engineering" industry/government (Dti) initiative.

References

Akalu M. (2001). Re-examining project appraisal and control: developing a focus on wealth creation. *International Journal of Project Management* **19** (7): 375–383.

Anderson D and Merna A. (2003). Project management strategy – project management represented as a process based set of management domains and the consequences for project management strategy. *International Journal of Project Management* **26** (3): 586–612.

Archer N and Ghasemzadeh F. (1999). An integrated framework for project portfolio selection. *International Journal of Project Management* **21** (4): 207–216.

Artto K and Dietrich P. (2004). Strategic business management through multiple projects. In: Morris PWG and Pinto J (eds). *The Wiley Guide to Managing Projects*. Wiley: Hoboken, NJ.

Artto K, Kujala J, Dietrich P, and Martinsuo M. (2008a). What is project strategy? *International Journal of Project Management* **26** (1): 4–12.

Artto K, Martinsuo M, Dietrich P, and Kujala J. (2008b). Project strategy: strategy types and their contents in innovation projects. *International Journal of Managing Projects in Business* **1** (1): 49–70.

Artto K, Martinsuo M, Gemünden H, and Murtoroa J. (2008c). Foundations of program management: a bibliometric view. *International Journal of Project Management* (in press).

Association for Project Management. (2004). *Directing Change. A Guide to Governance of Project Management*. Association for Project Management: High Wycombe.

Association for Project Management. (2006). *Body of Knowledge for Managing Projects and Programmes* (5th edition). Association for Project Management: High Wycombe.

Barney J and Clark D. (2007). *Resource-Based Theory: Creating and Sustaining Competitive Advantage*. Oxford University Press: Oxford.

Blyth A and Worthington J. (2001). *Managing the Brief for Better Design*. Spon: London.

Bossidy L, Charan R, Burck C, and Smith C. (2002). *Execution: The Discipline of Getting Things Done*. Crown Business: New York.

Boyatzis R. (1982). *The Competent Manager: A Model for Effective Performance*. John Wiley: New York.

Boyatzis R. (1993). Beyond competence: the choice to be a leader. *Human Resource Management Review* **3** (1): 1–14.

Bredillet C. (2007, 2008). Exploring research in project management: nine schools of project management research. *Project Management Journal* **38**, **39**.

Bridges W. (1996). *Managing Transitions: Making the Most of Challenges (People Skills for Professionals)*. Nicholas Brealey Publishing Ltd.: London.

British Standards Institution. (BSI) (2000). *Value Management, Standard. BS EN 12973:2000*. British Standards Institute: London.

Brown S and Eisenhardt K. (1997). The art of continuous change: linking complexity theory and time-paced evolution in relentlessly shifting organisations. *Administrative Science Quarterly* **42**:1–24.

Bryson J and Delbecq A. (1979). A contingent approach to strategy and tactics in project planning. *American Planning Association Journal* **45**:176–179.
Burgelman R and Doz Y. (2001). The power of strategic interaction. *Sloan Management Review* **42** (3): 28–38.
Cameron E and Green M. (2005). *Making Sense of Change Management: A Complete Guide to the Models, Tools and Techniques of Change Managemen*t. Kogan Page: London.
Christensen C. (1999). *Innovation and the General Manager*. Irwin McGraw Hill: Boston.
Christensen C and Raynor M. (2003). *The Innovator's Solution: Creating and Sustaining Successful Growth*. Harvard Business School Press: Cambridge, MA.
Cohen D. (2005). *The Heart of Change Field Guide*. Harvard Business School Press: Cambridge, MA.
Collis D and Montgomery C. (1995). Competing on resources: strategy in the 1990s. *Harvard Business Review* July–August: 118–128.
Cooke-Davies T. (2004). Project success. In: Morris PWG and Pinto J (eds). *The Wiley Guide to Managing Projects*. Wiley: Hoboken, NJ.
Cooper R, Edgett S, and Kleinschmidt E. (1998). *Portfolio Management for New Products*. Addison-Wesley: Reading, MA.
Cox A et al. (2003). *Supply Chain Management: A Guide to Best Practice*. Financial Times/Prentice Hall: London.
Crawford L. (2005). Senior management perceptions of project management competence. *International Journal of Project Management* **23** (1): 7–16.
Dallas M. (2007). *Value and Risk Management: A Guide to Best Practice*. Blackwell: Oxford.
Davis A, Hickey A, and Zweig A. (2004). Requirements management in the project management context. In: Morris PWG and Pinto J (eds). *The Wiley Guide to Managing Projects*. Wiley: Hoboken, NJ.
Dvir D, Raz T, and Shenhar A. (2003). An empirical analysis of the relationship between project planning and project success. *International Journal of Project Management* **21** (2): 89–95.
Eden C and Ackermann F. (1998). *Making Strategy: The Journey of Strategic Management*. Sage: London.
Eisenhardt K. (1989). Making fast strategic decisions in high-velocity environments. *Academy of Management Journal* **32** (2): 543–576.
Eisenhardt K and Martin J. (2003). Dynamic capabilities: what are they? In: *The SMS Blackwell Handbook of Organisational Capabilities*. Blackwell: Oxford.
ENAA. (2002). *P2M: A Guidebook of Project & Program Management for Enterprise Innovation: Summary Translation*. Project Management Professionals Certification Center (PMCC): Tokyo.
Englund R and Graham R. (1999). From experience: linking projects to strategy. *The Journal of Product Innovation Management* **16** (1): 52–64.
Flyvbjerg B, Bruzelius N, and Rothengatter W. (2003). *Megaprojects and Risk: An Anatomy of Ambition*. Cambridge University Press: Cambridge.
Foulkes J and Morris PWG. (2004). Pharmaceutical drug development project management. In: Morris PWG and Pinto J (eds). *The Wiley Guide to Managing Projects*. Wiley: Hoboken, NJ.
GAO. (2006). *Assessment of Selected Major Acquisition Systems*. *GAO-06-626* General Accountability Office: Washington, DC.
Gilbert X, Büchel B, and Davidson R. (2007). *Smarter Execution: Seven Steps to Getting Results*. Financial Times, Pearson: London.

Hamel G and Prahalad C. (1994). *Competing for the Future*. Harvard Business School Press: Cambridge, MA.
ICE. (1998). *RAMP: Risk Analysis and Management for Projects*. Institution of Civil Engineers and Institute of Actuaries: London.
INCOSE. (2004). *International Council on Systems Engineering Handbook*. International Council on Systems Engineering: Seattle, WA.
IPA. (2008). http://www.ipaglobal.com Accessed April 2008.
Jørgensen M and Moløkken-Østvold K. (2006). How large are software cost overruns? Critical comments on the standish group's CHAOS reports. *Information and Software Technology* **48** (4): 297–301.
Kaplan R and Norton D. (2004). *Strategy Maps: Converting Intangible Assets into Tangible Outcomes*. Harvard Business School Press: Cambridge, MA.
Kaplan R and Norton D. (2008). *The Execution Premium: Linking Strategy to Operations for Competitive Advantage*. Harvard Business School Press: Cambridge, MA.
Kotter J. (1996). *Leading Change*. Harvard Business School Press: Cambridge, MA.
Kotter J. (2000). *A Force for Change*. Free Press: New York.
Kruchten P and Kroll P. (2003). *The Rational Unified Process Made Easy: A Practitioner's Guide to the RUP*. Addison-Wesley: Reading, MA.
Kuhn T. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press: Chicago.
Laufer A and Hoffman E. (2000). The "real" success factors on projects. *International Journal of Project Management* **20** (3): 185–190.
Lechler T. (1998). When it comes to project management, it's the people that matter: an empirical analysis of project management in Germany. *IRNOP III. The Nature and Role of Projects in the Next 20 Years: Research Issues and Problems*. University of Calgary: Calgary.
Leffingwell D. (2007). *Scaling Software Agility*. Addison Wesley: Upper Saddle River, NJ.
Loch C. (2000). Tailoring product development to strategy: case of a European technology manufacturer. *European Management Journal* **18** (3): 246–258.
Luecke R. (2003). *Managing Change and Transition*. Harvard Business School Press: Cambridge, MA.
McKinlay M. (2007). Managing requirements. In: Turner J and Simister S (eds). *The Gower Handbook of Project Management* (4th edition). McGraw-Hill: Maidenhead.
Meier S. (2008). Best project management and systems engineering practices in pre-acquisition practices in the federal intelligence and defense agencies. *Project Management Journal* **39** (1): 59–71.
Miller R and Lessard D. (2001). *The Strategic Management of Large Engineering Projects: Shaping Institutions, Risks and Governance*. The MIT Press: Cambridge, MA.
Milosevic D. (1989). Systems approach to strategic project management. *International Journal of Project Management* **19**:437–443.
Milosevic D. (2002). Selecting a culturally responsive project management strategy. *Technovation* **22**:493–508.
Milosevic D and Srivannaboon S. (2006). A theoretical framework for aligning project management with business strategy. *Project Management Journal* **37** (3): 98–110.
Mintzberg H. (1987a). Crafting strategy. *Harvard Business Review* July–August **65** (3): 66–73.
Mintzberg H. (1987b). The strategy concept I: five P's for strategy. *California Management Review* **30** (1): 11–24.

Montgomery C. (2008). Putting leadership back into strategy. *Harvard Business Review* January **86** (1): 54–60.
Morris PWG. (1994). *The Management of Projects Thomas*. Telford: London.
Morris PWG. (2003). The (Ir) relevance of project management. *Proceedings of International Association for Project Management World Congress* June, Moscow. IPMA: Zurich.
Morris PWG et al. (2006). Exploring the role of formal bodies of knowledge in defining a profession – the case of project management. *International Journal of Project Management* **24**:710–721.
Morris PWG and Hough GH. (1987). *The Anatomy of Major Projects*. John Wiley & Sons: Chichester.
Morris PWG and Jamieson H. (2004). *Translating Corporate Strategy into Project Strategy*. Project Management Institute: Newtown Square, PA.
Morris PWG and Pinto J (eds). (2004). *The Wiley Guide to Managing Projects*. Wiley: Hoboken, NJ.
Munns A and Bjeirmi B. (1996). The role of project management in achieving project success. *International Journal of Project Management* **14** (2): 81–87.
OGC. (2002a). *Managing Successful Projects with PRINCE 2*. The Stationery Office: Norwich.
OGC. (2002b). *Management of Risk: Guidance for Practitioners*. The Stationery Office: London.
OGC. (2007). *Managing Successful Programmes*. The Stationery Office: Norwich.
Pendlebury J, Grouard B, and Meston F. (1998). *The Ten Keys to Successful Change Management*. John Wiley & Sons: New York.
Porter M. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press: New York.
Porter M. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press: New York.
Prahalad C. (1994). Corporate governance or corporate value added?: rethinking the primacy of shareholder value. *Journal of Applied Corporate Finance* **6** (4): 40–50.
Prasad B. (1996). *Concurrent Engineering Fundamentals*. Vols. (I and II). Prentice Hall: New York.
Project Management Institute. (2004). *A Guide to the Project Management Body of Knowledge* (3rd edition). Project Management Institute: Newtown Square, PA.
Project Management Institute. (2006a). *The Standard for Program Management*. Project Management Institute: Newtown Square, PA.
Project Management Institute. (2006b). *The Standard for Portfolio Management*. Project Management Institute: Newtown Square, PA.
Rajegopal S, McGuin P, and Waller J. (2007). *Project Portfolio Management: Leading the Corporate Vision*. Palgrave Macmillan: Basingstoke.
Reiss G et al. (2006). *The Gower Handbook of Programme Management*. Gower: Aldershot.
Richards K. (2007). *Agile Project Management: Running PRINCE2 Projects with DSDM Atern*. The Stationery Office: Norwich.
Sense A. (2008). Conceptions of learning and managing the flow of knowledge in the project-based environment. *International Journal of Managing Projects in Business* **1** (1): 33–48.
Shenhar A, Dvir D, and Levy O. (1997). Mapping the dimensions of project success. *Project Management Journal* **28**:5–13.
Simister S. (2007). Managing value. In: Turner J and Simister S (eds). *The Gower Handbook of Project Management* (4th edition). McGraw-Hill: Maidenhead.

Simon P, Hillson D, and Newland K. (1997). *Project Risk Analysis and Management (PRAM)*. Association for Project Management: High Wycombe.

Society of American Value Engineers. (1997). *Value Methodology Standard*. SAVE International: Northbrook, IL.

Srivannaboon S. (2006). Linking project management with business strategy. *Project Management Journal* **37** (5): 88–96.

Srivannaboon S and Milosevic D. (2004). The process of translating business strategy in project actions. In: Slevin D, Cleland D, and Pinto J (eds). *Innovations: Project Management Research 2004*. Project Management Institute: Newtown Square, PA, pp. 175–192.

Standish. (1994). http://www.cs.nmt.edu/~cs328/reading/Standish.pdf Accessed April 2008.

Standish. (2006). http:www.infoq.com/articles/Interview-Johnson-Standish-CHAOS. Accessed April 2008.

Stevens R, Brook P, Jackson K, and Arnold S. (1998). *Systems Engineering: Coping with Complexity*. Prentice Hall: Hemel Hempstead.

Thiry M. (2004a). Program management: a strategic decision management process. In: Morris P and Pinto J (eds). *The Wiley Guide to Managing Projects*. Wiley: Hoboken, NJ.

Thiry M. (2004b). Value management. In: Morris P and Pinto J (eds). *The Wiley Guide to Managing Projects*. Wiley: Hoboken, NJ.

Thomas J, Delisle C, and Jugdev K. (2002). *Selling Project Management to Senior Executives: Framing the Moves that Matter*. Project Management Institute: Newtown Square, PA.

Thomke S. (2003). *Experimentation Matters*. Harvard Business School Press: Cambridge, MA.

Thompson A and Strickland J. (1995). *Crafting and Implementing Strategy*. Irwin: Chicago.

Turner J. (2000). Projects for shareholder value. In: Turner J and Simister S (eds). *The Handbook of Project-Based Management* (3rd edition). McGraw-Hill: Maidenhead.

Turner J. (2006). Editorials. *International Journal of Project Management* **24**.

Wateridge J. (1995). How can IS/IT projects be measured for success? *International Journal of Project Management* **16** (1): 59–63.

Williams T. (2005). Assessing and building on project management theory in the light of badly over-run projects. *IEEE Transactions on Engineering Management* **52** (4): 497–508.

Winter M, Smith C, Morris P, and Cicmil S. (2006). Directions for future research in project management: the main findings of a UK government-funded research network. *International Journal of Project Management* **24** (8): 638–649.

Yeo K and Qiu F. (2003). The value of management flexibility – a real option approach to investment evaluation. *International Journal of Project Management* **21** (4): 243–250.

Youker R. (1993). Defining the hierarchy of objectives. In: *Proceedings of PMI 24th Annual Symposium, San Diego*. Project Management Institute: Newtown Square, PA.

4
Scenarios Planning

Kees van der Heijden

Introduction

This chapter discusses the contribution of scenario planning to project conceptualisation and definition. In essence, the methodology helps to improve judgements made under conditions of scant information. The focus is on the area of appreciative judgement (Vickers, 1995), based on making sense of diverse and often contradictory signals against a background of significant uncertainty. A general discussion of the topic is followed by case studies illustrating many of the points made.

Scenario planning and projects

Organising projects is a way of getting things done in the world. A project is an expression of a view taken of the future. However, the future is intrinsically uncertain. Projects therefore involve making a judgement about the future and involve risk, namely, that the future will be different. While this risk cannot be eliminated, it can be minimised by basing such judgement on relevant global knowledge. However, in the early stages of the project such information is scant. Scenario planning is a way of thinking about the future that aims to take account of the uncertainty in a situation, by accessing and developing the pool of knowledge, and by being upfront and specific on irreducible uncertainty. Scenario planning can be thought of as a type of research methodology.

A high degree of uncertainty leads to a high level of discounting of what will happen at a future time. A high level of future discounting makes investment difficult and slows down, or even stops, projects. Uncertainty is often associated with prevarication, and prevarication is the big enemy of project execution. That is why scenario planning does not always sit comfortably with project work once the project is underway, and commitment levels are increasing. When project implementers have made major commitments, and are in the process of building things, they do not always have much patience for someone who comes along to talk about what can go wrong (Emery and Trist, 1965).

However, at the "choosing the concept" phase of the project, scenario planning should fit naturally into the agenda. In the total life cycle of the project, scenarios belong primarily in the conceptualisation, definition and design phases, when uncertainty can and should be on the agenda as part of the process of deciding on committing time and resources to an irreversible process. Irreducible uncertainty is a part of life and part of any project. Generally it is preferable to be fully aware of the risks before launching the project. It is particularly important that there is an awareness of these in the early stages of the project, as in the later execution stages it will become increasingly difficult to allow for the fact that things may turn out differently from expected.

Scenario planning and uncertainty

What does it mean to try to account for uncertainty in the conceptualisation of a project? First is the need to increase awareness of how the future might deviate from present plans. By imaginative and creative thought, new knowledge can be tapped into, giving rise to fresh ideas of how things might develop. Of course, it is impossible to predict everything that might happen: there will always be unavoidable, unanticipated events and unexpected consequences. It is useful to consider all this. The reason is that allowing uncertainty to enter the thinking processes can give a new perspective. If uncertainty is ignored altogether, decision-making becomes a simple rational calculation process, based on the optimisation of a utility function. On the other hand, if uncertainty is consciously acknowledged, deciding becomes an ongoing process. What seems optimal today may look like a diversion tomorrow. An awareness of this produces a corresponding awareness of all the options that remain open, together with a resolution to keep these open for as long as possible.

Studying and understanding uncertainty can help more effectively distinguish between key and less important developments, enabling better anticipation and the ability to distinguish what matters from the daily background noise of the environment. This does not mean that scenario planners try to predict the future. The approach starts from the premise that there are irreducible uncertainties in the world. The aim is to understand these better, allowing genuine predetermined elements in the future to become more salient in our understanding and decision-making. Improving understanding of the business environment is a key objective of the scenario planner. This makes it easier to see predetermined elements more clearly against a background of irreducible uncertainty.

In order to do this, the scenario planner needs to be able to see and understand the whole interrelated system which gives rise to events in the relevant world. The pivotal element in this system is the human being in his/her individual manifestations and social interactions. The scenario planner is thus fundamentally a multi-disciplinary analyst. However, a lot of what the world knows about the business environment is encoded in disciplinary expert knowledge, stored in knowledge "silos". The scenarist needs to make effective connections between this expert knowledge and the specifics of the situation encountered so that it can be integrated into one overall assessment of the relevant future. A large part of the methodology that scenario planning brings to the decision-making process is related to the access, elicitation, capture and integration of expert knowledge to make it relevant to the particular business situation.

Principles of the scenario process

Scenario planning works from the outside in (Van der Heijden, 2005). The process starts in the distant contextual environment characterised by macro descriptions of amalgamated phenomena that are in constant motion in evolutionary processes. In this remote territory, individual players have disappeared in the background, and the situation is described in terms of factors. Scenario planning needs to explore these areas, as many of the driving forces that determine the project situation originate from here. Understanding the contextual environment means identifying these driving forces and developing a picture of how they mutually influence each other. The purpose of articulating and studying the contextual driving forces is to allow the exploration of relationships between these and the competitive forces acting on the players in the transactional field in which the project lies.

In a turbulent environment, these interrelations are numerous and intricate. The resulting system has many interconnections. The larger the number of interrelations, the bigger the chance that some of these constitute self-reinforcing loops that give rise to exponentially growing phenomena caused by tiny variations at sensitive points in the system. In our finite world, exponential growth cannot go on forever and is always caught somewhere along the time-line in saturation of some kind. This leads to a situation in which aspects of our environment become locked in. Prigogine calls these loops "dissipating structures" and suggests that these determine the way in which the structured world presents itself to us (Bernard, 2008). Specifically, they give rise to predetermined behaviour of great interest to the scenario planner and project developer. Understanding predetermined elements in the environment means understanding such systemic loops. Scenario planning is closely related to systems analysis.

However, there is a big difference between scenario planning and systems analysis. This is due to the fact that scenario planning, contrary to systems analysis, is essentially multi-disciplinary. Because of this, it needs a way of describing the world that has a degree of relevance and flexibility, connecting all disciplines. For this reason, scenario planning makes use of the narrative storyline as a descriptive tool. This requirement is not of particular relevance to systems descriptions, which are developed within a single disciplinary view of the world. The aim here is to develop state equations or relationships between variables that pertain over time.

Time and state domains

One can describe the behaviour of a system in the time domain or in the state domain. In the time domain, the description makes use of events and a narrative storyline to describe how the system behaves over a period of time, from a beginning to an end. In the state domain, the description uses variables, relationships and models that apply at any time. A typical example is a mathematical model based on the laws of nature, such as a climate change model. Expert knowledge makes use of such formal models (some more formal than others), the role of the expert being to develop understanding of the world in such terms.

Of course models can be used to simulate events in the world. Model builders try to develop and change the model until such a simulation reproduces actual developments over some period in history. Once this state of affairs has been achieved, the model has been "validated" and can be used to extrapolate behaviour into the future. State and time domain descriptions are not independent of each other. One can be transformed into the other, and each can be developed into an equally complete description of the system.

Knowledge development can proceed in two different ways. The first starts in the state domain by trying to develop understanding of the fundamental relationships between variables that underpin developments in the world. Once these have been hypothesised they can be put together in a model to show how events come to pass in the time domain, in this way validating (or falsifying) assumptions: this is the approach of science.

The scenario approach is essentially different in that it starts in the time domain while the state domain understanding of the situation is incomplete. The equivalent of the state domain's "theory" in the time domain is the "story". The role of the scenarist is to tell stories about developments in the relevant world. In the early stages of the process, the activity is unencumbered by the need to fully account for all underlying logic, other than in the most general cause/effect terms. This creates space for interconnecting multiple disciplines in an imaginative and creative way that, at this stage, is impossible to achieve in the state domain. Through storytelling the scenarist deals with the need for multi-disciplinary modelling and the creativity this inspires.

However, scenario planning will need to touch base with the state domain in order to tap into the pool of available expert knowledge. At this point, the scenario planner needs to work on the underlying causal logic of the scenario stories, which will give rise to an understanding of the state domain representation of the scenarios. It is here that scenario planners engage in interaction with experts to learn about the knowledge relevant to underpinning the scenario stories. These stories change, often significantly, as this process unfolds. For this reason, scenario planning is regarded as an iterative research process in which storytelling alternates with systems modelling with experts, until the understanding of the whole stabilises. The scenario process is outlined in more detail below.

The place of scenario planning in projects

Because of their emphasis on irreducible uncertainty and creativity, multidisciplinary scenarios belong primarily in the early stages of project conceptualisation and definition, when ideas are still fluid, the overall logic of the project is emerging, and minds are still open to different perspectives.

However, this does not mean that scenarios have no role to play later on. In fact the process of entering into commitments during a project's life cycle is a gradual one, and at any time there will be a balance between elements that have been pinned down and others that are still open for consideration. As long as there are questions to be resolved in the face of uncertainty, scenario planning can make a contribution. Even if its main potential reduces as things become pinned down over the life cycle of the project, it continues to offer its unique ability to bridge disciplinary boundaries and maintain a holistic perspective while interacting with the world of exper knowledge.

The scenario process

Context setting, the "business idea"

Scenario planning is most useful in the early stages of a project, when purpose and scope are still under development. The first question to be answered is why has the organisation decided to consider a project. This question relates to the overarching purpose of the organisation wishing to engage in the project and its underlying logic. Project conceptualisation is essentially the translation and extension of the organisation's business logic into one for the future.

Understanding this is particularly relevant in scenario planning, which needs a clear view of purpose. The world is very large, and everything is connected to everything else: it is easy to lose the way. Any successful scenario project requires the ability to understand and articulate the boundaries of the situation. These can only be delineated by considering the "success formula" of the organisation of which the prospective project is designed to become a part.

The tool proposed here for articulating the organisation's success formula, and its extension into a successful project, is referred to as the "Business Idea". It is a summary description of how the organisation (or project) can create value in the world, based on unique competencies and resources built by investing part of this value creation. An organisation's Business Idea is the representation of its success formula in three basic elements of specification:

- Value creation, i.e. what scarcities in the world are addressed by the products/services created and provided
- Value appropriation, i.e. what unique resources/competencies allow the organisation to provide a rare service and be rewarded for it
- Value renewal, i.e. how appropriated value is invested to ensure uniqueness in the future.

The resulting feedback loop is the fundamental force that drives the organisation forward towards increasing value creation. But it can also decline and spiral downwards, if momentum is lost and feedback loses strength. The fundamental purpose of any project reduces the chance of this happening, by reinvigorating elements of the loop and ensuring that it remains firmly in the growth mode.

A project's Business Idea is specified in similar terms, except that the present tense is replaced by the future tense:

- Value creation, i.e. which scarcities in the world will be addressed by the products/services created and provided
- Value appropriation, i.e. what unique resources/competencies will allow the project to provide a rare service and be rewarded for it
- Value renewal, i.e. how this appropriated value will be invested to ensure uniqueness in the future.

Success is fundamentally associated with distinctiveness. A distinctive resource or competence is one that cannot be emulated by competitors. This means that the only source of distinctiveness is building on already existing distinctiveness. Anything else can be copied. Assuming that a project emerges as a way to reinforce or extend an organisation's success formula, the project's Business Idea can only be based on unique aspects of the sponsoring organisation(s).

The better the organisation's business logic can be articulated, the more purposeful the project definition becomes. Unfortunately, in many cases this issue is left hanging and is not addressed to any degree of specificity. People often take the success of their organisation for granted, without feeling the need to spend time trying to articulate the "secret of its success".

Articulation of the Business Idea will help the project/scenario team to define criteria for overall project effectiveness. It can serve this purpose during the lifetime of the project but is particularly important during the concept/definition phases. In this first step, the boundaries of the project definition and, therefore, relevance are defined – a key requirement for a successful scenario project.

Case study: Example of a Business Idea for a public sector agency

A Business Idea of an organisation or project is often expressed in the form of an influence diagram, to articulate and highlight the key reinforcing loop that is the source of sustainability and ongoing success. In this example, we show such a diagram for a new public sector agency created to provide a new service for which there was a strong pent-up demand. The example focuses on the question of Value Appropriation. The other two elements of the Business Idea (Value Creation and Value Renewal) were less problematic in this case, where a new organisation, entering a field with strong demand, would continue to drive that demand into the future.

As soon as the management of the new agency was in place, they decided to spend time thinking about how they could safeguard their chances of success. The simplified diagram in Figure 4.1 summarises the result of these deliberations in the form of a Business Idea. The question of the nature of the scarcity they were alleviating was not

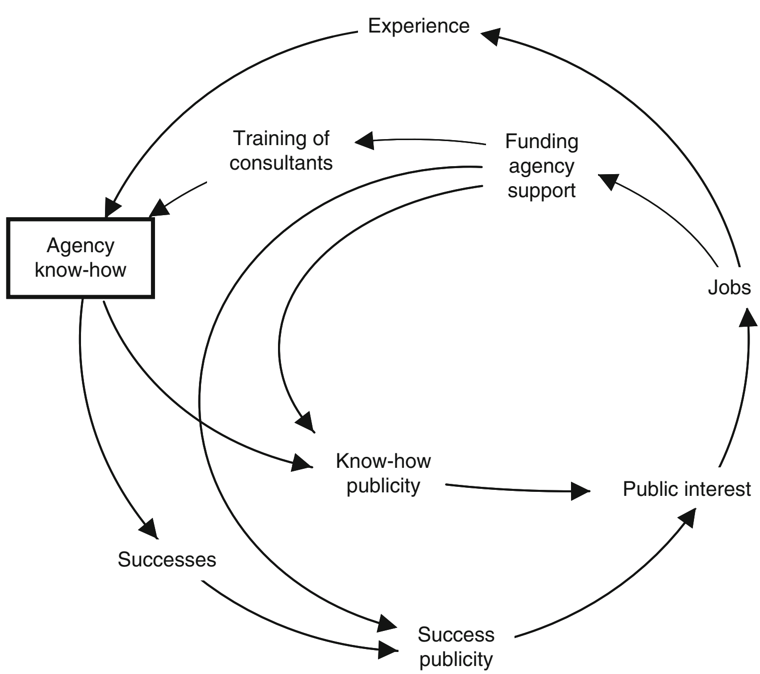


Figure 4.1 Example of a Business Idea

problematic. It was obvious why the agency had been created, and there could be no doubt that demand for this service would be strong for quite some time to come. The question of what distinctiveness this agency had to offer was less clear. Since the agency was new, it was not possible to reasonably argue that it had any distinctive competencies in the field. It was agreed that the only distinctiveness at this stage was the mandate granted by the government, who would be providing resources for a limited period of time. Since they were the only agency with this mandate, this seemed to be a unique feature of their organisation. However, the management team decided that while this might provide some strength in the near future, in the longer term they could not rely on the mandate being automatically renewed, if competing organisations came up with a superior offering. They therefore felt that the agency, rather than relying on its mandate, had to build up its own capabilities as soon as possible. This would result in a unique pool of working knowledge acquired during the short period that it had a degree of monopoly power to build the business. Growth and success would mean strong support from government, and the mandate would follow automatically. Growth would bring unique experience and know-how that, in turn, would be the basis for further growth.

Early growth became the key element around which the Business Idea was developed. The central question was how to kick-start this virtual loop around the idea of building know-how as quickly as possible. While training had a role to play, staff experience was the other key source of knowledge, and this necessitated building the business. While there was a demand for the service provided, there was also a need to build confidence among the public that this organisation was the one they wanted to do business with. It had to engage with the public. This is shown in the diagram as "know-how publicity" and "success publicity". It was decided that these were two distinct tasks, where the organisation would present itself at two different levels, in terms of its knowledge and its success story.

On the basis of this discussion, the management decided to launch three projects, "training", "know-how publicity" and "success publicity" and to give these projects priority over other developmental activity. Being aware of the reason for this priority greatly facilitated the conceptualisation and description of the projects, leading to a quick path towards design and implementation.

Conclusion: Explicit understanding of fundamental success drivers facilitates project conceptualisation and definition. The Business Idea is a helpful scenario planning tool for articulation and priority setting/management.

The strategic conversation

Scenario planning exploits the multiple perspectives that people have on a strategic situation, to give new and creative insights, through the juxtaposition and integration of different views and knowledge. Scenarios come in multiples, never singly, allowing space for alternative views. Underpinning the scenarios, and tapping into specialist knowledge for integration into holistic models of the environment in which the project will be realised, is largely a conversational process or "strategic conversation". The role of scenario planners is to design and facilitate this process.

In order to make this possible, the scenario facilitator will bring together a team of people representing the various project interests, enriched by external participants, the "scenario team", who are invited for their specific knowledge or perspective. It is the task of this team to carry out the knowledge activities that are intended to lead to new and relevant insights supporting the project design.

The scenario process consists of two types of activities which oscillate back and forth until diminishing returns indicate the approaching end of the exercise. The first task involves the scenario team in a scenario-building workshop. Within the boundaries of the Business Idea, stories of how the future might unfold are explored. The purpose of this is to identify driving forces and possible events that would be worth considering. By using general cause and effect reasoning, the team structures their insights in the form of a number of scenarios, in which the important forces and events are related to each other in internally consistent storylines. The overall scenario structure, determining how a limited set of scenarios covers the area, requires the ordering and prioritising of the main driving forces shaping the environment. Within this framework, scenarios are developed that give a narrative meaning to all events thought to be worth exploring. Internal consistency has to be born in mind by the scenario planners. Through this exercise, the group expresses their understanding of how things hang together in the business environment. Turbulent environments are characterised by the salience of such interconnections, and articulating these is an important contribution of scenario planning.

At this early stage, the result reflects the limited understanding of the team alone. The next task of scenario planning involves taking these insights into the world of experts, in order to test them against specialist knowledge and identify relevant areas that have so far been missed out. Conducting a productive conversation between project generalists and disciplinary experts is not a simple matter. Each group brings their own agenda and "language" to the conversation. Setting an agenda of topics for discussion is not easy, as neither party is yet aware of what potential contribution the other could make. In the absence of good research questions, such conversations tend to be less than productive. Scenarios make a major contribution here, in that they put one or more clear research questions on the table in the form of hypotheses. In essence, this boils down to presenting the scenarios to the experts and asking where they consider the stories to be incomplete, inconsistent, or plain wrong. During this conversation, the expert side will need to immerse themselves in the stories by exploring events and internal consistency and the underlying logic developed by the scenario team. They will then be in a position to juxtapose the scenarios with their particular expert knowledge. On the basis of this, it is now possible for them to articulate areas of contradiction and/or white space. This is what the scenario planners want to achieve. They are not there to defend their scenarios: the more contrary the information triggered, the more effective the learning exercise.

The scenario team will continue to consult a wide area of expertise within the boundaries of the project's Business Idea. The power of the scenarios is that they allow the team to access any relevant knowledge discipline, much of which will be new and unexpected.

When the scenario team returns from this research task, the situation will look very different from where they started. New insights will have been gained, not only on what might happen but also on how things hang together in the state domain. A big task now awaits the team: to order and catalogue the newly acquired knowledge, structure the systemic understanding gained, and, on the basis of this, repeat the scenario-building task. The scenarios now come into their own as a descriptive tool, providing a means of causally interrelating insights across many disciplines into one overall meaning framework. At the end of this phase, the team will have developed entirely new scenarios that now reflect a big chunk of knowledge about the area. These are known as "second generation" scenarios.

This processing of what has been learned is one of the key tasks of the scenario team. They must capture, articulate and operationalise the enhanced knowledge gained from the scenario planning activity, making it relevant to the project's decision-making process.

By putting all newly gained insights next to each other, and establishing systemic interrelationships, the team will soon become aware of gaps in their overall understanding that require further research. They may decide to engage in new discussions with the experts and their organisations, on the basis of the much better-informed second generation scenarios. The ensuing discussions will have a lot more depth, allowing the scenario team a deeper insight into the systemic relations across the working environment disciplines. At this stage, it is no longer just the scenario team which benefits, but the experts on the other side of the table often gain insights on the boundaries between their own and other areas of expertise, allowing them to see their own work in a new light.

Systems and predetermined elements

As Prigogine has suggested, the future structure of the business environment will be determined by the dense system of cross-disciplinary relationships and the reinforcing feedback loops this will create (Bernard, 2008). Being able to identify such a specific system earlier than one's competitors provides a unique anticipatory advantage to the organisation doing the scenario planning. It is obviously impossible to plan for such a breakthrough. However, the chance of having such a unique insight increases with the number of iterations, which in turn require more resource and investment.

Reporting

The end of the process approaches when the team finds that further rounds of scenario exercises will produce diminishing returns and decides to finalise the findings. Reporting the new insights to others who have not been part of the learning process is a major challenge. In order to do this effectively, the team will have to study the implications of the findings for the project that has triggered the exercise. This will require further workshops, which will ideally include members of the project management, whose decision-making will need to reflect the learning obtained. The implications workshops will take as a starting point the Business Idea, agreed at the outset of the exercise, and ask how this success formula would fare under the conditions of the various future outcomes expressed in the scenarios, with and without the project being implemented. In this way, the attractiveness of the project is assessed across the range of uncertainty pertaining.

The main purpose of the scenario planning is not to come to a yes/no decision on the project but to trigger ideas for improvement. The systemic understanding of the business environment will suggest changes to the project design that will improve performance and robustness. In particular, ideas about predetermined elements, triggered by the scenario analysis, will point the way towards a stronger and less vulnerable project.

Case study: predetermined elements in Indian agriculture

Introduction

The Indian economy is in a state of change, resulting from the country's rapid economic development and the opening-up of its markets to international competition. The agricultural sector is not isolated from this, and it seems plausible that traditional structures and ways of doing things will come under pressure. For many years, the World Bank has supported agricultural development in India, and this support continues. A recent scenario project undertaken by the World Bank aimed to explore evolving aspects of the agricultural sector in a broadly-based strategic conversation among stakeholders (Rajalahti et al., 2006).

Origins of the project

Traditionally the World Bank's support for the agricultural sector in India has taken the form of the National Agricultural Technology Project. Projects like these have been successful over the years in making India self-sufficient in food production. When this scenario project was launched, the government of India and the World Bank were involved in negotiating a new contract, the National Agricultural Innovation Project (NAIP). The purpose of the scenario project was to consider the wider context for the NAIP.

The agricultural sector in India is changing. Demand and the market are subject to changes such as population growth, increasing wealth and new dietary patterns. The industry is increasingly exposed to global market forces and international competition. The traditional regulatory regime is under increasing pressure, and new players are pushing for fundamentally different structures. In a growing sector, local constraints such as availability of land and water are increasingly felt. All these forces are linked in complex ways which are only partly understood. In these circumstances, it was felt that the NAIP might need to be different from previous government contracts, and the scenario project was intended to provide new insights which would inform the negotiations on the NAIP.

The World Bank and the government of India were the main stakeholders in the project. It was decided to include in the strategic conversation a wide range of other interests in agriculture and rural affairs, to ensure that the area would be properly covered.

The scenario project

The project was completed within 2 years, in parallel with the negotiation process of the NAIP. The procedure essentially followed the iterative approach towards scenario planning, in which categorising and interpretation of data alternates with research on the issues arising from the structure, as portrayed in Figure 4.2. When negotiations concerning the NAIP were brought to a successful conclusion, the formal part of the scenario project was also brought to an end. Since then, several stakeholders have used the results to explore future policy issues.

The core scenario-building team included some 30 participants, including private and public sector interests, farmers' union leaders, non-government organisations, donor representatives, experts on agricultural development and a number of "generalists". Their main task was to produce the first-generation scenarios. The approach was conventional and involved brainstorming for driving forces, followed by categorisation, prioritising and clustering, leading to the identification of two key scenario dimensions as the basis of a two-by-two matrix and allowing the development of storylines in each of the four resulting quadrants. The dimensions were chosen at a high conceptual level to ensure that scenarios expressed a wide-angle view

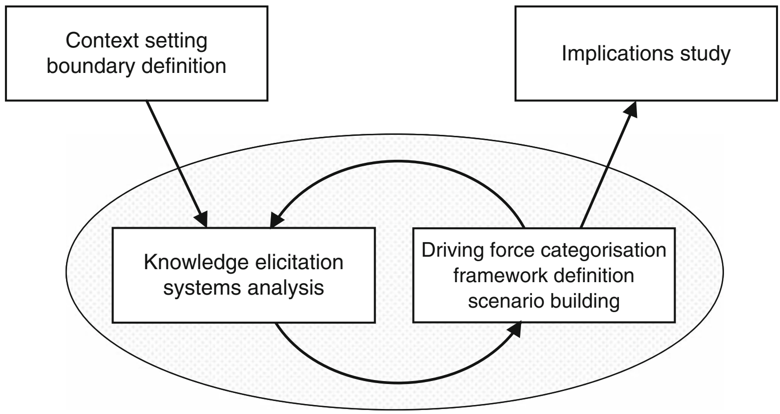


Figure 4.2 Stages in scenario planning

of the scene studied. The two-by-two matrix produced to characterise the scenarios represented economic forces ranging from "controlled" to "commercialised", and societal forces ranging from "cohesive/concern for equity" to "individualistic/personal resilience".

Participants then considered possible future events and allocated them to one of the scenarios in which they would be most salient. From these, they developed the four scenarios as storylines, providing a causal explanation in line with the logic of the relevant quadrant. In order to ensure internal consistency, causal loop influence diagrams (state domain descriptions) of the scenarios were drawn up to articulate the underlying dynamics of the stories and to highlight the key strands in each.

The research phase; emerging key issues; scale

In the iterative scenario approach, a wide range of experts and/or thoughtful individuals are invited to study the scenarios and comment on the logic presented, using their specialist knowledge and insights. These observations are carefully collected and used as the start of the new sequence of scenario-building.

During this consultation process, the team became conscious of a gradual but fundamental shift in Indian agriculture. While earlier projects had been focused on self-sufficiency, this was now becoming less relevant with the opening up of the Indian economy to global forces. Competitiveness was becoming more important than production levels, and this raised the issue of overall productivity in Indian agriculture.

Current productivity is low. McKinsey (Beardsley and Farrell, 2005) estimates that current levels are lower than 20% of those in Europe and the

USA. India's agricultural sector performs at less than half the productivity level of Brazil. This means that the sector depends heavily on government regulations to survive. India has a rather elaborate and complex set of rules regulating agriculture, built up over many years, and aimed at preventing small farmers dropping below the poverty line. McKinsey blames the poor performance of India's productivity on these "harmful government regulations". Poverty is a consequence of low productivity.

The next key point emerging from the consultation phase was low productivity being dominated by lack of operational scale and therefore of economies of scale. Farmland is privately owned in India. The average farm measures less than one and a half hectares, two thirds of farms being of one hectare or less. Year on year, average farm size is decreasing further, with some holdings being split when ownership moves from one generation to the next. The great majority of farmers are therefore poor and unable to invest in improving their productivity. With more and more farmers relying on the government for their survival, the system is stuck in a catch-22 situation, with poverty requiring regulation and regulation killing productivity and thus creating poverty.

This feedback challenged the scenario planners. While there was a dearth of ideas as to how this stalemate could be broken, most people agreed that if nothing changes in agriculture, India cannot continue with its current project of opening up to the rest of the world and taking off economically. This was considered an unpalatable fact by most commentators, driving home the conundrum facing the country.

The debate is an important one for the country, as the long-term national economic performance depends on it. As far as the NAIP is concerned, it was clear that innovation needed to refocus on the societal and economic aspect of economies of scale issues, rather than on technology, which had been the main attention of the earlier "self-sufficiency era".

Second generation

The scenario team adopted the key insights from this consultation and research phase and redesigned the scenarios to reflect the findings. The scenarios were named Valley, Edge, Mountains and Hills, metaphors of the different paths chosen towards the future. The need to improve productivity through scale enlargement as a pre-condition for economic take-off was placed centre stage and became a "predetermined element" in the scenarios. The set of four scenarios was reconfigured around this, exploring a different value set in each:

- **Valley**: justice-based ethics sacrificing wealth for equity
- **Edge**: utilitarian ethics sacrificing equity for self-reliance and personal choice

- **Mountains**: rights-based ethics centralising power in the light of major catastrophe
- **Hills**: communitarian ethics decentralising individual responsibility for the joint project.

This framework is shown against the original two scenario axes in Figure 4.3.

The resulting stories are a careful balance between the key societal and economic factors and relationships shaping the future and maintaining plausibility and social acceptability within the perceived limits of the current system. This does not offer much room for manoeuvre.

"Valley" is the scenario in which India fails to get to grips with the big issues in the agricultural scene. Justice-based ethics maintains current rigidities, scale consolidation does not happen, as a result of which productivity does not improve and economic growth fizzles out.

In contrast, the team then developed the three other scenarios to reflect three different ways in which the major scale consolidation issue is overcome, in line with the basic value set driving the storyline:

- Consolidation through market forces (utilitarian ethics) in the "Edge" scenario
- Consolidation through government dirigisme (rights-based ethics) in the "Mountains" scenario and
- Consolidation through enabling policies and investment climate (communitarian ethics) in the "Hills" scenario.

Each of these scenarios takes the storyline close to what the team felt was the limit of what would be considered plausible and acceptable in Indian

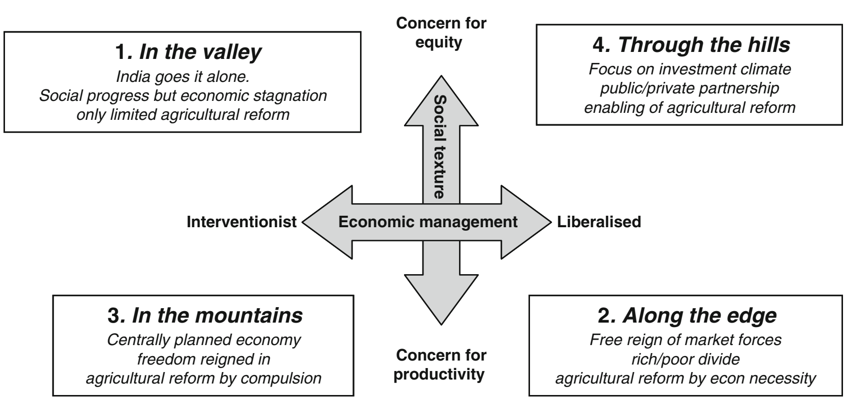


Figure 4.3 Scenario framework for Indian agriculture

society, both in terms of social equity and government intervention. These boundaries are mostly intuitive, are not well-defined and are difficult to get to grips with. People seem hard-put to venture into defining such limits without a suitable trigger. Presenting provocative stories triggers responses that provide insights as to where people stand. Listening carefully to comments on the first-generation scenarios, the team observed that there seemed to be some degree of consensus among the commentators on what were considered plausible and implausible stories.

None of these scenarios will develop in the real world. The four value systems are outcomes of possible but different evolutionary processes. Indian society will evolve its own set of institutionalised behavioural codes based on traditional Indian values but remoulded for the global environment of the 21st century. Accelerating this evolutionary process involves widening the range of what is perceived in the environment, by expressing developments and explanations in common terms, so that they become comparable. This requires an ability to look at an issue from various different perspectives simultaneously. It is not how the strategic conversation normally approaches issues, which is more akin to homing in on one specific way of viewing the situation as quickly as possible: this results in a narrowing of the observation field, leading to the locking into a single way of interpreting what is going on. This project demonstrates how scenarios can get round such a danger. The experience of the scenario group was that introduction of alternative scenario perspectives, through which various ways forward became visible, made it possible to introduce difficult issues in the strategic conversation, creating an in-depth dialogue and reaching conclusions that would otherwise be difficult to develop.

In his 2007 Independence Day speech, the Prime Minister of India cautioned his people against hubris. "India cannot become a nation with islands of high growth and vast areas untouched by development, where the benefits of growth accrue only to a few", said Mr Singh. The tone was far more cautious than the previous year. Instead of praising the erection of "vast industrial estates and special economic zones", as he did the year before, he simply advised that "we must not be over-confident". He promised new investments in rural development and pointed to the challenge of "how to industrialize and move people out of the lagging agricultural sector". He continued, "The transition from an agrarian society to an industrial economy has always been a difficult one, but industrialization offers new opportunities and hope, especially for people in rural areas displaced by agrarian change." (*New York Times*, 16th August 2007).

With the strategic conversation intermediating between a scenario project and possible outcomes, it will always be nigh impossible for scenario planners to causally trace direct or indirect results from their work. Even so, the scenario planning group here were struck by the sentiments expressed in this speech, which seemed to reflect the issues which had surfaced in the scenario conversation.

Conclusions

The various stages in scenario planning as discussed above are portrayed in Figure 4.2. It shows how the scenario-planning process can contribute to project conceptualisation and design. In summary:

Scenario planning should be seen as a methodology for analysing the business environment of a project, based on assimilating global knowledge. It is essentially multi-disciplinary. While it will contribute towards the assessment of the overall value of the project, its main aim is to improve the design and create a more robust project.

The methodology is primarily exploratory, although hypothesis-testing plays a role, where testable knowledge is elicited or developed. Scenario planning takes full account of multiple perspectives and irreducible uncertainty in any real-world project situation. The scenario planning process aims to reduce uncertainty by gaining knowledge about the business situation. It actively searches for predetermined elements in the causal systemic network in which the project is embedded.

Scenario planning has become very popular in the area of business strategy. Project managers should consider actively how it can help them, particularly in the early definition stages of a project.

References

Beardsley S and Farrell D. (2005). Regulation that is good for competition. *The McKinsey Quarterly* **2005** (2): 48–59.

Bernard M. (2008). New forms of coherence. In: *Business Planning in Turbulent Times, New Methods for Applying Scenarios*. Earthscan: London.

Emery F and Trist E. (1965). The causal texture of organizational environments. *Human Relations* **13** (1): 21–32.

Rajalahti R, Van der Heijden K, Janssen W, and Pehu E. (2006). *Scenario Planning to Guide Long-Term Investments in Agricultural Science and Technology*. The World Bank: Washington.

Sengupta, S. (2007). Economic boom fails to generate optimism in India. *New York Times*. 16th August 2007.

Van der Heijden K. (2005). *Scenarios, the Art of Strategic Conversation* (2nd edition). John Wiley & Sons, Ltd.: Chichester, UK.

Vickers G. (1995). *The Art of Judgment: A Study of Policy Making* (Centenary edition). Sage: Thousand Oaks.

5

Up-Front Assessment of Needs

Petter Næss

Introduction

This chapter discusses needs analyses in connection with front-end planning of large-scale public works projects. Relevant methods for needs analyses include normative, market-oriented and interest group based approaches. The analyses may be carried out within more or less objectivistic versus interpretative perspectives, reflecting different interpretations of the notion of need. Based on a review of different methods of analysis and experience of deficient and misleading needs analyses in large public investment projects, the chapter provides advice regarding the extent and demarcation of needs analyses, what documentation is needed, appropriate scheduling and organizational responsibility for the analyses.

A short story of exaggerated expectations

When an express railway connection to Oslo's new airport at Gardermoen was planned in the early 1990s, the need was justified by referring to a forecast showing that 60% of aeroplane passengers would use the express train between the airport and the city of Oslo. However, the traffic forecasting models focused on rail as the only public transport option, in spite of the fact that there was already a bus service along the route. Moreover, the forecasts did not consider how ticket price would influence the market share. As a result, the forecasts exaggerated the number of rail passengers by 67% and hence overestimated the need for the new railway line. Other alternatives for augmenting the public transport services between Oslo and the airport, such as improved, low-fare express bus services, were not investigated.

Needless to say, revenue from the railway line is substantially lower than what was predicted. Together with higher-than-forecast operational and construction costs, this implies that "the possibility for the rail line to become a profitable project, as presupposed to be in the Parliamentary decision on the development and financing of the airport and connecting surface transport services, can be entirely ruled out" (NOU, 1999).

Unfortunately, the above example is not a rare or anecdotal case. Experience from a number of large-scale investment projects has shown that the needs analyses on which decisions to implement the projects were based have often been insufficient and sometimes misleading.

Conceptualization

The notion of needs

Needs, biological and social, are a basic property of human beings. However, the notion of needs is diffuse and elastic. What are we actually in need of? Different disciplinary traditions conceive of the notion of needs in different ways. Within biology, the notion is used about deficits that threaten or weaken the existence of an organism. A need is said to exist when the organism is in a state that has to be changed (Østerberg, 1973), e.g. because it has a deficit or surplus of certain substances. In psychological research, it is common to consider needs as circumstances that appear to be necessary for a person's prosperity and well-being but cannot be explained physiologically (e.g. needs for stimulation and contact).

Both in biological and psychological research, the innate, static, objective and measurable nature of needs is often emphasized (Lian, 2000). Contrary to this, needs are conceived as purely subjective within economics, where market demand for commodities and services is considered an adequate indication of human needs or preferences. Preferences are considered to influence the way individuals allocate their income to different types of consumption. Distinct from this view, sociological research views needs as socio-culturally constructed (except for biological needs related to mere survival). Sociologists also distinguish between needs and wants. We may need something that we do not wish for, and we may wish for something we do not need. For example, a child may be in need of nutrition from vegetables and not from soda pops but still wishes to have soda pops for all meals and to be excused from eating vegetables. Needs in a sociological sense are related to accepted norms. A need may not necessarily be recognized by the individual in question (Lian, ibid.). Social and economic needs that cannot be met through individuals' purchase of commodities and services but are covered for society as a whole or for large groups jointly (e.g. needs for legal protection, traffic arteries, etc.) are termed collective needs.

Both wants and needs are always tied to value prioritizations – they are not value neutral. Needs evolve within certain historical and cultural contexts. There is a reciprocal influence between social development and needs. Apart from the physiological life necessities demanded by our biological constitution, prevailing opinions about essential needs are social constructs. What

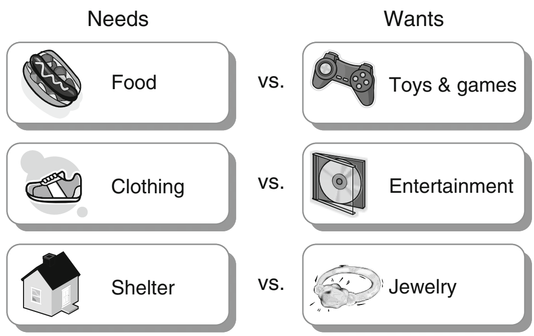


Figure 5.1 Examples of needs and wants (Based on an illustration from Monroe County Women's Disability Network, http://www.mcwdn.org/ECONOMICS/NeedWant.html)

are considered luxury needs in one historical period of time may at another point of time be considered necessary (Lian, ibid.).

Figure 5.1 shows some elementary examples of the difference between needs and wants. In practice, the distinction is often less clear-cut. For example, we may want expensive fashion dresses or suits, while what we need is just plain clothing. If we buy a fashion dress, we may satisfy a need for clothing, but this need could be met in a much less expensive way.

In connection with large-scale public work projects, the distinction between needs and wants may be much more complicated and blurred. McKillip (1987) defines a need as "the value-based finding that a population group has a problem that can be solved". The investment projects focused on in this book are physical structures and systems, e.g. transport infrastructure, building complexes or military infrastructure. These are essentially meant to respond to what is termed collective needs[1] above. The extent to which there is a need for an investment project can generally be assessed by comparing the anticipated effects of the project with the needs expressed by users and other affected groups. However, unlike many consumer goods, it is difficult to measure the effect of large-scale public works projects on the basis of individual or singular needs. This is partly due to the fact that such projects are supposed to meet collective as well as individual needs, often across several generations. Stakeholders affected by the projects frequently have different and conflicting views and priorities regarding needs (Røsnes, 1992). In practice, power relations determine which groups will have their needs catered for by society (Thomsen, 2000).

It may therefore be difficult to state with certainty to what extent there is a need for large public investment projects, or to make objective comparisons between alternative projects. The needs may be expressed indirectly through demands in the market, but, as mentioned above, needs and wants do not necessarily concur. People may not have the economic ability or purchasing power to express their needs in the form of market demand. Needs may also be expressed by initiatives from public authorities or as wishes put forward by different groups in society.

Needs at different levels

A distinction is commonly made between primary (basic) and secondary (acquired or derived) needs. Needs may be ranked hierarchically, where the most basic needs (physiological needs for food, drinking water and protection against pain, cold and heat) must be met first (Assiter and Noonan, 2007; Maslow, 1962; WCED, 1987). Needs at different levels in the hierarchy may be linked. For many people, the satisfaction of a need high up in the chain (e.g. happiness) depends on the presence of certain pre-conditions (e.g. material security, love and social esteem/contact). Obtaining the latter states may be considered "means" or "solutions" in order to meet the higher need of happiness. Material security, love and social esteem/contact may also be considered as needs in their own right but at a somewhat lower level in the hierarchy than the need for happiness. In order to meet the need for material security, employment may be a means or a solution. It may also have its own place in the hierarchy of needs.

Similarly, "improved transport between A and B" may be a need at one level, whereas "higher road standards between A and B" is a need at a lower level.[2] Typically, needs at higher levels can be met in different ways, where each of these means/solutions may appear as needs at a lower level. Thus, a ramification and specification of needs occurs when moving downward in the hierarchy. The less generally a need is defined, the stronger the ties established towards specific types of solutions.

Early-stage needs analysis versus needs analyses at later stages of a project

In order to ensure that investment projects fulfil high-priority societal needs, which may differ from those within a specific sector, or among the proponents of a certain technical solution, the analysis should not start at too low a level in the chain of needs. It must be carried out at the *concept* or *strategic* level, i.e. at the early stage of project planning, before decision-making on the choice of a concept solution. There may also be a requirement for more detailed needs analysis at later stages of project planning, but this cannot substitute the needs analysis required at the early stage of the project. If the needs analysis is conducted too late, the solution concept will be chosen without sufficient clarification of the needs existing in the situation from which the proposal for the investment project emanated. The needs analysis will then be reduced to an analysis of the demand for a certain technical solution (and any initiatives to remedy its side effects).

Interdependence between needs analysis, goal formulation and impact assessment

Needs analysis, goal formulation and impact assessment are separate activities but are mutually dependent. When there is a need for a project, the implication is that the project will have certain effects, and highlight particular needs. When formulating goals, a decision must be taken as to which needs among those identified should be fulfilled. Equally, goal formulation pre-supposes that a relevant initiative or project can be undertaken, with the desired effects. Moreover, goal formulations and identified needs have a bearing on the selection of the types of consequences on which the impact assessment should focus. These conditions underline the necessity of tight connections between needs analyses, goal formulations and impact assessments in project planning.

An essential feature in large-scale investment projects is their justification in relation to national political objectives within the sectors to which the projects belong, as well as across sectors. The political objectives indicate which needs to prioritize among all those that might be identified in a situation. In order to achieve the objectives, and meet the underlying needs, a range of conceivable solutions may be relevant.

Project-triggering needs, and needs related to the side effects of projects

In large investment projects, a distinguishing feature is between the need(s) forming the main motivation for a proposed project and other needs that also ought to be taken into consideration if the project is realized. For example, in a project to establish a new railway line, the immediate justification – or "project-triggering need" – might be to establish a fast and direct train connection between two cities, A and B. If this was the only need to consider, the best solution would be to build the track along a straight line between the two. However, if the shortest connection cuts across a natural recreation area of importance to local inhabitants and visiting tourists, such a solution would not be the best one (unless a tunnel was built under the natural area), because it would violate the need to keep the area free of technological encroachment. However, railway connections to minor settlements located between cities A and B might offer new employment opportunities for residents of these settlements (commuting to the cities would be easier) and also make these locations more attractive for businesses. If such regional policy considerations were defined as part of the need, the need fulfilment would perhaps be highest, if the line passed through these settlements. This would, however, be at the expense of the wish for the shortest possible travel time between cities A and B.

The example shows that initiatives arising from a given need may have positive as well as negative side effects. In a societal perspective, it will be necessary to obtain the largest possible positive, and the smallest possible negative side effects, while at the same time meeting the need that was originally used to justify the project. It is a matter of balancing between different needs, where the most favourable solution for society will often be different from that based solely on the "project-triggering" need.

Moreover, different needs will often be associated with different stakeholder groups. In the above example, the fastest possible train connection between A and B will serve those who travel the full distance between the two cities (including passengers getting on or off at stations beyond this route). On the other hand, the need for train stops between A and B will primarily be felt by those who live in, or have errands to, these local communities. The need for avoiding encroachment into the natural area will apply particularly to environmentalists and outdoor recreation enthusiasts. Outdoor recreation visitors may, at the same time, have a need for easy access to the area by rail. If a station is built close to the area, without the railway line intersecting it, both these needs may be met.

In this example, the project-triggering need was defined in relation to a specific mode of travel, i.e. rail. The need can, however, be defined at a more general level, e.g. in terms of a lower *friction of distance* between town A and town B. If the need is defined in this way, alternative solutions to a railway (e.g. an express bus line or maybe a highway) might contribute to meet the need.

Table 5.1 shows examples of project-triggering needs, together with needs related to side effects for three different concept solutions in the energy sector. The table does not offer any complete overview of relevant solution concepts nor of needs related to side effects. For example, demand-oriented interventions (e.g. initiatives to improve energy efficiency or reduce excessive electricity consumption) should also be considered as possible solution concepts in a situation where the use of electricity threatens to exceed the existing supply capacity if current trends of growth continue.

Apart from the frequent occurrence of different, often competing and conflicting needs related to public investment projects, some derived needs may also appear. Needs for interventions or installations in order for the proposed project to operate in a functional way, and to meet safety and environmental requirements, are examples of this category. In the above-mentioned case, there might be a derived need for a tunnel under the natural area, if it is considered politically unacceptable to let the railway track fragment it, and if the increase in travel time resulting from the track circumventing the area is also rejected. It is important to identify such derived needs in order to avoid too narrow a definition of the project. The likelihood of cost overrun and delays at the stage of implementation is high, if cost estimates and work schedules are based on a "stripped" project, later requiring additional grants in order to be fully functional or to meet safety and environmental standards.

Table 5.1 Some examples of needs that may manifest themselves in connection with large-scale energy supply projects, subdivided into project-triggering needs and needs in connection with positive and negative side effects.

Concept-level solution	Hydro-electric development	Development of gas-fired power plants	Development of windmill parks
Project-triggering need	• Need for an adequate level of energy supply	• Need for an adequate level of energy supply	• Need for an adequate level of energy supply
Needs related to the realization of positive side effects	• Need for realizing potential regional economic gains • Need for replacing/preventing electricity production based on fossil fuels	• Need for realizing potential regional economic gains and establishing permanent gas-related industrial jobs • Need for utilizing surplus heat for district heating • Need for replacing electricity import from more polluting coal or oil-fired power plants	• Need for realizing potential regional economic gains • Need for replacing/preventing electricity production based on fossil fuels
Needs related to the minimizing of negative side effects	• Need for avoiding/limiting: flooding of buildings due to dam construction, encroachments on valuable natural and outdoor recreation areas, negative visual landscape effects, encroachments on cultural heritage	• Need for avoiding/limiting: CO_2 emissions and other pollution from the production process, encroachments on valuable natural and outdoor recreation areas, negative visual landscape effects, encroachments on cultural heritage	• Need for avoiding/limiting: negative visual landscape effects, noise in inhabited areas

Methodology

Needs analyses may be based on different interpretations of the notion of need. They may also be carried out within more or less objectivistic versus interpretative perspectives (Hiebert and Smallwood, 1987). Methods for needs analyses can be classified into three main categories – *normative, market-oriented* and *interest group-oriented*.

Normative analyses are based on political objectives or experts' definitions of appropriate levels of service or performance. The use of quantitative norms related to specific technical solutions also belongs to this approach. *Market-oriented* analyses aim at measuring the demand for a planned facility, either in the form of income from tickets, user fees, etc., or from the number of users (e.g. the number of vehicles on a new road). Such analyses often use model simulations, perhaps based on investigations of willingness-to-pay. An alternative kind of demand-oriented needs analysis is the reference class prognosis method (cf. below). *Interest group orientated* analyses involve the collection of information about the needs of different interest groups in connection with a set of problems. These may be public authorities, groups within business life or groups within civil society.

Seen from a planner's perspective, needs analyses aim to identify needs that can be met by potential initiatives (McKillip, 1987). An analysis should therefore not focus on needs that the planning agency cannot be expected to meet. We do not analyze the "need" for sunshine on Midsummer Eve! That said, it is important to be aware that the availability of policy instruments is largely a political question. Needs that are considered to be politically important may induce politicians to accept policy instruments they would normally be reluctant to use.

The situation in which planning of large-scale governmental investment projects takes place is often characterized by ambiguity, dynamic environments, lack of clarity about the content and unclear means, ends and boundaries of the analysis. In particular, available information will usually be limited at the early stage of planning. Carrying out the needs analysis as a closed and purely "objectivistic" optimizing process will be ill-placed in this context. A more open and interpretative needs analysis would be more appropriate, involving various interest groups and utilizing alternative forms of knowledge together with the expert knowledge of the authorities. In order to balance between the needs and interests of different groups, these must be made known, and the various groups must therefore be given the chance to express their needs. Relevant citizen participation methods in connection with needs analyses include focus group discussions, nominal group sessions, public meetings, and public hearings (McKillip, 1987). However, the involvement of different interest groups does not prevent formal analyses based on logic, scientific rationality and the measurable from being included as part of the broad interpretative analysis (Hiebert and Smallwood, 1987).

Brinkerhoff (1986) distinguishes between four different definitions of needs, each giving rise to different analytical approaches:

- *Need understood as deviation* between the actual situation and an ideal, normative or expected standard. Using this definition makes possible a measurable, precise and specific identification of needs. The deviation-based need definition requires that both the present and the desired standards are measurable and that a desirable standard has been set in advance, i.e. something against which to benchmark.
- *"Democratic"*[3] *needs definition*. Needs are defined here as what most people prefer, choose or "vote for". Participating in the process of need definition invites acceptance and commitment and may thus contribute to stronger efforts being directed towards satisfying these needs.
- *Diagnostic needs* are defined through research and causal analyses. When such analyses show a causal relationship, e.g. that A contributes to obtaining B, then A is needed for B. The strength of this need definition is that it can discover "true" needs and show their validity.
- *"Analytic" needs* are discovered through intuition, insight, expert consideration or enlightenment. Because this need definition is not constrained by previously defined standards, majority opinions or established knowledge, "analytically" deduced needs may be ground-breaking and set new agendas.

Table 5.2 is an attempt to classify the four previously mentioned main approaches (demand analysis, the reference class prognosis method, normative needs analysis and interest group based needs analysis) in relation to Brinkerhoff's four need definitions and Hiebert and Smallwood's objectivistic – interpretative dimension.

In *deviation-based* needs analyses, the need for intervention is deduced from the deviation between the norm and the actual level of need satisfaction. This approach requires a normative need assessment that may be relevant in situations where publicly adopted norms for need satisfaction exist. There is a paternalist element in this approach. The norms set within different sectors of society have often been developed and defined by professional experts within their respective fields. The politicians subsequently adopting these norms may have limited insight into their justification and the impact their implementation would have on sectors of society outside the fields in question. This may also be beyond the knowledge of the experts who formulated the norms – or they may deliberately choose to ignore it. It is not uncommon among experts in a particular subject to consider the satisfaction of needs to be more important within their own field than in other sectors of society. If such norms are adopted without the politicians being able to understand the consequences, in terms of resource consumption or

Table 5.2 Main approaches to needs analyses, classified according to conception of needs and analytical perspectives.

Conception of needs	**Analytical perspective**	
	Objectivistic perspective	**Interpretative perspective**
Deviation-based	• Normative need analysis	—
"Democratic"	• Demand analysis (incl. the reference class prognosis method)	• Interest group based need analysis
Diagnostic	• Normative need analysis • Interest group based need analysis	• Normative need analysis • Interest group based need analysis
"Analytical"	—	• Interest group based need analysis

conflicts related to other societal interests, they may be used to legitimize the investment in specific projects. In so doing, they may also hinder politicians from making real prioritizations between different needs. Large-scale public works projects affect a broad spectrum of interests and concerns, not all of which can depend on quantified, adopted norms for need satisfaction.

The so-called *democratic* need definition represent a highly demand analyses or other forms of mapping preferences among the population. The pitfall of this approach is that needs may be confused with wishes, and societal needs about which there is little knowledge in the population may accordingly be ignored. Demand analyses belong to a quantitative, objectivistic tradition. In many cases, it is important to estimate the number of users of a proposed public facility (e.g. a new road link or public transport service). Such estimates are necessary to determine the size of the investment project in a reasonable way. Within the transport sector, traffic forecasting, using transport modelling tools, is widely used in such analyses.

Demand analyses have, however, shown a tendency to focus one-sidedly on the project-triggering needs, ignoring those connected with possible side effects of a project. There are also important methodological problems associated with this approach in situations where political goals have been adopted, aiming at a development different from the one indicated by present trends. Within the transport sector, the forecasting tools most commonly used have been criticized for grossly underestimating the traffic-generating effect of road capacity increases in congested urban areas (Kenworthy, 1990; Næss, 2006; Nielsen and Fosgerau, 2005), thus giving the misleading impression that road building in such areas can solve congestion problems. For rail projects and other project types where a high number of passengers/users are considered politically desirable, demand analyses often overestimate the actual demand (Flyvbjerg et al., 2003). The so-called reference class prognosis method does not elucidate the demand for a new project "from within" based on characteristics of the specific project but, instead, from a comparison of the project in question with a reference class of similar projects (Flyvbjerg et al., 2005). This reduces the scope for over-optimistic forecasts, but not even this method pays attention to possible side effects. Mapping wishes among the population may also be carried out by means of more qualitative and interpretative approaches. Interest group based needs analyses are usually conducted as qualitative studies based on a "democratic" definition of needs.

Distinct from the deviation-based conception of needs, the *diagnostic* need definition includes needs emanating from lacks as well as needs in order to sustain the strengths of a current situation in the future. It thereby avoids some of the pitfalls of needs analyses focusing solely on shortcomings. Diagnostic needs analyses make use of available scientific knowledge about the possible impacts of large-scale public works projects (and hence about any needs affected positively or negatively). Needs analyses based on this notion are likely to use objectivistic methods but may also be carried out within an interpretative approach. A diagnostic approach is usefully employed in normative needs analyses, taking publicly adopted goals as the point of departure. One of the advantages of this approach is that it focuses naturally on both the project-triggering needs and also on needs connected to side effects. Analyses based on a diagnostic notion of needs may also be used to identify important "stakeholders" in an interest group based needs analysis. Here, knowledge about causal relationships is used to identify need for initiatives that may satisfy the needs of different population groups. Interest group based needs analyses within a diagnostic perspective presuppose a close dialogue between professional experts and the involved interest groups.

Analyses based on the need definition termed by Brinkerhoff as "*analytic*" paves the way for an interpretative approach. Given its high degree of subjectivity, intuition and exploration of needs, in dialogue between participants, it is difficult to imagine this approach forming the basis for normative needs analyses, taking publicly adopted objectives as their point of departure. For interest group based needs analyses, such a combination of needs definition and analytical approach might sometimes be fruitful. It may, however, be difficult to establish the validity of the "analytically" deduced needs and justify that priority should be given to them. It is therefore hard to imagine that needs analyses rooted mainly in this perspective may form the basis for prioritization and decision-making about billion-Euro government investment projects.

Backcasting versus forecasting

Needs analyses often include predictions for the future. The investments made in order to meet needs ought not to become insufficient or superfluous a few years after their implementation. This future orientation may take two essentially different forms. The analyst may produce a trend prolongation and use it as a basis for assessing the need for a certain type of facility in, say, 15 years time. Available statistics and social indicators will offer relevant input data for an extrapolation of the development during recent years. This may, among others, envisage how the composition of population groups, with varying needs for a type of service or facility, is likely to develop within the geographical area covered by the analysis (McKillip, 1987).

However, a trend-based development is not always considered desirable. If public authorities have at their disposal policy instruments that may contribute to change a present, undesirable development, it would be inappropriate to base the needs analysis on a continuation of this undesirable development. Instead, the needs analysis may take as its departure a desirable future situation and "calculate backward" to the initiatives that should be implemented today, in order to start a development leading to the desired future situation. This latter technique is termed "backcasting" (Dreborg, 1996), as distinct from making forecasts based on observed traits of development. Backcasting is a declared normative method that presupposes political willingness and ability to intervene in the development. Forecasting involves adapting to the development rather than trying to change it. However, needs analyses based on trend prolongation are no more neutral than analyses presupposing trend breaks. The values supported by trend-adapting analyses are those supported by the present development and gaining benefits from its continuation.

Needs analyses within the transport sector may serve as an illustration of the difference between the two approaches. Whereas most European countries until the early 1990s attempted to adapt to expected traffic growth in metropolitan areas by increasing road capacity ("predict and provide"), several countries (including the UK and Norway) now aim at curbing this growth (Hine and Preston, 2003; Norwegian Ministry of the Environment, 1993; St.meld, 2001–2002). Within the latter approach, a projected continuation of metropolitan traffic growth would not be interpreted as an indication of a need for further road construction but rather as an indication of a need for intervention, e.g. road pricing or improved public transport, to prevent the forecast from coming true ("predict and prevent").

Forecasting the future situation from present traits of development, with subsequent identification of the need for projects in order to adapt to the expected development, is a typical approach within demand analyses. Whereas these are reactive, needs analyses based on backcasting are more proactive. Backcasting-based needs analyses may also make use of predictions based on present traits of development. Such a combined approach was employed in the Swedish transport research project "Travels in the Future" (Steen et al., 1997), which included both backcasting and forecasting and used a diagnostic (identifying causal factors influencing the development of transport) and a deviation-based notion of needs (identifying the deviation between an environmentally sustainable and a trend-based level of future CO_2 emissions).

Some of the problems associated with demand-based needs analyses (notably over-optimistic assessments of the demand) may be avoided by comparing the proposed project with a class of comparable projects, instead of calculating the demand "from within". However, in some cases it may be desirable to establish a different context for the project than the one applying to the reference projects (e.g. road pricing influencing the field of competition between personal car use and public transport). In such cases, the number of projects from similar contexts will often be too low to allow for statistical comparison. In order to apply the reference class prognosis method in such a situation, a so-called "analytic generalization" (Yin, 1994) from the reference projects to the proposed project must be made, using counterfactual reasoning to assess what the demand among the reference projects might have been if their contexts had been similar to the proposed project. This requires knowledge of cause-and-effect relationships between, respectively, the elements of the project context that deviate from the reference projects and the demand. Usually, only the project-triggering needs can be compared directly by means of the reference class prognosis method. In order to incorporate needs connected with side effects, and differences between population groups, an additional judgement about the geographical and social context of the proposed investment project must be made. The need for intervention in connection with side effects must be assessed using scientific and professional knowledge about causal relationships between different characteristics of the project in question and other categories that may be affected.

There is much to be said for combining different analytical methods and perspectives within the same needs analysis. In particular, this applies to projects of the magnitude dealt with in this book. By combining several methods, the deficiencies and weaknesses of one method may be compensated by the strengths of an alternative method.

Deficiencies in current practice

Experience from a number of large-scale investment projects has shown that the needs analyses on which decisions to implement the projects were based were often insufficient and even misleading. Sometimes there had been no publicly available needs analysis at all. Formal needs analysis are frequently absent in the early stage of planning. Research literature on deficiencies and distortions in needs analyses, in the context of large-scale government investment projects, shows that this is a serious and widespread phenomenon. The problem is partly due to defective qualitative surveying of the various needs occurring in the situation which a proposed investment project is supposed to improve and partly to misleading quantifying of the demand for a chosen project concept.

Defective qualitative surveying of needs

When needs analyses narrow the identification of needs down to the market demand for a specific solution, while ignoring the broader spectrum of societal needs that might justify, or be affected by, an investment project, there is a danger that the needs of the originators of the project are confused with society's needs. The wishes of narrow interest groups for economic benefits, prestige or ideologically preferred solutions might take precedence over national political objectives and the needs of broader groups in society.[4]

Lack of distinction between normatively/socially defined needs and market demand is especially problematic in situations where politically adopted goals seek a development different from that indicated by present trends and market demand. Depicting one particular technical solution as the need shows that the needs analyses were carried out at project level instead of at a strategic planning level, legitimizing a chosen solution, instead of exploring the project-relevant needs existing in the situation and identifying possible ways of meeting these needs. Missing or distorted assessment of needs in connection with side effects occurs when the needs analysis is reduced to an analysis of the demand for a proposed solution, and in cases where assumed positive side effects are included as part of the justification for a project, while ignoring negative side effects. Neglecting to survey and assess the needs of different stakeholders/population groups may result in undesirable distribution of benefits and burdens, protests and implementation problems. Delayed identification of needs for additional investment necessary for the installation to function in an effective or environmentally defensible way frequently leads to cost overrun and delayed completion.

Defective surveying of project-relevant needs sometimes results from lack of knowledge and narrow horizons among the professionals performing the analysis. If the needs analyses are only performed by professional specialists within a particular sector, there is a risk that side effects and alternative solutions will be ignored. Lack of cross-sector integration in public administration may be another cause of narrow needs analyses. The sectors represent different "cultures" in terms of dominating values, attitudes and opinions about which needs are the most important. Pressure from other sectors of society urging an agency to downsize traditional tasks may be perceived by the employees as a threat against their jobs. Disagreement over adopted objectives is another related cause of narrowness in needs analyses. If the professionals running the needs analysis believe that the decision-making bodies do not care much about some objectives affected by an investment project, they may deem it unnecessary to include needs related to these objectives in the analysis. Finally, there may be contested knowledge about the relationship between intervention and need fulfilment. Knowledge which is counter-intuitive, or is perceived as threatening to general lifestyles and customs, runs the risk of being rejected in policy-making. In practice, power often determines what will be accepted as truth.

Biased quantification

Misleading quantification of the demand for the chosen solution[5] usually implies that the need for a proposed project is exaggerated, either through an overestimation of the demand for the project itself, or by drawing too negative a picture of the consequences of suppressed demand in case of non-realization of the project. Overly optimistic demand analyses have been documented for railway projects, especially urban rail, but also occur quite frequently in connection with investment in roads or tourist facilities in remote areas. There are also examples of underestimation of demand in situations where growth is not considered desirable. This has occurred in connection with proposed road investment in urban areas, where reduction of car traffic is a goal. In addition, misleading analyses of what will happen if a proposed investment project is not implemented have sometimes given the impression that the proposed solution is necessary, in order to preclude a future situation few would wish for.

Professional insufficiency (defective methods of analysis, poor databases etc.) and unforeseen events may partially explain inaccuracy of quantitative need assessments. However, the various types of errors and deficiencies show a clearly non-random pattern. As a rule, they function to support a certain conclusion: that investment in the proposed technical solution is desirable and necessary. An alternative explanation of the widespread tendency of forecasts to give a more positive picture of projects than what turns out to be the case, may simply be a human tendency to wishful thinking. However, an analysis by Flyvbjerg et al. (2005) of opinions among project managers and evaluation researchers about causes of forecast inaccuracy and distortion in transport infrastructure projects indicates that biased assessments made by consultants or project promoters may be an important cause of this inaccuracy, especially for rail projects. The material supports a suspicion that deliberate distortion of the analysis in order to make the project appear in a favourable light may be a frequently occurring phenomenon. This conclusion is underpinned by comparison with American (Wachs, 1989), Norwegian (NOU, 1999) and Danish examples (Næss, 2006).

The suspicion that deliberate distortions are made in order to have projects implemented that would otherwise not gain political support, is supported by the fact that exaggerated forecasts of the number of rail passengers is often accompanied by grossly underrated construction cost estimates. For road projects, the deviation between forecasting and actual traffic shows a more complicated pattern. However, a comparison of traffic forecasting for Danish road projects commenced before 1980, with similar projects in the period 1980–1990, hints at strategically underestimated traffic forecasts in the 1980s, when Denmark placed stronger emphasis on reducing gasoline consumption than most other European countries (Næss et al., 2006).

Model computations are influenced by the individuals who construct the models and carry out the analyses, and the background, knowledge and attitudes of these individuals. The consultants often integrate their own political preferences into the framework of the forecasts. Project promoters often have an interest – economic or prestige-related – in presenting the project in as favourable a light as possible. In a situation where there are few sanctions against those who make inaccurate forecasts, accurate projections may be counter-productive, because the project runs the risk that other projects bolstered by over-optimistic analyses will be preferred (Flyvbjerg et al., 2003).

Many of the above-mentioned examples are from large transport investment projects. This mirrors the fact that research into the quality of need and demand analyses has frequently focused on this sector. Transport is a sector of society where political goals often point in a direction different from the trend development, and where a large proportion of projects have an institutional anchoring that may direct the focus towards a particular type of solution (new or extended roads), instead of the transport/accessibility needs to which road construction is but one among several possible answers. However, also in sectors where the political goals do not imply that current trends should be broken, and with less extensive project side effects, needs analyses confining themselves to project-triggering needs, or market demand, run the risk of overlooking important and relevant needs. Exaggerated demand analyses may, independent of project type, lead to misplaced investment, which might have been avoided, if the scaling of the project were based on more realistic assessments of the demand.

Conclusions and recommendations

Any needs analysis presupposes that assumptions are made about the effects of possible interventions. Often, the effects in question are changed activity patterns among potential users of the planned facilities. However, theoretical and empirical knowledge about such effects is often incomplete and nearly always context-dependent. When using such knowledge in needs analyses, it has to be interpreted, adapted and modified according to the situation at hand. Moreover, in addition to the impacts of the planned investment project, a number of other factors usually also influence the satisfaction of the needs addressed by the project. The relative influence of these factors may change as society changes. The context after the implementation of a large-scale public works project will necessarily be different from the situation at the time when the project idea was conceived. Together, these circumstances imply a considerable and unavoidable uncertainty about the level of need satisfaction after the realization of a large government investment project (Næss, 2004b). This uncertainty will be high at the early stage of planning, where the details of the project have not yet been designed and the available information is less than it will be at later stages of the planning process. This has implications in terms of appropriate levels of measurement in early-stage assessments of needs for large-scale investment projects. More specifically, the reasonableness of using sophisticated mathematical prediction models in such analyses could be questioned.

Based on the review of methods of analysis, and the experience of deficient and misleading needs analyses, the following guidelines are recommended regarding the extent and demarcation of needs analyses, requirements on the documentation material, appropriate scheduling and organizational responsibility.

Extent and demarcation

The needs analysis must cover all relevant societal needs, not only those expressed as willingness-to-pay. In particular, needs incorporated into national-level politically adopted government objectives must be given due consideration. Needs analyses at the early planning stage of large-scale government investment projects must be carried out at a strategic level, not at a project level. This implies that the analysis must focus on a higher level in the hierarchy of needs than the more narrowly defined needs directly tied to a particular type of technical solution. The needs analysis must not only elucidate the "project-triggering" needs but also those connected with negative and positive side effects.

In a situation where, for instance, a hydro-electric development project has been proposed, the analysis of the need for such a project must consider the need for an adequate level of electricity supply, the need for realizing potential environmental benefits (employment, renewable energy etc.), as well as the need to avoid or reduce negative side effects on natural recreation areas, cultural heritage sites etc. This implies that collective needs expressed in the political priorities of several ministries (notably the ministries of Petroleum and Energy, Local Government and Regional Development, Environment, and Agriculture and Food), as well as affected regional and local authorities, must be taken into account, along with consumer demand for electricity. The needs of population groups particularly affected must be identified. The needs analysis should therefore be carried out as a process involving a broad spectrum of affected parties.

The documentation material

Different methods of analysing and different perspectives should be combined in a needs analysis. In this way, the deficiencies and weaknesses of one method may be compensated by the strengths of an alternative method. The methods chosen should together enable identification of all relevant needs – "project-triggering" needs, as well as needs in connection with side effects – and assessment of their importance and relevance to different population groups.

Different needs should be measured at appropriate levels. Needs that can reasonably be expressed in economic terms may be "translated" into monetary units in order to facilitate subsequent economic analyses. When making such economic assessments of needs, due attention must be directed towards the fact that considerable uncertainty and disagreement usually exists about the assumptions on which the calculations are based. This should be exposed by indicating uncertainty margins around the estimates.

The importance of needs that cannot reasonably be expressed in numbers or monetary terms may be indicated on a crude scale. Needs already quantified and/or assessed in monetary units may preferentially be "translated" to this scale in presentations of the main conclusions of the needs analysis.[6] This makes it easier to compare between needs valuated in economic terms, those measured in other quantitative entities and those impossible to quantify. Table 5.3 shows schematically how an assessment scheme as indicated above might be designed.

Because of the societal nature of needs connected with large-scale government investment projects, the needs analysis should reflect general government objectives within relevant sectors of society – not only the sectors from which the "project-triggering" needs emanate but also sectors that might be positively or negatively affected by side effects. Such a normative approach should be combined with the involvement of different interest groups, including different sectors (and possibly levels of administration) within the civil service, as well as groups in civil society and business life. The needs analysis should not be based on pre-formulated quantitative technical standards.

In many cases it will be necessary to prepare forecasts about the future use of a large investment project, e.g. a railway line. Provided that data about comparable projects are available, the reference prognosis method should be preferred to forecasts based on an "inside" assessment of the technical features of the project and possible circumstances that may affect the demand.

In needs analyses taking higher-level government goals as their point of departure, forecasts based on present trends should not stand alone. By comparing the future situation resulting from present traits of development within a policy area with a normatively desirable situation, any

Table 5.3 Schematic example of a presentation table showing different project-relevant needs in connection with large-scale public works projects.

Needs	Target group A	Target group B	Target group C	Target group D	Total
Project-triggering needs:					
Project-triggering need 1	**	***	**	*	**
Project-triggering need 2	***	*	****	**	***
Etc.					
Needs in connection with positive side effects:					
Positive side effect-related need 1	*	**	*	*	**
Positive side effect-related need 2	**	*	**	*	**
Etc.					
Needs in connection with negative side effects:					
Negative side effect-related need 1	**	*	**	****	***
Negative side effect-related need 2	****	***	**	***	***
Etc.					

The importance of the need for the target group in question, and totals are indicated by asterisks: **** = highly significant need, *** = considerable need, ** = some need, * = insignificant or no need.

need for intervention to change the current development may be identified ("backcasting").

The considerable costs that transport modelling represents in many planning processes make the possibility of replacing such model computations with qualitative assessments of relevant factors that may contribute to increase or reduce traffic well worth considering. Transport modelling should only be included in needs analyses of large-scale government investment projects if the built-in assumptions of the model have been quality-controlled by independent experts from subject fields covering a wider range than that of the model makers.

Scheduling and organizational responsibility

The main needs analysis must be carried out at the early stage of project planning, before decision-making on the choice of a concept solution. There may also be a need for more detailed needs analyses at later stages of project planning, but these analyses cannot substitute those required at the early stage of the project.

The analysis should be conducted by an agency providing as "neutral a turf" as possible, e.g. a secretariat appointed by the affected ministries, with a broad, interdisciplinary staff. Public hearings, citizen juries, etc. should be organized to make it possible for special interest groups and civil society to express criticism or support. The analysis should be subject to independent scientific evaluation (peer review). Professional sanctions should be enforced against planners and forecasters who repeatedly produce seriously misleading predictions.

Notes

1. But not exclusively – the ambitions and wishes of individuals to realize projects that may serve as monuments to their work may also play a role. The same applies to the wishes of involved agents for economic gain.
2. Both needs in the example are situated some way down in a hierarchy of needs. For example, "good accessibility to workplaces and service facilities" would be at a higher level than "better transport opportunities between A and B".
3. The notion "democratic" here is based on an individual-oriented, liberal tradition of democracy, widespread in North America, as distinct from the European concept of democracy, which emphasizes the realization of political objectives that politicians, through democratic elections, have been given the task to promote.
4. See Næss (2004a, 2006) for a more elaborate account.
5. See Næss (2004a, 2006); Flyvbjerg et al. (2005); Næss et al. (2006).
6. This of course does not preclude quantified and valorized data from being utilized, when appropriate, in more detailed analyses and presentations.

References

Assiter A and Noonan J. (2007). Human needs: a realist perspective. *Journal of Critical Realism* **6**: 173–198.

Brinkerhoff R. (1986). Expanding needs analysis. *Training and Development Journal* **40**: 64–65.

Dreborg K. (1996). Essence of backcasting. *Futures* **28** (9): 813–828.

Flyvbjerg B, Bruzelius N, and Rothengatter W. (2003). *Megaprojects and Risk. An Anatomy of Ambition*. Cambridge University Press: Cambridge.

Flyvbjerg B, Holm M, and Buhl S. (2005). How (in)accurate are demand forecasts in public works projects? The case of transportation. *Journal of the American Planning Association* **71**: 131–146.

Hiebert M and Smallwood W. (1987). Now for a completely different look at needs analysis. *Training and Development Journal* **41**: 75–79.

Hine J and Preston J. (2003). Introductory overview. In: Hine I and Preston J (eds). *Integrated Futures and Transport Choices*. Ashgate: Aldershot, pp. 1–9.

Kenworthy J. (1990). Don't shoot me – I'm only the transport planner. In: Newman P and Kenworthy J (eds). *Transport Energy Conservation*. Murdoch University: Perth.

Lian O. (2000). En sosiologisk forståelse av behov. (A sociological understanding of needs.) *Tidsskrift for Den Norske Lægeforening* **120**: s11–s114.

Maslow A. (1962). *Toward a Psychology of Being*. Van Nostrand: New York.

McKillip J. (1987). *Need Analysis: Tools for the Human Services and Education*. In: *Applied Social Research Methods Series*. Vol. 10. Sage Publications: Newbury Park/Beverly Hills/London/New Delhi.

Ministry of the Environment (Norway). (1993). *National Planning Provisions for Coordinated Land Use and Transport Planning*. St.meld. nr. 26 (2001–2002): Oslo.

Næss P. (2004a). *Bedre Behovsanalyser (Better Needs Analyses)*. NIBR-rapport 2004:15. Norwegian Institute for Urban and Regional Research: Oslo.

Næss P. (2004b). Prediction, regressions and critical realism. *Journal of Critical Realism* **3**: 133–164.

Næss P. (2006). Cost-benefit analyses of transportation investments: neither critical nor realistic. *Journal of Critical Realism* **5**: 32–60.

Næss P, Flyvbjerg B, and Buhl S. (2006). Do road planners produce more "honest numbers" than rail planners? An analysis of accuracy in road-traffic forecasts in cities vs. peripheral regions. *Transport Reviews* **26**: 537–555.

Nielsen O and Fosgerau M. (2005). *Overvurderes Tidsbenefit Af Vejprojekter? (Are Time Benefits from Road Projects being Overestimated?)*. Paper for the conference *Trafikdage på Aalborg Universitet*, 22–23rd August 2005.

NOU. (1999:28). *Gardermoprosjektet. Evaluering av planlegging og gjennomføring (The Gardermoen Project. Evaluation of Planning and Implementation)*. Report from a committee appointed through a Royal Resolution of May 15, 1998. http://odin.dep.no/sd/norsk/publ/utredninger/NOU/028005-020002/hov001-nu.html

Østerberg D. (1973). *Behov (Needs)*. Article in the *Pax* encyclopedia. http://lotus.uib.no/norgeslexi/paxlex/alfabetet/b/b06.html

Røsnes A. (ed.) (1992). *Fysisk planlegging. Byer-tettsteder-spredtbygde områder (Physical Planning. Cities, Villages and Rural Areas)*. Kommuneforlaget: Oslo.

St.meld. nr. 26 (2001–2002). *Bedre kollektivtransport (Better Public Transport)*. Ministry of Transport: Oslo. http://odin.dep.no/sd/norsk/publ/stmeld/028001-040007/dok-bn.html

Steen P et al. (1997). *Färder i framtiden – Transporter i ett bärkraftigt samhälle (Transport in a Sustainable Society)*. KFB-Rapport 1997:7. Kommunikationsforskningsberedningen: Stockholm.

Thomsen J. (2000). *Magt og indflydelse (Power and Influence)*. Magtudredningen: Århus.

Wachs M. (1989). When planners lie with numbers. *Journal of American Planning Association* **55** (4): 476–479.

WCED. (1987). *Our Common Future*. Oxford University Press: Oxford/New York.

Yin R. (1994). *Case Study Research: Design and Methods* (2nd edition). In: *Applied Social Research Methods Series*. Vol. 5. Sage Publications: Thousand Oaks/London/New Delhi.

6
Front-end Alignment of Projects – Doing the Right Project

Terry Cooke-Davies

Introduction

Before starting significant work on a project or programme, its goals and objectives need to be made explicit and assessed to ensure that, if accomplished; they will further the sponsoring organization's chosen corporate strategy and contribute to its overall goals. Unfortunately, this is not always done and can result in significant underperformance compared to expectations. Challenges are created by the different nature of the tasks facing both the organization and the project, by the variety and differing worldviews of those involved and by the multiplicity of roles necessary to achieve front-end alignment. These challenges can be overcome through judicious use of a well-founded business case and by the skilful execution of a well-formulated engagement programme.

There is growing consensus among practitioners of project, programme and portfolio management that aligning projects to business strategy is a good thing to do. This is hardly surprising, since projects and programmes are means to achieve some end, rather than being simply ends in themselves. These "ends" are inevitably related to the strategic intentions of those involved in promoting and executing the projects.

Both programme management (APM, 2007; PMI, 2006a) and portfolio management (PMI, 2006b) are advanced as concepts to assist with the alignment of projects with business strategy. They are quick to describe and advocate generic processes to implement these concepts efficiently and effectively.

If alignment is important, it is important irrespective of the particular management approach that is adopted. This chapter will pose a number of questions that need to be answered for any organization undertaking a project, using any standards and will seek to provide helpful answers. These answers should be consistent with both accepted practitioner standards and results of current research and also allow a reflective practitioner to reach his/her own conclusions, independently of any given management approach.

The questions to be answered are as follows:

- What exactly does "front-end alignment" mean? Why is it challenging?
- To whom is front-end alignment important?
- What are the roles of people who play a part in front-end alignment?
- How can front-end alignment be achieved?

What exactly does "front-end alignment" mean? Why is it challenging?

Purpose and strategy

Every organization in every field of human endeavour exists for some purpose, irrespective of whether the purpose is explicit or implicit, conscious or completely outside the awareness of the people who constitute it.

This is not to say that people in an organization always share a common purpose or strategy – that is patently not the case. But by whatever process of accommodation, the organization will always *de facto* pursue some strategic intent to accomplish some purpose.

Among the more conscious and explicit purposes are those of commercial enterprises, competing in their chosen industry or market, in order to make money for their shareholders, serve society and provide satisfying employment for their staff. For example:

- Pharmaceutical organizations seek to develop, manufacture and market innovative medicines and medical devices;
- Transport companies seek to develop, manufacture and market transportation products and services;
- Telecommunications companies seek to develop, manufacture and market communications services.

The list is an extensive one and is not limited to commercial organizations in the private sector, even if public and third sector organizations do not share the common goal of making money for their shareholders, for example:

- Government agencies seek to develop and operate services to segments of the community, in order to further the legitimate business of government;
- Charitable organizations seek to finance and provide particular services for the good of society;
- Clubs and voluntary organizations seek to carry out their objectives for the benefit of their members or other beneficiaries.

Just as each organization exists for a particular purpose, so does it inevitably seek to accomplish its purpose through adopting some form of strategy, either explicitly or implicitly, and either inside or outside the awareness of its people. "Every firm competing in every industry has a competitive strategy, whether explicit or implicit" (Porter, 1980).

Two classes of activity

Accomplishing this strategy requires two different classes of activity to be carried out in parallel with each other (Figure 6.1).

The first is the day-to-day operation of the particular activity by which the organization accomplishes its purposes. Using the six examples already provided:

- Pharmaceutical organizations manufacture and distribute licensed medicines and devices to hospitals, clinics and patients;
- Transport companies move people and freight from place to place by road, rail, air or sea;
- Telecommunications companies manufacture and sell telephones and communication equipment and operate and maintain networks by which subscribers can communicate with each other;
- Government agencies collect taxes, pay benefits, issue licences, provide healthcare, educate children and adults and police society, etc.;
- Charitable organizations provide aid to disaster-hit communities, support the victims of particular diseases and other misfortune and raise and keep account of funds, etc.;
- Clubs and voluntary organizations support all aspects of membership and the specific activities they pursue, etc.

	Projects & programs	Business as usual
Organization	Temporary	Permanent
Challenge	Innovation: realizing envisioned future state.	Adaption: improving known current state.

Figure 6.1 Two classes of activity necessary to implement strategy

Generically, the sum of these activities that are the means by which each organization accomplishes its purposes on a day-by-day basis can be referred to as "business as usual". The theory and practice of managing these activities is the main focus of the discipline of "management" – not only general management but also the sub-disciplines of operations management, marketing management, finance management, human resource management, etc.

The second class of activity comprises those efforts that are undertaken to improve an organization's ability to accomplish its purpose. If it is to survive and flourish in the continually changing word in which we live, each organization needs to undertake efforts that are designed to:

- Improve the performance of "business as usual" activities;
- Develop new businesses, new products, new services or new markets;
- Introduce new technology, new processes or new ways of working;
- Build new physical infrastructure, acquire new assets, etc.

The sum of these efforts comprises the practices and disciplines associated with managing projects, programmes and project portfolios. The theory and practice that has developed through these efforts has been the main focus of, initially, project management, but, more recently, programme and project portfolio management.

What distinguishes this class of activity (projects and programmes) from the first (business as usual) is that it involves planning and then creating some product or service that, at the point of inception, exists only in the imagination of the person or people who are promoting it. The process of planning these activities, therefore, involves imagining a series of steps that may or may not work out as planned, each of which may have unforeseen consequences. It could be described as a process of "enfolding an envisaged future into a known present", and labelled as "innovation".

In contrast, business as usual is already known and experienced and thus forms a predictable basis from which to plan variations that can generally be relied upon to deliver the required improvement. This can conveniently be labelled "adaptation".

A second distinction between the two classes of activities lies in the nature of the organization that undertakes these activities. For all but the most short-lived or routine of adaptations, it is normal practice for an organization to create a "project team" or "programme organization", provided that the innovation is sufficiently important to require its own dedicated resources and structure for management. Such a team, existing solely for the purpose of accomplishing the particular activity, is by its nature temporary, regardless of whether the people who make up the team are employees of the permanent organization, employees of a supplier organization or self-employed contractors hired specifically for the duration of the project or programme (Turner and Müller, 2002). Business-as-usual activities, on the other hand, are, by their very nature, at the heart of the permanent organization that is seeking to accomplish its specific purpose through the strategy that it has adopted.

The nature of front-end alignment

These two different but complementary classes of activity represent the means by which strategy is accomplished and suggest two progressions. The first is that the permanent organization's strategy should be translated into the optimum mix of activity spent on business as usual and on the right projects and programmes (as discussed in Part 3 of this book). The second is that each of these projects and programmes should be fully aligned to the organization's strategy. Before starting significant work beyond simply assessing the feasibility of a project or programme, the goals and objectives of the temporary organization should be made explicit and assessed to ensure that, if accomplished, they will further the chosen strategy and contribute to the organization's overall goals. This is what is meant by "front-end alignment".

Recent research indicates that this is indeed the case in normal practice (Teague and Cooke-Davies, 2007) and is important not only to the organization but to each individual project as well. The same research, however, shows that, in nearly a half of the organizations surveyed, projects were not selected according to how they matched the total portfolio and strategic objectives, and less than half of all projects were approved on the basis of a well-founded business case linking the benefits of the project to explicit organizational/financial goals.

Why might this be? What are the challenges that prevent the front-end alignment of each project and programme? There are two distinct areas that provide some clues to the answer.

Differing worldviews

The first of these areas is the different worldviews of those people who have made each of the two classes of activity their particular specialization – general managers (and their specialized colleagues) and project managers. It has been argued that their respective worldviews are so different that communication between the two groups is problematical (Dinsmore and Cooke-Davies, 2005). It has also been shown that the systems and structures necessary to manage temporary organizations differ from those for permanent organizations (Turner and Keegan, 1999).

The second of these areas relates more closely to the topic of this book: limitations in the way the human brain works and how this relates to the difference between innovation and adaptation. Extensive research by Michael

Kirton over a period of more than 30 years has drawn attention to the different psychological profiles of people who are more attuned to solving problems using Adaptation or "Adaption" rather than innovation and vice versa (Kirton, 2003). There is even a psychometric instrument (the Kirton Adaption/Innovation Indicator, or KAI) that can be used to demonstrate the extent of preference for one or the other in any individual and that has been shown to correlate well to results obtained using other psychometric instruments, such as the MBTI (Higgins and Cougar, 1995).

The cognitive challenges faced by the human brain in contemplating and planning for innovations are discussed in Part 4 of this book.

Working in quite separate fields of study, Ralph Stacey (2003) and René Girard (1986) have each drawn attention to the socially determined nature of the individual, while Tom Wright (1997) has demonstrated the importance of worldview in forming, and being formed by, social customs.

It is quite feasible, therefore, when considering these two different avenues of exploration (worldviews and the relation of the individual to the social), to assert that groups of people in temporary organizations charged with the task of innovation are likely both to make decisions and to share practices that differ considerably from those charged with the task of adaptation in permanent organizations. This brings significant challenges to the task of front-end alignment between each project or programme, and to the strategy of the organization that is sponsoring it.

To whom is front-end alignment important?

Up to this point, the distinction has been drawn between the permanent organization and the project or programme as temporary organization. The explicit assumption has been that the project or programme is undertaken as an innovative activity to improve the permanent organization's ability to accomplish its purpose, in alignment with its chosen strategy.

On this basis, it is the permanent organization that is responsible both for choosing the strategy for accomplishing its purpose and also for ensuring that it invests resources only in projects and programmes that are aligned with this strategy. However, this does not always translate into practice. On closer examination, it turns out that there are challenges related specifically to the permanent organization, as well as those that can be anticipated within the project or programme itself.

Challenges related to the permanent organization

The difficulties of front-end alignment within the permanent organization are due to the fact that it consists of many individuals and groups of individuals, each of whom may well have a different understanding of, or commitment to, different aspects of the chosen strategy.

In proposing a specific method of mapping "stakeholders" in a project, Graham Winch (2004) identified the following roles within the organization that commissioned the failed TAURUS project for the London Stock Exchange:

- Client
- Sponsor
- Financiers
- Client's employees
- Client's customers
- Client's tenants
- Client's suppliers.

Each of these groups (and the variations within them) will have had their own understanding of the strategy, and therefore of the scope and objectives, of any project or programme to be aligned to it. This places great importance on the permanent organization's processes for both strategy formulation and project portfolio management, which can be effective only if all significant voices are taken into account, and the permanent organization is itself aligned behind both the chosen strategy and its translation into projects and programmes.

It also places a considerable onus on the project or programme team during the feasibility and conceptual stages to ensure that the plan reflects all the effort that will be necessary to work with the significant stakeholders in the permanent organization, who have both a high level of interest in the outcome of the project, and the power to influence this. Where the project or programme is undertaken by a number of organizations working together as an alliance, the challenges multiply, since the front-end alignment applies to groups of stakeholders in each of the organizations in the alliance.

Challenges related to the temporary organization

In his article about TAURUS referred to above, Graham Winch (2004) also identifies stakeholders involved with the "supply" side or the temporary organization. He lists:

- Consulting engineers
- Principal contractors
- Trade contractors
- Materials suppliers
- Employees of the above.

In many circumstances, this list can be extended considerably, as prime contractors and sub-contractors add both complications to the plans and complexity to the number and nature of human interfaces involved.

Since suppliers are usually concerned with their own strategies, involving not only the financial success or failure of the specific contract but their own corporate goals, any of these groups of stakeholders might contribute to a front-end misalignment, either through the potential for misunderstanding that is inherent in all communications between people, or through pursuing goals of their own, which only partially align with those of the project or programme.

What are the roles of people who play a part in front-end alignment?

Given the importance of front-end alignment, these challenges need to be overcome, and the remainder of this chapter will move from considering challenges to examining mechanisms for overcoming them. Three parties have roles which provide the key to effective front-end alignment: portfolio management, sponsor, and project or programme manager.

Portfolio management

The role described here as "portfolio management" refers to the group of people in the permanent organization who are responsible for approving or turning down requests that a specific project or programme should receive funds and/or other resources. This may be formalized into a hierarchy of governance committees, as it often is in pharmaceutical R&D organizations or major energy companies; or, in the case of highly devolved organizations, it may be decided by individual business units.

Whatever the organizational structure and style, what is important is that the process by which resources are allocated is effective in supporting only those projects and programmes that are fully aligned with the organizational strategy. Such processes are described by many authors in this field (Archer and Ghasemzadeh, 2004; Artto et al., 2001; Cleland and King, 1988; Cooper et al., 2001), but all agree that preconditions for a successful process include clarity about the organizational strategy and a well-founded business case for the project or programme.

Sponsor

As a senior executive with a foot in both the permanent and the temporary organization, the role of the sponsor has been under increasing scrutiny (Crawford and Cooke-Davies, 2005). In a forthcoming research monograph (Crawford et al., 2008), in-depth interviews with both sponsors and project managers on 32 projects in nine organizations highlighted inter alia the critical role that sponsorship plays in front-end alignment. It is a complex role that is often shared between several individuals and committees. It also straddles the two organizations (permanent and temporary) and is therefore uniquely placed to interpret each to the other.

No role occupies a better position to sense when potential front-end misalignment exists, regardless of whether this is due to the project or programme not being in line with the business strategy, or where there is an internal misalignment within the permanent organization. Unfortunately, it is a role the significance of which is not fully appreciated by large segments of all three sectors – private, public and voluntary. An emphasis on greater proficiency in the sponsorship role prior to projects receiving resources would undoubtedly help to improve front-end alignment.

Project or programme manager

The third of these pivotal roles is that of the project or programme manager, supported by his or her team. There is a great deal of work to be done at the front-end of the project, in order to be satisfied that

- the permanent organization's strategy has been expressed in the benefits that are to be delivered;
- the benefits to be delivered have been correctly expressed in the product or service specifications;
- the scope of work contained in the project plan is both necessary and sufficient to deliver the desired product or service in the most appropriate manner, after considering all sensible alternative approaches;
- the required resources to carry out the work have been correctly identified.

How can front-end alignment be brought about?

In the face of these challenges, and in the light of the multiplicity of groups of stakeholders, how can front-end alignment be accomplished? Two mechanisms have been developed for this purpose – the creation of a well-founded business case and the execution of an engagement programme.

Creating a well-founded business case

The artefact that provides the focus for all decisions concerning front-end alignment is the business case. This term can be used to denote a document that is produced at the beginning of a project to secure the necessary funds or to provide the complete rationale for allocating scarce resources to a particular project in the light of the permanent organization's strategic goals.

Used in either sense, the business case is critical for front-end alignment. Figure 6.7 below shows that for the particular 311 projects considered in Human Systems' research, the average gain in efficiency (ascertained by traditional measures of time, cost, scope, quality and safety) is 14% if the business case is well-founded, unlike those projects where it is largely, partially or not at all well-founded.

The business case is the sponsor's primary means of communicating the needs of the permanent organization to the project and vice versa. A "well-founded" business case, in this instance, is one that is capable of being understood by all, is acceptable to all stakeholders in the permanent organization, can be agreed upon by all members of the temporary team and is internally consistent.

To be internally consistent, it needs to be considered in conjunction with the project strategy and the scope of work included in the project plan so as to answer three fundamental questions about the project:

- Why are we doing the project? (Business case)
- How are we planning to implement the project? (Project strategy)
- What do we have to do to deliver the project and its benefits? (Project scope)

Only when all three questions have been answered satisfactorily is the project ready to be taken through the portfolio management process for final approval and for completion of the front-end alignment.

Skilfully executing a thoroughly prepared engagement programme

An "engagement programme" is the name given to a formal process of consultation with groups of stakeholders during the initial stages of a project, for the purpose of engaging their active support. Such activity is best conducted once the shape of the business case is understood so as to ensure that what they are being asked to engage with is the actual project that is to be undertaken. However, it should be before all the final details have been worked out and approved by the portfolio committee so as to ensure that the legitimate concerns and viewpoints of the stakeholders can be reflected in the final scope of the project.

The design of such a programme calls for an awareness of group dynamics and the structure of facilitative meetings, as well as political awareness of the factors at play among the different groups of stakeholders with a legitimate voice in the permanent organization's strategy. The stages in the design and execution of such a programme are typically as follows.

Programme design

This involves agreeing how many meetings need to be held with which groups of stakeholders, and designing each meeting so as to maximize the chances of either winning "hearts and minds" in support of the project, or of eliciting information that is in the possession of the attendees and is essential to the construction of a well-formed business case. The planning needs to take account of the number and diversity of attendees, their likely frame of mind at the start of the meeting, the information that they will need to be given before and during the meeting, the desired outcomes from the meeting and the "meeting process", in terms of what subjects will constitute the agenda and how each subject is best presented. Both the identification of a suitable venue and the appropriate mechanism for the issue of invitations should also be taken into account.

Meeting preparation

Once the programme has been designed, each meeting should be meticulously prepared, with the venue set up in the right physical format, and with appropriate refreshments to create an atmosphere conducive to the meeting purpose, with all input materials prepared and with meeting roles clarified in advance. Since facilitation plays such an important role in meetings of this sort, it is as well to make formal arrangements for facilitation of both plenary discussions and any small group discussions that are planned. It is useful to have a facilitator who is neither the project manager nor the sponsor, to allow these two important roles to enter into a dialogue with the stakeholders who will be present.

Meeting conduct

Each meeting represents a significant opportunity to influence and communicate with the particular group of stakeholders that has been invited to attend. Thus, if events do not work out as planned, it is important to be able to modify the meeting "on the wing" so as not to lose the opportunity. This calls for the facilitator and the project manager (who should be present at all such meetings) to be sufficiently flexible to modify either the process of the meeting or the agenda, in order to improve the likelihood of obtaining the desired outcomes. Resources should also be provided to capture the outcome from each meeting for the benefit of both the project team and the stakeholders.

Programme review

On the assumption that genuine dialogue is taking place between the project manager and the stakeholders, fresh information will emerge from each meeting that could influence the design of the engagement programme and/or the nature of the business case. In the light of this, the engagement programme should be reviewed after the first few meetings and suitable changes made if appropriate.

The active and supportive engagement of stakeholders with the project is likely to remain a major goal, together with the production of a well-formed business case. However, the means by which this goal is to be achieved may change from those envisioned at the outset. After all, engagement programmes, like the projects that they exist to align, are planned with "scant information", and thus need to be skilfully steered towards their desired goal.

The importance of front-end alignment

Evidence from a piece of empirical research conducted by Human Systems International (Teague and Cooke-Davies, 2007) suggests that the front-end alignment of projects with organizational strategy is of critical importance to both an organization's "agility" and to the success of the project as measured by traditional measures. The basis to this piece of research has not yet been fully published but warrants a somewhat fuller account here than has hitherto been provided.

The author's own empirical research (Cooke-Davies, 2001), combined with a desk review of the literature on project success factors for inclusion in the Wiley Guide to Managing Projects (Morris and Pinto, 2004), led to the conclusion that project success can be assessed at three different levels: project management success (was the project done right?), project success (was the right project done?) and consistent project success (were the right projects done right, time after time?) Published research has identified a number of different factors that could contribute to each of these levels of success and also the criteria by which this success could be assessed.

Taking this work as a basis, the author and his colleagues developed a web-based questionnaire that investigated both the extent to which each of the factors was present in any given organization (consistent project success), programme or project (project success), or project (project management success). The questions about each level of success were asked only of people who could reasonably be expected to have reliable knowledge about both the factors that were present and the degree of success achieved. Senior business managers, and those responsible for project management practices and processes throughout an organization, could answer all three sets of questions; project sponsors and programme managers could answer questions about project success and project management success; and project managers or team members could answer questions only about project management success. The data is continually being added to, but, at the time of writing, contains 350 sets of data in total: 46 about all three levels, 134 about project success and project management success and 170 only about project management success. Not everyone questioned answered all questions.

The two potential project success factors that have given rise to the conclusion that front-end alignment of projects matter are as follows:

#3Projects are selected based on their fit with the total portfolio and its strategic objectives [Alignment to Strategy];
#56Each project is approved on the basis of a well-founded business case, linking the benefits of the project to explicit organizational/financial goals [Business Case].

The first of these two factors was asked at the organizational level (of senior managers concerning consistent project success), while the second was asked at both the governance/sponsorship level (of project success) and also at the project level (of project management success). Respondents were asked to signify the extent to which the above statements were true of their organization, programme or project, using a 4-point Likert scale that ranged from completely true to completely false. Scores were then assigned of 1, 2, 3 or 4 (4 = completely true), and the numerical score was taken as a measure of the presence or absence of the specific factor (or, as it was referred to in the questionnaire, "capability").

Distributions of the answers to the two selected questions are shown in Figures 6.2 and 6.3 below.

At each level, performance was asked about a series of specific success criteria relevant to the level, and the mean of all these results

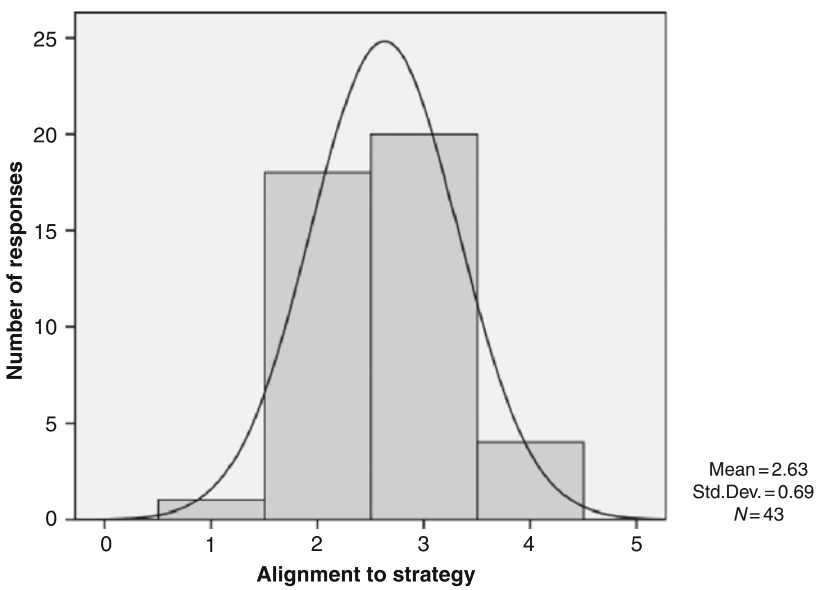


Figure 6.2 Range of responses to Question #3

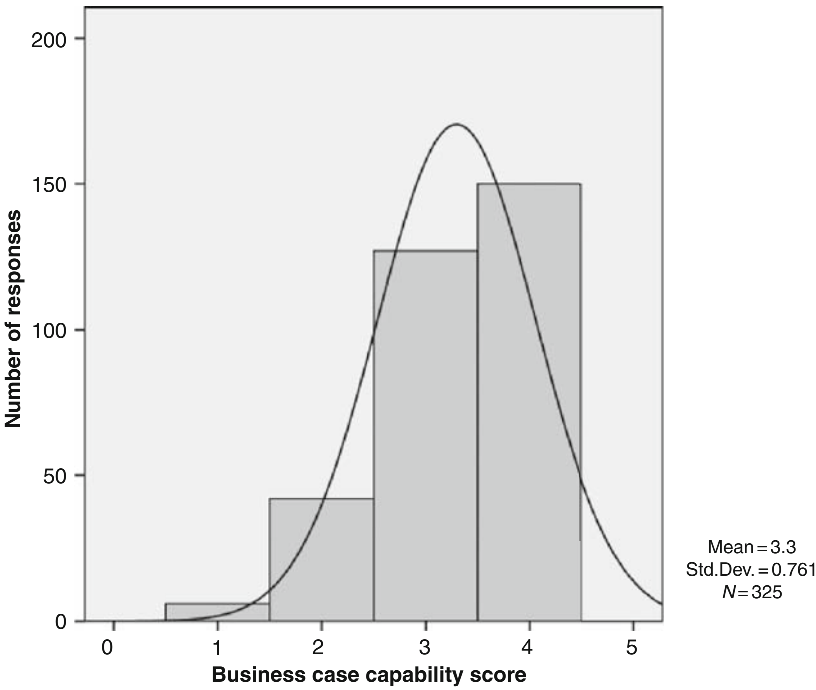


Figure 6.3 Range of responses to Question #56

was combined into a single measure known, respectively, as "agility" (consistent project success), "effectiveness" (project success) and "efficiency" (project management success). The measure for each of the individual criteria was chosen for its appropriateness to the specific criterion, e.g. percentage of benefits delivered, percentage variation from budget or customer satisfaction. In order to provide some comparability across these, the question always asked for a response within a specific range, whereby a score of 3 was equivalent to a result as anticipated, a score of 4 was more than 10% better than anticipated, a score of 2 was more than 10% worse than anticipated and a score of 1 was more than 25% worse than anticipated. This is somewhat crude but allows the results to be viewed for either single criteria or for overall success at a coherent level.

The ranges of results for agility and efficiency, respectively, are shown in Figures 6.4 and 6.5.

The ranges of results for effectiveness (project success) are not shown, because no correlation to questions #3 or #56 were found to

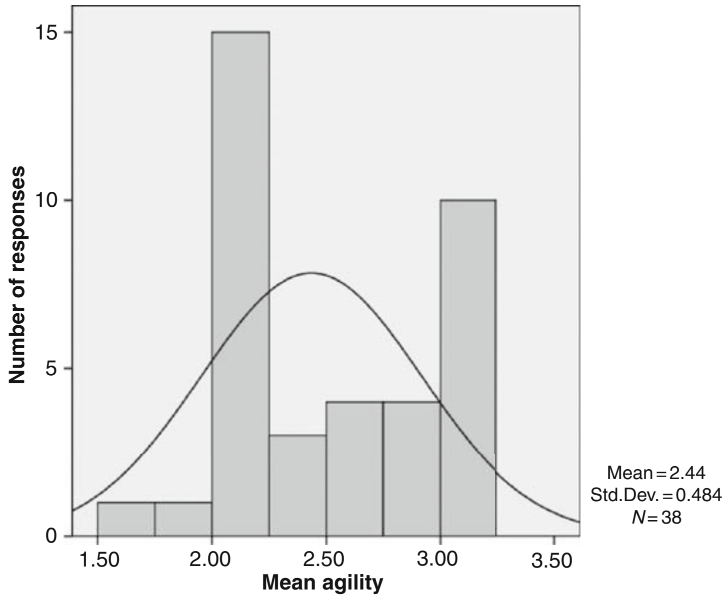


Figure 6.4 Range of results for agility (consistent project success)

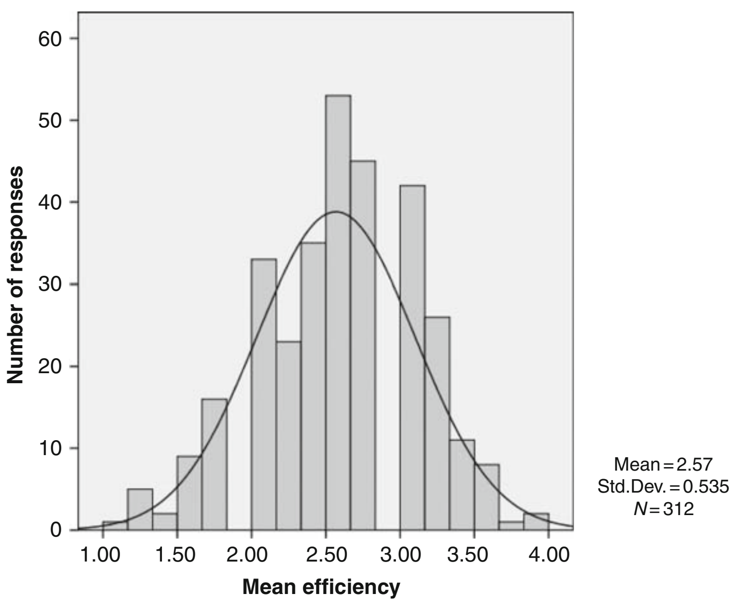


Figure 6.5 Range of results for efficiency (project management success)

be present. Analyses of the results were conducted using a variety of correlation techniques, including "categorization and regression trees" (C&RT), by which the population of results to questions about independent variables is divided into two, according to the variable that has the highest significance with regard to the dependent variable. The resulting populations are then further divided, according to the next most significant variable etc.

As can be seen from Figure 6.6, Question #3 is the second most significant factor correlating to an organization's "agility" – its ability to deliver consistently successful results from the right combination of projects.

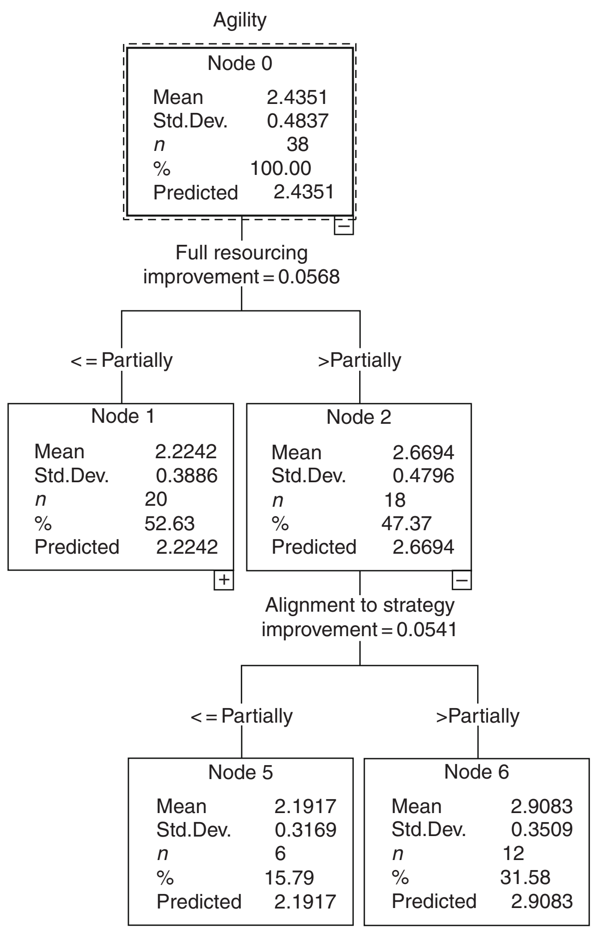


Figure 6.6 C&RT analysis of capabilities correlating to agility

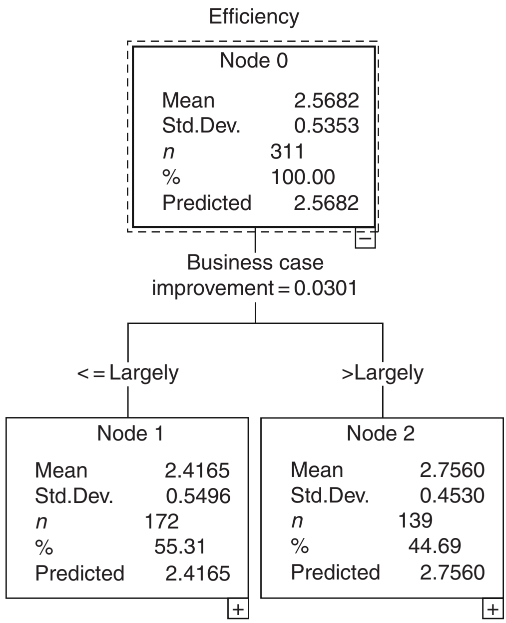


Figure 6.7 Partial C&RT analysis of factors correlating to efficiency

The practical mechanism by which projects are aligned to strategy is, by and large, the production of a well-formed "business case". A similar analysis of the factors that most correlate to "efficiency", shown in Figure 6.7 below, highlights the business case explored in Question #3, as being the most significantly correlated of all the factors under investigation.

The sample from which these results have been obtained is in no sense true for all projects, all of the time. The majority of the respondents are from the USA (67%) and UK (27%) and from industries newly adopting project management, such as financial services and manufacturing (60%), information, computers and telecommunications (24%) and bio-pharmaceuticals (14%).

On the other hand, the results, particularly for the correlation between an effective business case and project results as assessed by the classic measures of time, cost, scope and quality, are very strong statistically and make logical sense.

Aligning projects with organizational strategy at the front-end makes sense for both the project and the organization.

Conclusion

Achieving front-end alignment is challenging but important. It is important to the permanent organization if it is to succeed in its chosen strategy, and it is important to each project if it is to make efficient use of the resources allocated to it. It is, however, fraught with challenges. In this chapter, several of these have been examined: the innovative worldview embodied in the project, compared with the adaptive worldview of the permanent organization surrounding it; the varied interests and perspectives of different groups of stakeholders; and the complementary roles that must work together in order to bring about front-end alignment. Two mechanisms have been recommended for achieving successful front-end alignment in the face of these challenges: a well-planned business case, and well-formulated engagement programme. There is always a temptation to "get on with the project" rather than investing in all this front-end effort. This chapter has attempted to demonstrate that this is a temptation best resisted.

References

APM. (2007). In: Rayner P (ed.) *APM Introduction to Programme Management*. Association for Project Management: High Wycombe, UK.

Archer N and Ghasemzadeh F. (2004). Project portfolio selection and management. In: Morris P and Pinto J (eds). *The Wiley Guide to Managing Projects*. John Wiley and Sons: New York, pp. 237–255.

Artto K, Martinsuo M, and Aalto T. (2001). *Project Portfolio Management. Strategic Management Through Projects*. Project Management Association Finland: Helsinki, Finland.

Cleland D and King W. (1988). Project owner strategic management of projects. In: Cleland D and King W (eds). *Project Management Handbook*. John Wiley and Sons: New York, pp. 165–188.

Cooke-Davies T. (2001). *Towards Improved Project Management Practice: Uncovering the Evidence for Effective Practices Through Empirical Research*. Dissertation.com, USA.

Cooper R, Edgett S, and Kleinschmidt E. (2001). *Portfolio Management for New Products*. Perseus: Cambridge, MA.

Crawford L and Cooke-Davies T. (2005). Project governance: the pivotal role of the executive sponsor. In: *PMI Global Congress, North America*. Project Management Institute: Philadelphia.

Crawford L et al. (2008). *Situational Sponsorship: A Guide to Sponsorship of Project and Programs*. Project Management Institute: Philadelphia.

Dinsmore P and Cooke-Davies T. (2005). *The Right Projects Done Right. From Business Strategy to Successful Project Implementation*. Jossey Bass: San Francisco.

Girard R. (1986). *The Scapegoat*. Johns Hopkins University Press: Baltimore, Maryland.

Higgins L and Cougar J. (1995). Comparison of KAI and ISP instruments for determining style of creativity of IS professionals. *Proceedings of the 28th Annual Hawaii International Conference on System Sciences*. IEEE: New York.

Kirton M. (2003). *Adaptation-Innovation: In the Context of Diversity and Change*. Routledge: East Sussex, UK and New York.

Morris P and Pinto J. (2004). *The Handbook of Managing Projects*. Wiley: New York.

Porter M. (1980). *Competitive Strategy. Techniques for Analyzing Industries and Competitors*. The Free Press: New York.
Project Management Institute. (2006a). *The Standard for Program Management*. Project Management Institute: Newtown Square, PA, US.
Project Management Institute. (2006b). *The Standard for Portfolio Management*. Project Management Institute: Newtown Square, PA, US.
Stacey R. (2003). *Complexity and Group Processes. A Radically Social Understanding of Individuals*. Brunner-Routledge: Hove, UK.
Teague J and Cooke-Davies T. (2007). Developing organizational capability: pointers and pitfalls. In: *PMI Global Congress*. EMEA Project Management Institute: Philadelphia.
Turner J and Keegan A. (1999). The versatile project-based organization: governance and operational control. *European Management Journal* **17** (3): 296–309.
Turner J and Müller R. (2002). On the nature of the project as a temporary organization. *Proceedings of IRNOP V. 5th International Conference of the International Network of Organizing by Projects*. Erasmus University: Rotterdam.
Winch G. (2004). Managing project stakeholders. In: Morris P and Pinto J (eds). *The Wiley Guide to Managing Projects*. John Wiley & Sons: Hoboken, NJ, pp. 321–339.
Wright N. (1997). *The New Testament and the People of God*. SPCK: London.

7

Using Soft Systems Methodology to Structure Project Definition

Mark Winter

Introduction

According to a research report published in the 1990s, one of the areas where new and original research could provide the most practical benefit is the front-end of projects: "better understanding is needed of the 'soft' methodologies and their relevance and credibility". This chapter focuses on one of these "soft" methodologies, namely, soft systems methodology (SSM) and how it can be applied at the front-end. The chapter will demonstrate the relevance of SSM, together with its strong credibility, having been developed in real situations for more than thirty years. A real example of SSM in action is presented and, linked to this, a brief introduction to the approach. It is not intended to be a definitive guide to using SSM at the front-end: it merely presents one experience of using the approach in order to highlight its relevance and credibility and to illustrate the kind of role that SSM can play at this crucial stage.

Consider the following situation at the start of a project: some years ago, the author was asked to supervise a postgraduate consultancy project for a company that produces specialist plastics products for the car industry. The objective of this apparently straightforward project was to determine a new method or system for improving communication between two of the company's manufacturing sites. Upon visiting the first site, it was found that no one knew anything about the project and nor was communication between the two sites seen to be an issue. Similarly at the second site, communication was not seen to be an issue, and it also transpired that the managing director was unaware of the project's administrative fee that was coming out of his budget. Only a week into the project, and already there was a messy, problematical situation! Although just a small example, this kind of situation illustrates the differing interests and perspectives that are always present at the front-end of projects, as Professor Peter Morris (2002) points out:

> At the front end … we often have quite messy, poorly structured situations where objectives are not clear, where different constituencies

> have conflicting aims and where the way forward requires vision and leadership as well as hard analysis and design.

It is not just that front-end situations are messy and problematical; what is also significant is the limited intellectual support in this area, in the form of practical methodologies for tackling such situations. For some years now, this lack of intellectual support has been highlighted by various researchers, including some well-known academics whose observations were summarized in a report published in 1997 (Ashmore, 1997):

> One of the widest fields where new and original research could provide most practical benefit is within the front-end processes of a project. In this area, an entire range of tools could be developed . . . Better understanding is needed of the "soft" methodologies and their relevance and credibility.

By "soft" methodologies, the reference here is to a particular group of approaches developed over the last 30 years in the field of management science (Rosenhead and Mingers, 2001). Known more recently as problem structuring methods, or PSMs, such approaches are more concerned with *problem setting* than *problem solving*, thus making them highly appropriate for the front-end of projects, and hence the subject of increasing interest in project management research (Winter et al., 2006). One of these approaches in particular has been the subject of various papers over the years (Crawford et al., 2003; Green, 1994; Neal, 1995; Ramsay et al., 1996; Saunders, 1992; Yeo, 1993) and is now attracting growing interest in the field of project management. Known as soft systems methodology (SSM), the relevance of this approach to the front-end of projects is easily demonstrated, as this chapter will show; moreover, the credibility of SSM is also particularly significant in relation to the extract above.

Developed under the intellectual leadership of Professor Peter Checkland, SSM is the product of a rigorous research programme carried out over more than 30 years, involving real action experiences in organizations and the deliberate use of these experiences to help develop the approach. From its intellectual origins in systems engineering, SSM has steadily evolved through a cyclic process of practical application and theoretical refinement into what is now a highly credible approach, with wide applicability across many fields. Not only this, but the whole story of SSM's development has been meticulously documented and discussed in a series of books, including a 30-year retrospective (Checkland, 1999) and, more recently, a short definitive guide to SSM aimed at practitioners, teachers and students (Checkland and Poulter, 2006). Some would argue that Checkland's approach to developing SSM has been quite exceptional in a field that is often criticized for its lack of practical relevance and theoretical rigour.

The primary aim of this chapter is to introduce the methodology of SSM and show how it can be applied at the front-end of projects, whilst also recognizing this is not the only "soft" approach available to practitioners (Eden et al., 1998; Howick, 2003; Williams, 2002, 2003, 2004). It is not intended to be a definitive guide to using SSM at the front-end but merely presents one experience of using the approach in a real situation and, linked to this, offers a brief introduction to SSM. Above all, it seeks to highlight the relevance and credibility of the approach and to illustrate the kind of role that SSM can play at this crucial stage. To do this, the chapter takes a somewhat unconventional approach: rather than introduce the ideas of SSM first, followed by the example, the approach here is to introduce the case example first and then the main ideas of SSM. In short, experience suggests that in introducing SSM, it is more beneficial to start with a real example first and then introduce the ideas in the context of the example; doing this not only helps to introduce what the main ideas are but also how the approach can be used in practice. In summary, the aim of this chapter is not to provide an academic critique of SSM but to illustrate the kind of contribution that SSM can play in real situations at the front-end of projects.

Introduction to the case example: Tesco Stores Ltd

Tesco Stores Ltd is the UK's leading food retailer with operations now in Europe, Asia and North America. Building on the success of its Clubcard initiative in the early 1990s and other customer-focused initiatives, Tesco's growth strategy in the 1990s was to embark on a major retail development programme involving new store concepts such as Tesco Express, Tesco Metro and Tesco Extra. To help achieve this, a new strategic project called Branch Specific Ranging (BSR) was initiated by the Tesco Board in August 1995, aimed at ranging Tesco's stores much more specifically to the local areas in which the stores were trading. For each store, this would mean taking into account aspects such as local competition (current and anticipated) and local demographics, in order to identify the optimum non-core product range for each store. Also included were non-food items, such as clothes and white goods, and household services, such as dry-cleaning. Following the Board's approval in 1995, a cross-functional team was formed, with representatives from the various divisions within Tesco, including commercial, retail and IT; in short, since the project was going to affect nearly every part of Tesco's operation, it was necessary to have representation from all the relevant parts of the company. By early 1996, however, the project team were struggling to develop an agreed conceptual model of BSR. Two workshops had been held in November 1995 and February 1996, and there was a growing concern that the project was fast becoming an IT project. In today's language, the BSR project was essentially a business transformation programme which extended well beyond IT to most of the other operating divisions within

Tesco. Moreover, with another workshop planned for April 1996, and a one store trial arranged for June to September 1996, there was now an urgent need to develop a conceptual model for the one store trial. In SSM terms, this was a *problematical situation* at the front-end of the project.

Front-end situation: what should the conceptual model be?

In response to this situation, the lead facilitator on the project, known as RD from here on, felt that SSM could be a useful approach for developing the conceptual model of BSR, but neither he nor the team had experience of using SSM in a live project environment. This was a major strategic project for Tesco, and although RD had studied SSM at university, he felt he did not have the experience to tackle this particular situation. In February 1996, RD therefore decided to seek external help and approached two UK universities with specialist knowledge and experience in SSM. During the course of his evaluation, RD held several discussions with the present author (MW), who was duly selected to help facilitate the April workshop (Winter, 2002).

In planning the workshop, it soon became apparent from the documents produced so far that the two-day event needed to be very carefully planned. This was an important project for Tesco, and people were expecting significant progress on the development of the conceptual model. Indeed, this had been the main aim of the second workshop in February 1996, and yet very little progress had so far been made. In terms of the project's schedule, it was important for the April workshop to succeed in developing this model, for without it, the trial could not go ahead. Furthermore, many of the documents from the previous two workshops showed conflicting views on BSR, including differing ideas on how it should be implemented. A major challenge for the next workshop was how to bring these views together into an overall model that could be evaluated in the one store trial.

Against this background, MW proposed that the April workshop be organized using SSM but emphasized the need to use language familiar to the project team, rather than the technical language of SSM. This was to avoid discussion about the underlying approach and avert the risk of getting deflected from the primary task of developing the model. RD agreed with this and outlined several aims for the workshop, including the main requirement to produce *"a model agreed by all that can be taken forward"*. MW then set about designing a structure for the two days, based on the general process of SSM, and after several discussions with RD, a framework of activity was finally agreed, which is shown in Figure 7.1.

The morning of Day 1 was to be spent discussing people's views on a number of aspects relating to BSR (Questions A1–A5); the afternoon, together with the morning of Day 2 would be spent developing a prototype model of BSR; and finally, the afternoon of Day 2 would be allocated to reviewing the evaluation project and next steps. Prior to the workshop, MW would create

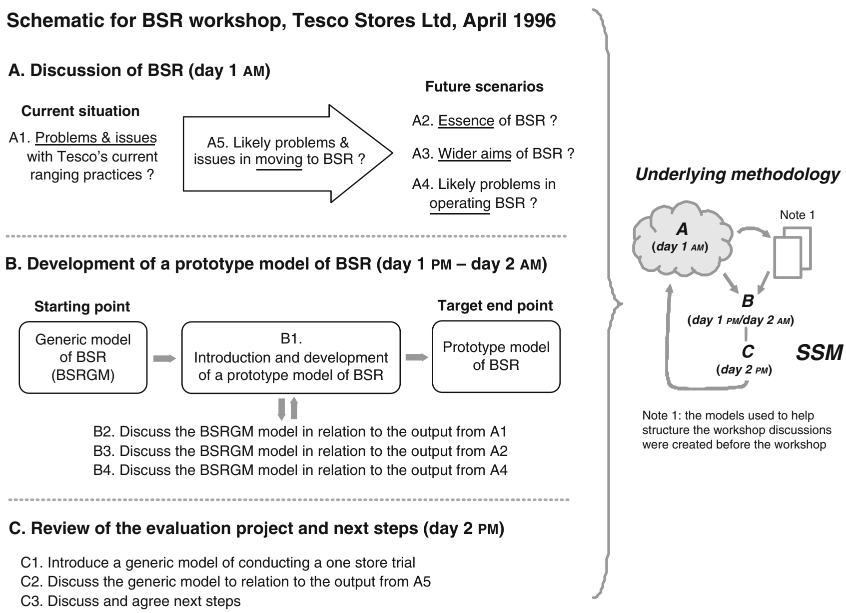


Figure 7.1 Schematic for BSR workshop in April 1996

a generic model of BSR to use as a starting point for the main discussion, the assumption being that, rather than trying to develop a prototype model from scratch, it would be better to start with a model based on the work done to date and develop this further in relation to the morning's discussion (Figure 7.1 (B2–B4)).

For those readers already familiar with SSM, Figure 7.1 shows how the three-part structure of the workshop (A–C) relates to the general process of SSM; for those new to the methodology, SSM will be introduced in the next section. Turning now to the events of the two days, the remaining sections briefly summarize the team's discussions and include a copy of the generic model produced by MW and the final version produced by the team. It is important to note, however, that these models are not meant to be understood: they are included only as examples to illustrate the use of SSM.

Discussion of BSR (Day 1 AM)

Twelve people from different divisions in Tesco attended the workshop, with MW and RD acting as facilitators for the two days. In opening the workshop, the background and aims were outlined by the project manager, and RD explained how MW had been chosen to help with the project. Following these introductions, MW introduced the framework for the two days by talking through the schematic shown in Figure 7.1 (without mentioning SSM).

He explained the three-part structure, and this was agreed by the team to be a sensible framework for the two days. The workshop then commenced with a general discussion of BSR, generating many observations about Tesco's current ranging practices, some possible difficulties in operating BSR and the likely problems in moving to BSR. All these points were recorded on flip charts and displayed on the walls of the room for later use. It was explained by MW that the purpose of this stage was simply to capture these points for discussion later in the afternoon. Other points were raised in relation to the aims and objectives of BSR, including the view that BSR should become a core competency for Tesco. To summarize this part of the workshop in terms of SSM, the morning was spent understanding how the different team members perceived Tesco's current ranging practices, where they wanted to be and why and what they saw as the problems and issues in getting there. Not surprisingly, given their different roles within the project, there were many different worldviews on all these aspects, and it was important to capture these for consideration later in the workshop.

Development of a prototype model of BSR (Day 1 PM – Day 2 AM)

After lunch on the first day, attention turned to the development of the prototype model, and MW explained the suggested process for developing this. A generic model would be introduced as the starting point for discussion, and the team would then develop this model through a structured comparison with the points raised in the morning session. For example, having identified a range of possible problems in operating BSR, the team might seek to develop the model further by designing in features to deal with some of these anticipated problems. The generic model was based on the following concept of BSR:

A generic concept of BSR

WHAT: to continuously (re)define the core and non-core range of Tesco's stores in accordance with the perceived and expected requirements of the stores' individual customers and customer groups.

HOW: through the collection and detailed analysis of various operational and environmental data about customers, competition, demographics, etc.

WHY: in order to satisfy more fully the various requirements of its customers (existing and potential), whilst at the same time increasing company profitability.

As a concept for what BSR *might* be, rather than what it *should* be, this particular notion was perceived by MW to be especially relevant to the current situation, because, according to his reading of the documents produced so far, it was this part that was presenting difficulties for the project team. [N.B. In SSM, the concept above is known as a *root definition*, which defines a particular concept of purposeful activity that can be used as the basis for discussion. In most applications of SSM, a number of concepts are suggested, but in this situation, this was the chosen concept.] In crafting this particular concept, MW developed a detailed conceptual model which not only reflected the concept but also incorporated certain aspects from the previous workshops, in order to maintain continuity with the work done so far (see Figure 7.2). MW talked through this generic model, and the team felt that it would be a good starting point for discussion. To help facilitate the discussion, RD had created a large wall chart version before the workshop, and this proved particularly useful in helping the team to discuss and adapt the model in relation to what had been covered in the morning session.

Throughout the afternoon and the next morning, various activities were added, changed and refined in relation to many of the points raised in the first session. For example, in relating the model to the perceived problems of operating BSR (Question A4), various changes were identified which the team had not foreseen, and this represented important learning for the project. Throughout this stage, a considerable amount of effort was spent developing the model, with everyone contributing ideas, counter ideas and on-the-spot evaluations of possible implications for Tesco's supply chain, etc. By lunchtime on the second day, the model was sufficiently detailed for a one store trial to proceed, based on the team's view that they had incorporated most of the relevant points from the initial discussion of BSR. A copy of the final model is shown in Figure 7.3.

Review of the evaluation project and next steps (Day 2 PM)

Following the development of the prototype model, the team's attention was directed towards the evaluation project and the next steps following the workshop. The important question now was: how should the prototype model be taken forward? To help answer this, MW had prepared another SSM activity model to use as a starting point, and a similar process was enacted, with MW and RD facilitating the team's discussion of what needed to be done to prepare for the trial. As with the BSR generic model, MW's initial model introduced a number of important activities that had not been considered. Many of these emerged by relating the model to the discussion on Day 1, specifically, Question 5. By this stage, all of the planned work had been done, and after the workshop a full set of notes was produced from the flip charts in preparation for the forthcoming trial. Electronic versions of the two models were also produced, and included in the workshop documentation was the following note written by the project manager to the team: "Thank you very much for all your input to the workshop. I thought the output from the 2 days was excellent, covering a lot of ground and giving us a very good start for mapping out the work required over the forthcoming months to implement the one store trial."

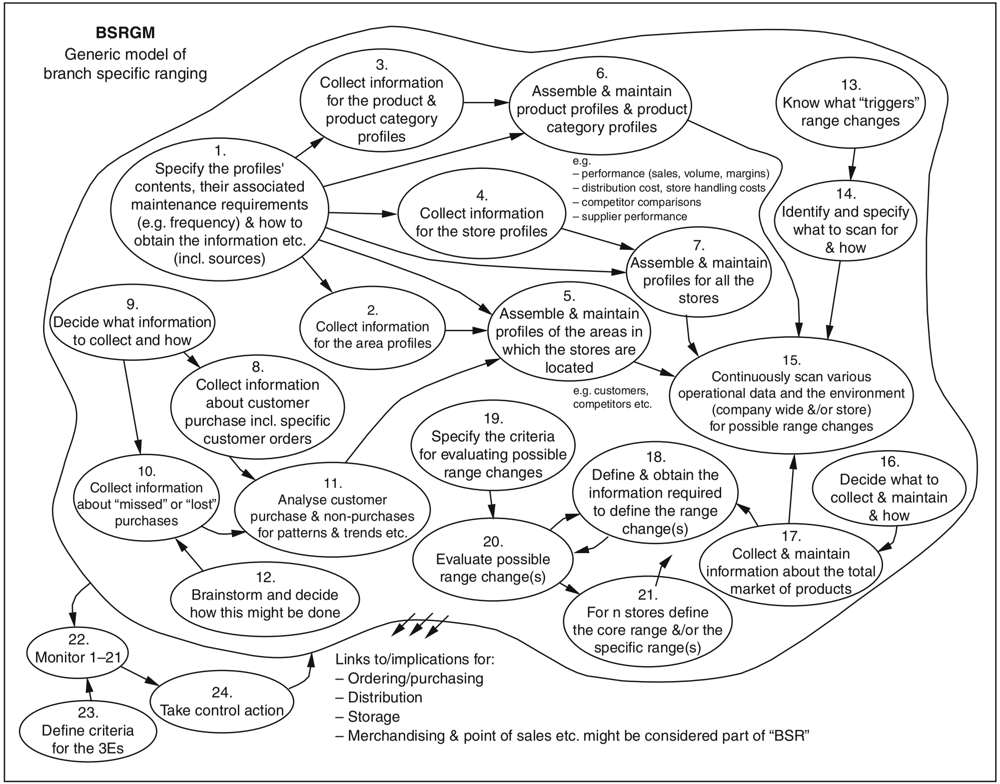


Figure 7.2 A generic model of BSR for use as a starting point for discussion

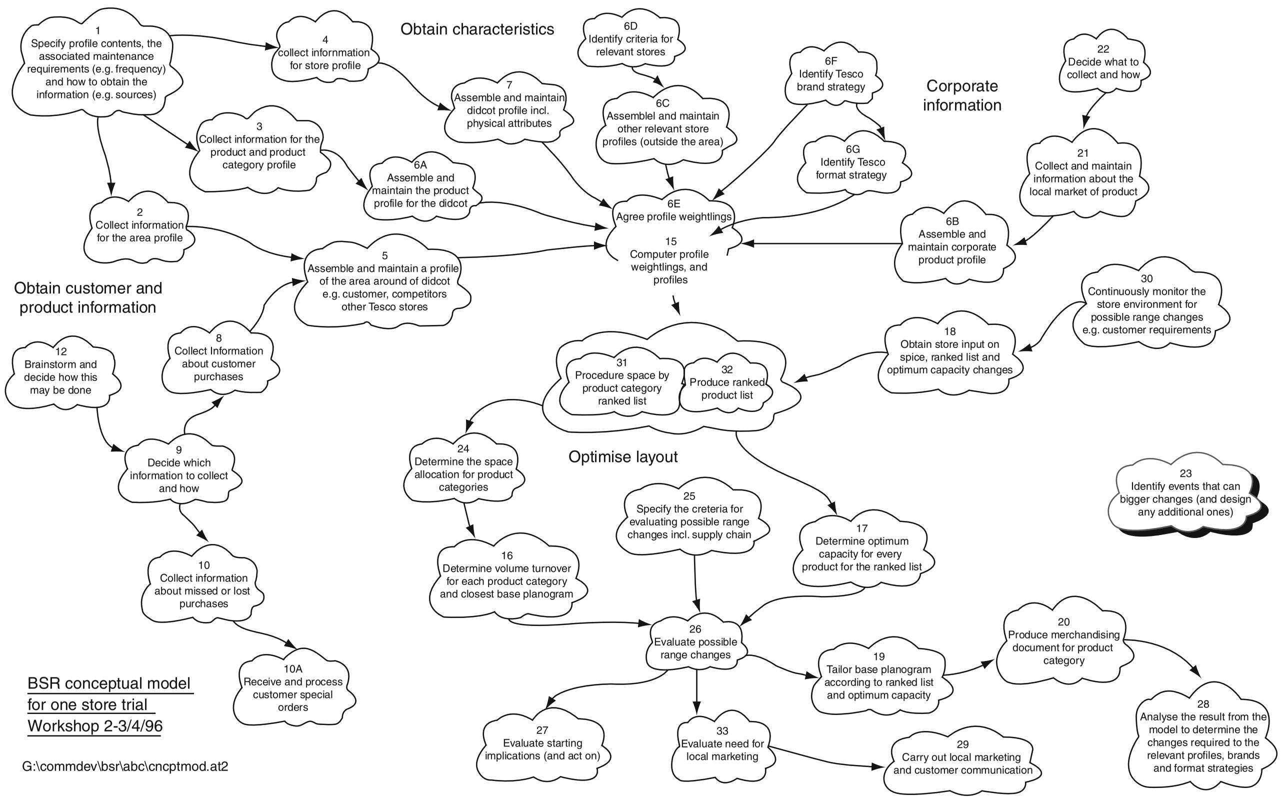


Figure 7.3 The final model of BSR produced by the project team

This trial was held between June and September 1996 and was subsequently extended to several other stores later that year. Following a series of successful trials, Tesco implemented BSR across a range of selected stores, and it has now become standard merchandizing practice across the company. Reflecting back on these events, it is clear that the April workshop made an important contribution to the project, much of this being due to the commitment and contributions from the project team, including the facilitation support provided by RD. The role played by SSM at this stage of the project was also significant, as the next section explains.

A brief introduction to soft systems methodology (SSM)

At the end of the April workshop, RD conducted a brief review of the two days, and one of the comments made was: "We seemed to have covered a lot of ground in this workshop; has there been some approach in the background?" This was a most perceptive comment, to which the short answer was yes, there had been an underlying approach to the workshop, namely soft systems methodology. So what is the approach of SSM and how was it used in the Tesco situation? As explained in the introduction, the Tesco example has been deliberately presented first, to help explain some of the main features of SSM. Four main features are briefly introduced here, based on the latest publication by Checkland and Poulter (2006): (i) the general process of SSM, (ii) the concept of problematical situations, (iii) the notion of "learning" in SSM and (iv) SSM as a methodology rather than a method. Special attention is given to these four features because of their relevance to the front-end of projects and their application within the Tesco project. Due to space constraints, however, only a brief introduction can be provided, and a more comprehensive discussion can be found in Checkland and Poulter (2006).

The general process of SSM

According to Checkland and Poulter (2006), "soft systems methodology (SSM) is an organized way of tackling perceived problematical (social) situations. It is action-oriented. It organizes thinking about such situations so that action to bring about improvement can be taken". How this is done is shown in the left-hand diagram in Figure 7.4. As this diagram shows, the general process of SSM consists of four main activities: (1) learn about the problematical situation, including the social and political aspects; (2) create some models of purposeful activity relevant to the situation; (3) discuss the situation, using the models to determine the action needed; and (4) define/take action to improve the situation. This essentially was the process used to structure the two-day workshop on Tesco: firstly, the morning of Day 1 was spent learning about the current situation from the perspective of the different team members; secondly, MW had created a couple of models of

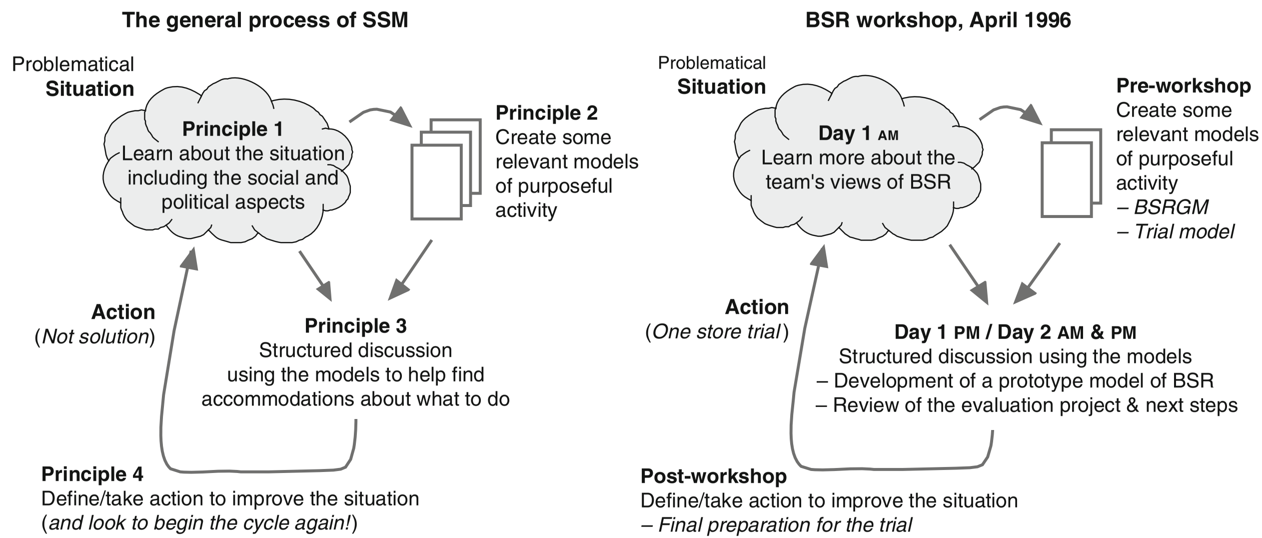


Figure 7.4 The general process of SSM and its use in structuring the BSR workshop

purposeful activity to help structure the team's discussions; thirdly, these models were then used to structure the team's discussions about the prototype model and the one store trial; finally, the resulting outputs were further defined for the trial.

As the right-hand diagram in Figure 7.4 shows, although the actual workshop process differed slightly from the general process of SSM, the basic inquiry process was essentially that of SSM, organized and enacted in a way to deal with the perceived reality of the current situation. On the latter point, the Tesco example also shows that the general process of SSM should always be adapted to the particular area of concern, which can often mean having to translate the terminology of SSM into language more familiar to the people involved. There are many different ways in which the general process of SSM can be used at the front-end of projects, the BSR project being just one example. In summary, whatever the situation might be, the general process of SSM is an organized and flexible framework for dealing with complex situations, which makes it particularly relevant to the front-end of projects.

The concept of problematical situations

Another important feature of SSM is the concept of problematical situations, which is very important for understanding the general process of SSM outlined in Figure 7.4. As the diagram shows, the starting point of SSM is always a perceived real-world problematical situation, in which some action is needed, as opposed to a defined problem, for which a solution needs to be found. In terms of a simple "A to B" formulation, SSM is not about getting from "problem to solution" but from "situation to action". Understanding this distinction is crucial to understanding SSM. The focus is always a perceived situation (see Figure 7.4). As Checkland and Poulter (2006) point out:

> Nothing is intrinsically "a situation"; it is our perceptions which create them as such, and in doing that we know that they are not static; their boundaries and their content will change over time. Some of the situations we perceive, because they affect us in some way, cause us to feel a need to tackle them, to do something about them.

Applying this to the Tesco example, the "perceived situation" in early 1996 was that the BSR project was fast becoming an IT project, and action was needed to help move it forward. This was certainly RD's perception of the situation, and was soon shared by MW, following his initial discussions with RD. This essentially was the starting point for the SSM-based intervention, where the primary aim was that of getting from the perceived situation to some agreed action for moving forward on the project.

The notion of "learning" in SSM

Consider the following two definitions of "learning": learning *about* something (e.g. how a car engine works) and learning *from* a real-world experience (e.g. a failed project). Neither of these definitions describes the concept of "learning" in SSM, which instead, means *learning what action to take in problematical situations*. In other words, it is the actual tackling of a situation and the emergent output, which constitutes the learning in SSM. As Checkland and Poulter (2006) point out, the general process of SSM is:

> a cycle of learning which goes from finding out about a problematical situation to defining/taking action to improve it. The learning which takes place is social learning for the group . . . though each individual's learning will be, to a greater or lesser extent, personal to them.

Linking this to the Tesco example, the "social learning for the group" was the BSR project team "learning" its way to an agreed prototype model for BSR, and the action was that of going forward into the one store trial. Furthermore, each member of the project team also "learnt" a lot more about their individual role within the project and how it was going to affect their particular division within Tesco. In practice, this does not mean having to use the term "learning" with the people involved but that the mindset in using SSM is essentially that of learning what action to take in problematical situations. In summary, as the Tesco example shows, SSM's notion of "learning for action" is clearly very relevant to the front-end of projects.

A methodology rather than a method

Completely unrelated to SSM, consider the following extract which seeks to "demystify" the general concept of "methodology" (Reiss, 2007):

> The term "methodology" is popular but inaccurate. The "ology" suffix should refer to the study of a topic; hence zoology, psychology and pharmacology. So consider the question: what is the difference between a project management method and a project management methodology? The answer, dear reader, is simply the price. The theory says that because the name is long and sounds complicated, more organizations will be happy to pay larger sums for a purpose-built, complex sounding methodology than they would for a plain old method. This book will firmly stick to the word "method" as a small voice of protest.

This unfortunate stance is misguided in two respects: firstly it misunderstands the power of language in directing people's actions. For example, calling an approach a "method" will influence how that approach is used (e.g. prescriptively). Secondly, it ignores the very pragmatic distinction between methodology and method which Checkland has consistently highlighted since 1981. Here is what he says about the distinction in his latest book:

> It is obvious . . . that any approach able to deal with the changing complexity of real life will have to be flexible. It could never be reduced to a sequence of steps . . . It needs to be flexible enough to cope with the fact that every situation involving human beings is unique. . . . This means that an approach to problematical human situations has to be a methodology rather than a method, or technique. A methodology, as the word indicates, is a logos of method; that is to say it is a set of ongoing principles which can be adapted for use in a way which suits the specific nature of each situation in which it is used. SSM provides a set of principles which can be both adopted and adapted for use in any real situation in which people are intent on taking action to improve it.

Linking this to the Tesco example, as the two diagrams in Figure 7.4 show, the basic principles of SSM were adapted to form a situation-specific framework, to help guide the development of a prototype model of BSR; the underlying approach was SSM, but the actual approach was the framework of activity shown in Figure 7.1. In other words, in developing the prototype model, the project team were not following SSM directly, as if it were a method; rather, they were being guided by a situation-specific framework that had been created from the general process of SSM. This leads to another important feature of SSM, and the subject of the next section. In summary, being a methodology rather than a method, SSM offers considerable flexibility and rigour in tackling problematical situations, which is yet another feature that makes it highly relevant to the front-end of projects.

Front-end situation: how should the BSR model be developed?

Returning to the Tesco project, this section briefly illustrates the final feature of SSM to be covered here, namely, two fundamentally different uses of SSM, known as SSMc and SSMp. In essence, SSMc is about the use of SSM to tackle the perceived content of a problematical situation (e.g. the example presented so far). SSMp is about using SSM to plan the process of activity in a situation, and is what this section seeks to illustrate. In other words, at the start of MW's intervention, the initial concern was not what needed to be in the BSR model, but how this model should be developed, given the overall situation within the project. In the language of SSMp, the question was: what should the process be for developing this model? As stated earlier, MW felt that the April workshop needed to be very carefully planned, not least because of its strategic importance, and the fact that people were expecting significant progress to be made in order for the one store trial to go ahead. It was against this background that MW consciously sought to "learn" what the workshop needed to achieve, how the programme for the two days should to be organized, and how he would work with RD in facilitating the workshop. The following discussion uses the general process of SSM to explain how the approach was used to plan the activity of the workshop.

SSMp Activity 1: learning more about the situation

To learn more about the current situation at the start, various documents were studied by MW, and it was clear from these that different team members had developed their own ideas, without integrating these into a shared model of BSR. There was also uncertainty about whether the BSR project represented one project or several projects. It seemed to MW that a more sensible approach would be to see it as a programme of projects which could develop as the work progressed. From this perspective, the two-day workshop could then be seen as part of a project to develop and evaluate the business case for BSR, which would include the one store trial. At their next meeting, MW and RD discussed these ideas and also RD's expectations for the workshop. RD agreed that the project would be better seen as a programme rather than one large project and confirmed that the next major objective was to evaluate the business case for BSR through the development of a detailed conceptual model and a one store trial. It was against this background that MW agreed to create a draft model of the workshop for discussion with RD.

SSMp Activity 2: creating models of purposeful activity

In planning the workshop, MW agreed to create three models of purposeful activity for discussion with RD: (i) a model of how the workshop could be organized using SSM, (ii) a generic model of branch-specific ranging and (iii) a model for thinking about the one store trial. This section deals only with the logic of the workshop model, because it was this model that represented the proposed activity for the two days. Given the different views on BSR, it seemed sensible to begin the workshop (after introductions etc.) with a discussion of how the different team members viewed BSR in relation to the following questions: (1) What were the current problems with Tesco's ranging practices? (2) What was the essence of BSR from each team member's point of view? (3) What were the team's views on the wider aims of BSR? (4) What problems might be encountered in operating BSR? (5) What problems might be encountered in moving to BSR?

After an initial discussion of these aspects, the group would be directed to develop the prototype model, the idea of a prototype being that further learning and development would take place through the trial. In order to initiate this, MW would introduce a generic model of BSR based on the work done to date, the logic here being that this would be a more sensible approach than trying to develop a model from scratch. The team would then consider how the model might be developed in relation to their answers to the five questions above and thus produce a model that could be used for the trial. Finally, the group would be directed to consider the trial itself, and here again the author would use an SSM-based activity model to initiate discussion amongst the team. This essentially was the proposed activity for the two days, but these were only ideas for *what might be*, not *what should be*: the next step was to discuss these ideas with RD. To illustrate this further, Figure 7.5 shows the original SSM activity model that was used to help "learn" what was needed for the April workshop.

SSMp Activity 3: structured discussion using the models

Given that the workshop was now only a week away, it was agreed that MW would fax RD the workshop model in Figure 7.5 for discussion over the phone. MW would then make any necessary refinements, fax them back, and RD would produce the materials for the workshop. In the event, only minor changes were made to the workshop model, and RD agreed to produce a schematic of this model for use as an agenda with the project team. Moreover, since the other two models (the BSRGM and the trial model) were only starting points for discussion, RD took the view that these were sufficiently detailed to enable a group discussion and agreed to produce electronic versions for use in the workshop. Had more time been available, RD might have suggested more changes but judged that ideas would be developed further as the workshop progressed.

SSMp Activity 4: define/take action to improve the situation

Having agreed on a framework of activity to guide the forthcoming workshop, MW and RD discussed how they would facilitate the event and agreed that MW should take the lead role. Also, as agreed, RD produced the workshop schematic shown in Figure 7.1 from the model shown in Figure 7.5, together with electronic versions of the other two models. In summary, by this stage, MW and RD had "learnt" their way to an agreed framework for the workshop, and the aim now was to help the BSR project team "learn" their way to an agreed prototype model for BSR. A first cycle of SSM had been enacted, namely SSMp; the next cycle to be enacted was SSMc.

SSMc and SSMp: two fundamentally different uses

As the Tesco example shows, the distinction between SSMc and SSMp is an important one, because it highlights two fundamentally different uses of the approach: the use of SSM for tackling the perceived content of situations (SSMc) and its use for planning the process of activity in situations (SSMp). As the case shows, both types of use were employed in the BSR project: firstly, SSM helped MW and RD "learn" their way to an agreed process framework for the workshop (SSMp); and secondly, it helped the BSR project team to "learn" their way to an agreed model of BSR, thus enabling the one store

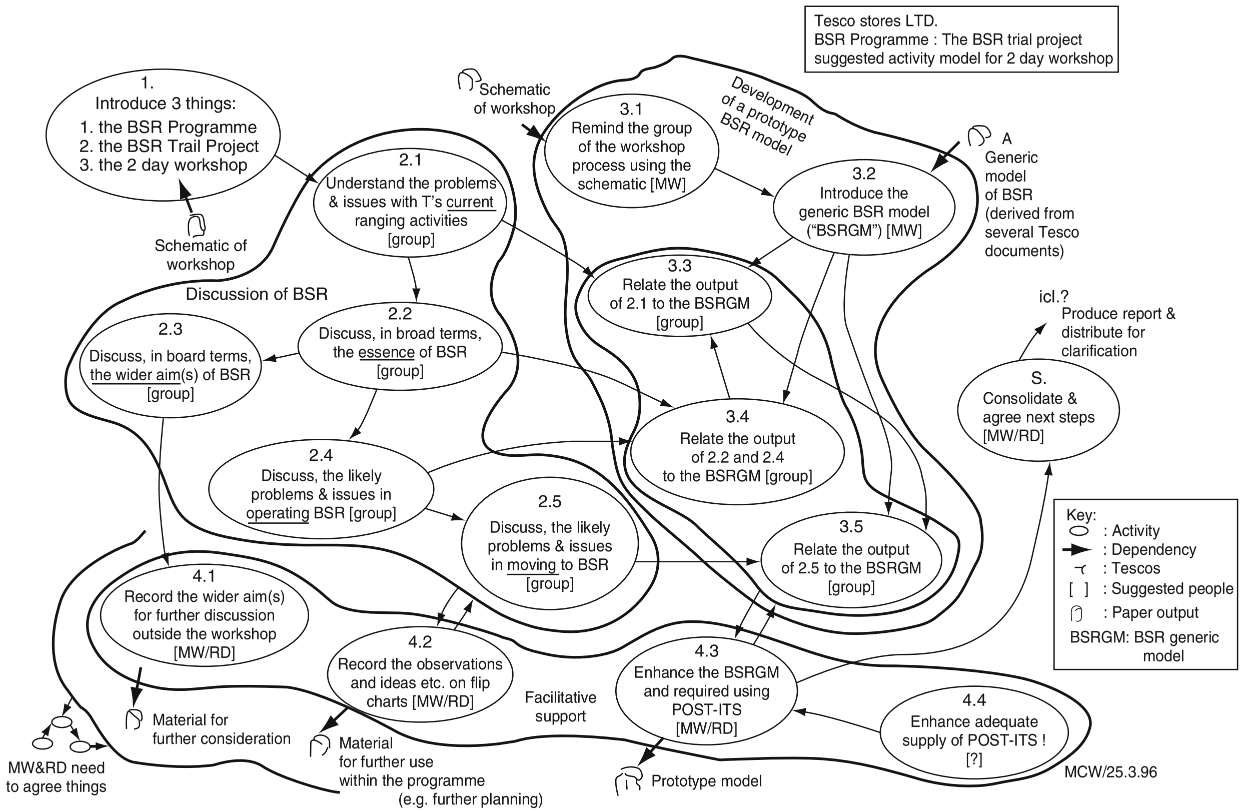


Figure 7.5 Suggested activity model for the 2-day workshop

trial to go ahead and the project to continue. The cyclic nature of SSM is also demonstrated here, in that its first use represented a cycle of getting from "situation to action", i.e. from the initial situation to an agreed framework for the workshop; the second use represented another cycle of getting from the new situation to an agreed model of BSR. This suggests that SSM is not just relevant to the front-end of projects: it is relevant at all stages of a project or programme.

Conclusion: SSM's underlying image of social reality

To conclude this chapter, it is important to briefly mention the underlying image of social reality upon which SSM is based. It helps to explain the process of SSM, and why, for example, the mindset of "learning" is so fundamental to the whole approach. Indeed, far from being an academic point, SSM's philosophical stance on the nature of social reality is arguably the most important idea of all, for it is this idea which underpins the whole methodology. Consider the following extract from Checkland (1981):

> It seems appropriate to try to answer explicitly the question about the nature of social reality implicit in soft systems methodology. The success of the methodology in real situations suggests the following answer: social reality is the ever-changing outcome of the social process in which human beings, the product of their genetic inheritance and previous experiences, continually negotiate and re-negotiate with others their perceptions and interpretations of the world outside themselves.

To illustrate this using the Tesco experience, when the author first became aware of the April workshop in 1996, it was far from clear what the workshop would involve and what would be needed to develop the concept of BSR. Indeed, as is typical at the front-end of projects, all that existed at the start were differing ideas and perspectives amongst the people involved. In other words, the "workshop" and "BSR" were not things "out there" that existed independently of the people involved: they were ideas and images in people's heads, which needed to be negotiated and re-negotiated through various discussions. SSM helped the team negotiate and renegotiate these differing ideas and images in an organized way, thus enabling purposeful action to be taken and the one store trial to go ahead. The central premise of SSM is that all human situations are subjectively experienced, and this subjectivity needs to be taken seriously if effective action is to be taken. Such a stance is clearly relevant to the concerns of this book, and it is important to remember that in "using SSM to structure project definition" (the title of this chapter), this will always mean different things in different contexts, because in real situations there is no one "best way" of using the approach. Such is the nature of soft systems methodology.

Acknowledgements

The author would like to acknowledge Richard Dodd (RD) of Tesco for his part in this work and the other members of the BSR project team who contributed to the workshop in April 1996.

References

Ashmore C. (1997). *BPRC Focus Group: Project Management, the Motivations and Conditions for Change*. Internet: http://bprc.warwick.ac.uk/focus12html Accessed 31 October 2001.

Checkland P. (1981). *Systems Thinking, Systems Practice*. Wiley: Chichester.

Checkland P. (1999). *Soft Systems Methodology: A 30 Year Retrospective*. Wiley: Chichester.

Checkland P and Poulter J. (2006). *Learning for Action: A Short Definitive Account of Soft Systems Methodology and its use for Practitioners, Teachers and Students*. Wiley: Chichester.

Crawford L, Costello K, Pollack J, and Bentley L. (2003). Managing soft change projects in the public sector. *International Journal of Project Management* **21** (6): 443–448.

Eden C, Williams T, and Ackermann F. (1998). Dismantling the learning curve: the role of disruptions on the planning of development projects. *International Journal of Project Management* **16** (1): 131–138.

Green S. (1994). Beyond value engineering: SMART value management for building projects. *International Journal of Project Management* **12** (1): 49–56.

Howick S. (2003). Using system dynamics to analyse disruption and delay in complex projects for litigation: can the modelling purposes be met? *Journal of the Operational Research Society* **54** (3): 222–229.

Morris P. (2002). Science, objective knowledge and the theory of project management. *Civil Engineering Proceedings of ICE* **150**:82–90.

Neal R. (1995). Project definition: the soft systems approach. *International Journal of Project Management* **13** (1): 5–9.

Ramsay D, Boardman J, and Cole A. (1996). Reinforcing learning, using soft systemic frameworks. *International Journal of Project Management* **14** (1): 31–36.

Reiss G. (2007). *Project Management Demystified* (3rd edition). Taylor and Francis: London.

Rosenhead J and Mingers J. (2001). *Rational Analysis for a Problematic World* (2nd edition). Wiley: Chichester.

Saunders R. (1992). Project management: a systems perspective. *International Journal of Project Management* **10** (3): 153–159.

Williams T. (2002). *Modelling Complex Projects*. Wiley: Chichester.

Williams T. (2003). Learning from projects. *Journal of the Operational Research Society* **54** (5): 443–451.

Williams T. (2004). Identifying the hard lessons from projects – easily. *International Journal of Project Management* **22** (4): 273–279.

Winter M. (2002). *Management, Managing and Managing Projects: Towards an Extended Soft Systems Methodology*. PhD Thesis, Lancaster University: UK.

Winter M, Smith C, Morris P, and Cicmil S. (2006). Directions for future research in project management: the main findings of a UK government-funded research network. *International Journal of Project Management* **24** (8): 638–649.

Yeo KT. (1993). Systems thinking and project management – time to reunite. *International Journal of Project Management* **11** (2): 111–117.

Part III

Generating Information

8

Optimism and Misrepresentation in Early Project Development

Bent Flyvbjerg

Pervasive misinformation about the costs, benefits and risks involved is a big problem in major project development. Consequences of misinformation are cost overruns, benefit shortfalls and waste. This chapter identifies optimism bias and strategic misrepresentation as main causes of misinformation. Bias and misrepresentation are problems throughout the project cycle, but the problems are greatest during early project development, because here measures to curb bias and misrepresentation are weakest. Measures for improving early project development are presented, emphasizing accountability and better methods.

Misinformation about costs, benefits and risks

Major projects generally have the following characteristics:

- Such projects are inherently risky, due to long planning horizons and complex interfaces.
- Technology and design are often not standard.
- Decision-making and planning are typically multi-player processes with conflicting interests.
- Often there is “lock in” with a certain project concept at an early stage, leaving alternatives analysis weak or absent.
- The project scope or ambition level will typically change significantly over time.
- Statistical evidence shows that such unplanned events are often unaccounted for, leaving budget and time contingencies sorely inadequate.
- As a consequence, misinformation about costs, benefits and risks is the norm.
- The result is cost overruns and/or benefit shortfalls with a majority of projects.

Within this chapter, data are presented on cost overruns and benefit shortfalls for large transportation infrastructure projects. However, comparative studies indicate that the problems identified apply across a wide range of project types, including transportation, public buildings, IT, power plants, dams, water projects, oil and gas extraction, aerospace, defence, development of new major products, plants and markets, and even large mergers and acquisitions (Flyvbjerg et al., forthcoming).

Table 8.1 shows the inaccuracy of construction cost estimates measured as the size of cost overrun. The cost study covers 258 projects in 20 nations on five continents. All projects for which data were obtainable were included in the study.[1] Nine out of ten projects in the study have cost overruns. For rail, the average cost overrun is 44.7% measured in constant prices. For bridges and tunnels, the equivalent figure is 33.8% and for roads 20.4% (Flyvbjerg et al., 2002). Overrun is calculated from the time of decision to build (typically the stage of the final business case). If overrun was calculated from earlier stages of project development, i.e. from the draft business case, or pre-feasibility study stages, overrun would typically be substantially higher. The large standard deviations shown in Table 8.1 are as interesting as the large average overruns. The size of the standard deviations demonstrates that uncertainty and risk regarding cost overruns can be very large. Overrun is found across the nations and is constant for the 70-year period covered by the study. Estimates of cost have not improved over time.

Table 8.2 shows the inaccuracy of travel demand forecasts for rail and road projects. The demand study covers 208 projects in 14 nations on five continents. All projects for which data were obtainable were included in the study.[2] Nine out of ten rail projects have overestimated traffic, whereas 50% of road traffic forecasts are wrong by more than ±20%; 25% of road traffic forecasts are wrong by more than ±40%. For rail, actual passenger traffic is 51.4% lower than estimated traffic on average. This means that less than half of the passengers forecasted actually showed up on the trains, often resulting in fiscal crises. For roads, actual vehicle traffic is on average 9.5% higher than forecasted traffic. Rail passenger forecasts are highly biased, whereas this is less so for road traffic forecasts. Again the standard deviations are

Table 8.1 Inaccuracy of transportation project cost estimates by type of project, in constant prices.

Type of project	**No. of cases (*N*)**	**Avg. cost overrun %**	**Standard deviation**
Rail	58	44.7	38.4
Bridges and tunnels	33	33.8	62.4
Road	167	20.4	29.9

Table 8.2 Inaccuracy in forecasts of rail passenger and road vehicle traffic.

Type of project	No. of cases (*N*)	Avg. inaccuracy %	Standard deviation
Rail	25	−51.4	28.1
Road	183	9.5	44.3

large, indicating a high level of risk and uncertainty. Inaccuracy of forecasts is constant for the 30-year period covered by the study (Flyvbjerg, 2005b; Flyvbjerg et al., 2005).

If techniques and skills for arriving at accurate cost and traffic forecasts have improved over time, these improvements have not resulted in an increase in the accuracy of forecasts.

Cost overruns and benefit shortfalls of the frequency and size described above are a problem because they (1) lead to a Pareto-inefficient allocation of resources, i.e. waste; (2) result in non-democratic decisions, because for democracy to work, it must be based on information, not misinformation; (3) lead to delays that in turn lead to further cost overruns and benefit shortfalls; (4) destabilize policy, planning, implementation and operations of projects. These problems are getting bigger, because projects are getting bigger.

From the data in Tables 8.1 and 8.2, it is clear that when cost and demand forecasts are combined, e.g. in the cost-benefit and environmental impact analyses that are typically used to justify large infrastructure investments, the consequence is inaccuracy to the second degree. Benefit-cost ratios are often wrong, not only by a few per cent but by several factors. In the worst cases, projects that were projected and promoted to benefit the economy turn out to detract from it instead, producing a negative input to GDP. As a case in point, the Channel Tunnel between France and the UK ended up costing approximately twice that forecast, with only half the revenue. In consequence, Anguera (2006) finds that the actual net present value of the Channel Tunnel to the British economy is negative, namely, –£10 billion and that "the British Economy would have been better off had the Tunnel never been constructed". It is a main task of front-end planning of major projects to avoid this type of situation and to identify projects that will contribute positively to the economy, the environment, etc. This can be done only if it is understood what causes underperforming projects to be chosen for implementation again and again and if this understanding is used to stop bad projects and promote good ones.

There are three main types of explanation that claim to account for inaccuracy in forecasts of costs and benefits in major projects: *technical, psychological and political-economic* explanations.

Technical explanations

Technical explanations account for cost overruns and benefit shortfalls, in terms of imperfect forecasting techniques, inadequate data, honest mistakes, inherent problems in predicting the future, lack of experience on the part of forecasters, etc. Technical error may be reduced or eliminated by developing better forecasting models, better data and more experienced forecasters. This was, until recently the most common type of explanation of inaccuracy in forecasts (Ascher, 1979; Flyvbjerg et al., 2002, 2005; Morris and Hough, 1987; Vanston and Vanston, 2004; Wachs, 1990).

However, the credence of technical explanations could mainly be upheld, because samples had been too small to allow tests by statistical methods. The data presented above, which come from the first large-sample study in the field, suggest that technical explanations of forecasting inaccuracy should be rejected, as they do not fit the data well.

Firstly, if misleading forecasts were truly caused by technical inadequacies, simple mistakes and inherent problems with predicting the future, there should be a less biased distribution of errors in forecasts around zero. In fact, for four out of five distributions of forecasting errors, the distributions have a mean statistically different from zero. Only the data for inaccuracy in road traffic forecasts have a statistical distribution that seems to fit with explanations in terms of technical forecasting error.

Secondly, if imperfect techniques, inadequate data and lack of experience were main explanations of inaccuracies, an improvement in accuracy over time might be expected, since in a professional setting, errors and their sources would be recognized and addressed through the refinement of data collection, forecasting methods, etc. Substantial amounts have, in fact, been spent over several decades on improving data and methods. Nevertheless, the data here show that this has had no effect on the accuracy of forecasts. Technical factors, therefore, do not appear to explain the data. It is not forecasting "errors" or their causes that need explaining. It is the fact that, in a large majority of cases, costs are underestimated and benefits overestimated. For technical explanations to be valid, they would have to explain why forecasts are so consistent in ignoring cost and benefit risks, from early project development, through approval, to implementation.

Psychological explanations

Psychological explanations account for cost overruns and benefit shortfalls in terms of what psychologists call "the planning fallacy and optimism bias". Such explanations have been developed by Kahneman and Tversky (1979a), Kahneman and Lovallo (1993) and Lovallo and Kahneman (2003).

In the grip of the planning fallacy, planners and project promoters make decisions based on delusional optimism, rather than on a rational weighting of gains, losses and probabilities. They overestimate benefits and underestimate costs. They involuntarily spin scenarios of success and overlook the potential for mistakes and miscalculations. As a result, planners and promoters pursue initiatives that are unlikely to come in on budget, or on time or to even deliver the expected returns. Over-optimism can be traced to cognitive biases, i.e. errors in the way the mind processes information. These biases are thought to be ubiquitous, but their effects can be tempered by simple reality checks, thus reducing the odds that people and organizations will rush blindly into unprofitable investments of money and time.

Psychological explanations fit the data presented above. The existence of optimism bias in planners and promoters would result in actual costs being higher and actual benefits being lower than those forecasted. Consequently, the existence of optimism bias would be able to account, in whole or in part, for the peculiar bias found in the data. Interestingly, however, when forecasters are asked about causes for forecasting inaccuracies in actual forecasts, they do not mention optimism bias as a main cause of inaccuracy (Flyvbjerg et al., 2005). This could, of course, be because optimism bias is subconscious and thus not reflected by forecasters. After all, there is a large body of experimental evidence for the existence of optimism bias and the planning fallacy (Buehler et al., 1994, 1997; Newby-Clark et al., 2002). However, the experimental data are mainly from simple, non-professional settings. This is a problem for psychological explanations, because it remains an open question as to whether they are general and extend beyond such simple settings to include major projects as well.

Optimism bias would be an important and credible explanation of underestimated costs and overestimated benefits in infrastructure forecasting, if estimates were produced by inexperienced forecasters, i.e., persons who were estimating costs and benefits for the first time and who were thus ignorant of the realities of infrastructure building and were not drawing on the knowledge and skills of more experienced colleagues. Such situations may exist and may explain individual cases of inaccuracy. But, given the fact that in modern society it is a defining characteristic of professional expertise, and that it is constantly tested (through scientific analysis, critical assessment and peer review) in order to root out bias and error, it seems unlikely that a whole profession of forecasting experts would continue to make the same mistakes decade after decade, instead of learning from their actions. Learning would result in the reduction, if not elimination, of optimism bias and the planning fallacy, which should result in estimates becoming more accurate over time.

However, the data clearly show that this has not happened. The profession of forecasters would indeed have to be an optimistic, non-professional group to retain their optimism bias throughout the 70-year period covered by this study for costs and the 30-year period covered for patronage and not realize that they were deceiving themselves and others by underestimating costs and overestimating benefits. This would account for the data, but does not seem a credible explanation. Therefore, on the basis of the data, optimism bias should be rejected as the only, or primary, cause of cost underestimation and benefit overestimation in major projects.

Political-economic explanations

Political-economic explanations see planners and promoters as deliberately and strategically overestimating benefits and underestimating costs, when forecasting the outcomes of projects. They do this in order to increase the likelihood of their project, and not the competitor's, gaining approval and funding. Political-economic explanations have been set out by Flyvbjerg et al. (2002, 2005) and Wachs (1989, 1990).

According to such explanations, planners and promoters purposely spin scenarios of success and gloss over the potential for failure. Again, this results in the pursuit of ventures that are unlikely to come in on budget, or on time, or to deliver the promised benefits. Strategic misrepresentation can be traced to political and organizational pressures, being in competition for scarce funds or jockeying for position, and it is rational in this sense. If a lie can be defined as making a statement intended to deceive others (Bok, 1979; Cliffe et al., 2000), it is clear that deliberate misrepresentation of costs and benefits is lying. Furthermore, lying pays off, or at least political and economic agents believe it does. Where there is political pressure, there is misrepresentation and lying, but misrepresentation and lying can be moderated by measures of accountability.

Political-economic explanations and strategic misrepresentation account well for the systematic underestimation of costs and overestimation of benefits found in the data. A strategic estimate of costs would be low, resulting in cost overrun, whereas a strategic estimate of benefits would be high, resulting in benefit shortfalls. A key question for explanations in terms of strategic misrepresentation is whether estimates of costs and benefits are intentionally biased to serve the interests of promoters in getting projects started. This question raises the difficult issue of lying. Questions about lying are notoriously hard to answer, because a lie is making a statement intended to deceive others, and in order to establish whether or not lying has taken place, one must know the intentions of the participants. If promoters and planners have intentionally skewed estimates of costs and benefits to get a project started, they are unlikely to formally tell researchers or others that this is the case, for legal, economic, moral and other reasons. Despite such problems, two studies exist that succeeded in getting forecasters to talk about strategic misrepresentation (Flyvbjerg and Cowi, 2004; Wachs, 1990).

Flyvbjerg and Cowi (2004) interviewed public officials, planners and consultants who had been involved in the development of large UK transportation infrastructure projects. The response of a planner with a local transportation authority typifies how respondents explained the basic mechanism of cost underestimation:

> You will often as a planner know the real costs. You know that the budget is too low but it is difficult to pass such a message to the counsellors [politicians] and the private actors. They know that high costs reduce the chances of national funding.

Experienced professionals like the interviewee know that out-turn costs will be higher than estimated costs, but because of political pressure to secure funding for projects, they hold back this knowledge, which is seen as detrimental to the objective of obtaining funding. Similarly, an interviewee explained the basic mechanism of benefit overestimation:

> The system encourages people to focus on the benefits – because until now there has not been much focus on the quality of risk analysis and the robustness [of projects]. It is therefore important for project promoters to demonstrate all the benefits, also because the project promoters know that their project is up against other projects and competing for scarce resources.

Competition between projects and authorities creates political and organizational pressures, which in turn create an incentive structure that makes it rational for project promoters to emphasize benefits and de-emphasize costs and risks. A project that looks highly beneficial on paper is more likely to get funded than one that does not.

Specialized private consultancy companies are often engaged to help develop project proposals. In general, the interviewees found that consultants showed high professional standard and integrity. But interviewees also found that consultants appeared to focus on justifying projects, rather than critically scrutinizing them. A project manager explained:

> Most decent consultants will write off obviously bad projects but there is a grey zone and I think many consultants in reality have an incentive to try to prolong the life of projects which means to get them through the business case. It is in line with their need to make a profit.

The consultants interviewed confirmed that appraisals often focused more on benefits than on costs. But they said this was at the request of clients and that, for specific projects discussed, "there was an incredible rush to see projects realized".

One typical interviewee saw project approval as "passing the test" and precisely summed up the rules of the game as follows:

> It's all about passing the test [of project approval]. You are in, when you are in. It means that there is so much focus on showing the project at its best at this stage.

Even though this and the previous quotes are about the approval stage in project development, it is not difficult to imagine strategic behaviour in early project development as well. In fact, the UK study indicates that without counter-incentives, the project development process easily degenerates into a justification process, where a project concept chosen at an early stage, i.e. when uncertainty was highest and is given an increasingly positive presentation through pre-feasibility study, draft business case, final business case and approval. The needs which the project is intended to meet, and the alternative ways of meeting these needs, are ignored or under-analyzed. The initial concept survives intact and often turns out to be the final choice and is implemented. This is a recipe for disaster, but a recipe that is often followed.

The UK study shows that strong interests and incentives exist in project development to present projects as favourably as possible, i.e. with benefits emphasized and costs and risks de-emphasized. Local authorities, developers and land owners, local trade unions, politicians, officials, MPs and consultants all stand to benefit from a project that looks favourable on paper, and they have little incentive to actively avoid bias in estimates of benefits, costs and risks. National bodies, e.g. some sections of the Department for Transport and the Ministry of Finance, who fund and oversee projects, may have an interest in more realistic appraisals, but so far they have had little success in achieving such realism. The situation may, however, be changing, with the initiatives to curb bias set out in HM Treasury (2003) and in the UK Department for Transport (2006).

Wachs (1986, 1990) found similar results for transit planning in the USA. Taken together, the UK and US studies account well for existing data on cost underestimation and benefit overestimation. Both studies falsify the notion that in situations with high political and organizational pressure, the underestimation of costs and overestimation of benefits is caused by non-intentional technical error, or optimism bias. Both studies support the view that in such situations, promoters and forecasters intentionally use the following formula in order to secure approval and funding for their projects:

Underestimated costs
\+ Overestimated benefits
= Project approval

Using this formula and thus "showing the project at its best", as one interviewee said above, results in an inverted Darwinism, i.e. the "survival of the un-fittest". It is not the best projects that get implemented, but the projects that look best on paper. And the projects that look best on paper are the projects with the largest cost underestimates and benefit overestimates, all things being equal. But these are the worst, or "un-fittest", projects, in the sense that they are the very projects that will encounter most problems during construction and operations, in terms of the largest cost overruns, benefit shortfalls and risks of non-viability.

Reference class forecasting

When contemplating what planners can do to improve project development, there is a need to distinguish between two different situations: (1) planners and promoters consider it important to obtain accurate forecasts of costs, benefits and risks; and (2) planners and promoters do not consider it important to get forecasts right, because optimistic forecasts are seen as a necessary means of getting projects started. The first situation is the easier one to deal with, and here better methodology will go a long way to improve planning and decision-making. The second situation is more difficult. Here, changed incentives are essential in order to reward honesty and punish deception: today's incentives often do the exact opposite.

Thus two main measures of improvement are as follows: (1) better forecasting methods; and (2) improved incentive structures, the latter being the more important. Better forecasting methods are covered in this section, better incentives in the next.

If planners genuinely consider it important to obtain accurate forecasts, it is recommended that they use a new forecasting method, "reference class forecasting", to reduce inaccuracy and bias. This method was originally developed to compensate for the type of cognitive bias in human forecasting that Princeton psychologist Daniel Kahneman found in his Nobel prize-winning work on economic forecasting and decision making (Kahneman, 1994; Kahneman and Tversky, 1979a). Reference class forecasting has proved to be more accurate than conventional forecasting and the method has been endorsed by the American Planning Association (2005):

> APA [the American Planning Association] encourages planners to use reference class forecasting in addition to traditional methods as a way to improve accuracy. The reference class forecasting method is beneficial for non-routine projects such as stadiums, museums, exhibit centres and other local one-off projects. Planners should never rely solely on civil engineering technology as a way to generate project forecasts.

For reasons of space, only an outline of the method is presented here, based mainly on Lovallo and Kahneman (2003) and Flyvbjerg (2006).

Reference class forecasting consists in taking an “outside view” on the particular project being forecast. The outside view is established on the basis of information from a class of similar projects. The outside view does not try to forecast the specific uncertain events that will affect the particular project but instead places the project in a statistical distribution of outcomes from this class of reference projects. Reference class forecasting requires the following three steps for the individual project:

1. Identify a relevant reference class of past projects. The class must be broad enough to be statistically meaningful, but narrow enough to be truly comparable with the specific project.
2. Establish a probability distribution for the selected reference class. This requires access to credible, empirical data for a sufficient number of projects within the reference class to make statistically meaningful conclusions.
3. Compare the specific project with the reference class distribution, in order to establish the most likely outcome for the specific project.

Daniel Kahneman relates the following story about curriculum planning, to illustrate reference class forecasting in practice (Lovallo and Kahneman, 2003). Some years ago, Kahneman was involved in a project to develop a curriculum for a new subject area for high schools in Israel. The project was carried out by a team of academics and teachers, who began to discuss how long the project would take to complete. Everyone on the team was asked to write on a slip of paper the number of months needed to finish and report the project. The estimates ranged from 18 to 30 months. One of the team members – a distinguished expert in curriculum development – was then posed a challenge by another team member to recall as many projects as possible that were similar to theirs and remember these projects as they were at a similar stage as their own project. “How long did it take them from that point to reach completion?” the expert was asked. After a while he answered, with some discomfort, that not all the comparable teams actually completed their task. About 40% of them eventually gave up. Of those remaining, the expert could not think of any that completed their task in less than 7 years or more than 10 years. The expert was then asked whether he had reason to believe that the present team was more skilled in curriculum development than the earlier ones had been. The expert said no, he did not see any relevant factor that distinguished this team favourably from the teams he had been remembering. His impression was that the present team was slightly below average in terms of resources and potential. The wise decision at this point would probably, according to Kahneman, have been for the team to break up. Instead, the members ignored the pessimistic information and proceeded with the project. They finally completed the project 8 years later, and their efforts were largely wasted – the resulting curriculum was rarely used.

In this example, the curriculum expert made two forecasts for the same problem and arrived at very different answers. The first forecast was the inside view; the second was the outside view, or the reference class forecast. The inside view is the one that the expert and other team members adopted. They made forecasts by focusing tightly on the case in hand, considering its objective, the resources they brought to it and the obstacles to its completion. They constructed in their minds scenarios of anticipated progress and extrapolated current trends into the future. The resulting forecasts, even the most conservative, were overly optimistic. The outside view is that provoked by the question put to the curriculum expert. It ignored the details of the project in hand and involved no attempt at forecasting the events that would influence the project's future course. Instead, it examined the experiences of a class of similar projects, laid out a rough distribution of outcomes for this reference class and then positioned the current project in that distribution. The resulting forecast, as it turned out, was much more accurate.

Similarly, to take an example from practical infrastructure planning, city planners preparing to build a new subway would first establish a reference class of comparable projects. This could be the relevant rail projects from the sample used for this article. Through analyses, the planners would establish that the projects included in the reference class were indeed comparable. Secondly, if the planners were concerned with getting construction cost estimates right, they would establish the distribution of outcomes for the reference class regarding the accuracy of construction cost forecasts. Figure 8.1 shows what this distribution looks like for a reference class relevant to building subways in the UK, developed by Flyvbjerg and Cowi (2004) for the UK Department for Transport. Thirdly, the planners would compare their subway project to the reference class distribution. This would make it clear to the planners that unless they have reason to believe they are substantially better forecasters and planners than their colleagues who did the forecasts and planning for projects in the reference class, they are likely to grossly underestimate construction costs. Finally, planners would then use this knowledge to adjust their forecasts for more realism. Figure 8.2 shows such adjustments for the UK situation and that, for a forecast of construction costs for a rail project, this forecast would have to be adjusted upwards by 40%, if investors were willing to accept a risk of cost overrun of 50%. If investors were willing to accept a risk of overrun of only 10%, the uplift would have to be 68%. For a rail project initially estimated at, say £4 billion, the uplifts for the 50 and 10% levels of risk of cost overrun would be £1.6 billion and £2.7 billion, respectively.

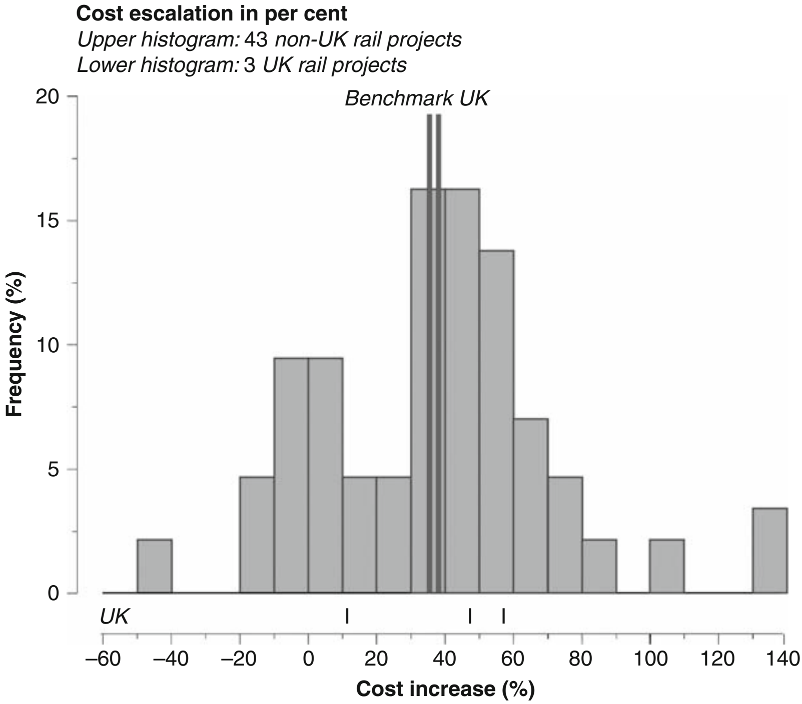


Figure 8.1 Inaccuracy of construction cost forecasts for rail projects in reference class. Average cost increase is indicated for non-UK and UK projects separately. Constant prices

The contrast between inside and outside views has been confirmed by systematic research (Gilovich et al., 2002). The research shows that when people are asked simple questions requiring them to take an outside view, their forecasts become significantly more accurate. However, most individuals and organizations are inclined to adopt the inside view in planning major initiatives. This is the conventional and intuitive approach, for lay people and experts alike. The traditional way to think about a complex project is to focus on the project itself and its details, to bring to bear what one knows about it, paying special attention to its unique or unusual features and trying to predict the events that will influence its future. The thought of going out and gathering simple statistics about related cases seldom enters a planner's mind. This is usually the case, according to Lovallo and Kahneman (2003). It is certainly the case for cost and benefit forecasting in large infrastructure projects. The first reference class forecast in practical policy and planning was carried out for the Edinburgh Tram in 2004 (Flyvbjerg, 2006).

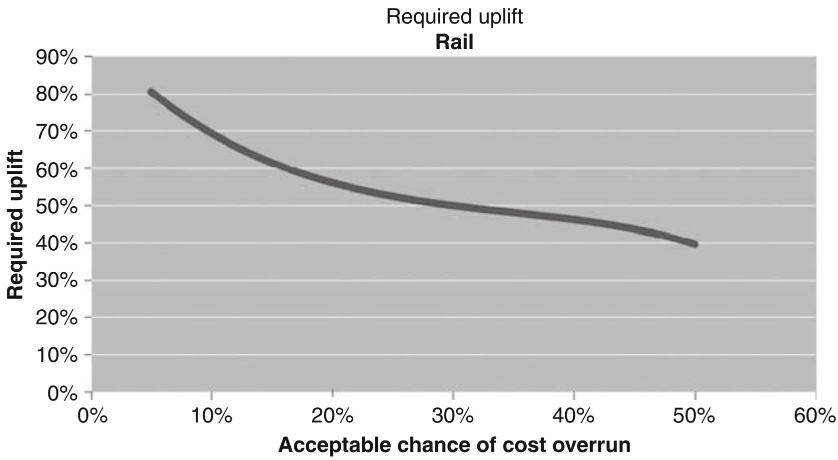


Figure 8.2 Required adjustments to cost estimates for UK rail projects as function of the maximum acceptable level of risk for cost overrun. Constant prices

While understandable, planners' preference for the inside view over the outside view is unfortunate. When both forecasting methods are applied with equal skill, the outside view is much more likely to produce a realistic estimate. This is because it bypasses cognitive and political biases, such as optimism bias and strategic misrepresentation and cuts directly to outcomes. In the outside view, planners and forecasters are not required to make scenarios, imagine events or gauge their own and others' levels of ability and control, so they cannot get all these things wrong. The outside view, being based on historical precedent, may fail to predict extreme outcomes, i.e. those that lie outside all historical precedents, but, for most projects, it will produce more accurate results. In contrast, a focus on inside details is the road to inaccuracy.

The comparative advantage of the outside view is most pronounced for non-routine projects – that planners and decision-makers in a certain locale have never attempted before, e.g. building an urban rail system in a city for the first time, or a new major bridge or an opera house where none existed before. It is in the planning of such new efforts that the biases towards optimism and strategic misrepresentation are likely to be largest. Choosing the right reference class of comparative past projects may be more difficult when planners are forecasting initiatives for which precedents are not easily found, e.g. the introduction of new and unfamiliar technologies. However, most large infrastructure projects are both non-routine locally and use well-known technologies. Such projects are, therefore, particularly likely to benefit from the outside view and reference class forecasting. The same applies to concert halls, museums, stadiums, exhibition centres and other local one-off projects.

How to avoid lock-in and premature closure in early-stage development

The typical reference class forecast, as described above, is based upon data on performance for a minimum of 10–15 comparable projects. But for a project in early-stage development, data for this many projects will not often have been collected. Nevertheless, an outside view in early-stage development may help planners be both more realistic from the outset about risks and avoid lock-in. Taking an outside view in early-stage development may be done by carefully selecting a restricted sample of projects and information, which would then constitute a set of "reference points", if not a reference class. The reference points may later form part of a reference class and be used for reference class forecasting.

Kahneman and Tversky (1979b) argue that the major source of error in planning and forecasting future prospects is the prevalent tendency with planners – be they experts or lay people – to underweigh or ignore distributional information. Planners "should therefore make every effort to frame the [planning and] forecasting problem so as to facilitate utilizing all the distributional information that is available", say Kahneman and Tversky (1979b: p. 316). This may be considered the single most important piece of advice regarding how to get better performance in planning and forecasting through improved methods. Incentives should accordingly be established to reward planners who frame planning and forecasting in terms of all available distributional information and to punish planners who ignore such information.

For early-stage planning, this advice, which is firmly grounded in Nobel Prize-winning theories of decision making, is particularly important. This is so because early-stage planning is particularly vulnerable to planners and promoters underweighing and ignoring distributional information. No other stage in the project cycle is more susceptible to premature closure, lock-in, path dependence, anchoring, overconfidence, group think and similar problematic behaviour that all result in ignorance of relevant distributional information and thus in inadequate project preparation (Flyvbjerg et al., forthcoming).

Enforcing the outside view is an antidote to such behaviour. The outside view may be implemented by reference class forecasting, as we saw above. In early-stage development, where a full reference class forecast is often not possible, a simplified version of the outside view may be implemented by employing maximum variation or extreme case sampling techniques as described by Flyvbjerg (2007). The basic idea is to capture the range of variation that is relevant to the decision at hand and thus not to discard pertinent distributional information.

For instance, an outside view of early-stage development of a given project could be established by collecting data on just two comparable projects, one of which was a significant success and one of which was a significant disaster. This would help early-stage planners get an initial idea of the following:

1. the range of risk involved, from success to disaster
2. the possible causes of success and failure
3. the potential benefits from considering alternatives.

As a case in point, consider planners and promoters in early-stage deliberation over whether to build a particular example of world-class architecture, e.g. a new opera house, concert hall, museum or skyscraper designed on the dictum of signature architecture, "build it and they will come". This has recently been the situation for more and more cities, as they jockey for position in an increasingly international and fiercely competitive urban hierarchy. Such planners and promoters could benefit from considering as reference points the extreme cases of the Sydney Opera House and the Guggenheim Museum in Bilbao and from understanding how the first became a complete disaster in terms of delays, cost overruns (a record-setting 1400%) and destroying the career of its architect, Jørn Utzon, while the second, which is no less complex and innovative, was built on time and budget and catapulted its architect, Frank Gehry, into a career of international stardom (Flyvbjerg, 2005a). Other examples could be used, but the basic idea should be clear: the outside view and a small set of reference points may help establish a much-needed reality check on the project concept in early-stage development, in terms of risk and robustness and thus help prevent lock-in and premature closure.

Public and private sector accountability

This section covers the situation where planners and promoters do not find it important to get forecasts right and therefore do not help to clarify and mitigate risk but, instead, generate and exacerbate it. Here planners are part of the problem, not the solution. This situation needs some explanation, because it may sound like an unlikely state of affairs. After all, it is generally agreed that planners ought to be interested in being accurate and unbiased in forecasting. The requirement for accuracy is even stated as an explicit requirement in planners' code of ethics, e.g. the code of the American Institute of Certified Planners (AICP), which states that "A planner must strive to provide full, clear and accurate information on planning issues to citizens and governmental decision-makers" (American Planning Association, 1991). The British Royal Town Planning Institute (RTPI, 2001) has laid down similar obligations for its members. The literature is replete with things planners and planning "must" strive to do but which they do not. Planning must be open and communicative but is often closed. Planning must be participatory and democratic but is often an instrument of domination and control. Planning must be about rationality but is often about power (Flyvbjerg, 1998; Watson, 2003). This is the "dark side" of planning and planners, identified by Flyvbjerg (1996) and Yiftachel (1998) and which is remarkably under-explored by planning researchers and theorists.

Forecasting, too, has its dark side. It is here that "planners lie with numbers", as Wachs (1989) has aptly put it. Planners on the dark side are busy, not with getting forecasts right and following the AICP or RTPI Code of Ethics but with getting projects funded and built. Accurate forecasts are often perceived to be an ineffective means for achieving this objective. Indeed, accurate forecasts may be seen as counter-productive, whereas false forecasts may be seen as effective in competing for funds and securing the go-ahead for construction. "The most effective planner is sometimes the one who can cloak advocacy in the guise of scientific or technical rationality" (Wachs, 1989). Such advocacy would stand in direct opposition to AICP's ruling that "the planner's primary obligation [is] to the public interest" (American Planning Association, 1991).

Nevertheless, seemingly rational forecasts that underestimate costs and overestimate benefits have long been an established formula for project approval, and if members of the AICP or RTPI have been excluded for producing inaccurate and biased forecasts, this is seldom broadcast. Enforcement of planners' code of ethics is lax compared with other professions. If, for instance, medical doctors repeatedly offered diagnoses and prognoses for their patients that were as inaccurate as planners' forecasts of costs and benefits, these doctors would quickly be struck off and out of work. For planners, forecasting is too often not a professional activity but just another kind of rent-seeking behaviour, resulting in a make-believe world of misrepresentation which makes it extremely difficult to decide which projects deserve undertaking and which do not. The consequence is, as even one of the industry's own bodies, the Oxford-based Major Projects Association, acknowledges, that too many projects go ahead which should not. Conversely, many projects that probably should do not proceed, had they not lost out to projects with "better" misrepresentation (Flyvbjerg et al., 2002).

In this situation, the question is not so much about what planners can do, both in early project development and later on, to reduce inaccuracy and risk, but what others can do to impose on planners the checks and balances that might give them the incentive to stop producing biased forecasts and begin to work according to their code of ethics. The challenge is to change the power relations that govern project development and forecasting. Better methods and appeals to ethics are insufficient; institutional change, with a focus on transparency and accountability, is necessary.

Flyvbjerg et al. (2003) argue that two basic types of accountability define liberal democracies: (1) public sector accountability through transparency and public control and (2) private sector accountability via competition and market control. Both types of accountability may be effective tools to curb planners' misrepresentation in forecasting and to promote a culture which acknowledges and deals effectively with risk. In order to achieve accountability through *transparency and public control*, the following would be required as practices embedded in the relevant institutions:

- National-level government should not offer discretionary grants to local infrastructure agencies for the sole purpose of building a specific type of infrastructure. Such grants create perverse incentives. Instead, national government should simply offer "infrastructure grants" or "transportation grants" to local governments and let local political officials spend the funds however they choose to but make sure that every pound they spend on one type of infrastructure reduces their ability to fund another.
- Forecasts should be made subject to independent peer review, e.g. by national or state accounting and auditing offices.
- Forecasts should be benchmarked against comparable forecasts, e.g. reference class forecasting, as described in the previous section.
- Forecasts, peer reviews and benchmarking should be made available to the public as they are produced, including all relevant documentation.
- Public hearings, citizen juries and the like should be organized, to allow stakeholders and civil society to voice criticism and support.
- Scientific and professional conferences should be organized, where planners present and defend their forecasts in the face of colleagues' scrutiny.
- Projects with inflated benefit-cost ratios should be reconsidered and stopped, if re-calculated costs and benefits do not warrant implementation. Projects with realistic estimates of benefits and costs should be rewarded.
- Professional and occasionally even criminal, penalties should be enforced for planners and forecasters who consistently and knowingly produce deceptive forecasts (Garett and Wachs, 1996). Malpractice in planning should be taken as seriously as it is in other professions.

In order to achieve accountability in forecasting via *competition and market control*, the following would be required, again as practices that are both embedded in, and enforced by, the relevant institutions:

- The decision to go ahead with a project should, where possible and appropriate, be made contingent on the willingness of private financiers to participate, without a sovereign guarantee for at least one third of the total capital needs. This should be required whether projects are subsidized or not. Private lenders, shareholders and stock market analysts would produce their own forecasts or would critically monitor existing ones. If they get the forecasts wrong, they and their organizations would be hurt. The result would be more realistic forecasts and reduced risk.
- Forecasters and their organizations must share financial responsibility for covering cost overruns and benefit shortfalls resulting from misrepresentation and bias in forecasting.
- The participation of risk capital should not mean that government gives up or reduces control of the project. On the contrary, it means that government can more effectively play the role it should be playing, namely, as the ordinary citizen's guarantor for ensuring concerns about safety, environment, risk and a proper use of public funds.

Whether projects are public, private, or public-private, they should be vested in one and only one, project organization, with a strong governance framework. The project organization may or may not be a company, may be public or private, or a mixture. What is important is that this organization enforces accountability vis-à-vis contractors, operators, etc. and that, in turn, the directors of the organization are held accountable for any cost overruns, benefits shortfall, faulty designs, unmitigated risks, etc. that may occur during project planning, implementation and operations.

If the institutions with responsibility for developing and building major infrastructure projects were to effectively implement, embed and enforce such measures of accountability, then the misrepresentation in cost, benefit and risk estimates, which is widespread today, could be mitigated. If this is not done, misrepresentation is likely to persist and the allocation of funds for infrastructure would continue to be wasteful and undemocratic.

Clearly, all the measures of accountability mentioned above cannot and will not be in place in early project development. Their implementation is a gradual process that takes place throughout the project cycle. Nevertheless, it is important that measures of accountability are implemented as early as possible and that planners and promoters with early-stage responsibilities know that their actions will eventually be judged by these measures. If accountability is not enforced, the high frequency of project failure documented above is likely to continue.

Signs of improvement

Fortunately, after decades of widespread mismanagement of major project development, signs of improvement have recently appeared. The conventional consensus, that deception is an acceptable way of getting projects started, is under attack, as will be apparent from the examples below. This is in part because the largest projects are now so big in relation to national economies that cost overruns, benefit shortfalls and risks from even a single project may destabilize the finances of a whole country or region. This occurred with the 2004 Olympics in Athens and the new international airport in Hong Kong (Flyvbjerg, 2005a). Law-makers and governments begin to see that national fiscal distress is too high a price to pay for the conventional way of developing large projects.

Moreover, with private finance in major projects on the rise, capital funds and banks are increasingly gaining a voice in the project development process. Private investors place their own funds at risk, as opposed to governments, who place the taxpayer's money at risk. Capital funds and banks do not automatically accept at face value the forecasts of project planners and promoters but typically bring in their own advisers to do independent forecasts and risk assessments – an important step in the right direction.

Finally, democratic governance is generally getting stronger around the world. The Enron scandal and its successors have triggered new legislation and a war on corporate deception that is spilling over into government with the same objective: to curb financial waste and promote good governance. Although progress is slow, good governance is gaining a foothold even in major project development. The main drivers of reform come from outside the agencies and industries conventionally involved in major project development, which increases the likelihood of success.

In 2003, the Treasury of the United Kingdom required, for the first time, that all ministries develop and implement procedures for large public projects to curb what the Treasury calls "optimism bias". Funding will be unavailable for projects that do not take this bias into account, and methods have been developed for dealing with this (Flyvbjerg and Cowi, 2004; HM Treasury, 2003; UK Department for Transport, 2006). In the Netherlands, the Parliamentary Committee on Infrastructure Projects conducted extensive public hearings for the first time in 2004. These were to identify measures that would limit the misinformation about large infrastructure projects given to Parliament, the public and media (Tijdelijke Commissie Infrastructuurprojecten, 2004). In Boston, the government has sued to recoup funds from contractor overcharges for the Big Dig related to cost overruns. More countries and cities are likely to follow the lead of the UK, the Netherlands and Boston in coming years; Switzerland and Denmark are already doing so (Danish Ministry for Transport and Energy, 2006; Swiss Association of Road and Transportation Experts, 2006). It is too early to tell whether the measures they implement will ultimately be effective. It seems unlikely, however, that the forces that have triggered the measures will be reversed, and it is those forces that reform-minded groups need to support and work with, in order to curb deception and waste. This is the "tension-point" where convention meets reform, power-balances change, and new things may happen.

The key weapons in the war on deception and waste are accountability and critical questioning. The professional expertise of planners, engineers, architects, economists and administrators is indispensable in constructing the infrastructures that make society work. However, the claims relating to costs, benefits and risks made by these groups cannot always be trusted and should be carefully examined by independent specialists and organizations. The same applies to claims made by project-promoting politicians and officials. Institutional checks and balances, including financial, professional or even criminal penalties for consistent and unjustifiable bias in cost estimation, benefits and risks should be employed. The key principle, often violated today, is that the cost of making a wrong forecast, or a wrong decision, should fall on those making that forecast or decision.

The culture of misinformation in major project development has historical roots which are deeply ingrained in professional and institutional practices. It would be naive to think that this culture is easily toppled. Given the stakes involved – saving taxpayers from billions of dollars of waste, protecting citizens' trust in democracy and the rule of law and avoiding the destruction of spatial and environmental assets – it should not deter us from trying.

Notes

1. For details on data and methodology, see Flyvbjerg et al. (2002).
2. For details on data and methodology, see Flyvbjerg (2005b).

References

American Planning Association. (1991). *AICP Code of Ethics and Professional Conduct*. Adopted October 1978, amended October 1991. http://www.planning.org

American Planning Association. (2005). *JAPA Article Calls on Planners to Help End Inaccuracies in Public Project Revenue Forecasting*. http://www.planning.org/newsreleases/2005/apr07.htm Accessed 20 November 2008.

Anguera R. (2006). The channel tunnel: an ex post economic evaluation. *Transportation Research Part A* **40**:291–315.

Ascher W. (1979). *Forecasting: An Appraisal for Policy-makers and Planners*. The Johns Hopkins University Press: Baltimore.

Bok S. (1979). *Lying: Moral Choice in Public and Private Life*. Vintage: New York.

Buehler R, Griffin D, and MacDonald H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and Social Psychology Bulletin* **23** (3): 238–247.

Buehler R, Griffin D, and Ross M. (1994). Exploring the "planning fallacy": why people underestimate their task completion times. *Journal of Personality and Social Psychology* **67**:366–381.

Cliffe L, Ramsey M, and Bartlett D. (2000). *The Politics of Lying: Implications for Democracy*. Macmillan: London.

Danish Ministry for Transport and Energy. (2006). *Aktstykke om nye budgetteringsprincipper (Act on New Principles for Budgeting)*. Aktstykke 16: Finansudvalget, Folketinget, October 24.

Flyvbjerg B. (1996). The dark side of planning: rationality and realrationalität. In: Mandelbaum S, Mazza L, and Burchell R (eds). *Explorations in Planning Theory*. Center for Urban Policy Research Press: New Brunswick, NJ, pp. 383–394.

Flyvbjerg B. (1998). *Rationality and Power: Democracy in Practice*. University of Chicago Press: Chicago.

Flyvbjerg B. (2005a). Design by deception: the politics of megaproject approval. *Harvard Design Magazine* **22** (Spring/Summer): 50–59.

Flyvbjerg B. (2005b). Measuring inaccuracy in travel demand forecasting: methodological considerations regarding ramp up and sampling. *Transportation Research A* **39** (6): 522–530.

Flyvbjerg B. (2006). From Nobel prize to project management: getting risks right. *Project Management Journal* **37** (3): 5–15.

Flyvbjerg B. (2007). Five misunderstandings about case-study research. In: Seale C et al. (eds). *Qualitative Research Practice: Concise Paperback Edition*. Sage: London and Thousand Oaks, CA, pp. 390–404.

Flyvbjerg B, Bruzelius N, and Rothengatter W. (2003). *Megaprojects and Risk: An Anatomy of Ambition*. Cambridge University Press: Cambridge, UK.

Flyvbjerg B and Cowi. (2004). Procedures for dealing with optimism bias. In: *Transport Planning Guidance Document*. Department for Transport: London, UK.

Flyvbjerg B, Massimo G, and Lovallo D. *Delusion and Deception in Large Infrastructure Projects: Two Models for Explaining Executive Disaster*, forthcoming.

Flyvbjerg B et al. (2002). Underestimating costs in public works projects: error or lie? *Journal of the American Planning Association* **68** (3, Summer): 279–295.

Flyvbjerg B et al. (2005). How (In)accurate are demand forecasts in public works projects? The case of transportation. *Journal of the American Planning Association* **71** (2, Spring): 131–146.

Garett M and Wachs M. (1996). *Transportation Planning on Trial: The Clean Air Act and Travel Forecasting*. Sage: Thousand Oaks, CA.

Gilovich T, Griffin D, and Kahneman D (eds). (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press: Cambridge, UK.

Kahneman D. (1994). New challenges to the rationality assumption. *Journal of Institutional and Theoretical Economics* **150**:18–36.

Kahneman D and Lovallo D. (1993). Timid choices and bold forecasts: a cognitive perspective on risk taking. *Management Science* **39**:17–31.

Kahneman D and Tversky A. (1979a). Prospect theory: an analysis of decisions under risk. *Econometrica* **47**:313–327.

Kahneman D and Tversky A. (1979b). Intuitive prediction: biases and corrective procedures. In: Makridakis S and Wheelwright SC (eds). *Studies in the Management Sciences: Forecasting*. Amsterdam: North Holland, p. 12.

Lovallo D and Kahneman D. (2003). Delusions of success: how optimism undermines executives' decisions. *Harvard Business Review* (July): 56–63.

Morris P and Hough G. (1987). *The Anatomy of Major Projects: A Study of the Reality of Project Management*. John Wiley and Sons: New York.

Newby-Clark I, McGregor I, and Zanna M. (2002). Thinking and caring about cognitive inconsistency: when and for whom does attitudinal ambivalence feel uncomfortable? *Journal of Personality and Social Psychology* **82**:157–166.

Royal Town Planning Institute. (2001). *Code of Professional Conduct*. As last amended by the Council on 17 January 2001. http://www.rtpi.org.uk

Swiss Association of Road and Transportation Experts. (2006). *Kosten-Nutzen-Analysen im Strassenverkehr*. Grundnorm 641820, valid from August 1. Swiss Association of Road and Transportation Experts: Zürich.

Tijdelijke Commissie Infrastructuurprojecten. (2004). *Grote Projecten Uitvergroot: Een Infrastructuur voor Besluitvorming*. Tweede Kamer der Staten-Generaal: The Hague.

Treasury HM. (2003). *The Green Book: Appraisal and Evaluation in Central Government*. Treasury Guidance, TSO: London.

UK Department for Transport. (2006). *Changes to the Policy on Funding Major Projects*. Department for Transport: London.

Vanston J and Vanston L. (2004). Testing the tea leaves: evaluating the validity of forecasts. *Research-Technology Management* **47** (5): 33–39.

Wachs M. (1986). Technique vs. advocacy in forecasting: a study of rail rapid transit. *Urban Resources* **4** (1): 23–30.

Wachs M. (1989). When planners lie with numbers. *Journal of the American Planning Association* **55** (4): 476–479.

Wachs M. (1990). Ethics and advocacy in forecasting for public policy. *Business and Professional Ethics Journal* **9** (1–2): 141–157.

Watson V. (2003). Conflicting rationalities: implications for planning theory and ethics. *Planning Theory and Practice* **4** (4): 395–408.

Yiftachel O. (1998). Planning and social control: exploring the dark side. *Journal of Planning Literature* **12** (4): 395–406.

9

Decision Behaviour – Improving Expert Judgement

Geir Kirkebøen

> Expert judgments have been worse than those of the simplest statistical models in virtually all domains that have been studied.
>
> (Camerer and Johnson, 1991)

> In nearly every study of experts carried out within the judgement and decision-making approach, experience has been shown to be unrelated to the empirical accuracy of expert judgements.
>
> (Hammond, 1996)

> Professionals are not stupid, but they are human.
>
> (Bazerman, 2006)

In this chapter, common biases in professionals' judgement and decision-making, and how these deviations can be corrected, are explained. A principal reason for these decision biases is human beings' cognitive limitations. Due to these limitations, simplifying strategies are often used, producing predictable biases. Emotional and motivational factors may also contribute to reduced decision quality. Often, experience does not improve the quality of professionals' judgement and decision-making, because they do not receive accurate and timely feedback. No easy recipe for eliminating decision biases exists. However, simple cognitive strategies like "take an outsider's view" and "consider the opposite" are efficient in many circumstances. Research findings suggest a wide use of technological decision support within many professions. In particular, professionals' judgements in repeatable decision-making situations should be automated to a far greater extent than is the case today.

Introduction

People's judgements are not perfect. Even experts make mistakes. Interestingly, mistakes in judgement and decision-making are not random. Professionals deviate systematically from normative standards. Per definition, systematic biases result in reduced decision quality, and accordingly, may have serious consequences. However, the fact that judgement biases are systematic also gives hope. If the biases and their causes can be identified, it should be possible to correct them.

The typical biases in our decision behaviour are well described in the psychology of judgement and decision-making (henceforth JDM). This chapter first gives an overview of central findings and theories within JDM. Typical biases in judgement and decision-making are then presented. Finally, some acknowledged methods for improving professionals' judgements and decision-making are discussed.

Human decision-making – central findings

Decision-making is an essential part of all projects in life. Decisions often involve risk or uncertainty with respect to the outcome. A central topic within JDM is how individuals behave when faced with a risky choice. Historically, the descriptive study of decision behaviour is closely linked to the normative question of how we should behave. In their monumental work *Theory of games and economic behavior*, Von Neumann and Morgenstern (1947) put forward a very influential answer to this normative question. In 1947, a central assumption in economics had, for a long time, been that economic agents make rational decisions. But what does that mean? What is rational decision behaviour?

Von Neumann and Morgenstern put forward a mathematical theory as an answer to this question. The essence of their so-called Expected Utility (EU) theory is a small set of simple axioms. Von Neumann and Morgenstern prove that if you make choices in accordance with these basic axioms, then it is possible to attach a particular utility (or personal value) to each possible outcome of the choice, in such a way that one alternative is preferred to another, if, and only if, the expected utility (EU) of that alternative is higher than the expected utility of the other alternatives. The expected utility of each choice alternative under consideration is calculated by multiplying the (estimated) utility of each of the potential consequences of outcomes by their probabilities of occurrence and then adding up all the component products.

The EU-theory is a model of decision under risk, i.e. the theory assumes that the probability for each outcome is objective or given. Savage's (1954) *Subjective Expected Utility* (SEU) theory extended expected utility from (objective) risk to subjective (or personal) probabilities of outcomes. Thus, SEU is the generalization of EU from risk to uncertainty in general.

Bounded rationality

The (S)EU theory was soon subject to debate. Is this or that axiom of the theory reasonable as a normative principle? Do people make their choices in accordance with the principle? Do people make decisions in accordance with what the SEU-theory as a whole prescribes?

According to the SEU theory, to make a rational choice requires one to first map out all the alternatives, find the possible outcomes of each of them, estimate the probability for each of these outcomes, make clear their consequences, estimate the utility of the consequences of each outcome relative to each other, calculate the expected utility of each alternative and finally select the alternative with the highest expected utility. Do people perform all this work when they make their decisions?

In the 1950s, Herbert Simon, a subsequent Nobel Prize winner in Economics, was the most pronounced critic of the assumption that the SEU theory gives a relevant description of human rationality. According to Simon, people rarely have access to all the information they need in order to make choices in the way the SEU theory prescribes. In addition, the ability to process information is usually far too limited to follow the SEU theory's prescriptions. Motivated by these insights, Simon (1955) launches the concept of bounded rationality. He argues that "the task is to replace the global rationality of economic man with a kind of rational behaviour that is compatible with the access to information and the computational capacities that are actually possessed [by humans]". A year later, Simon follows this up with another central idea in modern decision psychology. He claims, "However adaptive the behaviour of organisms in learning and choice situations, this adaptiveness falls far short of the ideal of maximizing in economic theory. Evidently, organisms adapt well enough to 'satisfice'; they do not, in general, optimize" (1956).

Simon's bounded rationality perspective sets the agenda for later decision psychology, namely, to uncover what particular "satisficing" strategies are made use of in different choice situations and how decision-makers, as a consequence, deviate from normative models, such as the SEU theory.

Expert judgement versus formula

The first quotation in the chapter introduction summarizes one of the most spectacular findings in JDM research: simple statistical models integrate information systematically better than intelligent and experienced human experts. Paul Meehl (1954) was the pioneer behind this discovery. He reviewed research comparing clinicians' judgement or interpretation of data with simple regression equations integration of the same data. (Meehl drew a sharp distinction between collecting data and interpreting data.) The equations outperformed the clinicians in most of the 20 studies reviewed by Meehl. In none of the studies did the clinicians perform better than the equation.

One reason why Meehl's finding was surprising is that his linear equations did not incorporate what clinicians believed characterize their own judgement, namely that the weight they put on a particular variable's value depends on the values of the other variables involved. Still, the equations systematically outperformed or equalized with the clinicians' "holistic" judgements.

In the comparisons undertaken by Meehl, the models and the clinicians had access to the same information. However, even if the human judge has access to more information than the equation, the equation often outperforms the judge. For example, it is common practice in the USA that a committee of experienced educational professionals is involved in the admission process to college. These committees normally have access to the student's grades, a letter of recommendation from the student's high school, and information on the quality of that particular high school. The committee also undertakes a lengthy interview with each candidate. Dawes (1971) worked out a regression equation, based on registered data, that predicted how well students would perform in college. The only variables in that model were the student's Grade Point Average (GPA), a letter of recommendation (scored 1 to 5) and the quality of the high school the student came from (also scored 1 to 5). Dawes applied the regression equation on 384 new applicants. He found that the model eliminated 55% of the weakest students, without eliminating any of those accepted by the committee. In addition, the model predicted more accurately than the committee how the students would perform at college. Dawes also found that only the students' GPA was sufficient to outperform the committee's predictions.

Since 1954, hundreds of studies have been conducted in many different professional domains, under the conditions set by Meehl for comparing expert judgement and equations. In hardly any study has the expert beaten the model. It is now a well-established fact that simple models consistently outperform intelligent and experienced professionals' judgements (e.g. Dawes et al., 1989). Why is it so?

Heuristic and biases

> *The prevalence of breast cancer is 1% for woman over the age of 40. A widely used test, mammography, gives a positive result in 10% of women without breast cancer, and in 80% of women with breast cancer. What is the probability that a woman in this age bracket who tests positive actually has breast cancer?*

David Eddy (1982) asked experienced doctors this question. The majority of the doctors (95%) estimated the probability for breast cancer to be between

70 and 80%. The correct answer (according to Bayes' formula) is c. 7.5%. Why do highly educated and experienced doctors get the answer so wrong to a kind of question they routinely answer in their work?

Kahneman and Tversky's influential *heuristic and biases* perspective in JDM answers this question. The rules of probability prescribe the optimal way to reason with uncertainty. However, to follow these rules can easily become too complicated. It thus becomes necessary, in Simon's terminology, to "satisfice". According to Kahneman and Tversky, reasoning with uncertainty is ruled by a number of satisficing strategies, or "judgement heuristics". The doctors in Eddy's study were asked to estimate how probable it is that a woman who has a positive mammography test, also has breast cancer. Kahneman and Tversky demonstrated that, when asked to judge whether a person (event or object) belongs to a particular category, "representative heuristics" is often used. The answer is based on how typical this person is judged to be of the category in question. According to the heuristic and biases perspective, the doctors in Eddy's study confused the probability for breast cancer with how likely it is that a woman with breast cancer also has a positive mammography test (as is the case in 80% of woman with cancer).

The doctors in Eddy's study deviated strongly from Bayes' formula. This formula states that judgement of the probability of breast cancer when the mammography test is positive, has to take into account the relation between the base rates, i.e. the relation between the frequency of breast cancer and positive mammography respectively, among 40-year-old women. This relation is little less than one to ten, i.e. almost ten times more women have a positive test than the number of women who have breast cancer. When the relation among the base rates is uneven, the use of representative heuristics in judgement of category-belonging, results in systematic biases in respect to the rules of probability.

What percentage of the housework do you contribute to at home? Or, what percentage of the total work did you perform in the last project you participated in? When spouses or project participants are asked such questions separately, the total adds up to far more than 100%. The use of "availability heuristics", another strategy that Tversky and Kahneman (1974) claim influences our probability judgements, can account for these phenomena. They suggest that judgement of how frequently something (X) takes place (relative to something else), is influenced by how easily available X is, i.e. how easy it is to notice, remember or imagine X. It is easier to register and remember one's own work than the work of others. For several reasons, one's own work is far more "available" in retrospect. It is therefore easy to overestimate how much work has been done personally.

The heuristic and biases perspective has been very influential, particularly after Kahneman and Tversky managed to incorporate this perspective on probability judgements into a descriptive alternative to the SEU theory.

Prospect theory

Imagine that the United States is preparing for the outbreak of an unusual Asian disease that is expected to kill 600 people. Two alternative programmes to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programmes are as follows:

> *If Programme A is adopted 200 people will be saved.*
>
> *If Programme B is adopted, there is a one third probability that 600 people will be saved and a two-thirds probability that no people will be saved.*

Which of the two programmes would you favour?

Tversky and Kahneman (1981) gave subjects this choice. Most chose Programme A (72%). They gave others the following option:

> *If Programme A′ is adopted, 400 people will die.*
>
> *If Programme B′ is adopted, there is a one third probability that no one will die and a two-thirds probability that 600 people will die.*

Most people (78%) then chose Programme B′.

Close examination of the two sets of programmes shows that they are objectively the same. Programme A has the same consequence as Programme A′, and Programme B has the same expected effect as Programme B[1]. Nevertheless, most individuals chose programme A in the first set and programme B′ in the second set. The subjects in this study were psychology students. The same study on professionals (doctors) gave similar results (McNeil et al., 1982). How are these findings explained?

Before the Asian disease study, decision psychology had largely been ignored by economists. Simon's bounded rationality perspective had not changed the economists' core assumption that economic actors act rationally. The kinds of decision behaviour that Simon considered as evidence for bounded rationality were typically explained away as rational adaptations to the cost of searching for more information. The Asian disease findings were not so easily explained away. It is a central assumption in the SEU theory that preferences should not be influenced by irrelevant aspects, e.g. the way in which identical alternatives are described. This is an essential aspect of the concept of rationality in economic theory. Kahneman and Tversky's Asian disease study demonstrated beyond doubt that this assumption does not hold. Positive and negative presentations of objectively identical alternatives can have a dramatic impact on decision behaviour.

A couple of years before the Asian disease study, Kahneman and Tversky (1979) published a psychological theory on decision-making that predicted the findings of this study. Their *Prospect theory* specifies how decision-makers systematically deviate from the SEU theory. Many of the individual elements in the theory had been known a long time before they published the

Prospect theory. Kahneman and Tversky's great contribution was that they were able to incorporate these findings in a precisely formulated theoretical alternative to the SEU theory. Kahneman described their strategy thus:

> The theory that we constructed was as conservative as possible ... We did not challenge the philosophical analysis of choices in terms of beliefs and desires that underlies utility theory, nor did we question the normative [status of] models of rational choice ... The goal ... was to assemble the minimal set of modifications of expected utility theory that would provide a descriptive account.
>
> (Kahneman and Tversky, 2000)

The main reason why Kahneman and Tversky were "as conservative as possible" was that they considered economists to be their main target. It was important to maintain most of the established approach, so that economists could easily recognize their alternative theory. They did. The paper in which the Prospect theory was first described is the second most cited paper in economics (Laibson and Zeckhauser, 1998). In 2002 Kahneman got the Nobel Prize in Economics, mainly due to the Prospect theory.

The central components in the Prospect theory are the value function, the decision weight function, and the idea of a variable reference point that tells what the decision-maker (in choice situations) experiences as the division between losses and gains. The value function specifies how decision-makers typically assess objective values, or more precisely, changes in objective values. The value function has four major characteristics. Firstly, the theory incorporates an insight that goes back to David Bernoulli's (1738) *law of diminishing returns*, i.e. the idea that fixed increments in cash (objective value) lead to ever smaller increments of perceived wealth (or utility). Secondly, the value function describes a similar declining curve related to experienced losses. Thirdly, the function shows that losses are twice as painful as gains are pleasurable. Fourthly, the Prospect theory assumes that when decisions are made, gains and losses are experienced relative to a subjective reference point, which is usually the status quo. The theory also assumes that this individual reference point is easily changed.

The Prospect theory's assumption on a subjective and changeable reference point explains the findings in the Asian disease study. If the subjects accept Programme A in the first choice, 200 people are definitely saved. This outcome usually becomes the subjects' reference point. When they then consider Programme B, they will experience a risk of losing these 200 lives. In the second choice, subjects typically frame the choice so that no death becomes the reference point. Decision-makers then experience that they may avoid the loss of 400 lives, by choosing alternative B′. The value function suggests that losses feel twice as painful as gains feel pleasurable. This, together with the change in reference point, explains why subjects typically choose Programme A and Programme B′ in the Asian disease study.

Thaler's (1980) discovery of the *endowment effect* was another early example on how people systematically deviate from the SEU theory in the way predicted by the Prospect theory. Thaler showed that the maximum amount people were willing to pay for particular goods was far less than the minimum amount they demand for selling the same goods. Buying price typically surpasses the selling price by a factor somewhat larger than two, exactly as Prospect theory's value function predicts.

Prospect theory is, like the SEU theory, a theory on choice under uncertainty. The SEU theory does not distinguish between probabilities, and how probabilities or uncertainties are experienced. Prospect theory's decision weight function specifies the systematic biases in people's (subjective) experience of objective uncertainty. The function suggests that people have a strong tendency to attribute far too much weight to small probabilities. In particular, they are very sensitive to the difference between impossibility and a tiny possibility. This sensitivity is named the *possibility effect*. There is usually a good fit between objective probabilities around 0.2 and experienced uncertainty. However, people are insensitive to differences in intermediate probabilities, with a tendency to underestimate probabilities in the interval 0.2 to 0.8. Moreover, they are very sensitive to changes from something almost certain, to completely certain, and vice versa. For example, when the objective probability of an outcome changes from certain (1.0) to 0.99, this has a far greater psychological impact than a similar change in probability from 0.7 to 0.69, a bias named the *certainty effect*.

Prospect theory's decision weight function predicts several typical deviations from what the SEU theory prescribes. For example, the possibility effect can account for people's willingness to participate in lotteries, a phenomenon that for a long time had been a paradox, when considered from a normative perspective on decision behaviour. This is one example of many that shows the relevance of the Prospect theory in the real world (e.g. Camerer, 2001).

Emotions

In the 1980s it was quite obvious that the SEU theory was not able to account for descriptive decision-making. Prospect theory was one response to this observation. As we saw above, Kahneman and Tversky were "as conservative as possible" when they constructed the theory. One consequence of their conservatism was that emotions, which had no explicit role in the SEU theory, were not incorporated into the Prospect theory either. An alternative strategy to explain deviations from the SEU theory was to supply the SEU framework with emotional variables. This was done by Loomes and Sugden (1982) in their *Regret theory*. They included expected or anticipated emotions (e.g., regret and rejoice) in the possible decision outcomes when explaining decision-makers' estimation of future expected utilities. They hoped to thereby account for people's systematic deviations from the SEU theory. Even if the Regret theory was not very successful, it is now beyond doubt that affective phenomena strongly influence our decision behaviour (e.g. Loewenstein and Lerner, 2003).

Much research evidence indicates that every stimulus evokes an affective evaluation (e.g. Zajonc, 1980). On meeting a person, viewing a house, or reading the first chapter of a book, the mind is immediately made up, as to whether one likes it or dislikes it. This first, affective (good/bad) reaction is swift, and is not always conscious. It may be difficult to explain, but the feeling is there, and it will often influence one's judgement of the person, the house, or the book. Another example: when the boss has decided to employ a new person, he will probably claim that, after serious examination, he has chosen the best qualified applicant. Psychological research shows that the person who finally gets the job is not necessarily the best qualified, but is very often one the boss immediately liked. Decisions to appoint new employees, as well as other decisions, are, to a large extent, influenced by affective feelings.

In a paper based on his Nobel lecture, Kahneman (2003) considers the idea of an *affect heuristic* (Slovic et al., 2002), i.e. that basic affective reactions replace more complex evaluations, as the third main heuristic (together with representative and availability heuristics) in judgement and decision-making. He also offers a new understanding of the process of heuristic reasoning, namely that, without noticing it, people simply replace a difficult question with an easier one. For example, when asked what the probability is that X belongs to the category Y (X could be a woman with a positive mammography test, and Y the category breast cancer) they instead answer the question: How representative is X (positive test) to Y (breast cancer)? This utilizes representative heuristic. In a similar way, the use of affect heuristic involves the unconscious replacement of a complicated question such as, "Who is best qualified?" with the far easier question, "Who do I like best?"

Intuition versus analysis

Quickly answer this question:

> *A bat and a ball cost $1.10 in total.*
> *The bat costs $1 more than the ball.*
> *How much does the ball cost?*

It is now widely accepted that the use of heuristics is distinctive for *System 1* reasoning, one of two main kinds of reasoning characterized as associative, affective, largely automized, implicit (inaccessible to introspection) fast (information-processing in parallel) and requiring little cognitive effort. *System 2* reasoning, the other main kind of reasoning, is affectively neutral, analytic, deliberate, slow (serial information-processing), controlled, rule-governed, flexible, and requires large cognitive efforts. One main function of System 2 is to monitor the quality of fast System 1 responses.

By the way, what was the price of the ball above? If you came up with the wrong answer "10 cents", you are not alone. Almost everyone reports an initial impulse to give this answer, and 50% of Princeton students actually replied, "10 cents" (Kahneman, 2003). The high rate of errors in this easy problem illustrates how lightly the output of System 1 is monitored by System 2. Deliberate thought is often too demanding, and it is easier to trust the seemingly plausible answers spontaneously supplied by System 1. From this System 1/System 2 perspective, intuitive judgements may be said to correspond with judgements not modified by System 2.

In the examples above, intuition is linked to bad performance. However, intuitive thinking may also be forceful and precise. In "kind" learning environments (cf. the conditions necessary for learning from experience below), a high level of accurate, fast and effortless judgemental performance can be acquired through praxis (e.g. Hogarth, 2005). Grand masters in chess are one obvious example. The quality of their play in speed chess is very high, even if their thinking is largely automatized. Some studies even indicate that experienced decision-makers in specific professions perform better when they trust their intuition than when they enter into deliberate analyses (Klein, 1998).

Decision biases – overview

A decision bias can be defined as a way of thinking that contributes to a systematic deviation from rational or normative decision-making. Decision biases thus contribute per definition to reduced decision quality. The number of demonstrated decision biases is large, and rapidly growing. It is not possible to give a full overview here, but some common biases are described and classified below.

Biases in information processing

Examples of a few kinds of decision biases have already been given. The doctors in Eddy's study (1982) systematically overlooked the relevant base rates when estimating the probability for breast cancer based on mammography test results. *Insensitivity to base rates* is one of many systematic biases in probability judgement (of category membership) due to the use of the representative heuristic. Another bias typically caused by representative thinking is the *conjunction fallacy*, where the events or properties A & B are judged to be more probable than A alone. In a classic example (Tversky and Kahneman, 1983) subjects were asked to look at a description of a person named Linda.

They were then asked to range the most probable statement of, among others, "Linda is a bank teller", and "Linda is a bank teller and active in the feminist movement." The majority judged the second statement to be more probable than the first, presumably because the description of Linda was representative of a typical feminist.

The use of availability heuristic also gives rise to predictable biases in probability judgement. Bazerman (2006) gives an example where subjects, very few of whom, when asked what they judged as the most probable cause of death in the USA between drug use, guns, traffic accidents, bad diet/physical inactivity and tobacco, came up with the correct answer. The correct answer lists the causes in the opposite order of that listed above. It is common to overestimate the frequency of the causes that are most easily noticed. How easy or difficult a phenomenon is to imagine or remember, often leads to biases in frequency judgement due to the use of availability heuristics.

Availability and representative heuristics are strategies that simplify the processing of information when probabilities are estimated. Tversky and Kahneman (1974) named the third satisficing strategy "anchoring and adjustment". Estimates are often made in two steps. The starting point is a more or less justified value (*the anchor*), based on easily accessible information. This anchor value is then adjusted to a more plausible estimate. Even anchors that have no relevance to the judgement being made can have a huge impact on the final estimate. In a classic study, Tversky and Kahneman asked subjects to estimate what percentage of the UN member states are African. Before the participants put forward their estimates, they were asked to throw a roulette ball. The roulette was manipulated so that it stopped on 65 for one group of subjects, and on 10 for the second group. The average answer given by the group who saw the roulette ball stopping on 65 was 45%; but 25% by the group where the ball stopped on 10. Even though the anchor in this case was completely irrelevant, it still strongly influenced the subjects' final estimates. Anchoring is a very robust phenomenon that systematically influences professionals' estimates too (Chapman and Johnson, 2002).

A decision bias that is often a contributory cause when projects are not finished on time, is the general tendency to overestimate the probability for conjunctive events, i.e., the probability that, for example, all the events A *and* B *and* C will take place, when the probability is known that each of them will take place separately. Similarly, there is a tendency to underestimate the probability for disjunctive events, i.e. the probability that A *or* B *or* C will take place. Therefore the probability that complex systems, like human bodies or nuclear plants, will break down is often underestimated. Even if the probability is tiny that each essential component in the system will break down, the probability that the whole system will break down can still be large, if many essential components are involved.

Preference reversals and biases associated with the presentation of data

Some examples have already been given of biases associated with how information is presented. In the Asian disease study, the subjects' decision behaviour was strongly influenced by the way choice alternatives were described in terms of gains or losses. This is one example of *preference reversal*, namely, that alternative A is preferred to alternative B, when the alternatives are presented in one way, but that preferences change when A and B are presented in a different (although from a normative perspective, equivalent) way. Tversky et al. (1988) demonstrated another common kind of preference reversal, between choice and matching. Participants were presented with the following scenario:

> *About 600 people are killed each year in Israel in traffic accidents. The ministry of transportation investigates various programmes to reduce the number of casualties. Consider the following two programmes, described in terms of yearly costs (in millions of dollars), and the number of casualties per year that is expected, following the implementation of each programme:*

Programme	*Expected number of casualties*	*Cost*
X	500	\$55 M
Y	570	\$12 M

When subjects were asked to *choose* which programme they favoured, 67% favoured Programme X, which saved more lives, but at a higher cost per life saved, than Programme Y. Subjects in the "matching condition" were given the same description of the two programmes but with one of the numbers missing. They were asked to *match* the two programmes, i.e. to fill in the missing number so that the two programmes would be equally desirable. Only 4% of the subjects then favoured program X, i.e. filled in a number equal to or larger than \$55 million when the cost of Programme X was missing.

Tversky et al. proposed the *prominence hypothesis* to explain this finding. The hypothesis simply says that the more prominent attribute will weigh more heavily in choice than in matching. Subjects regard casualties as more important, or *prominent*, than cost. In making a choice, the most important attribute is often the main consideration, while in matching, the different attributes are given more equal consideration. The reason is probably that a choice will often have to be justified to others (or ourselves) later on, and when justifying choice the most prominent attribute is typically in focus. An overview of other preference reversals is given by Hsee et al. (2004).

Motivational causes to decision biases

So far, mainly cognitive reasons have been given for the deviation from rational thinking. For example, the tendency to overestimate personal contributions to projects participated in has been explained in purely cognitive terms. It is easier to notice and remember one's own work than that of other participants. However, not only is there a tendency to take a disproportionately large share of the credit for collective successes, there is also an inclination to accept too little responsibility for collective failures. It is harder to explain this last fact in purely cognitive terms.

One non-cognitive reason for systematic deviation from rational thinking is the consistent motivation to come to a desired conclusion (e.g. Kunda, 1990). Reasons are sought to support one's own views, and counterarguments that support opposing conclusions are neglected. There is a tendency to evaluate ambiguous information in a way that is beneficial to one's own interests. Such *motivated reasoning* can explain both the tendency to claim responsibility for successes rather than failures, and the reasoning behind other so-called *"self serving" biases*.

It may not be surprising that judgements are often influenced by self-interest. More surprising, and more problematic, is the tendency to ignore certain important information when motivational or affective factors are apparently not involved.

Confirmation trap

> You will be given three numbers which conform to a simple rule that I have in mind. Your aim is to discover this rule by writing down sets of three numbers. After you have written down each set, I shall tell you whether your numbers conform to the rule or not. You should try to discover this rule by citing the minimum sets of numbers. When you feel highly confident that you have discovered [the rule], and not before, you are to write it down.

Wason (1960) gave these instructions to 29 college students, along with the sample set of numbers: [2, 4, 6]. What is the rule? Which sequence of three numbers would *you* like to test out? Think about it before proceeding.

The rule Wason had in mind was as follows: "Three numbers in increasing order of magnitude". Only six subjects discovered the correct rule without first naming an incorrect one. The others suggested more complex rules than the correct one. Commonly proposed rules included "Numbers that go up by two" and "The difference between the first two numbers equals the difference between the last two numbers." Wason found that subjects tried to confirm the rule they assumed the experimenter had in mind, far more often than they tried to disconfirm the rule. However, to find the rule requires the accumulation of disconfirming, rather than confirming, evidence. Wason concluded that few had the attitude necessary to succeed: "a willingness to attempt to falsify hypotheses, and thus to test those intuitive ideas that so often carry the feeling of certitude”.

So, do *you* have this attitude? Were the three numbers you would have tested out, an attempt to invalidate your own hypothesis on the rule in the experimenter’s mind?

Wason’s study illuminates a strong tendency everyone has to look for information that supports their beliefs. The belief that one should own a particular kind of car, employ a particular person, etc. leads to a strong tendency to seek out information that strengthens that belief. This tendency is known as the *confirmation bias*.

To look for information that can eventually invalidate what one believes in, is often the most effective way to test out both the weaknesses and the strengths of personal opinions. It is far more beneficial to hire a consultant company that acts as the devil’s advocate, and comes up with the best reasons to give up or change project plans, than one that, for the most part, advocates existing ideas.

Overconfidence

In 1957, the cost of the Sydney Opera House was estimated at seven million dollars with a completion date in 1963. An adapted version of the planned building was finished in 1973, at a cost of 102 million dollars. This is a prime example of the *planning fallacy*, i.e. the tendency to be over-optimistic with regard to when planned projects will be finished (Buehler et al., 2002). In a more mundane example, Buehler et al. asked students in the final stage of their psychology study when they “realistically” expected to submit their theses. They were also asked when they would submit them if “everything went as poorly as it possibly could”. The students’ “realistic” predictions were overly optimistic. Only 30% of them finished the project by the predicted time. On average, the students took 55 days to complete their theses, 22 days longer than they had anticipated, and seven days longer than the average worst-case prediction.

The planning fallacy is an example of the *overconfidence bias*, i.e. the strong tendency to be more certain about one’s judgements and conclusions than one has reason to be. Overconfidence depends on the difficulty of the judgemental task. Tasks resulting in correct answers of about 75% or less tend to produce overconfidence, whereas easier tasks tend to produce underconfidence. For example, in a series of experiments conducted in the 1970s, Lichtenstein and Fischoff found that people were 65–70% confident on being right, when they were actually correct about 50% of the time (Hoffrage, 2004).

No bias in judgement and decision-making is more prevalent, and more potentially catastrophic, than overconfidence. Historians have for a long time emphasized overconfidence as a substantial cause of war. In wars since 1500 AD, 50–75% of the attacking sides have lost. There is also a likelihood that the winning side in a war will find the victory to be more costly than expected before the war started. Moreover, if the belligerent parties' judgement as to the chances of victory before beginning a war is summarized in percentage terms, the numbers usually add up to far more than hundred (Johnson, 2004).

Hindsight bias

The fall of the Berlin Wall in 1989; the planes crashing into the World Trade Center on 11 September 2001: these are events that no-one could have failed to notice. But try to imagine the time before these events took place. How probable was it that the Wall would fall, and terrorists fly into the WTC? Can knowledge of past events be overlooked when giving an answer?

After an event has taken place, there is a strong tendency to overestimate the extent to which it could have been foreseen. This bias is called the *hindsight bias* (Fischhoff, 1975), and is often explained as an anchoring effect. Knowledge on an event easily becomes an anchor on how we believe that we judged the probability for that event in advance. Adjustments related to such anchors have a tendency to be insufficient. So, knowledge we receive after an event has taking place influence how we remember that we judged the probability of that event in advance.

An unfortunate consequence of the hindsight bias is that it reduces the possibility to learn from misjudgements. The bias detracts from mistakes made in the past, and thus contributes to overconfidence in future predictions.

Closely related to the hindsight bias is a bias known as the *curse of knowledge* (Camerer et al., 1989). In the business world it is important to be able to foresee the market. Imagine knowing exactly what information other participants in the market have. Camerer et al. demonstrated in several experiments that the more information about the market that is acquired beyond that, the harder it is to predict accurately how these other actors will behave, even if with the knowledge that they do not have the additional information. Better informed agents are unable to ignore private information, even when it is in their interest to do so. More information is not always better. Once something is known, it is hard to imagine what it was like not to know it: hence the "curse" of knowledge.

Why do people reason in ways that lead to systematic biases?

So far, the focus has been on the decision biases resulting from satisficing strategies. However, from a bounded rationality perspective, such non-optimal strategies also have a positive side. They contribute to decision efficiency, and decision-making in a complex world often requires the use of such simplifying strategies. Many of the satisficing strategies used contribute to adequate rather than inadequate decisions, and the loss of decision quality will usually be counterbalanced by increased efficiency. Tversky and Kahneman (1974) warned, "In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors". For example, the use of representative heuristic is a good strategy when there is no access to base rate information, as is often the case.

Correspondingly, availability heuristic works well when the correlation is strong between availability and frequency. This heuristic was probably more useful in a pre-modern world, when what was important in order to survive, was also easy to notice and to remember. In our modern world however, statistical information is often far more reliable than easily accessible personal experiences. Availability heuristic contributes to the strong tendency to put exaggerated weight on personal experiences. Instead of thinking in statistical terms, the inclination is to think "dramatically", in personal terms.

Seen from a bounded rationality perspective, ways of thinking that systematically lead to decision biases simultaneously contribute to decision efficiency. However, so far, the main focus within decision psychology has been on the biases, and not on the efficiency. One exception is Gigerenzer et al. (1999), who, through simulations, have demonstrated that, under certain circumstances (large uncertainty, little knowledge etc.), some very simple heuristics (satisficing strategies) often give results as good as more optimal methods.

The other ways of thinking, mentioned above, that lead to systematic decision biases, also have positive functions. For example, a marked tendency to seek confirmation, contributes to the strengthening and stabilizing of beliefs and opinions. Doubt is certainly not always a good thing. In many circumstances, it will be almost impossible, or at least not very practical, to go after information that contradicts one's beliefs. Moreover, many beliefs are hard to disconfirm.

It is well known that positive illusions and unrealistic optimism strengthen our health, creativity and performances in both physical and mental exercises (Taylor, 1989). Overconfidence as a positive illusion can undoubtedly have positive functions. A doctor who is overconfident in the treatment she provides will (because of placebo effects etc.) sometimes get better treatment results than a more realistic doctor. It is also "overconfident" editors and contributors who manage to publish books like this one. If we had had a realistic picture of our time schedules when we were asked to contribute to this book, then many of us would probably have answered, "No" . . .

How decision-making can be improved – some efficient methods

It is a common belief that experience improves the quality of judgements and decision- making. But this is far from the truth. In fact, "experience has been shown to be unrelated to the empirical accuracy of expert judgments" (Hammond, 1996).

Why does experience (often) not improve the quality of professionals' judgements? The short answer is that professional work is usually performed in "wicked" learning environments, where the two necessary conditions for learning from experience are not present (Hogarth, 2005). The first condition is that immediate, unambiguous and consistent feedback is offered where mistakes have been made. The second condition is that the content of the feedback gives a clear understanding of precisely what was done wrong. One reason why these two conditions are seldom present is that professionals usually perform their judgement and decision-making in probabilistic contexts, i.e. in circumstances where the same kind of judgement or choice can give different feedback. (In deterministic contexts, on the other hand, the same kind of judgement/choice will always elicit the same kind of feedback.) This means that professionals can make the right (over time) decision, but still occasionally achieve negative outcome feedback (results). Conversely, wrong choices sometimes may lead to positive results.

Another reason why decisions do not improve with experience is that often there is no feedback on the decisions per se. Frequently the only feedback given is on the decision in conjunction with the actions that follow the decision. For example, imagine a psychiatrist who decides to commit a person to a mental hospital. If the patient has calmed down after a week, how will the psychiatrist evaluate his decision to commit her? What if, after a week, the patient has turned completely crazy? The psychiatrist will, in both cases, probably interpret positively the outcome feedback on his decision to commit the patient.

So, it is well documented that experience is often not sufficient to improve judgements and decisions. How then to improve decision behaviour?

Debiasing – some main strategies

In the JDM literature several methods are suggested to improve or *debias* judgement and decision-making. Larrick (2004) distinguishes between three main kinds of debiasing strategies: motivational, cognitive and technological. Motivational strategies are based on the critical assumption that people possess the necessary normative strategies ("cognitive capital"), and will use them when the benefits exceed the costs. There are two main ways to motivate decision-makers: incentives to reward decision quality, and accountability, i.e. demanding that decision-makers will have to explain and defend their decisions to others.

Research shows that incentives motivate people to put more effort into their decision-making. It is thus natural to believe that incentives also improve decision quality. However, there is hardly any evidence to show that incentives consistently improve mean decision performance. A recent review study on the effect of incentives, concludes, "there is no replicated study in which a theory of rational choice was rejected at low stakes in favour of a well-specified behavioural alternative, and accepted at high stakes" (Camerer and Hogarth, 1999).

The principal reason why incentives do not usually have any effect, is simply that people do not possess the cognitive capital necessary to make better decisions. For example, if good decisions require knowledge of Bayesian reasoning, incentives will be of no help if the subject has not heard of Bayes' formula. Incentives thus contribute to what Larrick (2004) names the "lost pilot" effect: "I don't know where I'm going but I'm making good time".

Like monetary incentives, accountability motivates subjects to put more effort into their decision-making, but this does not necessarily lead to better decisions. In addition, accountability often evokes a strong need to look consistent to others. This can both improve and deteriorate the quality of decision-making. Accountability often leads to biases, because decisions are so easily adapted to suit the audience. Another problem is that accountability is likely to strengthen reliance on easily justified aspects of the decision.

Cognitive strategies to improve decision-making include: drawing people's attention to decision biases; training in normative decision rules, e.g. Bayes' formula; learning to re-formulate decision problems; the use of other special strategies. To make people aware of decision biases, e.g. by reading chapters like this one, seems to have little debiasing effect (Fischhoff, 1982). Training on normative rules in logic and probability theory has, in certain circumstances, some effect (Larrick, 2004). To re-formulate decision problems can, in some cases, have a dramatic effect. For example, the doctors faced with Eddy's (1982) question presented above on breast cancer formulated in frequency terms instead of probability terms, i.e. 10 out of 1000 women aged 40 have breast cancer, 8 out of 10 women with cancer test positively, 1 out of 10 women without cancer also get a positive test result. More than half of the doctors then answered correctly the question on the probability that a woman who has a positive test also has breast cancer (Gigerenzer, 1996). However, such re-formulations require quite a lot on behalf of the decision-maker.

Contrary to motivational and cognitive strategies, "technological" strategies (in Larrick's use of the term) involve improving decision-making by going beyond the individual decision-maker, either by incorporating a group process, or by using different kinds of technological means. A large number of methods have been developed, both to support decision-makers in different aspects of the decision-making process, and to replace the decision-maker for part of the decision process. From a JDM perspective, there is unanimity with regard to one characteristic of decision support systems: they are grossly under-used.

Before focusing on two different technological strategies, a couple of simple but efficient cognitive strategies are suggested below.

Take an outsider's view!

When planning a new project, a major concern is how long it will take and how much it will cost. Estimates are usually far too optimistic. Kahneman and Lovallo (1993) claim that one reason for this is the neglecting of past statistics, and the tendency to consider new projects as unique. An "inside" view, is taken, rather than an "outside" view.

In 1976, Kahneman was involved in a project designed to develop a curriculum for the study of JDM for high schools. He asked each group member to indicate their best estimate of the number of months that would be needed to finish the project. The estimates ranged from 18 to 30 months. Kahneman then asked one member of the group, a distinguished expert in curriculum development, "We are surely not the only team to have tried to develop a curriculum when none existed before. Please try to recall as many such cases as you can. Think of them as they were in a stage comparable to ours at present. How long did it take them . . . to complete their projects?" (Kahneman and Lovallo, 1993). The experienced curriculum developer estimated that 40% of such projects were eventually abandoned, and that no such curriculum, as far as he knew, had been completed in less than 7 years. The team took 8 years to complete the JDM curriculum project.

In this example, all the participants in the meeting, Kahneman and the curriculum expert included, spontaneously adopted an inside view of the problem. This is typical. The inside view is overwhelmingly preferred in intuitive forecasting. Kahneman encouraged the curriculum expert to adopt an outside view. This ignores the details of the case in hand, and instead, focuses on the statistics of a class of cases similar to the present one.

Even if there is massive evidence that the outsider makes better estimates and decisions than the insider, decision-makers tend to believe and act on the insider view. However, it is not uncommon to switch between an inside and an outside view. Estimates based on detailed knowledge of a particular project are usually thought to be quite accurate (inside view), but at the same time there may be an awareness that most similar projects usually take a much longer time (outside view). For example, someone who builds their own house may have a particular opinion on what the price of the finished house will be but at the same time acknowledge that their own estimate is probably far too low. They then use the knowledge of other house builders who had underestimated the costs of their finished houses. In general, a way to proceed is to make an "inside" estimation of a project's costs first, and then correct this estimate by taking an outside view, discovering the ratio of actual spending to planned costs across other similar projects. Finally, the inside view estimate should be multiplied by this ratio.

Consider the opposite!

"Consider the opposite" is another efficient strategy for improving decision-making, and is closely related to "take an outsider's view". The strategy consists of simply asking the question, "What are some of the reasons that my initial judgement might be wrong?" Just to ask oneself this question has shown to effectively reduce overconfidence, hindsight biases and anchoring effects (e.g. Larrick, 2004).

The strategy is efficient because it counters the strong tendency to seek for information that confirms one's own opinions. The consideration of alternative options directs the attention to contrary evidence that would not otherwise be considered. The information base for decision-making is thus extended and made more representative. In this way, the fundamental problem with associative System 1 thinking is counteracted, namely that too narrow a selection of information is taken into account.

Group decision-making

In 1906, the multi-scientist Francis Galton, a half cousin of Charles Darwin, visited a cattle show, where a huge bull was being exhibited. The visitors were invited to guess the weight of the bull. A prize was put up for the best bet. After the exhibition, Galton collected all the bets on the weight of the bull. He calculated the average guess to be 1197 pounds. The weight of the bull was, in fact, 1198 pounds. The bets of the cattle experts were all far off the precision of the average bet of the unskilled mob. This is an early example of an acknowledged strategy to improve judgement, by combining the independent estimates of each individual member of the group.

Another example of the efficiency of this strategy was shown in an analysis of the quality of advice received by participants in the TV show, "Who wants to be a millionaire?". In this show the participants are able to ask for advice. They can place a call to a person who, they believe, knows a lot about the topic, or they can poll the studio audience. The analysis of the advice showed that the "experts" offered the right answer approximately 65% of the time. The advice from the audience, which was the multiple choice option voted by the majority in the TV studio, chose the right answer 91% of the time (Surowiecki, 2004).

In both of these examples, the best judgements are a result of combining the judgement of each individual member of a group. When groups reach a conclusion through deliberation, the quality of the conclusion is often far from impressive. Deliberation often produces a series of unfortunate results: group polarization, amplification of errors, "cascade" effects etc. (e.g. Sunstein, 2005).

Surowiecki emphasizes four requirements to create a judicious group. The first requirement is *diversity of opinion*: each person should have private information on the case discussed. The second is *independence*: people's opinions must not be determined by the opinions of those around them. The third requirement is *decentralization*: people must be able to specialize and draw on local knowledge. The fourth is *aggregation*: some "mechanism" has to be used for turning private judgements into a collective decision.

The *Delphi technique*, an interactive method for obtaining forecasts from a panel of independent experts, satisfies these requirements. This is a structured group process, in which individuals are required to give numerical judgements over a number of rounds. After each round, a facilitator provides an anonymous summary of the forecasts from the previous round, together with the reasons the experts provided for their judgements. Participants are thus encouraged to revise their earlier answers in light of the replies of other members of the group. During this process, the range of the answers will usually decrease, and the group will converge towards the "correct" answer. The process comes to an end after a pre-defined stop criterion (e.g. number of rounds, stability of results etc.), and the final aggregate is taken as the process output.

The Delphi method was developed by the RAND Corporation in the 1950s, and has been applied extensively. In a recent evaluation, Rowe and Wright (2001) found that the method gave more accurate estimates than traditional groups in five of their studies, less in one of them, and equalized in two. Overall, they found that the Delphi method improved accuracy in 71% of the cases, and reduced it in 12%.

The idea that the average of several independent estimates is better than the individual best estimate, is the fundamental concept behind an increasingly important information-aggregation tool, known as *prediction markets*. These are "opinion markets", and work in a similar way to commodity markets. Prediction markets channel inputs from all traders into a single dynamic stock price. Instead of determining the value of a particular good, a prediction market is used to determine the probability of a particular event occurring. The participants trade in contracts whose pay-off depends on unknown future events. The price of these contracts can be directly interpreted as a market-generated forecast of some unknown quantity. Prediction markets are extremely useful for estimating the market's expectation of the probability of a particular event occurring. For example, they have yielded very accurate predictions for elections, and outperformed the major pollsters. At Hewlett-Packard, a prediction market produced more accurate forecasts of printer sales than the firm's internal processes. Google utilized prediction markets successfully to obtain better forecasting of what the company and its competitors were planning to do (e.g. Wolfers and Zitzewitz, 2004).

Automate judgements

As we saw above, simple linear models systematically outperform experts' judgements, under certain circumstances. This robust finding indicates that professionals' judgements in repeatable decision-making situations should be automated. One reason for automation is to improve judgement and decision quality. Another good reason is to reduce costs. In the study discussed above, Dawes (1971) estimated that if his model had been used in the admission of students to college, not only would the normal admission quality have been better, but the USA public sector would also have saved 18 million dollars annually. Bazerman (2006) suggests that, adjusted for the current number of graduate-school applications and today's dollar, this number would exceed 500 million dollars.

However, even where a particular decision is automated, it is still necessary to depend on experienced professionals. A classic study by Einhorn (1972) highlights both the strength and weakness in professionals' judgemental abilities. Einhorn studied physicians who coded biopsies of patients with Hodgkin's disease, and then made an overall rating of severity. The individual ratings had no predictive power of the survival time of the patients, all of whom died of the disease. In contrast, a multiple regression analysis based on the nine biopsy characteristics scaled by the doctors succeeded, to some extent, in predicting how long the patients would live. The general point from this is that experts knew what information to consider, but that a linear model combines this information in a way that is superior to the global judgements of these very same experts.

Meehl (1954) compared clinicians' judgements with actuarial or "*proper*" equations, i.e. the regression weights (coefficients) in the equations were based on statistical techniques that optimize prediction. However, even when the models are "*improper*", this means that the variable weights are determined by non-optimal means (e.g. unity or random weights, by "bootstrapping" etc.), the models' integration of data normally outperform professionals' judgements of the same data. For example, a thousand applicants have to be judged and ranked, based on the same kind of information. If a hundred of the cases are judged by a human judge, and then a regression analysis is performed on these judgements, the resulting linear model (which is a bootstrapping model of those judgements) will almost invariably outperform the same human judge on the last nine hundred applicants.

The overall conclusion is that if the variables relevant for making a particular judgement are known, the weights given to the variables in the linear equations are of less importance. How then should professionals in repeatable judgemental situations be replaced? Dawes and Corrigan's (1974) much-quoted recipe is, "the whole trick is to decide what variables to look at and then to know how to add".

Conclusion

People's decision-making behaviour deviates systematically from normative models. Professionals are no exception. Their decision-making shows the same kind of biases as lay people's decision-making, and simple models consistently outperform intelligent and experienced professionals' judgements. A principal reason for biases in judgement and decision-making is limited cognitive capacity, which often makes it necessary to use simplifying reasoning strategies. Motivational and emotional factors can also create biases.

The use of simplifying strategies that sometimes lead to decision biases also contributes to decision efficiency. The loss of decision quality will often be counterbalanced by increased efficiency. One main problem is a lack of awareness of what kind of reasoning strategies have been applied, and how these may negatively influence decision-making. No distinction is made between situations where particular reasoning strategies are advantageous and situations where the same strategies are potentially damaging. A better understanding of this division is a key to improved decision-making.

The quality of professionals' judgement and decision-making is rarely improved through experience, mainly because professionals do not receive accurate and timely feedback. No simple recipe for eliminating decision biases exists. However, simple cognitive strategies like "take an outsider's view" and "consider the opposite" are efficient in many circumstances. Findings in decision psychology suggest a far wider use of technological decision support within many professions. In particular, professionals' judgements in repeatable decision-making situations should be automated to a far greater extent than is the case today.

References

Bazerman M. (2006). *Judgment in Managerial Decision Making* (6th edition). Wiley: Hoboken, NJ.

Bernoulli D. (1738/1954). Exposition of a new theory on the measurement of risk. *Econometrica* **22**:23–36.

Buehler R, Griffin D, and Ross M. (2002). Inside the planning fallacy: the causes and consequences of optimistic time predictions. In: Gilovich T, Griffin D, and Kahneman D (eds). *Heuristics and Biases. The Psychology of Intuitive Judgment.* Cambridge University Press: Cambridge, pp. 250–270.

Camerer C. (2001). Prospect theory in the wild: evidence from the field. In: Kahneman D and Tversky A (eds). *Choices, Values, and Frames*. Cambridge University Press: Cambridge, pp. 288–300.
Camerer C and Hogarth R. (1999). The effects of financial incentives in experiments: a review and capital-labor-production framework. *Journal of Risk and Uncertainty* **19** (1–3): 7–42.
Camerer C and Johnson E. (1991). The process-performance paradox in expert judgment: how can experts know so much and predict so badly? In: Anders Ericsson K and Smith J (eds). *Toward a General Theory of Expertise*. Cambridge University Press: Cambridge, pp. 195–217.
Camerer C, Loewenstein G, and Weber M. (1989). The curse of knowledge in economic settings – an experimental-analysis. *Journal of Political Economy* **97** (5): 1232–1254.
Chapman G and Johnson E. (2002). Incorporating the irrelevant: anchors in judgments of belief and value. In: Gilovich T, Griffin D, and Kahneman D (eds). *Heuristics and Biases. The Psychology of Intuitive Judgment.* Cambridge University Press: Cambridge, pp. 120–138.
Dawes R. (1971). A case study of graduate admissions: application of three principles of human decision making. *American Psychologist* **26**:180–188.
Dawes R and Corrigan B. (1974). Linear models in decision making. *Psychological Bulletin* **81**:95–106.
Dawes R, Faust D, and Meehl P. (1989). Clinical versus actuarial judgment. *Science* **243**:1668–1673.
Eddy D. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities. In: Kahneman D, Slovic P, and Tversky A (eds). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press: Cambridge, England, pp. 249–267.
Einhorn H. (1972). Expert measurement and mechanical combination. *Organizational Behavior and Human Performance* **7**:86–106.
Fischhoff B. (1975). Hindsight foresight: the effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance* **1**:288–299.
Fischhoff B. (1982). Debiasing. In: Kahneman D, Slovic P, and Tversky A (eds). *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge University Press: Cambridge, England, pp. 422–444.
Gigerenzer G. (1996). The psychology of good judgment: frequency formats and simple algorithms. *Medical Decision Making* **16**:273–280.
Gigerenzer G, Todd P, and The ABC Research Group. (1999). *Simple Heuristics that Make Us Smart*. Oxford University Press: New York.
Hammond K. (1996). *Human Judgment and Social Policy. Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. Oxford University Press: New York.
Hoffrage U. (2004). Overconfidence. In: Pohl RF (ed.) *Cognitive Illusions*. Psychology Press: New York, pp. 235–254.
Hogarth R. (2005). Deciding analytically or trusting your intuition? The advantages and disadvantages of analytic and intuitive thought. In: Betsch T and Haberstroh S (eds). *The Routines of Decision Making*. Lawrence Erlbaum Associates, Inc.: Mahwah, NJ, pp. 67–82.
Hsee C, Zhang J, and Chen J. (2004). Internal and substanstive inconsistencies in decision making. In: Koehler D and Harvey N (eds). *Blackwell Handbook of Judgment and Decision Making*. Blackwell: Malden, MA, pp. 360–378.
Johnson D. (2004). *Overconfidence and War: The Havoc and Glory of Positive Illusions*. Harvard University Press: Cambridge, MA.

Kahneman D. (2003). A perspective on judgment and choice. Mapping bounded rationality. *American Psychologist* **58**:697–720.

Kahneman D and Lovallo D. (1993). Timid choices and bold forecasts – a cognitive perspective on risk-taking. *Management Science* **39** (1): 17–31.

Kahneman D and Tversky A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* **47**:263–291.

Kahneman D and Tversky A. (2000). *Choices, Values, and Frames*. Cambridge University Press: New York.

Klein G. (1998). *Sources of Power. How People Make Decisions*. MIT Press: Cambridge, MA.

Kunda Z. (1990). The case for motivated reasoning. *Psychological Bulletin* **108**: 480–498.

Laibson D and Zeckhauser R. (1998). Amos Tversky and the ascent of behavioural economics. *Journal of Risk and Uncertainty* **16**:7–47.

Larrick R. (2004). Debiasing. In: Koehler D and Harvey N (eds). *Blackwell Handbook of Judgment and Decision Making*. Blackwell: Malden, MA, pp. 316–337.

Loewenstein G and Lerner J. (2003). The role of affect in decision making. In: Davidson R et al. (eds). *Handbook of Affective Sciences*. Oxford University Press: Oxford New York, pp. 619–642.

Loomes G and Sugden R. (1982). Regret theory – an alternative theory of rational choice under uncertainty. *Economic Journal* **92** (368): 805–824.

McNeil B, Pauker S, Sox H Jr., and Tversky A. (1982). On the elicitation of preferences for alternative. *New England Journal of Medicine* **306**:1259–1562.

Meehl P. (1954). *Clinical Vs. Statistical Prediction*. University of Minnesota Press: Minneapolis.

Rowe G and Wright G. (2001). Expert opinions in forecasting: the role of the Delphi technique. In: Armstrong J (ed.) *Principles of Forecasting*. Kluwer: Boston, pp. 125–144.

Savage L. (1954). *The Foundations of Statistics*. Wiley: New York.

Simon H. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics* **69**:99–118.

Simon H. (1956). Rational choice and the structure of the environment. *Psychological Review* **63**:129–138.

Slovic P, Finucane M, Peters E, and MacGregor D. (2002). The affect heuristic. In: Gilovich T, Griffin D, and Kahneman D (eds). *Heuristics and Biases. The Psychology of Intuitive Judgment*. Cambridge University Press: Cambridge, pp. 397–420.

Sunstein C. (2005). Group judgments: statistical means, deliberation, and information markets. *New York University Law Review* **80** (3): 962–1049.

Surowiecki J. (2004). *The Wisdom of Crowds*. Doubleday: New York.

Taylor S. (1989). *Positive Illusions*. Basic Books: New York.

Thaler R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization* **1**:39–60.

Tversky A and Kahneman D. (1974). Judgment under uncertainty: heuristics and biases. *Science* **185**:1124–1131.

Tversky A and Kahneman D. (1981). The framing of decisions and the psychology of choice. *Science* **211**:453–463.

Tversky A and Kahneman D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgments. *Psychological Review* **90**:293–315.

Tversky A, Sattath S, and Slovic P. (1988). Contingent weighting in judgment and choice. *Psychological Review* **95** (3): 371–384.

Von Neumann J and Morgenstern O. (1947). *Theory of Games and Economic Behavior*. Princeton University Press: Princeton, NJ.
Wason P. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology* **12**:129–140.
Wolfers J and Zitzewitz E. (2004). Prediction markets. *Journal of Economic Perspectives* **18** (2): 107–126.
Zajonc R. (1980). Feeling and thinking: preferences need no inferences. *American Psychologist* **35**:h151–h175.

10
Useful Heuristics

Benjamin Scheibehenne and Bettina von Helversen

Decision-making is one of the core tasks in project management. Traditionally, optimization methods have been developed to support managers in finding the best solutions. Alternatively, decisions can be based on a simple rule of thumb or on heuristics. Even though simple heuristics only require little in the way of time and information, they have been shown to outperform optimization methods in complex decision tasks across a wide range of situations. This chapter outlines relevant decision heuristics commonly used, demonstrates situations in which they outperform more complex decision algorithms and explains why and when simple heuristics provide powerful decision tools.

Introduction

Imagine two managers who want to develop a real estate project. At some point during the planning phase it becomes crucial to predict the future sales prices of houses in a certain area. The first project leader, Mr O, approaches this task by searching exhaustively for all available pieces of information that he knows will influence the selling price, such as property tax, lot size, total living space, age of the house, number of bathrooms and so on. Based on his past experience, he weighs all that information according to its importance and then integrates it to predict the selling price of each house, using some statistical software. The second manager, Mr F, makes a fast decision, relying on a simple strategy based on just one single piece of information that he regards as most important, such as total living space. Which of these two managers will make a more accurate forecast? Many people, researchers and lay persons alike, suppose that the outcome of a decision can be improved by (a) an exhaustive search for information, and the integration of many pieces of information, (b) having more time to think, and calculate possible outcomes, or (c) having more computational power and the use of complex forecasting software or decision tools. This seems to indicate that Mr O will make the better decision. However, this chapter demonstrates that less can sometimes be more, and that a strategy relying on very few pieces of information, and quickly deriving a decision based on a simple algorithm, may well outperform more sophisticated, supposedly rational decision strategies.

To this end, the topic of "fast and frugal heuristics" is introduced (Gigerenzer et al., 1999). This investigates when and how simple decision strategies can be used to make sound decisions. The evidence shows that heuristics work well in many real-world situations that are characterized by uncertainty and scant information. Moreover, simple heuristics can be seen to outperform more complex decision models. Before turning to the science of heuristics however, the status-quo of current decision research is examined and shown to consist mainly of optimization models.

Optimization models of decision-making

The unboundedly rational decision-maker

In his attempt to predict sale prices of houses, Mr O in the example above endorses a complex strategy that incorporates as much information as possible. In abstract terms, the underlying decision model could be described as an attempt to find the optimal solution, i.e. the solution that maximizes a given criterion measured by an objective function. In this case, the criterion was the probability of making a correct forecast; in other cases it might be the expected monetary gain or the number of people who receive a benefit. Optimization models commonly try to find a solution by searching for all relevant information, weighing it according to its importance and combining it into an overall score. However, optimization implies that a well-defined criterion exists that can be used to calculate the function that is maximized. This further requires that the decision-maker has full knowledge of the decision task, can acquire all relevant information and that the algorithm that leads to the optimal solution is known. It further requires that the decision-maker has the computational abilities and capacities to process all the information or else has a decision tool readily available that fulfills these criteria, respectively.

This "unboundedly rational" view on decision-making is highly unrealistic for almost every decision task in real-world environments, like predicting future house prices. First of all, decision-makers are not omniscient: even for the simplest decision (let alone the more complex decision tasks faced by project managers) not all relevant information is known. On the contrary, painstaking effort may be required to gather all relevant information, which may be expensive or impossible to achieve. Hence, neither the time nor the money to search for and integrate the information may be affordable or worth the possible increase in decision quality.

Furthermore, in many situations optimal solutions are not attainable. Even in well-defined decision problems with a limited number of options, such as chess, or the travelling salesman problem, optimal solutions are computationally intractable or NP-hard (Reddy, 1988). A problem is called NP-hard if its solution cannot be found in polynomial time. Roughly speaking, this means that no machine or mind can find the best solution in a reasonable amount of time, such as a millennium.

In many cases, including the evaluation of major projects, there are multiple criteria to measure success, and the importance assigned to these criteria may vary dynamically over time and between different stakeholders. Yet, in cases where the criterion is ill-defined, when more than one goal exists (e.g. minimizing time, cost and risk of failure) or when stakeholders differ in their priorities, finding the optimal solution is utterly impossible (Gigerenzer, 2004a).

Optimization under constraints

In an attempt to keep up with the gold-standard of optimality, some decision models try to take these restrictions into account. Aiming for "optimization under constraints", these models integrate search or information costs, by assuming that the decision-maker conducts a cost benefit analysis (CBA), and only continue to search for further information until the costs incurred in this search outweigh the benefits gained from the additional information. Even though these models try to impose constraints on decision-making, they often make even higher demands on the amount of information and computation necessary to conduct them. That is because they still require a lot of information, and, in addition to that, they also assume that the cost of searching for the information and its potential benefits can be known or at least accurately estimated to make an optimal decision. As such, they are simply another form of unbounded rationality (Arrow, 2004).

Decision tools widely used in project management, such as CBA (Pearce and Nash, 1981), multi-attributive utility theory (MAUT), multi-attributive value theory (MAVT), analytical hierarchy process (AHP) or decision trees, are examples of models that aim for optimization with or without constraints: they share a goal of finding the solution that maximizes (i.e. optimizes) the expected utility of stakeholders (Kiker et al., 2005). In line with the notion of optimization under constraints, these decision models further require a mathematical representation and integration of all possible information about risks and benefits. For instance, CBA tries to quantitatively evaluate the benefits, e.g. the expected profit or public utility, and the disadvantages, e.g. costs or risks involved in a decision, by expressing them in a single currency – money. The value of a decision is determined by the expected benefits minus the expected costs. Similarly, MAUT and AHP employ numerical scores to convey the merit of one option compared with others on a single "utility" scale.

Some assumptions made by these models can be problematic. For example, CBA makes the somewhat unrealistic assumption that objective prices and probabilities can be determined for any cost or benefit, including such things as human lives (Adams, 1995). MAUT and MAVT take into account that stakeholders differ in their preferences and in the utilities that they assign to different outcomes. However, as with CBA, these models assume that such interests can be represented numerically and traded-off against each other. Furthermore, all these methods are information-hungry and time-intensive.

The accuracy of these optimization methods hinges on the degree to which they manage to achieve an adequate representation of the decision problem. Yet, for several reasons, this representation can be difficult or impossible to attain. Project management decisions are often characterized by a high degree of foreseeable and unforeseeable uncertainty (De Meyer et al., 2002), which can make it too costly or impossible to gain a representation of the project environment that would allow the calculation of an optimal, or even close to optimal, course of action. Furthermore, in many projects, important decisions need to be made fast. For example, for many consumer goods the right timing is crucial, and in rapidly developing markets, technology product cycles may be very short. Here, an extensive and exhaustive search for information – which might be necessary to reach an accurate picture of the decision problem – can lead to a state of analysis-paralysis, where an opportunity is missed because of the sheer amount of information that needs to be processed.

Last but not least, even if an accurate representation could be reached quickly, more time and more information will not always lead to a better decision. Decisions in project management are commonly made in a dynamically changing environment. For instance, the importance assigned to benefits or the probability of risks is not stable but will depend on the political and economic environment. This means that decision strategies need to be robust to perform well in changing environments and be able to adapt to new circumstances. Moreover, complex decision algorithms are highly susceptible to change in their parameters and can therefore lead to sub-optimal outcomes (Brighton and Todd, 2007; Todd and Gigerenzer, 2007).

To summarize, there is little doubt that optimization methods such as CBA or MAUT are useful tools for project managers, and can lead to informed and accurate decisions, particularly when the structure of a project is clear, and there is little uncertainty. However, in project decisions where uncertainty is high, when risks and consequences are difficult to estimate or even foresee, and the time available for analysis limited, they may feign a degree of certainty that is by no means substantiated by reality (De Meyer et al., 2002). In such situations, simple decision strategies can be a useful alternative, as they allow for fast and robust decisions in dynamically changing environments.

Heuristics for decision-making

Most decision scientists acknowledge that heuristics are widely used, and yet there is controversy over why people employ heuristics. Some scholars see heuristics as a crutch that is necessary, because the memory and processing capacities of the human mind are severely limited. This impedes the use of more complex strategies and leads to the failure to consider all important information. From this perspective, heuristics are often seen as a flaw that will lead to second-best outcomes, systematic biases and a distorted evaluation of confidence in the decision (Kiker et al., 2005; McDaniels et al., 1999; Tversky and Kahneman, 1974, 1979). As a consequence, decision-makers would be well-advised not to use them but to rely on more complex strategies or decision support tools instead.

So are heuristics only for those who cannot appreciate the blessings of modern technology and thus remain irrational and biased? Or can there also be such a thing as a *useful* heuristic? Intuitively, one might think that a decision based on a simple strategy that uses little information, cannot be as sound as a decision in which many pieces of information have been evaluated, weighted and integrated. Yet, as it turns out, the chances are that this intuition is wrong.

The research on “fast and frugal heuristics” (Gigerenzer et al., 1999) embraces heuristics as adaptive responses to our environment and emphasizes their ability to ignore information as a strong virtue. Rather than considering the use of heuristics a flaw of human decision-making, Gigerenzer argues that heuristics are decision mechanisms, which evolved to enable humans to make quick and accurate decisions in uncertain environments. As the proverb “time is money” indicates, our world is competitive. Any time spent on decision-making keeps us away from other activities which could increase our chances to outperform our competitors. Because it takes time to search for, process and assess information, decision mechanisms which rely on scant information or computation have the advantage of speed. Further, simple mechanisms can be more accurate than complex mechanisms, because they are more robust. Robustness is an important feature of decision mechanisms in uncertain and dynamic environments. Because the same decision is rarely encountered twice, and decision environments are unstable, decision mechanisms need to adapt well to new circumstances (Brighton and Todd, 2007; Todd and Gigerenzer, 2007). Lastly, simple heuristics are highly transparent decision strategies. Because of their simplicity, they can be easily communicated and are thus more likely to gain public acceptance.

According to Gigerenzer et al. (1999), the mind consists of an “adaptive toolbox” of fast and frugal heuristics (Gigerenzer and Selten, 2001), which evolved to solve certain decision problems, such as a choice between two or more options. These heuristics are adapted to specific information structures, which they exploit to allow for fast and accurate decisions. The research on fast and frugal heuristics pursues a descriptive and a normative goal. It is descriptive, in that it tries to specify which heuristics decision-makers employ in a given situation, and it is normative, by showing that these heuristics are, in fact, adaptive solutions to specific decision environments.

Some of these fast and frugal heuristics are described in the following sections. It will be seen that they are widely adopted and can be as good as – and sometimes even better than – "optimal" decision strategies.

Relying on just one good reason: the take-the-best heuristic

In the introductory example, Mr F only relied on one good reason to predict the selling price of houses. He did not weigh or trade-off different pieces of information but chose the best option for the aspect that he regarded as most important. This particular simple heuristic is called take-the-best, because it gambles on the most valid piece of information (Gigerenzer and Goldstein, 1996). It belongs to a class of heuristics known as "lexicographic" (LEX), because they resemble a lexicon, in which words are strictly sorted by the order of their first letters (Payne et al., 1993).

A typical LEX strategy works as follows: In the first stage of the information search, the single most important aspect (i.e. the most valid cue) is taken into consideration. If one option is shown to be better than all the others with regard to this cue, the search is stopped and the option with the highest cue value is chosen. If several options are equally good for the most valid cue, the next best aspect is considered as a tie-breaker until a decision can be made (Figure 10.1). In many practical contexts, options are never completely similar with regard to any given aspect. In these cases, a threshold can be set that determines the point at which two options are still regarded as being equal.

LEX is a non-compensatory heuristic, which means that an advantage on an important aspect cannot be compensated by one (or a combination of) less important aspects. Because less important aspects are seldom needed to derive a decision, LEX requires little information search. But how good is LEX compared with other, more complex decision strategies?

Keeney and Raiffa (1993) argued that one-reason decision-making "is more widely adopted in practice than it deserves to be", is "naively simple" and "will rarely pass a test of reasonableness" (pp. 77–78). However, while they never provided such a test, others did.

In a series of computer simulations based on 20 real-world data sets, including the prediction of house-sale prices, Czerlinski et al. (1999) compared the LEX heuristic with information-hungry forecasting methods, such as multiple linear regression. In the case of fitting the data sets, the more flexible multiple regression analysis led to the highest percentage of correct decisions. However, if the criterion for comparison was a more realistic prediction task, where the models had to predict the future, rather than merely

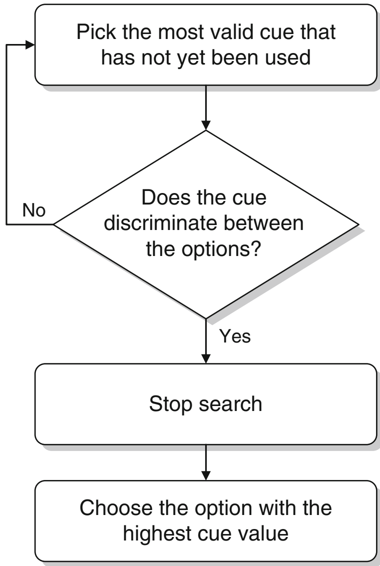


Figure 10.1 Flow chart of a simple lexicographic decision rule

post-predicting the past, they found to their surprise that LEX repeatedly outperformed its competitors.

The effectiveness of simple decision heuristics is not limited to the data sets analyzed by Czerlinski et al. or comparisons to multiple regression analysis. Brighton (2006) showed that, across a number of inductive inference tasks, LEX also outperformed a wide selection of state-of-the-art machine learning algorithms, including neural networks, exemplar models and elaborate classification trees. As a performance measure, Brighton used the models' ability to accurately predict new data in real-world environments, as well as the "minimal description length" criterion (Pitt et al., 2002), which indicates how well a model can compress the data. With regard to both criteria, in most situations LEX did as well as or even better than its allegedly more powerful competitors.

Relying on the first reason that comes to mind: the take-the-first heuristic

The usefulness of simple heuristics is not constrained to computer simulations and macro-economic data but can be applied to many real-life contexts, including the domain of professional sports. Imagine a handball player in a fierce game who has to decide on the next move. As handball is a fast sport, decisions have to be made within fractions of a second. Wrong decisions may cause the loss of the ball and sometimes even the whole game. In such a situation, the classical rational choice model would predict that having more time to think before passing the ball (or being able to process more information in the same time) would give players a competitive advantage, because they have more time to evaluate their options, estimate the probabilities that a certain move will be successful and predict the behaviour of the opponent. Yet, it turns out that this is not the case. In a series of studies on handball players, Johnson and Raab (2003) found that those players who had little time to think made better decisions than those who were given more time. Similarly, Beilock et al. (2004) showed that professional golfers were more accurate, if they had less time to consider how to approach the task.

After thorough analyses of the data, it appeared that players relied on the "take-the-first" heuristic, i.e. when time-pressure was tight, they chose the first move or pass that came to mind. As it turned out, for these highly experienced players, the options that came up first were usually the better ones. Even though more time enabled them to generate more options, these were shown to decrease in quality. Eventually, the players with more time to consider ended up with more low-quality options on the table, which, in turn, increased the chances that one of the low-quality options was selected.

Equal spread among all options: the 1/*N* heuristic

It could be argued that the success of simple heuristics might only apply to small or somewhat less important decisions, in which people avoid the additional effort required by a more thorough strategy. This argument can be countered by studying the more "important" case of financial asset allocation. To find out how people in the real world go about investing their money, Huberman and Jiang (2006) analyzed the records of more than half a million people who participated in pension plans (401(k) plans). They found that, when deciding in which funds to invest, the majority of people chose a small number of funds and then allocated their contributions evenly across them (Benartzi and Thaler, 2001; Loomes et al., 1991). This strategy is known as the 1/*N* heuristic, which can, indeed, be traced back to the 4th century, where Rabbi Isaac Bar Aha in the Talmud recommended splitting one's wealth equally among several investments.

From the perspective of neoclassical economic theory, the 1/*N* heuristic constitutes poor reasoning, sometimes mocked as "couch potato investments" or "coward portfolios", because investors ought rather to base their decisions on sophisticated statistical methods, such as probabilistic scenario analyses, that aim to optimize the mean variance, or the portfolio risk-return profile (Huberman and Jiang, 2006).

However, in a thorough comparison of fourteen optimization models for portfolio choice across several empirical data sets, De Miguel et al. (2007) found that none was consistently better than the apparently naïve 1/*N* heuristic, in terms of Sharpe ratio, certainty-equivalent return or turnover.

They showed that in order for the "optimal" portfolio strategies to achieve a higher certainty-equivalent return than the 1/*N* heuristic, the optimal models would need a portfolio with only 25 assets to have 291 years' worth of stock-market data for parameter estimation. For a portfolio with 50 assets, this period increases to more than 500 years, which is in sharp contrast to the common practice in which the parameters of these models are typically estimated using only 5 to 10 years of data (De Miguel et al., 2007). Thus, while the optimal models might theoretically lead to a superior outcome, for any real-world situation, in which people commonly do not have the patience to wait for 500 years, applying the 1/*N* heuristic is a sensible alternative.

Given these results, it is little surprise that the 1/*N* heuristic is not only used by laypeople but also by designated experts in the field. The Nobel laureate, Harry Markowitz, who developed a model for optimal portfolio selection (Markowitz, 1952), reported that he used the 1/*N* heuristic for his own, private investments (Gigerenzer, 2008). Moreover, the tendency to equally spread one's investments is not confined to financial decisions, and reliance on the 1/*N* heuristic can be observed in many other circumstances. For example, parents usually try to invest equally in their children (Hertwig et al., 2002), and across many cultures, equal splitting is a widely accepted norm for resource allocation (Leman et al., in press).

Simple heuristics in project management

In a thorough analysis of the role of uncertainty in project management, Pich et al. (2002) acknowledge that project teams often use heuristics to generate their policies, suggesting that the proliferation of heuristics is also apparent in management and business contexts, especially when it comes to project scheduling problems (Davis and Patterson, 1975; Hartman and Kolisch, 2000; Russell, 1986).

One recent and well-documented example of how heuristics can be successfully applied in project management stems from a study published in the renowned journal *Management Science* by Astebro and Elhedhli (2006). They provided evidence that experts' decisions were essentially made by the use of a heuristic strategy and that the use of this simple heuristic outperformed a sophisticated log-linear multiple regression analysis in predicting the future commercial success of R&D projects at an early stage.

In their study, Astebro and Elhedhli used records from 561 randomly selected R&D projects. Outcomes of R&D projects are notoriously uncertain, while at the same time, the potential feedback is patchy, and decision-relevant information is not easily quantified – hence, they chose a situation where accurate predictions are especially hard to come by. For each project, they acquired information indicating whether the product successfully reached the market (the criterion that had to be predicted), as well as 37 independent pieces of information (e.g. cost of production, tooling costs, existing competition, etc.), that could already be assessed at an early stage of the projects and thus could be utilized as cues for forecasting. Based on qualitative interviews, Astebro and Elhedhli found that, in order to predict project success, many expert analysts used a strategy described as a "non-compensatory tallying heuristic", which consisted of a search and a decision phase. Firstly, the experts evaluated the projects on 37 different criteria or cues, including aspects such as technical feasibility, expected costs and market development. For each cue, the experts rated if the project would do well, if it would do poorly or if would be critically flawed. Next, they added up the number of "good" and "bad" cues separately. If no critical flaws existed, if the number of good cues exceeded a fixed threshold *g*, and the number of bad cues was lower than a fixed threshold *b*, they predicted a success; if not, they predicted a failure. It turned out that this heuristic outperformed a competing logistic regression model which incorporated all available cues in out-of-sample prediction. It also did better than a stepwise log-linear model using backwards variable elimination.

The tallying model that mimicked experts' forecasting rules still used 33 out of the 37 possible cues. Interestingly, further analyses by Astebro and Elhedhli showed that experts actually used considerably more information than necessary and that they could have improved their predictive accuracy by using only 21 cues. Zacharakis and Meyer (2000) showed a similar result, where venture capitalists improved their predictions when provided with less information about the venture.

Similar results showing that less information and less computational complexity can lead to better outcomes were reported as long ago as 1979 by Makridakis and Hibon, who compared the accuracy of several statistical forecasting models in predicting future data ("out-of-sample" accuracy). The comparison was based on 111 real-life time series, covering a wide range of contexts, including business, industry and macro-economic data. What they found was that simple forecasting methods, e.g. calculating an ordinary moving average, repeatedly outperformed sophisticated ones, e.g. ARIMA models or multiple linear regression. Even though they were not the first to find such results (Dawes, 1979; Newbold and Granger, 1974; Reid, 1969, 1975), their findings provoked strong objections from statisticians and other decision scientists. To meet these objections, Makridakis et al. (1982) launched a number of so-called "M-Competitions", in which they invited their critics to let their models compete in a new set of time series. While the competing models were even more sophisticated than before (e.g. decomposition models, expert systems and neural networks), the results were strikingly similar to those in the earlier study. Complex methods did not provide more accurate forecasts than simpler ones (Makridakis and Hibon, 2000).

This is a small sample from countless well-documented cases in which lay people, as well as experts, employ simple heuristics to accomplish their goals across a wide range of situations (Gigerenzer, 2004b; Gigerenzer et al., 1999). This suggests that (a) heuristics are not an exception but rather the rule when it comes to decision-making in dynamic real-world environments that are characterized by uncertainty and information complexity, and that (b) experts who rely on simple heuristics can be as good as or even better at predicting uncertain outcomes than machines relying on complex statistical models.

Why do project managers need to know how people reason?

With regard to the descriptive aspect of heuristics research, it might be argued that knowledge about human decision-making strategies in real-world situations is a purely academic question that lacks practical importance. On the contrary, it can be argued that the success of many projects crucially depends on project managers' understanding of how people actually reason, and that, in many situations, people employ rather fast and frugal heuristics to derive their decisions.

In 1998, the German government made a law that allowed private households to freely choose their electricity provider. Similar to other attempts of market liberalization, one of the underlying rationales of this policy was to increase competition and lower prices. Yet, a couple of years later, it turned out that only very few consumers had switched their provider, even though most of them could have saved money by switching. From the perspective of humans as fully rational agents, who strive towards utility-maximizing, a convincing explanation for this apparent inertia to change was that people simply did not know that they had a choice or that they did not know the alternatives. Consequently, in 2002, a major German electricity producer launched a large-scale advertising campaign worth €22m. informing private households about their opportunity to switch. To their surprise, by the end of the campaign as few as 1100 customers had made use of this offer, resulting in a net cost of €20,000 for each new customer – rather an expensive advertising campaign.

From the perspective of decision research, one explanation for this miscarriage is that managers had the wrong representation of humans as utility-maximizers who trade-off costs and benefits and who would switch their provider as soon as there is a marginal monetary incentive to do so.

The assumption that people trade-off aspects against each other, and that they choose options with the highest net outcome, is not unique to German managers. When it comes to tools used to analyze and describe human decision-making, it is the rule rather than the exception. One prominent representative of such a tool, which remains to be widely used, is conjoint analysis. The term "conjoint analysis" stems from "consider-jointly" and describes a collection of sophisticated methods that assume decision-makers weigh the pros and cons (i.e. the risks and benefits) of each expected outcome, in order to derive an overall expected utility for each option. This feature of conjoint analysis is similar to the notion of multi-attributive utility theory and is shared by the method of cost benefit analysis, where money, rather than expected utility, depicts the common currency.

In sharp contrast to this notion, a growing body of literature shows that human reasoning might better be described and predicted by heuristic processes. In a review of 45 studies that investigated people's decision strategies in a variety of contexts, including choices between apartments, microwaves and birth control methods, Ford et al. (1989) found that decision-makers often evaluate surprisingly small amounts of information, and in many situations do not trade-off the pros and cons of the options but rather rely on a lexicographic heuristic.

Along the same lines, a recent study by Yee et al. (2007) found that the simple LEX strategy predicted people's decisions better than a state-of-the-art conjoint analysis using highly parameterized linear models and Bayesian estimation algorithms. In their experiment, participants chose from an assortment of smart phones that differed in design, functionality, operating system and other attributes. The results were subsequently confirmed by another group of researchers, who studied choices between laptop computers (Kohli and Jedidi, 2007).

Insofar as humans tend to make decisions without considering all information, and without making many trade-offs, decision tools which assume a fully rational and compensatory decision process, can grossly misrepresent the actual decisions and preferences of the target group. This can lead to wrong predictions of market potential, as is illustrated by the example above.

Moreover, decision-makers can also be influenced by procedures to elicit their preferences. For instance, Wilson and Schooler (1991) and Wilson et al. (1993) showed that people chose different options depending on whether they were instructed to make a decision based on their intuition or on more informed consideration. More importantly, after 2 weeks, those who had relied on their gut feelings were more satisfied with their choices than those who had made a considered choice. This indicates that procedures such as MAUT, which enforce a compensatory decision strategy, might lead to inferior and less satisfying decision outcomes.

Why do heuristics work?

It has been shown that decision-makers use heuristics, and these can outperform even vastly complex and information-hungry decision tools. Knowing this, the next challenge is to explain why and in which situations heuristics are useful. The key to answer these questions are two principles that are outlined in more detail below – domain-specificity and robustness.

Heuristics are domain specific. Heuristics are capable of ignoring a lot of "noise", and thus make robust predictions, because they can exploit certain patterns of information in the environment. For example, when trying to predict the outcome of the Men's Singles competition at Wimbledon, one could employ the "recognition-heuristic" (Goldstein and Gigerenzer, 2002). For predicting Wimbledon outcomes, this heuristic works as follows: if a player whose name is recognized competes against an unknown player, the prediction is that the known player will win the match. This essentially simple heuristic yields a surprisingly high accuracy, that can be as good as expert predictions or those based on official ATP rankings (Scheibehenne and Bröder, 2007; Serwe and Frings, 2006). As the heuristic relies on partial ignorance, it can only be used for those cases in which one of the players is unknown. If neither player is known, it does not apply. The success of the recognition heuristic depends on certain information structure in the environment, namely, that there is a systematic relationship between name recognition and success. Here, successful players gain more media attention in the time before the competition, and those players also have a higher probability of winning a match. The degree to which name recognition is linked to a certain criterion (e.g. success at Wimbledon) is known as "recognition validity". The recognition heuristic is domain specific, in that it works well in environments where the recognition validity is high.

There are a number of studies showing that recognition validity is surprisingly high across a wide range of situations, yet it is important to stress that it is not always the case (Pohl, 2006). This leads to an important aspect about why and when heuristics work. Their success depends on how well they fit the structure of the environment. Thus, heuristics are not rational per se; rather they are "ecologically rational". The Nobel laureate Herbert Simon (1990) illustrated this principle using the analogy of a pair of scissors with two blades. One blade depicts the decision strategy (i.e. the heuristic), while the other depicts the structure of the environment (i.e. the situation or circumstances). The scissors only cut when the two blades match up.

To illustrate this principle, the findings by Johnson and Raab (2003) showed that time-pressure led to better decisions for highly trained handball players. These players could benefit from the take-the-first heuristic, because the options that first came to their minds were usually the better ones, and the quality of additional options that they generated decreased over time. With this in mind, and in line with the idea of ecological rationality, Johnson and Raab further showed that, unlike experienced handball players, unskilled players did not benefit from time-pressure. For them, probably due to their lack of experience, good options did not necessarily leap out but rather had to be searched for. In this case, having more time increased their chances of finding a good option. These results suggest that a heuristic such as take-the-first is not universally recommended but rather depends on a certain environment – in this case, the individual player's experience.

Similarly, with respect to the 1/*N* investment heuristic outlined earlier, De Miguel et al. (2007) identified circumstances in which the heuristic underperformed against optimization models, namely, in situations where the number of investible assets is small, the assets differ widely in their return and the estimation window is long.

These examples show that the usefulness of a heuristic is context-dependent and that there is no such thing as a heuristic that is useful for everyone or in every situation. Rather, heuristics are tools that are tailored specifically to a given situation or problem. Just as a screwdriver or pair of pliers are made to solve a specific task, and fit to their respective environment, i.e. the head of the screw, each heuristic evolves to solve a specific decision task and achieves this by exploiting the structure of information in the environment. To continue with the metaphor of heuristics as specific tools, the repertoire of heuristics that decision-makers can choose from is referred to as the "adaptive toolbox" (Gigerenzer and Selten, 2001).

It might be asked, though, how people go about picking the right tool from the toolbox? There are many possible ways. Probably the most important one is that of learning through feedback. For example, in an experiment by Rieskamp and Otto (2006), minimal outcome feedback was apparently sufficient to learn which heuristic worked best in a given situation. Alternatively, to select the right heuristics from the toolbox, decision-makers can also resort to imitation of the successful (Garcia-Retamero et al., 2006) explicit instructions or advice from more experienced people, to name but a few.

Heuristics are robust. In hindsight, everything else being equal, a more complex model will always outperform a simple one. Yet for most real-world situations, what matters is out-of-sample accuracy – the prediction of the future. For example, after a project is finished, it may be easy to identify the reasons and the chain of events that eventually led to its success or failure. While this might be interesting information in itself, what one really wants to know is if these reasons will still apply for predicting the success of future projects. However, predicting the future is much harder than merely giving ex-post explanations of things that have already happened.

When attempting to predict upcoming future developments, available information can be divided into two parts: one part provides relevant data, the other part is irrelevant, or "random noise", that should best be ignored. The tricky part is to decide which particular piece of information is relevant and which is not. In general, the more uncertain the environment, the less information can be used to predict the future and the more data should be ignored (Gigerenzer, 2008). For example, when predicting the outcome of the stock market in the future, most past events are irrelevant, because they are already fed into the current stock prices. In this situation, a sophisticated model that incorporates lots of this past information will be quite capable of giving an accurate description of available data, but it will perform poorly in predicting an uncertain future, where new events will occur that do not necessarily resemble the past. This discrepancy between the ability to describe readily available data and to predict new data is known as "overfitting".

The reason why simple models often work better than more complex ones is because they ignore less important pieces of information and tend to focus on just a few, valid cues. For example, heuristics like LEX and take-the-first only use one very important cue and ignore all the remaining information. Simple heuristics can reduce the risk of overfitting, a property sometimes referred to as "robustness" and one that is especially valuable in uncertain environments.

Intuitively, one might think that the more uncertain an environment, the more information one should gather, and yet, as outlined above, a large body of evidence in such diverse fields as artificial intelligence, linguistics, psychology, economics and operations research suggests that it is precisely the other way round: the more uncertain the outcome, the less information should be employed to make a choice.

Conclusion

In uncertain and dynamic environments, and for many relevant project decisions, an optimization approach is often difficult and sometimes virtually impossible. In such situations, heuristics provide a feasible way to make decisions. Contrary to the common view of heuristics as second-best solutions that people only use because of their limited information-processing capabilities, the research on fast and frugal heuristics has provided substantial evidence that heuristics often achieve an astonishingly high performance, using just a fraction of the time and the amount of information required by standard decision strategies. Heuristics achieve this success for two reasons: firstly, they are ecologically rational, i.e. they are adapted to a specific task environment; secondly, they are robust, because they ignore much information, which makes them less vulnerable to random noise. By focusing only on relevant pieces of information, heuristics are well able to adapt to new situations.

To summarize, even though research on simple heuristics continues to emerge, the insights gained thus far already disclose a promising approach towards understanding and improving decision-making in project management, for instance, when allocating resources, or in predicting future outcomes at an early stage.

References

Adams J. (1995). *Risk*. UCL Press: London.

Arrow K. (2004). Is bounded rationality unbounded rational? Some ruminations. In: Augier M and March J (eds). *Models of a Man: Essays in Memory of Herbert A. Simon*. MIT Press: Cambridge, MA.

Astebro T and Elhedhli S. (2006). The effectiveness of simple decision heuristics: forecasting commercial success for early-stage ventures. *Management Science* **52**:395–409.

Beilock S, Berenthal B, Mccoy A, and Carr T. (2004). Haste does not always make waste: expertise, direction of attention, and speed versus accuracy in performing sensimotor skills. *Psychonomic Bulletin and Review* **11**:373–379.

Benartzi S and Thaler R. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review* **91**:79–98.

Brighton H. (2006). Robust inference with simple cognitive models. In: Lebiere C and Wray B (eds). *Between a Rock and a Hard Place: Cognitive Science Principles Meet AI-Hard Problems. Papers from the AAAI Spring Symposium*. AAAI Press: Menlo Park, CA.

Brighton H and Todd P. (2007). Situating rationality: ecologically rational decision making with simple heuristics. In: Robbins P and Ayedede M (eds). *Cambridge Handbook of Situated Cognition*. Cambridge University Press: Cambridge.

Czerlinski J, Gigerenzer G, and Goldstein D. (1999). How good are simple heuristics? In: Gigerenzer P and The ABC Research Group (eds). *Simple Heuristics that Make Us Smart*. Oxford University Press: New York.

Davis E and Patterson J. (1975). A comparison of heuristic and optimum solutions in resource-constrained project scheduling. *Management Science* **21**:944–955.

Dawes R. (1979). The robust beauty of improper linear models in decision making. *American Psychologist* **34**:571–582.

De Meyer A, Loch C, and Pich M. (2002). Managing project uncertainty: from variation to chaos. *Mit Sloan Management Review* **43**:60–67.

De Miguel V, Garlappi L, and Uppal R. (2007). Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*. Forthcoming, doi:10.1093/rfs/hhm075

Ford J et al. (1989). Process tracing methods: contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes* **43**:75–117.

Garcia-Retamero R, Takezawa M, and Gigerenzer G. (2006). How to learn good cue orders: when social learning benefits simple heuristics. In: Sun R and Miyake N (eds). *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Erlbaum Mahwah: New Jersey.

Gigerenzer G. (2004a). Striking a blow for sanity in theories of rationality. In: Augier M and March J (eds). *Models of a Man: Essays in Memory of Herbert A. Simon*. MIT Press: Cambridge, MA.

Gigerenzer G. (2004b). Fast and frugal heuristics: the tools of bounded rationality. In: Koehler D and Harvey N (eds). *Blackwell Handbook of Judgment and Decision Making*. Blackwell Publishing: Oxford.

Gigerenzer G. (2008). Why heuristics work. *Perspectives on Psychological Science* **3**:20–29.

Gigerenzer G and Goldstein D. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review* **103**:650–669.

Gigerenzer G and Selten R. (2001). *Bounded Rationality: The Adaptive Toolbox*. MIT Press: Cambridge, MA.

Gigerenzer G, Todd P, and The ABC Research Group. (1999). *Simple Heuristics that Make Us Smart*. Oxford University Press: New York.

Goldstein D and Gigerenzer G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review* **109**:75–90.

Hartmann S and Kolisch R. (2000). Experimental evaluation of state-of-the-art heuristics for the resource-constrained project scheduling problem. *European Journal of Operational Research* **127**:394–407.

Hertwig R, Davis J, and Sulloway F. (2002). Parental investment: how an equity motive can produce inequality. *Psychological Bulletin* **128**:728–745.

Huberman G and Jiang W. (2006). Offering versus choice in 401(k) plans: equity exposure and number of funds. *Journal of Finance* **61**:763–801.

Johnson J and Raab M. (2003). Take the first: option-generation and resulting choices. *Organizational Behavior and Human Decision Processes* **91**:215–229.

Keeney R and Raiffa H. (1993). *Decisions with Multiple Objectives*. Cambridge University Press: Cambridge.

Kiker G et al. (2005). Application of multicriteria decision analysis in environmental decision making. *Integrated Environmental Assessment and Management* **1**:95–108.

Kohli R and Jedidi K. (2007). Representation and inference of lexicographic preference models and their variants. *Marketing Science* **26**:380–399.

Leman P, Keller M, Takezawa M, and Gummerum M. (in press). Children's and adolescents' decisions about sharing money with others. *Social Development*. Doi 10.1111/j.1467-9507.2008.00486.

Loomes G, Starmer C, and Sugden R. (1991). Observing violations of transitivity by experimental methods. *Econometrica* **59**:425–439.

Makridakis S and Hibon M. (1979). Accuracy of forecasting: an empirical investigation. *Journal of the Royal Statistical Society* A **142**:97–145.

Makridakis S and Hibon M. (2000). The M3-competition: results, conclusions and implications. *International Journal of Forecasting* **16**:451–476.

Makridakis S et al. (1982). The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting* **1**:111–153.

Markowitz H. (1952). Portfolio selection. *Journal of Finance* **7**:77–91.

McDaniels T, Gregory R, and Fields D. (1999). Democratizing risk management: successful public involvement in local water management decisions. *Risk Analysis* **19**:497–510.

Newbold P and Granger C. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society* **137**:131–165.

Payne J, Bettman J, and Johnson E. (1993). *The Adaptive Decision Maker*. Cambridge University Press: New York, USA.

Pearce D and Nash C. (1981). *The Social Appraisal of Projects*. Macmillan: London.

Pich M, Loch C, and De Meyer A. (2002). On uncertainty, ambiguity, and complexity in project management. *Management Science* **48**:1008–1023.

Pitt M, Myung I, and Zhang S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review* **109**:472–491.

Pohl R. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making* **19**:251–271.

Reddy R. (1988). Foundations and grand challenges of artificial intelligence: AAAI presidential address. *AI Magazine* **9**(4):9–21.

Reid D. (1969). *A Comparative Study of Time Series Prediction Techniques on Economic Data*. PhD Thesis, Department of Mathematics, University of Nottingham: Nottingham.

Reid D. (1975). A review of short term projection techniques. In: Gordon H (ed.) *Practical Aspects of Forecasting*. Operational Research Society: London.

Rieskamp J and Otto P. (2006). SSL: a theory of how people learn to select strategies. *Journal of Experimental Psychology-General* **135**:207–236.

Russell R. (1986). A comparison of heuristics for scheduling projects with cash flows and resource restrictions. *Management Science* **32**:1291–1300.

Scheibehenne B and Bröder A. (2007). Predicting Wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting* **23**:415–426.

Serwe S and Frings C. (2006). Who will win Wimbledon? The recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making* **19**:321–332.

Simon H. (1990). Invariants of human behavior. *Annual Review of Psychology* **41**:1–19.

Todd P and Gigerenzer G. (2007). Environments that make us smart: ecological rationality. *Current Directions in Psychological Science* **16**:167–171.

Tversky A and Kahneman D. (1974). Judgment under uncertainty: heuristics and biases. *Science* **185**:1124–1131.

Tversky A and Kahneman D. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* **47**:263–291.

Wilson T and Schooler J. (1991). Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology* **60**:181–192.

Wilson T. et al. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin* **19**:331–339.

Yee M, Dahan E, Hauser J, and Orlin J. (2007). Greedoid-based noncompensatory inference. *Management Science* **26**:532–549.

Zacharakis A and Meyer G. (2000). The potential of actuarial decision models: can they improve the venture capital investment decision? *Journal of Business Venturing* **15**:323–346.

11
Expert Judgement of Probability and Risk

George Wright, Fergus Bolger and Gene Rowe

This chapter reviews extant research on the quality of expert judgement of probability and risk. Both types of judgement are of great relevance to the concerns of this book, because expert judgement under uncertainty is a key component in the making of decisions with scant information.

In the first section, we focus on the conditions under which experts have produced high quality assessments of probability – in terms of both the coherence and validity of assessed probabilities. We conclude that invalid probability judgements are common when features of the judgement domain are unfavourable and "learnability" is low – by which we mean that there is a lack of relevant background data and/or outcome feedback upon which to base and revise domain models. We identify judgement situations when such conditions are likely to be prevalent. We also show that use of ill-matched probability elicitation methods are a common operational flaw. We propose a task taxonomy that can be used to identify when expert probability judgements are likely to be well-made.

Next, we consider whether expert judgements of risk – by which we mean judgements of the likelihood of occurrence of hazardous events – are of higher quality than those made by members of the general public. We evaluate the nine empirical studies that have been conducted on expert versus lay judgements of risk, and find that there is little empirical evidence for the propositions (1) that experts judge risk differently from members of the public or (2) that experts are more veridical in their risk assessments. We discuss the nature of expertise and consider whether the commonsense assumption of the superiority of expert risk assessors in making magnitude judgements is, in fact, sensible. We end with a discussion of future research directions.

Expert judgement of probability

Expert judgement about uncertainty can be expressed and elicited in many ways – for example, as verbal probability expressions, numerical probabilities, odds, certainty factors, fuzzy sets and second order probabilities. The literature on behavioural decision-making has focused almost entirely on subjective probability because of the philosophical attractiveness of the axiom base which leads, logically, to expected utility as a choice principle. Expected utility underpins decision analysis (Goodwin and Wright, 2004), while subjective probabilities are used in forecasting techniques such as cross-impact analysis (Dalkey, 1972; Wright and Ayton, 1987). In expert system approaches to decision support, subjective probabilities are also a widely used way of representing the degrees of belief of experts (Kanal and Lemmer, 1986; Lemmer and Kanal, 1988; Shafer, 1987; Tonn et al., 1992). However, what do we know about the quality of *expert* probability judgement? If the probability assessments elicited from domain experts are poor in some respect, then the decisions of any decision-aiding technique or system which uses these probabilities will also be poor. Alternatively, if experts are capable of providing high quality probability assessments then poorly designed elicitation techniques may degrade this ability. Ensuring the quality of expert probability judgement is, therefore, of major practical importance to the implementation of decision support.

Two of the most commonly used approaches to the assessment of the quality of probability judgements are *coherence* and *calibration*. By coherence, we mean the extent to which a probability assessor's forecasts conform to the axioms of probability theory. For instance, according to one axiom of probability theory – additivity – the component probabilities of a set of mutually exclusive and exhaustive events should add up to one. Another axiom is the intersection law, which states that the probability of event *A* and event *B* both happening is the product of the probability of event *A* multiplied by the probability of event *B*, given that event *A* has happened. More formally, $p(A \text{ } and \text{ } B) = p(A)P(B|A)$. Other probability laws determine other relationships between such "complex" probabilities and their "simple" components of marginal and conditional probabilities. In studies of coherence, probability judgements are assessed in terms of the extent to which they conform to axioms, which basically means that the judgements are consistent or reliable (see Goodwin and Wright, 2004; Yates, 1990). In studies of calibration, probability judgements are validated against an external objective standard. For perfect calibration, a set of events all allocated a 0.XX probability of occurrence by a forecaster should, in actual fact, occur on XX% of occasions. For example, if we examined all the days where a perfectly calibrated weather forecaster assessed the probability of rain to be 0.7, then we should find the true proportion of rainy days also to be 0.7 (see Lichtenstein et al., 1982;

Yates, 1990 for reviews). Calibration is, therefore, a measure of the validity of a set of subjective probability assessments.

It has been argued that coherence and calibration are logically interrelated in the same way that reliability and validity are related in measurement theory (Wallsten and Budescu, 1983; Wright and Ayton, 1987). Reliability is usually regarded as a necessary but not sufficient prerequisite for validity. In terms of analogy, valid judgements must be reliable, but reliable judgements are not necessarily valid (Carmines and Zeller, 1979). For example, a metre rule which is actually 99 cm long will give reliable but non-valid measurements. However, an elastic rule which changes length from time to time is both unreliable and, generally, invalid. Thus, a coherent judge has the potential to be well-calibrated but an incoherent judge is, logically, incapable of being systematically well-calibrated.

Research on calibration and coherence

Research on the quality of calibration performance of experts' probability assessments – usually with respect to forecasting performance – has been found, in several instances, to be very good; for example, Kabus (1976) (financial interest rates); Hoerl and Fallin (1974) (horse racing); Keren (1987) (the card game, Bridge); and, as we shall discuss in detail later, most strikingly in weather forecasting – Murphy and Brown (1985). Conversely, in several instances poor calibration has been found e.g. Oskamp (1965) (clinical psychologists); Wallace (1923) (maize judges). One explanation for those instances, where poor expert calibration has been found, might be lack of coherence (recall the argument relating reliability and validity).

A number of studies have demonstrated incoherence in *students'* probability judgement using paper-and-pencil tasks (e.g. Kahneman et al., 1982; Yates, 1990), although we have located only a few studies of (in)coherence in experts. For example, Eddy (1982) reviewed medical literature on the diagnosis of breast cancer from X-rays and found that physicians misunderstood the relationship between marginal probabilities (e.g. the probability of cancer, the probability of positive test, etc.) and conditional probabilities (e.g. the probability of a positive test given cancer). However, Schafer et al. (1977) found that, in two out of three tests, people who self-rated as more knowledgeable than their peers in soccer and statistics were slightly more coherent than those who rated themselves as less knowledgeable. In a study of coherence of self-rated experts in snooker, we found self-rated experts more coherent in their judgements of the probability of the union of two events than self-rated novices (Wright et al., 1994). Yet again, in a study of professional restaurant managers, Dube-Rioux and Russo (1988) found failure to conform to the additivity axiom with respect to the disjunction and conjunction of probable causes of restaurant failure. Finally, DuCharme and Peterson (1968) and Youseff and Peterson (1973) found that the failure to revise probabilities in the light of new information required by Bayes' theorem was less marked in real-world tasks than had been previously found in the laboratory (e.g. Phillips and Edwards, 1966; Phillips et al., 1966).

Thus, as with calibration, the picture with respect to coherence is unclear. Experts sometimes demonstrate incoherence of the sort found in naive judges, but there is a suggestion that specific knowledge/expertise may, in some cases, lead to a reduction in the extent of this incoherence. Given this tension in the literature, it remains to be demonstrated that it is incoherence which is responsible for poor calibration. For example, it might plausibly be argued that incoherence and miscalibration are both *symptoms* of poor probabilistic judgement. In fact, there are a number of psychological explanations for poorly calibrated probability judgement (see, e.g. McClelland and Bolger, 1994). We shall examine some of these other possible causes of poor calibration performance in probabilistic judgement shortly. However, it should be noted that there is a sizeable literature that shows random error in probability judgements can explain much of the observed miscalibration, although not all (see Ayton and McClelland, 1997 for a review).

Decomposition-recomposition

Most procedures for producing coherent probability judgements involve a technique called decomposition-recomposition. For example, consider the abstract example of assessing the probability of drawing two consecutive aces from a pack of 52 cards. This probability (without replacement) is

$$p(2 \text{ aces}) = 4/52 * 3/51 = 12/2652 = 0.00452.$$

Intuitively, it would seem that most people could, with a little thought, accurately assess the probability of drawing an ace on the first draw and the probability of a subsequent second ace. It also seems intuitively reasonable that most people could not make an accurate mental computation of the probability of the intersection of the two events. On the basis of the observation that most incoherence is manifest when judges attempt to revise or combine probabilities, it has been assumed that it is difficulties in computation (not probabilistic estimation *per se*) which lead to errors. Thus, it has been proposed that judgement problems should be broken up (decomposed) into small elements for which judges supply probability estimates. Probability assessments are then reproduced by mechanistically combining the individual component judgements on the basis of the laws of probability theory (a process known as recomposition).

Decomposition-recomposition has been shown to produce more coherent complex probabilities than those assessed by judges holistically (e.g. Edwards et al., 1968), but are the probabilities more valid, for example, in terms of calibration? Unfortunately, there is little empirical evidence available to answer this question. Wright et al. (1988) found that calibration of unions, intersection and disjunctions of events which were mechanically recomposed from marginal and conditional assessments were no better calibrated than holistically judged forecast probabilities for the same events. Further, no significant correlation was found between the coherence and the subsequent calibration of the assessed probabilities. Wright et al.'s (1994) study also found no relationship between an assessor's degree of incoherence and his or her degree of (mis)calibration, thereby also suggesting that *increasing* a person's degree of coherence will not necessarily increase that person's calibration. Why should this be the case?

Greater reliability does not necessarily imply greater validity, although greater validity does imply greater reliability. Thus, poor calibration may be due to factors other than poor coherence. Suggestions include memory problems, use of heuristics, insensitivity to task difficulty and confirmation biases (see e.g. Compte and Postlewaite, 2004; Dougherty et al., 1999; Ferrell and McGoey, 1980; Griffin and Tversky, 1992; Tversky and Koehler, 1994). It, therefore, follows that the success of the decomposition-recomposition approach lies in the assumption that the "simple" probabilities elicited for recomposition (often marginal and conditional probabilities, as in the cross-impact technology) are themselves free of bias. If systematic error exists in these simple assessments then recomposition will magnify this bias.

Some influences on the assessed quality of judgement

Good calibration has been demonstrated in a number of instances but most notably in weather forecasting. We (Bolger and Wright, 1994) proposed that a skilled judge *can* give valid probability estimates *if* the task, elicitation and assessment procedures are amenable. Specifically, we proposed that invalid probability judgements arise when:

1. Valid probability judgement cannot easily be learned. This may be due to such influences as the amount and complexity of information, the degree to which events to be judged are related to an underlying domain model (e.g. likelihood of rain to a model of the weather) and (perhaps most importantly) the lack of outcome feedback in the task domain upon which to base and revise judgement. We term these task influences *learnability*.
2. The judge is unskilled due to lack of knowledge about the task domain and/or probability laws.
3. Probability estimates are elicited in a manner which makes them unrepresentative of the judges' true feelings of subjective probability; for example, by asking him or her to respond in an unfamiliar metric (such as odds to a non-betting person).

One instance where judgemental probability forecasts are routinely generated is weather forecasting. The official forecasts issued by the National Weather Service in the United States are subjective probability forecasts. Murphy and Brown (1985) have evaluated these subjective forecasts and found that, for certain categories of weather, they were more accurate than the available objective statistical techniques. In this case, the experimental forecasters have a very large amount of information available, including the output from statistical techniques (cf factor 2). They also receive detailed feedback (cf factor 1) and have the opportunity to gain experience of making forecasts under a wide range of meteorological conditions. Furthermore, they have considerable practice in quantifying their internal state of uncertainty (cf factor 3). These circumstances may well be ideal for the relatively successful application of judgemental compared to purely quantitative forecasting.

In our view, performance-demonstrated experience in probability judgements is underpinned by practice and regular performance feedback. However, as Einhorn and Hogarth (1978) have argued, most judgements are made without the benefit of accurate feedback. They identified three main problems:

1. The lack of search for and use of disconfirming evidence.
2. The use of unaided memory for coding, sorting and retrieving outcome information.
3. When people take an action based on a forecast in order to facilitate or avoid possible futures, they can often only observe feedback associated with the action taken, and not the action not taken.

To illustrate (3), Einhorn (1980) gives the following example:

> Imagine that you are a waiter in a busy restaurant and because you cannot give good service to all the people at your station, you make a judgement regarding which people will leave good or poor tips. You then give good or bad service depending on your judgement. If the quality of service, in itself, has an effect on the size of the tip, outcome feedback will "confirm the predictions" ("They looked cheap and left no tip – just as I thought"). The extent of such self-fulfilling prophecies is much greater than we think and represents a considerable obstacle to learning from outcome feedback.

This third feedback problem is, of course, immaterial in contexts such as weather forecasting where actions cannot be taken to increase or reduce the likelihood of the forecast event. Unconfounded feedback in such circumstances is likely to prove more useful for the improvement of forecasting ability.

Wagenaar and Keren (1986) further point out that feedback must be attended to in order for it to be of any use in improving judgement. They show that blackjack dealers were no better calibrated than "lay" people for judgements about the frequency of occurrence of certain types of hand, despite being exposed to thousands of examples each working week. They suggest that this finding is due to lack of motivation on the part of blackjack dealers to attend to the outcome feedback available to them. It is now well known that attention to the target stimuli is a necessary condition for learning about the frequency of occurrence of those stimuli.

Murphy and Brown (1985) have argued that the presence of actual or potential users of judgemental weather forecasts provides the forecasters with a strong motivation for conducting the forecasting process in an efficient and more effective manner. Moreover, feedback from users of forecasts frequently contains information regarding possible improvements. The use of judgement in real-world forecasting thus contrasts strongly with the study of judgement in the psychological laboratory, where calibration feedback and incentives for good performance have seldom been utilized.

The metric in which probability responses are elicited can take a number of different forms (e.g. percentages, point probability estimates, odds, relative frequency, etc.). Depending on which metric is used, the judge's task of turning subjective feelings of uncertainty into measurable/usable numeric estimates can be either helped or hindered. For example, Wright et al. (1988) found that for a short-term forecasting task of impersonal events (e.g. "will the pound fall below one dollar in the next 2 months?"), 29 out 36 students were better calibrated on point probabilities than odds. This experiment, therefore, gives some empirical support for the view that point probability estimates, not odds, should be elicited from untrained forecasters.

So far in this discussion of elicitation and assessment effects, we have not differentiated between the sorts of probability estimates that are being elicited (e.g. marginal or conditional? simple or complex? intersections or disjunctions?) because calibration studies have not tended to differentiate either. However, it seems to us that an important research question is the extent to which it is more natural to make some sort of probability assessments than others.

Decomposition implies that "simple" probabilities, such as marginals and conditionals, are easier for judges to assess than "complex" probabilities, such as intersections, disjunctions and unions. However, our earlier discussion of decomposition-recomposition found little evidence of a distinction between simple and complex assessed probabilities in terms of calibration performance. One possible reason for this rather surprising finding is that the problem decomposition used may not have been appropriate for the judges. In other words, the judges may have been *framing* the problem differently to the experimenter so that the decomposition did not result in easier-to-assess probabilities. Of course, a characteristic of expertise is surely that experts can discriminate questions of a type they can answer from those they cannot (see Shanteau, 1992).

Section conclusion

In order to reduce potential invalidity due to features of the task domain, we propose that a thorough task analysis should be performed before probabilities are elicited. Questions to be addressed include the following: has the judge had the opportunity to attain good probability through experience of feedback? How long is the feedback loop? Is the feedback loop sensitive to treatment effects? To what extent are items/events related? and is it possible to validate judgements against some external standard? If the conclusion of the task analysis is that the conditions for learnability are not present in the task domain, and/or there are no available objective criteria for validating probabilistic judgement (perhaps because the events to be forecast are in the far future), then the only strategy is to ensure judgements are coherent. As we have argued, in *the absence of validity measures* the decision analyst can *only* ensure such coherence.

One practical step that can be taken to ensure that the elicitation process influences the validity of probabilistic judgement as little as possible is to encourage judges to decompose the problem in their *own* preferred way, using their *own* preferred response. In some forecasting situations, it may be that the assessor will feel happier assessing marginal and conditional probabilities than compound probabilities (cf the playing card example, earlier). However, as we have argued, in other situations the forecaster may feel more comfortable assessing compound probabilities directly. Overall, perhaps the most flexible approach to debiasing subjective probability forecasts would be to adopt what Keren (1990) calls *structure-modifying* techniques, where the user is forced/encouraged to understand the internal logic of a particular debiasing technique rather than follow a procedure blindly.

Expert judgement of risk

In a pioneering paper, Lichtenstein et al. (1978) investigated how well people (students and convenience samples from the lay population) could estimate the frequency of the lethal events that they may encounter in life.

In their study, Lichtenstein et al. (1978) found that although their subjects exhibited some competence in judging such frequencies – frequency estimates increased with increases in true frequency – the overall accuracy of both (1) paired comparisons of the relative frequency of lethal events and (2) direct estimates of frequencies of individual events were poor. In a comment on the Lichtenstein et al. study, Shanteau (1978) argued that if respondents had had more experience with the lethal events the validity of the required estimates may have shown improvement. He concluded that “It might also be of some value to investigate judgement of lethal events, using subjects who have direct knowledge and exposure to such events (such as life insurance analysts)" (1978: p. 581).

Since the 1978 paper, research on risk judgements has led to the generally accepted conclusion that expert judgements are, indeed, more veridical than those of the general public (e.g. Slovic, 1987, 1999). One basis for this argument is the work by Slovic et al. (1985). In this study, the authors utilized samples of the US League of Women Voters, university students, members of the US Active Club (an organization of business and professional people devoted to community services activities) and a group of professional "experts". Perceptions of risk were measured by asking participants to order the 30 hazards from least to most risky (in terms of the "risk of dying (across US Society as a whole) as a consequence of this activity or technology") (1985: p. 116). Participants were told to assign a numerical value of 10 to the least risky item and to make other ratings relative to this value. Since these instructions called for a risk assessment, rather than a (relative) frequency estimate (cf Lichtenstein et al., 1978), the avenue was open – for both experts and nonexperts – for qualitative risk attributes, such as the voluntary nature or controllability of the risk, to enter into these global risk judgements.

Slovic et al. (1985) concluded that the judgement of their experts differed substantially from non-expert judgement primarily because the experts employed a much greater range of values to discriminate among the various hazards that they were asked to assess, which included motor vehicles, smoking, alcoholic beverages, hand guns, surgery, X-rays and nuclear power. Additionally, Slovic et al. (1985) concluded that their obtained expert-lay differences were "because most experts equate risk with something akin to yearly fatalities, whereas lay people do not" (1985: p. 95). This conclusion is founded on the fact that the obtained correlations between perceived risk and the annual frequencies of death were 0.62, 0.50 and 0.56 for the League of Women Voters, students and Active Club samples, respectively. The correlation of 0.92 obtained within the expert sample is significantly higher than those obtained within each of the lay samples. However, Slovic et al. (1985) also found that both the lay and expert groupings viewed the hazards similarly on qualitative characteristics such as voluntariness of risk, control over risk and severity of consequences – when asked *directly* to do so (see Rowe and Wright, 2001, for a full discussion). It would seem that when asked for a "risk" estimate, Slovic et al.'s experts viewed this as a magnitude estimation task rather than a qualitative evaluation task. Additionally, an artificial ceiling *may* have been placed on the evaluation of the veracity of magnitude estimates of risk made by the *lay* samples, *if* members of the lay groupings were more likely to view the task of making a "risk" estimate as one of qualitative evaluation.

Since Slovic et al.'s (1985) study of expert-lay differences in risk judgement, several other papers have taken a similar theme. These have used expert samples of toxicologists (Kraus et al., 1992; Slovic et al., 1995), computer scientists (Gutteling and Kuttschreuter, 1999), nuclear scientists (Flynn et al., 1993), aquatic scientists (McDaniels et al., 1997), loss prevention managers in oil and gas production (Wright et al., 2000) and scientists in general (Barke and Jenkins-Smith, 1993). These studies concluded that there are substantial differences in the way that experts and samples of the lay population judge risk. Generally, experts perceive the risks as less than the lay public with regard to the questions asked and the substantive domains. The two exceptions are the studies by Wright et al. (2000) – where experts and members of the lay public shared similarities in risk perception of hazardous events in oil and gas production in the North Sea, and Mumpower et al. (1987), where the rating of the political riskiness of countries by undergraduate students closely paralleled the ratings of professional analysts. Both these sets of results contrast sharply with results of Slovic et al. (1985), described earlier, where the experts saw 26 out of 30 activities/technologies as *more* risky than each of the three lay groupings. However, in all studies, except for the latter study, the relative validity of expert versus lay risk assessments (in terms of the veracity of frequency estimates) *has not* been measured – hence, the commonly accepted view about expert-lay differences in risk judgements rests on the results of a single study that used just 15 experts and which compared their judgements of "risk" with those of groups of lay persons on a task where the validity standard (mortality rates) was not made salient to the lay group. Further, it would seem highly unlikely that the experts who took part in the Slovic et al. study could have had substantive expert knowledge in all of the variety of hazards that were utilized (including mountain climbing, nuclear power and spray cans), which begs the question: Were they truly experts? This might also, in part, explain why the results from this expert sample were inconsistent with the results from expert samples in the other studies. In a review of these studies, Rowe and Wright (2001) concluded that, contrary to received wisdom, there is little empirical evidence for the proposition that experts are more veridical in their risk assessments than members of the public.

More widely, Bolger and Wright (1994) and Rowe and Wright (2001) have argued that in many real-world tasks, apparent expertise (as indicated by, for example, status) may have little relationship to any real judgement skill at the task in question. In Bolger and Wright's review of studies of expert judgemental performance they found that only six had showed "good" performance by experts, while nine had shown poor performance. As we have seen in section 1 of this chapter, Bolger and Wright analysed and then interpreted this pattern of performance in terms of the "ecological validity" and "learnability" of the tasks that were posed to the experts. To reprise, by "ecological validity" we mean the degree to which the experts were required to make judgements inside the domain of their professional experience and/or express their judgements in familiar metrics. By "learnability" we mean the degree to which it is possible for good judgement to be learned in the task domain. That is, if objective data and models and/or reliable and usable feedback are unavailable, then it may not be possible for a judge in that domain to improve his or her performance significantly with experience. In such cases, Bolger and Wright argued, the performance of novices and "experts" is likely to be equivalent, and they concluded that expert performance will be largely a function of the interaction between the dimensions of ecological validity and learnability – if both are high then good performance will be manifest, but if one or both are low then performance will be poor.

From the perspective of Bolger and Wright's analysis, it is by no means certain that expert risk assessors will be better at judging the veridical risks of hazards than lay persons, and the limited empirical evidence cannot be considered compelling (Rowe and Wright, 2001). This has important implications for the communication of *judgements* of risk. As Rowe and Wright (2001) have argued, in hazard evaluations where the hazardous events happen rarely, if at all, then learnability will be low, and the veridicality of judgements of the magnitude of risks by experts will be suspect. For example, consider the validity of expert predictive judgements about the likelihood magnitude of human infection by "mad cow disease" resulting from eating beef from herds infected with Bovine Spongiform Encephalopathy (BSE) in the early 1990s and the subsequent, poorly predicted, mortality rates (Maxwell, 1999). In this instance, UK politicians selectively used expert predictions to reassure a frightened general public.

Wright et al. (2002) considered the issue of expert-lay differences in frequency, and relative frequency, judgements of lethal events using a sample of professional risk assessors. They extended and developed the study of Lichtenstein et al. (1978) and followed up the suggestion in Shanteau's (1978) commentary on that paper. They utilized a sample of life underwriters, of varying degrees of experience, and a task requiring assessment of a varied set of potentially lethal events.

The results from their study revealed that although both lay and expert groups showed relatively good performance in terms of the ordering of the absolute likelihood (marginal probabilities) and lethality (conditional probabilities) of events, as demonstrated by significant obtained correlations, they also showed similar and systematic bias in terms of overestimating these values. Such overestimation was almost uniform over the hazards for the direct marginal judgements, although less so for conditionals. The student group was no worse at direct marginal or direct conditional estimation than the experts.

Because the *direct* estimation of risks associated with potentially lethal events is an unusual task, even for the experts (at least for marginal estimates, although for conditional estimates the Chief Underwriter stated that this assessment mode captured the essence of his work-a-day task), we also obtained marginal and conditional estimates in a second, *indirect* way, namely, through pairwise comparisons. Correlational analysis revealed a trend that the experts were indeed better at the task, in terms of identifying which events of the *pairs* led to more deaths (marginals) and were more lethal (conditionals), although these correlations were not significantly different from those of the lay group. However, further analysis revealed that the experts *did* make significantly better judgements than lay person on marginal estimates in terms of ratios (i.e. the number of times one event was more likely to cause death than another) and conditionals (i.e. the number of times an event was more likely to cause death than the other, given that the event happens to someone). In spite of this, both lay persons and experts made the same general errors in the pairwise comparison tasks – namely, in underestimating the ratio of more-to-less ubiquitous and fatal hazards by overly compressing their ranges of estimates.

Section conclusion

As we have reviewed, the evidence for experts being better at the judgements of risks is not strong (see Rowe and Wright, 2001 for a review) and yet has been so readily accepted that there has been no apparent effort to research the topic further. For "true" expertise to be manifest (expertise related to performance, as opposed to social and political imperatives), Bolger and Wright (1994) have argued that the expert must perform a task that is ecologically valid, and the task must also be learnable. Wright et al. (2002) attempted to ensure that their expert-task match was as strong as possible (given experimental limitations), and that ecological validity was high, and yet still obtained expert performance that was not much better than lay person performance. This result suggests that the underwriting task is not truly "learnable", i.e. it is not one for which there is regular feedback on the correctness or otherwise of judgements. Indeed, in the training of underwriters, performance is assessed according to the similarity of junior underwriters' judgements to those of their seniors (Bolger et al., 1989). Once "trained", underwriters receive infrequent performance-related, objective feedback about the correctness of their judgements, and indeed it would be difficult to provide such feedback, given that a "poor" judgement might turn out to be insuring an applicant who subsequently died of a condition after perhaps 20 years of a 25-year policy.

We infer that the tasks performed by *other* professional risk assessors may also be unlearnable. For example, in the case of major hazards in the nuclear industry there may be no risk/judgement feedback at all. From this, we suggest that expert-lay differences in the accuracy of such risk judgements, or in the nature of such judgements (given that the biases evidenced in the Wright et al. (2002) study were similar across lay and expert groups), cannot be assumed. Further, even if experts are significantly more accurate than lay people, it may still be that differences in accuracy are small, as demonstrated in the present study. Perhaps the commonsense assumption of the superiority of expert risk assessors in making risk judgements is ill-founded. Certainly, future research needs to pay more attention to the *de facto* nature of the learnability of tasks performed by professional risk assessors.

Advice giving and changes in judgement: directions for future research

Our general conclusion is that experts can make valid judgements of probability – if the task conditions are amenable. To make valid judgements, we contend that a prediction/outcome feedback loop must be in place to enable learnability. Also, the judgement task itself should be matched with the expert's knowledge base, and the metric used to elicit probability judgements should be both familiar and acceptable to the expert. Such conditions prevail in meteorologists' predictions of weather events – where excellent calibration is the rule. Similarly, in the domain of risk judgement, such task conditions must, we contend, also prevail for valid judgements to be elicited from experts. So, we conclude that the ideal situation for the expression of valid judgement by experts is when (i) the ecological validity of both the judgement task and the elicitation metric are both high and (ii) when the task itself is truly learnable. In many situations, of course, the second condition may not hold and expert judgement will, we contend, likely be as poor as that of the lay population and subject to similar heuristics and biases. In these situations, there will exist no track-record of prior judgements or associated hit-rate. Here, the judgements made are likely to be one-off or unique. In such situations, the expert has only access to his/her heuristics and the advice of other people – perhaps also experts. The advisors may disagree, so how are/should such disagreements resolved?

Rowe and Wright (1999) studied change in expert opinion amongst members of such expert groups and found that the experts who held the more accurate opinions changed their opinions less than experts who held less accurate opinions over rounds in a mediated group process called "Delphi". One avenue for future research is, therefore, the exchange of knowledge and opinion between experts, but this research area is, as yet, under-explored.

For example, Brockriede and Ehninger (1960) have shown that only a limited number of argument types are, in principle, available to people advocating specific propositions or claims – arguments of parallel case, analogy, motivation and authority.

In *Analogous reasoning*, the reason given makes use of our general knowledge of relationships between two events in dissimilar situations. For example, if someone is trying to estimate the time it will take to drive to a nearby airport, an advisor may reason that "the airport is roughly the same distance away as the shopping mall. Therefore, the time it will take to get to the airport will be approximately the same as it is to travel to the shopping mall – about 30 minutes."

Parallel case reasoning involves making use of our knowledge of a previous experience of a near identical situation. For example, if someone is trying to estimate the time it will take to drive to a nearby airport an advisor may reason that, "it will take about 30 minutes to drive to the airport because it took me 30 minutes at the same time of day last month."

Authoritative reasoning involves making use of substantive knowledge. For example, "the radio announcer has said that traffic to the airport is heavy today and so I estimate that you should add 20 minutes to your journey time."

Motivational reasoning involves making use of specific insights about people's motivations or desires. For example, "since you will be in a hurry, then I reckon that you can cut five minutes off your usual journey time."

Research has shown that these argument types can be persuasive in some circumstances (McCroskey, 1969; Smith, 1972; Stanchi, 2006), but outside a legal context, no research has been conducted to explore the persuasiveness effects of different forms of argument structure on opinion change in experts. Thus, many crucial questions remain to be explored and answered. For example, what components of advice-giving cause opinion-change in individual experts? How is advice evaluated and under what conditions will advice be assimilated or discounted? When one expert defers in his or her own opinion to the well-argued opinion of another, is this an indicator of the presence of valid advice that will improve validity in (the revised) judgemental prediction? In our view, the study of the use of argument in advice-giving will become a major topic in investigations focused on understanding and improving expert judgement of probability and risk in unique, or one-off, situations where the expert cannot utilize learnability, and the expert is aware that his/her own judgement may be influenced by inappropriate heuristics that could lead to bias.

References

Ayton P and McClelland AGR. (1997). How real is overconfidence? *Journal of Behavioral Decision Making* **10**:279–285.

Barke RP and Jenkins-Smith HC. (1993). Politics and scientific expertise: scientists, risk perception, and nuclear waste policy. *Risk Analysis* **13** (4): 425–439.

Bolger F and Wright G. (1994). Assessing the quality of expert judgement: issues and analysis. *Decision Making Systems* **11**:1–24.

Bolger F, Wright G, Rowe Gammack J, and Wood RJ. (1989). Lust for life: developing expert systems for life assurance underwriting. In: Shadbold N (ed.) *Research and Development in Expert Systems, VI*. Cambridge University Press: Cambridge.

Brockriede W and Ehninger D. (1960). Toulmin on argument: an interpretation and application. *Quarterly Journal of Speech* **46**:44–53.

Carmines EG and Zeller RA. (1979). *Reliability and Validity Assessment*. Sage University Papers Series: Beverley Hills, CA.

Compte O and Postlewaite A. (2004). Confidence-enhanced performance. *American Economic Review* **94**:1536–1557.

Dalkey N. (1972). An elementary cross-impact model. *Technological Forecasting and Social Change* **3**:341–351.

Dougherty MRP, Getys CF, and Ogden EE. (1999). MINERVA-DM: a memory processes model for judgements of likelihood. *Psychological Review* **106**:180–209.

Dube-Rioux L and Russo JE. (1988). An availability bias in professional judgement. *Journal of Behavioural Decision-Making* **1**:233–237.

Ducharme WM and Peterson CR. (1968). Intuitive inference about normally distributed populations. *Journal of Experimental Psychology* **78**:269–275.

Eddy DM. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities. In: Kahneman D, Slovic P, and Tversky A (eds). *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press: Cambridge.

Edwards W, Phillips LD, Hays WL, and Goodman BC. (1968). Probabilistic information processing systems. *IEEE Transactions on Systems Science and Cybernetics* **4**:248–265.

Einhorn HJ (1980) Learning from experience and sub-optimal rules in decision making. In Wallsten T (ed.) Cognitive Processes in Choice and Decision Behavior. Hillsdale, N.J.: Erlbaum.

Einhorn HJ and Hogarth RM. (1978). Overconfidence in judgement. Persistence of the illusion of validity. *Psychological Review* **85**:394–476.

Ferrell WR and McGoey PJ. (1980). A model for calibration of subjective probabilities. *Organizational Behavior and Human Decision Processes* **26**:32–53.

Flynn J, Slovic P, and Mertz CK. (1993). Decidedly different: expert and public views of risks from a radioactive waste repository. *Risk Analysis* **13** (6): 643–648.

Goodwin P and Wright G. (2004). *Decision Analysis for Management Judgement*. Wiley: Chichester.

Griffin D and Tversky A. (1992). The weighting of evidence and the determinants of overconfidence. *Cognitive Psychology* **24**:411–435.

Gutteling JM and Kuttschreuter M. (1999). The millennium bug controversy in the Netherlands? Experts' views versus public perception. In: Goossens LHJ (ed.) *Proceedings of the 9*th *Annual Conference of Risk Analysis: Facing the Millennium*. Delft University Press: Delft, Netherlands, pp. 489–493.

Hoerl A and Fallin HK. (1974). Reliability of subjective evaluation in a high incentive situation. *Journal of the Royal Statistical Society* **137**:227–230.

Kabus I. (1976). You can bank on uncertainty. *Harvard Business Review*, May–June, **54**:95–105.

Kahneman D, Slovic P, and Tversky A. (1982). *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press: Cambridge.

Kanal LN and Lemmer JF (eds). (1986). *University and Artificial Intelligence*. Elsevier: Amsterdam.

Keren G. (1987). Facing uncertainty in the game of bridge: a calibration study. *Organisational Behaviour and Human Decision Processes* **39**:98–114.

Keren G. (1990). Cognitive aids and debiasing methods. In: Caverni JP et al. (eds). *Cognitive Biases*. Elsevier: Amsterdam.

Kraus N, Malmfors T, and Slovic P. (1992). Intuitive toxicology: expert and lay judgements of chemical risks. *Risk Analysis* **12** (2): 215–232.

Lemmer JF and Kanal LN (eds). (1988). *Uncertainty and Artificial Intelligence 2*. Elsevier: Amsterdam.

Lichtenstein S, Fischhoff B, and Phillips LD. (1982). Calibration of probabilities: the state of the art to 1980. In: Kahneman D, Slovic P, and Tversky A (eds). *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press: New York.

Lichtenstein S, Slovic P, Fischoff B, Layman M, and Combs B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory* **4**:551–578.

Maxwell RJ. (1999). The British government's handling of risk: some reflections on the BSE/CJD crisis. In: Bennett P and Calman K (eds). *Communications and Public Health*. Oxford University Press: Oxford, pp. 94–107.

McClelland AGR and Bolger F. (1994). The calibration of subjective probabilities: theories and models 1940–94. In: Wright G and Ayton P (eds). *Subjective Probability*. Wiley: Chichester.

McCroskey JC. (1969). Toward an understanding of the importance of "evidence" in persuasive communication. *The Pennsylvania Speech Annual* **23**:65–71.

McDaniels TL, Axelrod LJ, Cavanagh NS, and Slovic P. (1997). Perception of ecological risk to water environments. *Risk Analysis* **17** (3): 341–352.

Mumpower JL, Livingston S, and Lee TJ. (1987). Expert judgments of political riskiness. *Journal of Forecasting* **6**:51–65.

Murphy AH and Brown BG. (1985). A comparative evaluation of objective and subjective weather forecasts in the United States. In: Wright G (ed.) *Behavioural Decision Making*. Plenum: New York.

Oskamp S. (1965). Overconfidence in case-study judgements. *Journal of Consulting Psychology* **29**:261–265.

Phillips LD and Edwards W. (1966). Conservatism in a simple probabilistic inference task. *Journal of Experimental Psychology* **72**:346–354.

Phillips LD, Hays WL, and Edwards W. (1966). Conservatism in complex-probabilistic inferences. *IEEE Transactions on Human Factors in Electronics* **7**:7–18.

Rowe G and Wright G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting* **15**:353–375.

Rowe G and Wright G. (2001). Differences in experts and lay judgements of risk: myth or reality? *Risk Analysis* **21**:341–356.

Schafer RE, Borcherding K, and Laemmerhold CL. (1977). Consistency of future event assessments. In: Jungermann H and de Zeeuw G (eds). *Decision Making and Change in Human Affairs*. Reidel, Dordrecht: The Netherlands.

Shafer G. (1987). Probability judgement in artificial intelligence and expert systems. *Statistical Science* **2**:3–44.

Shanteau J. (1978). When does a response error become a judgement bias? Commentary on "judged frequency of lethal events". *Journal of Experimental Psychology: Human Learning and Memory* **4** (6): 579–581.

Shanteau J. (1992). The psychology of experts: an alternative view. In: Wright G and Bolger F (eds). *Expertise and Decision Support*. Plenum: New York.

Slovic P. (1987). Perception of risk. *Science* **236**:280–285.

Slovic P. (1999). Trust, emotion, sex, politics and science: surveying the risk-assessment battlefield. *Risk Analysis* **19** (4): 689–701.

Slovic P, Fischhoff B, and Lichtenstein S. (1985). Characterizing perceived risk. In: Kates RW, Hohenemser C, and Kasperson JX (eds). *Perilous Progress: Managing the Hazards of Technology*. Westview: Boulder, CO, pp. 91–125.

Slovic P, Malmfors T, Krewski D, Mertz CK, Neil N, and Bartlett S. (1995). Intuitive toxicology II. Expert and lay judgements of chemical risks in Canada. *Risk Analysis* **15** (6): 661–675.

Smith TJ. (1972). *The Effects of Truth and Desirability Evidence on Judgements of Truth and Desirability of a Proposition*. Unpublished Master's thesis. Michigan State University: East Lansing.

Stanchi KM. (2006). The science of persuasion: an initial exploration. *Michigan State Law Review I* **52**:1–45.

Tonn BE, Goeltz RT, and Travis C. (1992). Eliciting reliable uncertainty estimates. *Expert Systems* **9** (1): 25–33.

Tversky A and Koehler DJ. (1994). Support theory: a nonextensional representation of subjective-probability. *Psychological Review* **101**:547–567.

Wagenaar WA and Keren GB. (1986). Does the expert know? The reliability of predictions and confidence ratings of experts. In: Hollnagel E et al. (eds). *Intelligent Decision Support in Process Environment*. Springer-Verlag: Berlin.

Wallace HA. (1923). What is the corn Judge's mind? *Journal of the American Society of Agronomy* **15**:300–304.

Wallsten TS and Budescu DV. (1983). Encoding subjective probabilities: a psychological and psychometric review. *Management Science* **29**:151–173.

Wright G and Ayton P. (1987). *Judgemental Forecasting*. Wiley: Chichester.

Wright G, Bolger F, and Rowe G. (2002). An empirical test of the relative validity of expert and lay judgments of risk. *Risk Analysis* **22** (6): 1107–1122.

Wright G, Pearman A, and Yardley K. (2000). Risk perception in the UK oil and gas production industry: are expert loss-prevention managers' perceptions different from those of members of the public? *Risk Analysis* **20**:681–690.

Wright G, Rowe G, Bolger F, and Gammack G. (1994). Coherence, calibration and expertise in judgemental probability forecasting. *Organisational Behavior and Human Decision Processes* **57**:1–25.

Wright G, Saunders C, and Ayton P. (1988). The consistency, coherence and calibration of holistic, decomposed and recomposed judgemental probability forecasts. *Journal of Forecasting* **7**:185–199.

Yates JF. (1990). *Judgement and Decision Making*. Prentice-Hall: Englewood Cliffs, NJ.

Youseff ZI and Peterson CR. (1973). Intuitive cascaded inferences. *Organisational Behaviour and Human Performance* **10**:349–358.

12
Evaluation of Risks in Complex Problems

Gregory S. Parnell

Major investment projects are essential for companies and governments to develop new products and services to maintain the growth of the company or the viability of the government agency. Many investment decisions are complex problems. Complex investments are successful, if they provide the planned benefits to stakeholders (including consumers) for the proposed costs. Successful investment management depends on timely identification, evaluation and management of risks. Risk identification requires interaction with critical stakeholders. Once risks have been identified, risk evaluation provides essential information on investment risk and important dependencies that enable risk-informed management. Several qualitative and quantitative risk analysis techniques are surveyed. Early in the investment life cycle there is less information, and qualitative risk techniques (e.g. the risk matrix) will be very useful. Since well-managed investment programmes quantify their benefits and costs early in the life cycle, the use of Monte Carlo simulation with these benefit and cost models can provide useful insights to investment managers and key stakeholders. A complex investment project – a large data centre design and location decision – is used to illustrate some of the risk techniques.

Introduction

Major investment projects are essential for companies and governments to develop new products and services to maintain the growth of the company or the viability of the government agency. Early in an investment's life cycle, the identification and evaluation of risk is a critical project management task that will have a direct impact on the future success of the project. In fact, the key to successful project management of high technology projects is the identification, evaluation and management of risks. Unfortunately, decisions have to be made with little information in the front-end decision-making phase of major investment projects. This chapter focuses on the process of developing and assuring the quality of the risk evaluation in the early-stage concept development for major projects. Several qualitative and quantitative risk analysis techniques are surveyed. A complex investment project, a large data centre design and location decision is used to illustrate some of the risk techniques.

Major investment projects are complex problems

Major investments are usually defined in proportion to the size of the enterprise. For example, the United States Department of Defense (DoD) budget was about $440 billion[1] in 2007. DoD defines a major defence acquisition programme as a project with the eventual expenditure for Research, Development, Test and Evaluation (RDT&E) of more than $365 million (Fiscal Year 2000 constant dollars), or procurement of more than $2.19 billion (Fiscal Year 2000 constant dollars) or specially designated programmes.[2]

Major investment projects are complex problems involving difficult performance, schedule and cost challenges. Since much has been written about problem complexity and wicked problems (Rittel and Webber, 1973), it is useful to define the dimensions of complexity. Conklin (2001) and Ritchey (2008) have described wicked problems. According to Ritchey, the most evident and important wicked problems are complex, long-term social and organizational planning problems. Examples are the following:

- How should we fight the "War on Terrorism?"
- How do we get democracies to emerge from authoritarian regimes?
- What is a good national immigration policy?
- How should scientific and technological development be governed?
- How should we deal with crime and violence in our schools?

Clemens (2008) describes the characteristics of technical problems and complex problems. His complex problem characteristics are similar to the wicked problem characteristics of Ritchey. Table 12.1 provides the definition of problem complexity used in this chapter. A complex problem is defined as medium complexity, and a wicked problem is high complexity. The table identifies ten dimensions of problem complexity. Each dimension ranges from low to medium to high complexity. The problem **boundary** can range from isolated and defined to interconnected and defined to no defined boundary. The problem **type** can range from similar to solved problems to a problem with several unique features with new constraints developing over time to a unique and unprecedented problem. The number of **stakeholders** can range from a few homogeneous stakeholders to multiple stakeholders with conflicting views and interests to hostile or alienated stakeholders with mutually exclusive interests. The problem **challenges** can

Table 12.1 Dimensions of problem complexity.

	Complexity Level		
Problem dimension	**Low (Technical problem)**	**Medium (Complex problem)**	**High (Wicked problem)**
Boundary	Isolated, defined boundary	Interconnected, defined boundary	No defined boundary
Type	Similar to solved problems	Several unique features and new constraints will occur over time	Unique or unprecedented
Stakeholders	Few homogeneous stakeholders	Multiple with different/or conflicting views and interests	Hostile or alienated stakeholders with mutually exclusive interests
Challenges	Technology application and natural environment requirements	New technology development, natural environment, and intelligent adversaries (competition or terrorist)	No known technology, hostile natural environment, and constant threats
Parameters	Stable and predictable	Parameter prediction difficult or unknown	Unstable or unpredictable
Use of experiments	Multiple low-risk experiments possible	Modelling and simulation can be used to perform experiments	Multiple experiments not possible
Alternative solutions	Limited set	Large number are possible	No bounded set
Solutions	Single optimal and testable solution	Good solutions can be identified and evaluated objectively and subjectively	No optimal or objectively testable solution
Resources	Reasonable and predictable	Large and dynamic	Not sustainable within existing constraints
End state	Optimal solution clearly identified	Good solutions can be implemented but additional needs arise from dynamic needs	No clear stopping point

range from technology application and natural environment requirements to new technology development, national environmental threats and intelligent adversaries (competition or terrorists) to no known technology, hostile natural environment and constant threats. The problem **parameters** can range from stable and predictable parameters to those that are difficult to predict or unknown to unstable and unpredictable parameters. The **use of experiments** can range from multiple low-risk experiments to using modelling and simulation to perform experiments to no ability to perform multiple experiments. This could be due to lack of theory to develop models and simulations. The **alternative solutions** range from a limited set of solutions to a large number of possible solutions to no bounded set of solutions. The potential **solutions** range from a single optimal and testable solution to good solutions that can be identified by objective and subjective means to no optimal or objectively testable solutions. The solution **resources** range from reasonable and predicable resources to large resources and dynamic changes in resource to resources not sustainable within existing constraints. The **end state** ranges from an optimal solution that can be clearly identified to good solutions that can be implemented, but additional needs will arise from changing needs over time to no clear stopping point.[3]

The columns can be viewed collectively: a problem with all low complexity levels is a technical problem, a problem with all medium levels is a complex problem and a problem with all high settings is a wicked problem. Another way to categorize a complex problem is that most of the levels have medium complexity but with one or two low and one or two high complexity levels. The complex problems are some of the most challenging problems faced in project investment decision-making. Keeney (2004) estimates that 50 out of 10,000 decisions are complex enough to warrant a complete decision analysis. Fortunately, not all our major investments have all the attributes of the complex problem.

Key risk terminology

The beginning of wisdom in any discussion of risk is a clear definition of risk. It is important to begin with a clear understanding of the terminology of risk evaluation. Included here are definitions of risk, the major components of risk analysis and the elements of risk assessment (evaluation).[4]

The Society for Risk Analysis is a multidisciplinary, interdisciplinary, scholarly international society that provides an open forum for all those who are interested in risk analysis. They define **risk** as the potential for realization of unwanted, adverse consequences to human life, health, property or the environment.[5] A more general definition of risk for investment programmes is the likelihood of adverse consequences. There are two essential features of risk embedded in these definitions. Firstly, the consequences that contribute to risk are negative. Secondly, the potential scenario (sequence of future states) and its associated consequences are uncertain. We use the term likelihood (rather than probability) to acknowledge that risk can be represented qualitatively or quantitatively.

Risk assessment (risk evaluation) is a systematic analytic process for identifying risks and describing the nature and magnitude of the risk associated with a scenario, including consideration of relevant uncertainties. When we consider natural threats (e.g. hurricanes), engineered systems hazards (e.g. system failure, nuclear reactor accidents) and intelligent adversaries (e.g. competition, hackers, and terrorism) the risk assessment objective is to provide, as far as is practical, a scientific and analytically sound basis for answering the following questions:

- What can go wrong?
- How can it happen?
- How likely is it to happen?
- What are the potential consequences if it does happen?

Risk assessment usually includes at least three central elements – threats, vulnerabilities and consequences.

A **threat** is an event or scenario (group of events) that could result in loss or harm. The term "hazard" encompasses unintentional events, such as natural phenomena (e.g. hurricanes), accidents (e.g. nuclear power plant failures) and intentional threats (e.g. competition, hackers and terrorists). Risk analysis should include all hazards.

A **vulnerability** is an attribute of a system that can result in adverse consequences caused by natural phenomena, an accident or an intelligent adversary that seeks to cause harm. An example of vulnerability is an information technology system that allows unauthorized individuals to access information at rest (e.g. information being stored in a database) or during transmission (e.g. private communications).

A **consequence** is an adverse outcome. Consequences include the tangible and intangible and the quantifiable and unquantifiable: mortality, morbidity, economic loss, psychological and societal damage and myriad other forms of loss and harm. Consequences can cascade through interdependent infrastructures and can persist, or even increase, far into the future. For investment programmes, the consequences can include project consequences (e.g. performance, schedule and cost outcomes), as well as enterprise outcomes (e.g. increase in equity or loss of equity).

Risk management is the process of constructing, evaluating, selecting, implementing and monitoring actions to alter risk to acceptable levels. The goal of risk management is scientifically sound and cost-effective, integrated actions (including providing information, i.e. risk communication) that avoid, transfer, mitigate or accept risks, while taking into account natural environmental, technological, organizational, social, cultural, ethical, political, economic, security and legal considerations. Risk management addresses the following questions:

- What can we do?
- What should we do?
- What are the results of our actions?
- What can be done to account for the response of an adaptive, intelligent adversary?

The last question is an important addition to the risk management questions and is highly relevant to investment decisions.

Risk is a major consideration in complex problems

There are many sources of risk in complex problems: marketing risk, management risk, natural environment, engineered systems and intelligent adversaries. The natural environment includes hurricanes, tornados, snow, floods, hot temperatures, cold temperatures, etc. Engineered systems are another source of risk. Bridges can collapse, software can fail, buildings can have structural failures, aeroplanes can crash and ships can sink. Clearly the environment also impacts engineered systems. For example, the Titanic sank after hitting an iceberg, and the levees in New Orleans failed during a hurricane. Management risk usually includes performance (sometimes called technical), schedule and cost risk. The risk from intelligent adversaries is a relatively new concept in risk analysis. Intelligent adversaries can respond and adapt to planned risk management actions. Examples of intelligent adversaries are competitors, hackers and terrorists. All three have significant incentives to adapt to risk management actions and change their plans and actions.

Many sources of risk in complex investment problems

Project management risk includes performance, schedule and cost. Browning (1998) introduces three additional types of product development risk. The United States Department of Defense (2006) offers a comprehensive list of sources of 16 risks for large investment programmes. Building on these frameworks, Table 12.2 provides a list of the sources of investment risk used in this chapter. The first column of Table 12.2 lists the source of risk. The second column is the major question defining the risk. The third column lists some of the major challenges for this risk source. The challenges are meant to be illustrative and not all-inclusive.

Table 12.2 Sources of investment risk.

Source of risk	**Major question**	**Challenges**
Business	Will political, economic, labour, societal, or other factors adversely affect business environment?	Economic disruptions (e.g. recession) Changes to law Disruptive technologies Adverse publicity
Market	Will there be a market if the product or service works?	Consumer demand Threats from competitors (quality and price) and adversaries (hackers and terrorists) Continuing stakeholder support
Performance (Technical)	Will the product or service meet the desired performance?	Defining future requirements in dynamic environments Understanding the technical baseline Technology maturity to meet performance Adequate modelling, simulation, test and evaluation capabilities
Schedule	Can the product or service be delivered on time?	Concurrence in development Time and resources to resolve technical and cost risks
Development and Production Cost	Can the product or service be delivered within the budget?	Technology maturity Stability of the design Hardware and software design processes Industrial capabilities Manufacturing processes Production/facilities capabilities
Management	Does the organization have the people, processes and culture to manage a major investment?	Organization culture Investment management experience and expertise Mature baselining (technical, cost, schedule) processes Cost estimating processes

Operations and Support Cost	Can the owner afford to provide the product or service?	Increasing support (e.g. resource or environmental) costs
Sustainability	Will the product or service provide sustainable future value?	Future threats from the natural environment and intelligent adversaries

Overview of risk analysis techniques

Several risk analysis techniques have been developed and applied to complex investment problems. Books presenting risk analysis techniques include Henley and Kumamoto (1992), Ayyub (2003) and Haimes (2004). Many of the quantitative techniques require the assessment of probabilities from subject matter experts. Kahneman and Tversky (1982) describe the heuristics for judgements under uncertainty. Ayyub (2001) provides a summary of the challenges and techniques of expert opinion elicitation. Table 12.3 provides a list of common risk analysis techniques. For each technique, the table provides the name(s), whether the technique is primarily qualitative or quantitative, a brief description and one or more references. Many of the quantitative techniques begin with a qualitative framework. For example, multi-objective decision analysis begins with a value hierarchy and then uses an additive value model to evaluate candidate solutions.

The next sections will examine a couple of the most commonly used qualitative and quantitative risk analysis techniques. But first the risks to be evaluated must be identified.

Stakeholder interaction is key to risk identification

Risk evaluation is critically dependent on the identification of the risks that require evaluation. Understanding the sources of risk described in Table 12.1 is an important first step. However, when the focus is on the specific risks of a unique investment project, stakeholder interaction is the key to the identification of risks. There are several important stakeholder categories (Parnell et al., 2008). Some definitions are helpful. The decision authority is the stakeholder(s) with ultimate investment decision authority. The client is the person or organization funding the risk evaluation effort. The owner is the person or organization responsible for system operation and maintenance. The user is the person or organization accountable for system operation. The consumer is people using system products or services. In the private sector, the consumer pays directly for the goods and services. In the

Table 12.3 List of risk analysis techniques.

Technique	Qualitative or quantitative	Description	Reference
Bayesian networks	Quantitative	Graphical structure and set of conditional and unconditional probability distributions, as well as an algorithm for updating uncertainty based on evidence received over time	Pearl (1988); Schum (1994)
Cognitive maps and concept maps	Qualitative	Individual mind maps (cognitive map) or a mapping of a group's thoughts (concept map) about a particular situation or problem of interest	Tolman (1948)
Event trees, probability trees and decision trees	Quantitative	Diagram of a sequence of uncertain events where the chain or path through the tree represents a particular scenario. Mathematically solved using probability and utility theory	Clemen (1996); Paté-Cornell (1984)
Exercises	Qualitative	Event using role playing, physical system interactions and decision-making in controlled environments intended to reflect possible real-world scenarios	Perla (1987)
Failure mode and effect analysis	Quantitative	Diagram that identifies sequences of events that lead to system failure, consequences of the failure modes and mitigating actions or counter-measures	Henley and Kumamoto (1996)
Fault trees, success trees and attack trees	Quantitative	Graphical hierarchical tree structure where an undesirable event (called the top event) and the possible means for this top event to occur are analyzed using probabilities and Boolean algebra	Ayyub (2003); Haimes (2004); Schneier (1999)

Influence diagrams and decision diagrams	Quantitative	Graphical representation and algorithm for analyzing the probabilistic relationships among factors relevant to a decision. Solved using graph and utility theory	Howard (2004); Shachter (1986, 1988)
Monte Carlo simulation	Quantitative	Software that uses probability distributions, mathematical functional relationships and simulation to obtain a probability distribution of system or process outputs	Ragsdale (2008); Goldsim Monte Carlo Software (2008)
Multi-objective decision analysis	Quantitative	Mathematical technique for analyzing complex problems with multiple, conflicting objectives under uncertainty	Kirkwood (1997); Parnell (2007)
Probabilistic risk assessment	Quantitative	Graphical and mathematical techniques using probability distributions to assess system behaviour and estimate the likelihood of unfavourable consequences (includes all probabilistic risk methods)	US Nuclear Regulatory Commission (1995); Parnell et al. (2005)
Project scheduling using PERT and CPM	Quantitative	Graphical scheduling methods to plan, schedule and control complex project. CPM (Critical Path Method) and PERT (Programme Evaluation Review Technique) are two common techniques. Both can be done with Monte Carlo analysis	Marold (2004); Ragsdale (2008)
Risk matrix (Likelihood-consequence matrix)	Qualitative	Matrix that categories risk using two dimensions: likelihood of occurrence and the potential consequences, given occurrence of the event	DoD (2003)

Table 12.3 (Continued)

Technique	Qualitative or quantitative	Description	Reference
Scenario analysis	Qualitative	Study that develops and analyzes uncertain future events (scenarios) using an internally consistent story about how events might unfold over time	Kirkwood (1997)
System dynamics	Quantitative	A simulation methodology for studying and managing complex dynamic feedback systems	Forrester (1961); Senge (1990)

public sector, the consumer pays indirectly (i.e. via taxes) for the goods and services. Of course, other stakeholders may also be involved.

The future risks can be influenced by several environmental factors (Parnell et al., 2008):

Cultural. Many products and services are designed for national and international customers. Cultural factors can pose a risk of the consumer not accepting the product or service.

Economic. Economic factors are almost always a major investment risk. Most programme managers have a budget to manage. Major changes in resource costs can impact investment decisions, by changing the cost of ownership.

Political. Political factors come into play for many complex investment decisions. Many stakeholder groups (e.g. lobby groups) exist to impact decisions by private or public organizations. Many public decisions require approval by government agencies (e.g. drug approval by national authorities) and/or legislatures. Approval delay or disapproval pose potential risks.

Emotional. Sometimes decision-makers or key stakeholders have emotional issues about some investment decisions. For example, nuclear power is an emotional issue for some stakeholders. A history of unsuccessful similar investments can pose risk to stakeholder support for the investment.

Historical. Historical issues can impact investment decisions. Many communities have historical preservation societies that are concerned with changes that impact historical landmarks and facilities. These organizations can delay or disapprove implementation.

Legal. Products and services must comply with national, regional and community legal requirements. Regulation changes can pose risks for investment decisions.

Moral/Ethical. Often moral or ethical issues arise in investment decisions. Unethical or questionable acquisition practices have impacted investment decisions.

Natural Environment. Systems can have significant consequences on the natural environment. Environmental impacts can lead to cost and schedule risk, potential loss of community support and political risks.

Organizational. Decisions are made within organizations. The lack of mature investment management processes and experienced managers can create investment risk.

Security. Systems must be secure. System owners, users and consumers want to be sure that their system and their products and services are secure against potential threats. Physical and information security are risks that must be evaluated.

Social. Investment decisions can also have social implications. The construction of a major facility in a community can dramatically change that community. The community may view this as an opportunity or a threat.

Technological. Technologies perform functions for users and consumers. Some technologies are developed and available. New technologies may involve technical, cost and schedule risk for the system. In addition, the consequences of technologies are not always understood, e.g. the health consequences of construction with asbestos or the environmental impact of gasoline. A major system failure can delay a system for many years, e.g. the Challenger failure.

The list of environmental factors has several uses. It also provides a useful framework to identify risk from different stakeholder perspectives.

We will illustrate the risk techniques using an illustrative investment decision to determine the size and location of a large data centre for a government organization with vast amounts of sensitive data that must be processed and analyzed. This data centre is a very large facility with state of the art computers and communication equipment and very large megawatt and cooling requirements. The investment cost of the facility is estimated to be hundreds of millions of dollars. Table 12.4 lists some potential risks for the data centre design and location illustrative example.

In addition to the matrix in Table 12.4, a matrix could be developed using the sources of investment risk for the rows and the investment stakeholders for the columns. This matrix could also be useful in identifying potential risks.

Table 12.4 Risk identification matrix.

	Risk identification matrix	Decision-makers/Stakeholders					
		Decision authority	Client: investment manager	Owner	User	Consumer	Other: local communities
Environment factors	**Cultural**					Acceptance	Acceptance
	Economic	Budget		Resource (power and water) interruptions and costs			Recession changes data needs
			Component schedule slips			Power or water outages	Local support/ skilled labour
				Communications interruptions			
	Emotional			Dissatisfied employee(s)		Acceptance	Acceptance
	Historical	Past investment experience	Past investment experience	Past investment experience			Past experience with system
	Legal	Compliance	Compliance	Supplier loses licence (e.g. nuclear power)			Compliance

Moral/Ethical		Compliance with acquisition ethics			Authorized use	
Natural environment		Environmental requirements	Environmental threats	Environmental threats	Mission data limits	Local weather Natural hazards at site location
Organizational		Acquisition funding	Support funding Strike			
Political	Political opposition	Funding				Political protests due to mission
Security			Information security Terrorist attack		Mission data limits	Local facilities
Social				Accidents		
Technological		Performance risk	Future data needs	Mechanical failures		

Once the risks have been identified, we can consider using qualitative or quantitative risk evaluation techniques.

Qualitative evaluation of risk

The most common qualitative risk analysis technique is the risk matrix. The risk matrix has two dimensions. Since there are three risk dimensions – threat, vulnerability and consequences, there are two possible ways to produce a two-dimensional risk matrix. The first approach is the likelihood-consequence matrix. The likelihood is an assessment of the relative frequency of the threat, and the consequences are the potential consequences, given the known vulnerabilities. This approach is useful when there are reasonable estimates of likelihood, perhaps based on historical frequency of occurrence (e.g. hurricanes, floods, etc.). A second approach is to develop a vulnerability-consequences matrix. In this approach, it is not the likelihood of the risk event that is assessed but, instead, the vulnerabilities, should the event occur, together with how the hazards exploit those vulnerabilities. This approach is more useful when there are no historical data, or when an intelligent adversary has incentives to take advantage of vulnerabilities to impose consequences. Two examples are information technology vulnerabilities and terrorist attacks. In both cases, the hacker and the terrorist will try to exploit the vulnerabilities that can provide the consequences they desire.

The likelihood-consequences risk matrix has two dimensions: the likelihood of the event and potential consequences, given that the event occurs. Clearly defined levels must be specified for each dimension that allows the risk analyst to place the hazard in one cell in the matrix. Each cell in the risk matrix must then be labelled as one of several risk categories, e.g. low, medium and high. The risks with low likelihood and low consequences are low risk, the high likelihood and high consequences hazards are high risks and the medium risks are in the middle of the matrix (see Figure 12.1). Risk usually decreases lower down and to the left of the matrix. The risk matrix is used after the potential risks have been identified. Each risk is uniquely placed in one cell of the matrix. The evaluation of risks is an input to risk managers, who must determine how to deal with the risks.[6]

The United States Department of Defense has used the risk matrix in Figure 12.1 to examine the risk of major investment programmes. Likelihood is categorized by five levels, from remote to near-certainty. The consequences are composed of four consequence factors: technical performance, schedule, cost and other impacts.

According to Ozug (2002), an effective risk ranking matrix should have several characteristics. Three characteristics are most fundamental. Firstly, it should be simple and easy to understand. Secondly, it should have consistent

Level	What is the likelihood the risk event will happen?
a	Remote
b	Unlikely
c	Likely
d	High likely
e	Near certainly

Process variance refers to deviation from best practices Likelihood/probability refers to risk events.

Assessment guide

Likelihood					
e	M	M	H	H	H
d	L	M	M	H	H
c	L	L	M	M	H
b	L	L	L	M	M
a	L	L	L	L	M
	a	b	c	d	e

Consequence

Risk assessment

H High – Unacceptable. Major disruption likely. Different approach required. Priority management attention required.

M Moderate – Some disruption. Different approach may be required. Additional management attention may be needed.

L Low – Minimum impact Minimum oversight needed to ensure risk remains low.

Level	Technical performance	and/or	Schedule	and/or	Cost	and/or	Impact on other teams
a	Minimal or no impact		Minimal or no impact		Minimal or no impact		None
b	Acceptable with some reduction in margin need dates		Additional resources require; able to meet		<5%		Some impact
c	Acceptable with significant reduction in margin		Minor slip in key milestones; not able to meet need date		5–7%		Moderate impact
d	Acceptable; no remaining margin		Major slip in key milestone or critical path impacted		7–10%		Major impact
e	Unacceptable		Can't achieve key team or major program milestone		>10%		Unacceptable

Figure 12.1 United States Department of Defense acquisition risk matrix

likelihood and consequence ranges that cover the full spectrum of potential scenarios. Thirdly, the risk categories must be clearly defined and meaningful to risk management decision-makers.

Mission data centre design and location is a complex problem

Data centres are critical to the missions of organizations that use large computing power. Many corporations need large data centres, e.g. Google, Yahoo and Amazon. In addition, many large government agencies require large-scale computing power, e.g. the processing of large amounts of geospatial or intelligence data. A data centre is a facility used to house computer systems (large special purpose system or server farms) and associated components, such as data storage systems, telecommunications, power systems and cooling systems. It generally includes redundant or back-up power supplies, redundant data communications connections, environmental controls (air conditioning, fire suppression, etc.) and security (information and physical) systems. New data centres can occupy large facilities, use tens of megawatts of power and require investments in the hundreds of millions. Many communities encourage organizations to site data centres in their locations. Competition from competing communities can be intense.

The following is a list of some of the important mission data centre attributes (I/O Data Centres):

- **Floor Space** – The computers and associated equipment require significant floor space. The data centre facility is a structure featuring high ceilings, substantial load capacity and a large roof space and acreage designed to effectively accommodate mechanical and other equipment. Additional floor space should be available for IT support staff and or disaster recovery.
- **Power** – Data centres require high megawatts. Affordable and redundant power is essential to provide mission assurance.
- **Water** – The large electrical and mechanical data centre systems require cooling. Access to affordable and redundant water supplies is essential.
- **Communications** – The data centre should be served by major data and telecommunications carriers, enabling access and, preferably, multiple carrier options.
- **Location and Geography** – Location should consider distance from mission activities, employee safety, site security and ease of accessibility. Regionally, natural disasters are a factor. In addition, large cities might be a target of terrorist activity.

- **Security, Fire and Life-Safety** – Security, fire and life-safety systems are designed, built and operated to support the important objective of data preservation. All such systems may be backed up. In addition, advanced technologies such as digital video, electronic access control, biometric security and pre-action fire suppression can be used.

The mission data centre design and location can be used to illustrate risk evaluation of a complex problem. According to the ten characteristics of a complex problem, the data centre boundary is defined but clearly connected to mission and support providers. Data centres have unique attributes, and new constraints will occur over time from mission users and support providers. Multiple stakeholders have conflicting views and interests, and only some stakeholders (e.g. contractors and local communities) can be selected. New data centre technology continues to develop, and threats from the natural environment and intelligent adversaries (competition or terrorist) are certainly considerations. Parameter prediction (e.g. future megawatt requirements due to mission demand and computer system power usage) is difficult to predict. Modelling and simulation can be used to perform layouts and assess designs. There is a large number of potential locations and very large number of potential designs. Good data centre solutions can be identified and evaluated objectively (e.g. megawatts) and subjectively (e.g. location security). Data centres require large investment resources. Finally, good data centre solutions can be implemented, but additional needs will arise from dynamic mission needs and technology developments. Clearly, large data centre design and location is a complex problem.

Illustrative risk matrix for data centre problem

For illustrative purposes, Table 12.5 displays the use of a risk matrix using the format in Figure 12.1. The risks identified in the upper right hand of the matrix are coloured dark gray for high risk. The risks in the lower left hand corner of the matrix are light gray for low risk. The medium risks in the middle are coloured gray.

For each of the risks, especially the high and medium risks, risk management actions need to be considered. For example, the risk of a power blackout can be mitigated by alternative power sources, back-up generators and batteries. Likewise, the risk of dramatic cost growth during development can be mitigated by assigning experienced project managers and focusing leadership attention on key decision milestones.

Quantitative evaluation of risk

Well-managed investment programmes have clearly defined benefits and costs. The benefits are provided to the stakeholders (e.g. clients, owners,

Table 12.5 Risk matrix for data centre problem.

	Minimal	Acceptable with some impact	Acceptable with significant reduction in margin	Acceptable with no margin	Unacceptable
Near certainty	Dissatisfied employee	Some component schedule slips	Some technical issues	Increase in future data processing needs	Interruption in communica-tions
Highly likely	Mechanical failure in redundant systems	Periodic hot weather	Short power outage	Major technical problem	Dramatic cost growth during development
Likely	Minor accidents	Periodic water outage	Significant change in requirements	Significant increase is resource cost	Power blackout
Unlikely	Major accidents	Earthquake or hurricane slows construction	Worker strike	Nuclear power plant loses license	Recession dramatically changes data centre need
Remote			Resource supplier goes bankrupt	Political protests stop construction	Terrorist attack

H
M
L

users, consumers). Likewise the costs are born by the stakeholders. Systems engineers and engineering managers develop performance, schedule and cost requirements to achieve the benefits for the budgeted cost. For investment programmes, risk is defined as the likelihood of adverse consequences. Therefore, the adverse consequences of investment programmes usually include a combination of lack of projected benefits, delayed benefits or increased costs.

Complex investment problems usually require some form of benefit cost analysis to justify the investment and to manage the acquisition. For private companies, investments can be evaluated using net present value to assess the return on the investment (Clemen, 1996). For public programmes, multi-objective decision analysis is a common technique to assess the benefits when there are multiple, conflicting objectives of diverse stakeholders (Kirkwood, 1997; Parnell et al., 2008). Both approaches allow the explicit consideration of uncertainty using probabilities.

Figure 12.2 provides an example of an illustrative multi-objective value hierarchy that could be the foundation for a multi-objective decision analysis. The fundamental objective is to provide the best large data centre design for future mission support. Use of "future" is very important, since the current communications, floor space, processing power, cooling and security can change with changes in mission computing requirements, new technologies and the economy. The next level in the value hierarchy is functions. The functions, defined by verb-object combinations, describe what the system has to do. The third level, the objectives, describes the goals for each function. The fourth level (not shown in Figure 12.2) would be the value measures that define how well each objective is attained. For example, a value measure for floor space would be square footage and a value measure for mission power capacity would be megawatts.

Multi-objective decision analysis (MODA) uses many mathematical equations to evaluate alternatives (Kirkwood, 1997). The simplest and most commonly used model is the additive value model. This model uses the following equation to calculate each alternative's value:

$$v(x) = \sum_{i=1}^{n} w_i v_i(x_i)$$

where

$v(x)$ is the solution's value
$i = 1$ to n is the number of the value measure
x_i is the solution's score on the ith value measure
$v_i(x_i)=$ is the single dimensional value of a score of x_i
w_i is the weight of the ith value measure
and $\sum_{i=1}^{n} w_i = 1$ (all weights sum to one)

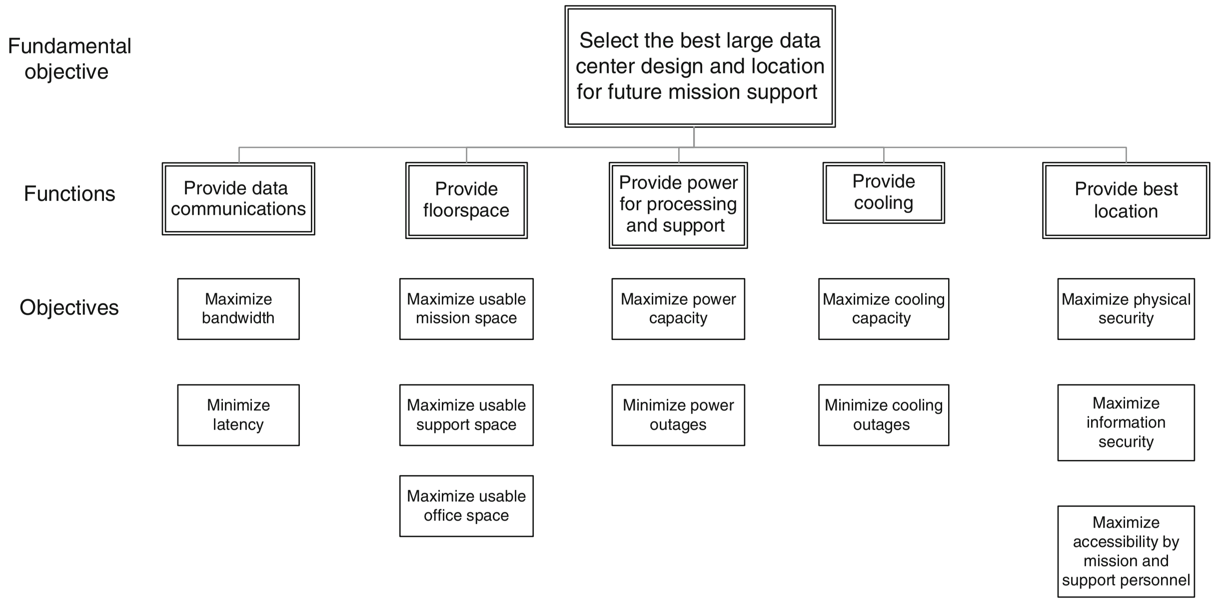


Figure 12.2 Illustrative large data centre value hierarchy

The value functions, $v_i(x_i)$, measure the returns to scale on the range of the value measures. The weights, w_i, measure the importance of the range of the value measure scale compared to the range of the other value measure scales.

The cost of an investment is best measured by its life cycle cost (Parnell et al., 2008). The life cycle cost includes the total costs of acquisition, operation and retirement. The best practice is to develop a life cycle cost model that supports budget development and investment decision evaluation.

In the illustrative data centre design and location, a life cycle cost model for the total cost of ownership of the physical infrastructure for network rooms and data centres is required for return-on-investment analysis and other business decision processes (APC, 2005). In addition, an understanding of life cycle cost drivers provides insight into opportunities to control costs. Many users are surprised when they consider that the life cycle cost for physical infrastructure may be comparable to, or larger than, the life cycle cost of the supported IT equipment (APC, 2005).

When the multi-objective decision analysis and the life cycle cost analysis have been completed, decision-makers and stakeholders can be provided with the information presented in Figure 12.3. This chart provides several important insights. Firstly, solutions Alpha and Delta are dominated by Charlie and Foxtrot, respectively. Secondly, only the decision-maker can determine which of the non-dominated solutions (Bravo, Foxtrot and Charlie) are the best. Thirdly, if the decision-maker expects continued growth in

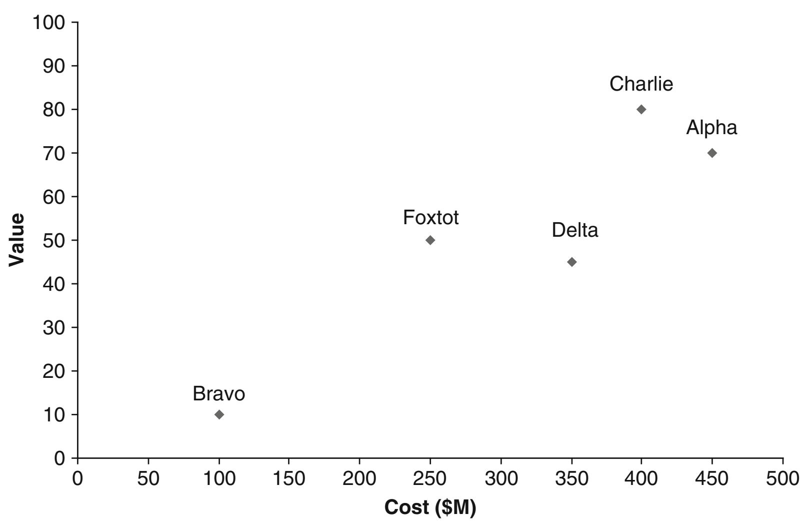


Figure 12.3 Value and cost comparison of solutions

data centre mission requirements, and there is a budget of over $400M, then solution Charlie may provide the best value.

The benefit and the cost risk can be assessed using decision trees, influence diagrams or Monte Carlo simulation. The use of decision trees with net present value models is a common decision analysis technique (Clemen, 1996). After the decision tree analysis is complete, risk can be explicitly shown by examining the cumulative distribution of net present value or multi-objective value. In fact, Clemen (1996) calls this the risk profile. The analyst can directly read of the risk as the probability that a threshold value will not be achieved. Perhaps more usefully, the risk profile helps to identify the lowest value that could occur, if all uncertainty nodes in the decision tree (or influence diagram) attain the worst possible outcomes. In addition, decision analysis provides very useful sensitivity analysis tools, including value of information, value of control, Tornado diagrams (to show the impact of one variable sensitivity), two-way sensitivity analysis and three-way sensitivity analysis.

Decision trees can also be used with multi-objective decision analysis models (Parnell, 2007). The multi-objective risk profile shows the cumulative distribution of multi-objective value. The same sensitivity analysis techniques apply.

One of the most powerful techniques for quantitative risk evaluation is Monte Carlo simulation. Monte Carlo simulation can be used with any model that has multiple variables and multiple outcomes. Therefore, it can

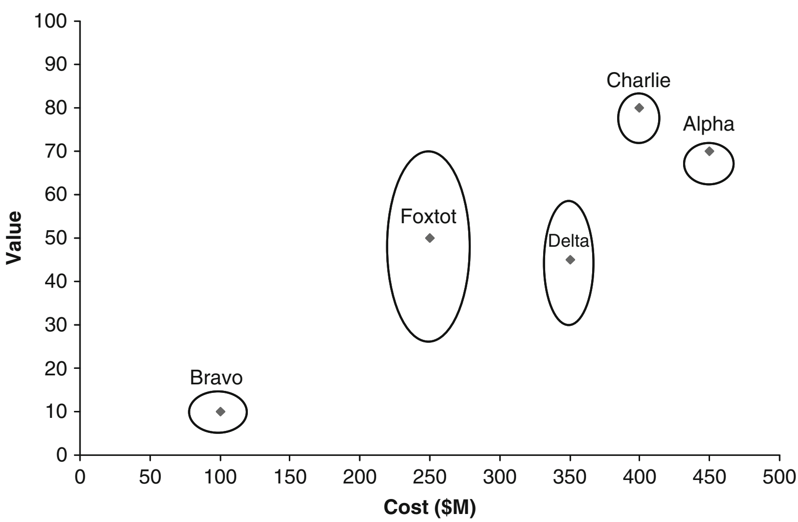


Figure 12.4 Value and cost comparison of solutions with Monte Carlo simulation

be used with any net present value or multi-objective decision analysis model (Parnell et al., 2008). Using commonly available Excel addins (Crystal Ball, @Risk, etc.), a simulation can be performed on any spreadsheet model of benefits and/or costs.

For the illustrative data centre design and location, Monte Carlo simulation can be performed on the value model and the cost model. Figure 12.4 displays a notational cost versus value plot when the risk of achieving the value and cost estimates is considered. The risk plot would not change the fact that Alpha and Delta are dominated. However, the analysis shows that Charlie has much less value and cost variance compared to Foxtrot.

Comparison of qualitative and quantitative risk evaluation techniques

One of the key questions for decision-makers is what risk evaluation techniques to use. Different techniques have different features and different resource requirements. Table 12.6 summarizes some of the important comparisons. First and foremost, risk evaluation is only as good as the risk identification inputs, which totally depend on access to key experts and stakeholders. Without good risk identification, the techniques do not matter. Secondly, a good risk matrix can provide very useful insights for risk managers. However, quantitative models are required to obtain additional insights about the most important variable and about benefit and cost dependences. Also, quantitative models can help identify resource requirements. For example, a total cost of ownership model can identify acquisition

Table 12.6 Comparison of risk evaluation techniques.

	Qualitative techniques	**Quantitative techniques**
Access to experts and stakeholders	Essential	Essential
Risk identification	Critical	Critical
Risk evaluation	Good first order insights	Provide estimates of benefit and cost risk. Additional insights about key variables and dependencies
Resources	Less resources but less information	More resources to develop models[7] and obtain quantitative data from experts and stakeholders

and operations costs which can be used to obtain funding from the organization. Thirdly, quantitative models require more resources to develop models and obtain the necessary data. Techniques like sensitivity analysis and Monte Carlo simulation can be used to determine the most important data for the data collection efforts. Fourthly, qualitative and quantitative techniques are not mutually exclusive. A qualitative technique can be used to identify the major risks and a quantitative technique can be used to obtain additional insights.

Risk evaluation of investment programmes is essential

The success of large complex investment programmes depends on providing the benefits for the proposed costs. Successful investment management requires identification, evaluation and management of risks. Access to experts and subject matter experts is required for risk identification. Risk evaluation provides essential information on investment risk factors and dependencies that enable risk-informed management.

Qualitative and quantitative techniques are not mutually exclusive. A qualitative technique can be used to identify and evaluate the risks and a quantitative technique can be used to obtain additional insights. Early in the investment programme life cycle stages, there may be less information, and so it might be expected that qualitative risk techniques, like the risk matrix, will be very useful. Since well-managed investment programmes will quantify their benefits and costs early in the programme life cycle, the use of Monte Carlo simulation with these benefit and cost models can provide very useful insights for investment managers and key stakeholders.

Notes

1. http://www.whitehouse.gov/omb/budget/fy2007/defense.html Accessed 27 January 2008.
2. Department of Defense Instruction 5000.2, Operation of the Defense Acquisition System, USD(AT&L), 12 May 2003.
3. Modified and expanded from The Art of Complex Problem Solving, Marshall Clemens, http://www.idiagram.com/CP/kindsofproblems.html Accessed 27 January 2008.
4. In this paper risk evaluation and risk assessment are equivalent.
5. Society for Risk Analysis definition of risk, http://www.sra.org/resources_glossary_p-r.php, Accessed 27 January 2008.
6. Department of Defense (DoD) Risk Management Guide For DoD Acquisition, August 2003, Defense Acquisition University, p. B-16.
7. One of the major benefits of Monte Carlo simulation is the use of existing models to evaluate risk.

References

American Power Conversion (APC). (2005). Determining Total Cost of Ownership for Data Center and Network Room Infrastructure. www.apc.com Rev 2005-3.

Ayyub B. (2001). *Elicitation of Expert Opinions for Uncertainty and Risks: Theory, Applications and Guidance*. CRC Press: Boca Raton, Florida.

Ayyub B. (2003). *Risk Analysis in Engineering and Economics*. Chapman and Hall, CRC: Boca Raton, Florida.

Browning T. (1998). Modeling and analyzing cost, schedule, and performance in complex system product development. *Technology, Management and Policy*. Boston: Massachusetts Institute of Technology.

Clemen R. (1996). *Making Hard Decisions: An Introduction to Decision Analysis* (2nd edition). Duxbury Press: New York.

Clemens M. (2000). The Art of Complex Problem Solving. http://www.idiagram.com/CP/kindsofproblems.html Accessed 27 January 2008.

Conklin J. (2001). Wicked problems and social complexity. *CogNexus Institute*. http://cognexus.org/wpf/wickedproblems.pdf Accessed 27 January 2008.

Crystal Ball. (2008). http://www.decisioneering.com Accessed 30 January 2008.

Department of Defense (DoD). (2003). *Risk Management Guide For DoD Acquisition, August 2003*, Defense Acquisition University.

Department of Defense (DoD). (2006). *Risk Management Guide For DoD Acquisition, August 2006* (6th edition) (Version 1.0). Defense Acquisition University.

Forrester J. (1961). *Industrial Dynamics*. Pegasus Communications: Waltham, MA.

Goldsim Monte Carlo Software. (2008). http://www.goldsim.com Accessed 30 January 2008.

Haimes Y. (2004). *Risk Modeling, Assessment, and Management*. John Wiley and Sons Inc.: Hoboken, NJ.

Henley E.J. and Kumamoto H. (1992). *Probabilistic Risk Assessment: Reliability Engineering, Design, and Analysis*. IEEE Press: New York.

Henley E. and Kumamoto H. (1996). *Probabilistic Risk Assessment: Reliability Engineering, Design, and Analysis* (2nd edition). IEEE Press: New York.

Howard R. (2004). Precise decision language. *Decision Analysis* **1** (2): 71–78.

I/O Data Centers. (2008). Mission Data Center Attributes. http://www.iodatacenters.com/data-center.htm Accessed 1 February 2008.

Kahneman D and Tversky A. (eds). (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press: NY.

Keeney R. (2004). Making better decision makers. *Decision Analysis* **1** (4): 193–204.

Kirkwood C. (1997). *Strategic Decision Making: Multi-Objective Decision Analysis with Spreadsheets*. Duxbury Press: Belmont, CA.

Marold K. (2004). A simulation approach to the pert/cpm time-cost trade-off problem. *Project Management Journal*. *Project Management institute* **35** (1): 31–37.

Marshall C. (2008). The Art of Complex Problem Solving. http://www.idiagram.com/CP/kindsofproblems.html Accessed 27 January 2008.

Ozug H. (2002). Designing an Effective Risk Matrix. An ioMosaic Corporation Whitepaper, archives1.iomosaic.com/whitepapers/risk-ranking.pdf Accessed 29 January 2008.

Parnell G. (2007). Chapter 19. Value-Focused Thinking Using Multiple Objective Decision Analysis. In: Rainey L and Loerch A (eds). *Methods for Conducting Military Operational Analysis: Best Practices in Use Throughout the Department of Defense*. Military Operations Research Society, pp. 619–656.

Parnell G., Dillon R. and Bresnick T. (2005). Integrating risk management with homeland security and antiterrorism resource allocation decision-making. In: Kamien D (ed.) *The McGraw-Hill Homeland Security Handbook*. McGraw-Hill: NY.

Parnell G., Driscoll P. and Henderson D. (eds). (2008). Decision making for systems engineering and management. In: Sage A (ed.) *Wiley Series in Systems Engineering*. Wiley & Sons Inc.: Hoboken, NJ.

Paté-Cornell M. (1984). Fault trees vs. event trees in reliability analysis. *Risk Analysis* **4** (3): 177–186.

Pearl J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Networks of Plausible Inference, Morgan Kaufmann: San Mateo, CA.

Perla Peter P. (1987). "War Games, analyses, and exercises". *Naval War College Review*. Newport, Rhode Island, 40 (Spring): 44–52.

Ragsdale C. (2008). *Spreadsheet Modeling & Decision Analysis: A Practical Introduction to Management Science* (Revised 5th edition). Cengage Learning: Florence, KY.

@Risk. (2008). Risk analysis and Monte Carlo simulation. http://www.palisade.com Accessed 30 January 2008.

Ritchey T. (2008). *Wicked Problems: Structuring Social Messes with Morphological Analysis*, Swedish Morphological Society, www.swemorph.com Accessed 19 May 2008.

Rittel H. and Webber M. (1973). Dilemmas in a general theory of planning. *Policy Science*. Vol. 4. Elsevier Scientific Publishing Company Inc: Amsterdam, pp. 155–169.

Schneier B. (1999). Attack trees: modeling security threats. *Dr. Dobbs Journal of Software Tools* **24** (12): 21–29.

Schum D. (1994). *The Evidential Foundations of Probabilistic Reasoning*. Wiley: New York.

Senge P. (1990). *The Fifth Discipline: The Art & Practice of the Learning Organization*. Doubleday Currency: New York.

Shachter R. (1986). Evaluating influence diagrams. *Operations Research* **34** (6): 871–882.

Shachter R. (1988). Probabilistic inference and influence diagrams. *Operations Research* **36** (4): 589–604.

Tolman E. (1948). Cognitive maps in rats and man. *Psychological Review* **55**: 189–208.

U.S. Nuclear Regulatory Commission (USNRC). (1995). Use of probabilistic risk assessment methods in nuclear activities: final policy statement. *Federal Register* **60**:42622.

13
Obtaining Distributions from Groups for Decisions Under Uncertainty

Roger M. Cooke

This paper considers the problem of obtaining group distributions from the standpoint of fundamental decision theory. Decision theory requires a probability distribution over possible states of the world and a value or utility function over possible outcomes of alternative actions. In a group context, this means that there should be a distribution over value functions which characterize a population of stakeholders and a method for combining the subjective uncertainty distributions from qualified experts. This paper focuses on both problems. Regarding combining experts' uncertainty distributions, a rich literature and body of experience is available. Techniques for capturing distributions over value functions are under development, and promising techniques are on the horizon.

Introduction

Decision theory provides a model for rational decision-making under uncertainty. The model involves

- defining the decision space of possible actions
- quantifying uncertainty regarding the true, but unknown, state of the world
- quantifying the values of possible outcomes of actions.

With this input, the expected value of each possible action can be assessed and the action with the highest expectation chosen.

In the realm of personal decision-making, the application of this model for rational decision under uncertainty is relatively clear. In dealing with highly structured decisions involving different groups and different stakeholders, two profound problems are confronted in applying the rational decision model:

1) a "group uncertainty distribution" over possible states of the world must be defined
2) a "group valuation" over outcomes must be defined.

Defining a group uncertainty distribution is the province of structured expert judgement. In many structured decision contexts, people are prepared to nominate a set of experts, whose judgements of uncertainty will form the basis for uncertainty quantification. How exactly this should be done is subject to discussion, but that it should be done is indisputable (Budnitz et al., 1998; Cooke, 1991; French, 1985; Genest and Zidek, 1986; O'Hagan et al., 2006; Winkler et al., 1995). The reason for this is that assessments of uncertainty tend to converge as more observations are performed, and experts are better able to anticipate future observations from a perspective of knowledge.

With regard to valuation of outcomes, the situation is fundamentally different. Values of diverse stakeholders may be conflicting. Moreover, there is nothing corresponding to "updating valuations based on observations". There is no such thing as a community of experts who can advise the various stakeholders on where their values should lie or what is really best for them. The modelling of valuations for structured group decision problems should aim to find a distribution over value functions which best reflects the values of the stakeholders.

Whereas methods for structured expert judgement have been developed over the last 20 years and applied extensively, the situation with regard to quantifying distributions over stakeholder valuations is relatively new. Techniques of "random utility models" are described in economics literature for "discrete choice" problems, but existing methods make very restrictive assumptions and do not directly aim at estimating a distribution over utility or value functions.

This paper reviews work on structured expert judgement and indicates a new approach to quantifying stakeholder values, based on a technique called probabilistic inversion. This technique has only been applied a few times but would seem to hold some promise.

Stakeholder preference

A brief account of stakeholder preference modelling with probabilistic inversion is given here. A more detailed discussion can be found in Neslo et al. (2008); Train (2003); Anderson et al. (1996) and Cooke and Misiewicz (2007).

Suppose the aim is to model stakeholder preferences for a set of *N* alternatives, and these utilities can be scaled so that they have the same "0" and

"1", whereas all other utilities fall between these bounds. The distribution of stakeholder utilities can be expressed as a distribution over $[0, 1]^N$; or the alternatives may be scored on a set of K criteria, and a distribution over the criteria weights sought, so that distribution of the weighted sum of criteria scores reflects the distribution of stakeholder preferences.

Various "discrete choice" methods or "random utility" models have been developed for this purpose. Probabilistic inversion has the advantage that no assumption is made regarding the dependences or interactions between the utilities of the various alternatives; rather the interaction structure is inferred from the stakeholder preference data with probabilistic inversion (Csiszar, 1975; Deming and Stephan, 1944; Du et al., 2005; Kurowicka and Cooke, 2006).

This is illustrated with a recent study on the valuation of threats to marine coastal ecosystems (Neslo et al., 2008). Sixty-four stakeholders were presented with 30 threat scenarios and asked to rank the top five. Each scenario was described by values on five criteria. This produced a set of probabilities that each of the 30 scenarios could be ranked in position 1, . . . ,5. The exercise was designed so that some rankings were inconsistent with the multi-criteria model, in the sense that some scenarios were dominated by others on each criterion. The relatively low probabilities for inconsistent rankings gave a rough validation of the multi-criteria model.

To find a distribution over criteria weights reflecting the discrete choice preference data, a non-informative distribution over the criteria weights is made, and this distribution is adapted so that:

- the probability of drawing a vector of weights which ranks scenario n in position j is as close as possible to the percentage of stakeholders ranking scenario n in position j; $1 \leq n \leq 30$; $1 \leq j \leq 4$.
- the resulting distribution is minimally informative with respect to the initial non-informative distribution.

The model was fit on the first four rankings and used to predict the fifth rankings. Figure 13.1 compares the predicted and observed percentages of rankings. The model is first used to "retrodict" or "recover" the first four rankings. These are the data actually used to fit the model, so this comparison is a check of model fit, rather than model prediction. Using the model, the percentages of experts ranking the various scenarios in the fifth position can be predicted. These percentages were not used in fitting the model and testing the ability of the model to predict preferences of the population of stakeholders. Of course, it is hoped that the predictions and retrodictions show agreement with the observed rankings.

The retrodictions are shown as diamonds and the predictions are shown as squares. The percentages along the horizontal axis correspond to rankings that were inconsistent with the multi-criteria model. Figure 13.1 shows

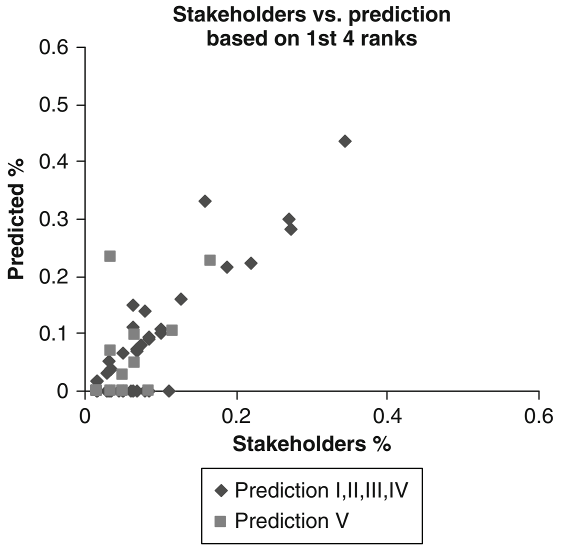


Figure 13.1 Predicted and observed percentages of stakeholder rankings in marine coastal ecosystem study

that the weight distribution predicts stakeholder preferences reasonably well. Except for the inconsistent rankings, the predicted percentages of stakeholders assigning a given rank to a given scenario agree reasonably with the observed percentages. Moreover the pattern of agreement for retrodictions and predictions is similar. This justifies the use of this model to predict other unobserved rankings of the population. Thus, if new scenarios need to be evaluated, there is no need to reconvene the 64 stakeholders and repeat the whole exercise; instead, the model can be used to assess the new scenarios, together with the original 30 scenarios.

Structured expert judgement

Expert judgement is sought when substantial scientific uncertainty impacts on a decision process. Because there is uncertainty, the experts themselves are not certain and hence will typically not agree. Informally soliciting expert advice is not new. "Structured" expert judgement refers to an attempt to subject this process to transparent methodological rules, with the goal of treating expert judgements as scientific data in a formal decision process. Standard sources include Cooke (1991), Cooke and Goossens (2000) and O'Hagan et al. (2006). A recent special issue of *Reliability Engineering and*

System Safety (Cooke, 2008) covers standard as well as emerging techniques. This section is based in part on Cooke and Goossens (2008).

The process by which experts come to agree is itself a scientific method. Structured expert judgement cannot pre-empt this role and therefore cannot have expert agreement as its goal. Broadly speaking, there are three different goals to which a structured judgement method may aspire:

- Census
- Political consensus
- Rational consensus.

A study aiming at **census** will simply try to survey the distribution of views across an expert community. An illustration of this goal is found in the Nuclear Regulatory Commission's *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts* (1997):

> To represent the overall community, if we wish to treat the outlier's opinion as equally credible to the other panelist's, we might properly assign a weight (in a panel of 5 experts) of 1/100 to his or her position, not 1/5.
>
> (NUREG/CR-6372: p. 36)

The goal of "representing the overall community" may, in this view, lead to a differential weighting of experts' views according to how representative they are of other experts. A similar goal is articulated in Winkler et al. (1995). The philosophical underpinnings of this approach are elaborated in Budnitz et al. (1998). Expert agreement on the representation of the overall community is the weakest, and most accessible, type of consensus to which a study may aspire. Other types of consensus are as follows:

- agreement on a distribution to represent a group
- agreement on a distribution
- agreement on a number.

Political consensus refers to a process in which experts are assigned weights according to the interests or stakeholders they represent. In practice, an equal number of experts from different stakeholder groups would form an expert panel and given equal weight in this panel. In this way, the different groups are included equally in the resulting representation of uncertainty.

Rational consensus refers to a group decision process. The group agrees on a method according to which a representation of uncertainty will be generated for the purposes of which the panel was convened, without knowledge of the result of this method. It is not required that each individual member adopt this result as his/her personal belief. This is a form of agreement on a distribution to represent a group. To be rational, this method must comply with necessary conditions devolving from the general scientific method. Cooke (1991) formulates the necessary conditions or principles which any method warranting the term "scientific" should satisfy:

- **Scrutability/accountability:** All data, including experts' names and assessments, and all processing tools are open to peer review, and results must be reproducible by competent reviewers
- **Empirical control:** Quantitative expert assessments are subjected to empirical quality controls
- **Neutrality:** The method for combining/evaluating expert opinion should encourage experts to state their true opinions so that they do not bias results
- **Fairness:** Experts are not pre-judged prior to processing the results of their assessments.

Thus, a method is proposed which satisfies these conditions and to which the parties pre-commit. The method is applied, and after the result is obtained, parties wishing to withdraw from the consensus incur a burden of proof. They must demonstrate that some heretofore unmentioned necessary condition for rational consensus has been violated. If unable to do this, their dissent is not "rational". Of course any party may withdraw from the consensus because the result is hostile to his or her interests. Since such withdrawal is based on interest rather than arguments, this is not rational dissent and does not endanger rational consensus.

The requirement of empirical control will strike some as peculiar in this context. How can there be empirical control with regard to expert subjective probabilities? To answer this, the question must be asked, "When is a problem an expert judgement problem?" There is no need for expert judgement to determine the speed of light in a vacuum. This is physically measurable and has been measured to everyone's satisfaction. Any experts queried would give the same answer. Neither is expert judgement consulted to determine the proclivities of a god. There are no experts in the operative sense of the word for this issue. A problem is susceptible for expert judgement only if there is relevant scientific expertise. This suggests that there are theories and measurements relevant to the issues at hand but that the quantities of interest themselves cannot be measured in practice. For example, toxicity of a substance for humans is measurable in principle but is not measured for obvious reasons. However, there are toxicity measurements for other species which might be relevant to the question of toxicity in humans. Other examples are given in Section 6.

If a problem is an expert judgement problem, then there will naturally be relevant experiments or measurements. Questions regarding such experiments can be used to implement empirical control. Studies indicate that performance on so-called almanac questions does not predict performance on variables in an expert's field of expertise (Cooke et al., 1988). The key question regarding seed variables is "Is performance on seed variables judged relevant for performance on the variables of interest?" For example, should an expert who gave very over-confident off the cuff assessments on the variables, for which we knew the true values, be equally influential on the variables of interest as an expert who gave highly informative and statistically accurate assessments? That is the choice that often confronts a problem-owner after the results of an expert judgement study are received. If seed variables in this sense cannot be found, then rational consensus is not a feasible goal, and the analyst should fall back on one of the other goals.

The above definition of "rational consensus" for group decision processes is evidently on a very high level of generality. Much work has gone into translating this into a workable procedure, which gives good results in practice. This workable procedure is embodied in the "classical model" of Cooke (1991) and is described in the following section.

Before going into details, it is appropriate to say something about Bayesian approaches. Since expert uncertainty concerns experts' subjective probabilities, many people believe that expert judgement should be approached from the Bayesian paradigm. This paradigm, recall, is based on the representation of the preference of a rational individual in terms of maximal expected utility. If a Bayesian is given experts' assessments on variables of interest and on relevant seed variables, then he/she may update his/her prior on the variables of interest, by conditionalizing on the given information. This requires that the Bayesian formulates his/her joint distribution over the following:

- the variables of interest
- the seed variables
- the experts' distributions over the seed variables and the variables of interest.

Issues that arise in building such a model are discussed in Cooke (1991). Suffice it to say here that a group of rational individuals is not itself a rational individual, and group decision problems are notoriously resistant to the Bayesian paradigm.

The classical model

The above principles have been operationalized in the "classical model", a performance-based linear pooling or weighted averaging model (Cooke, 1991; Goossens et al., 1998). This model has been applied in 45 contracted

Table 13.1 Summary of applications per sector.

Sector	# of experts	# of variables	# of elicitations
Nuclear applications	98	2203	20,461
Chemical & gas industries	56	403	4491
Groundwater/Water Pollution/Dike ring/Barriers	49	212	3714
Aerospace sector: Space debris/Aviation	51	161	1149
Occupational sector: Ladders/Buildings (thermal physics)	13	70	800
Health: Bovine/Chicken (Campylobacter)/SARS	46	240	2979
Banking: Options/Rent/Operational risk	24	119	4328
Volcanoes/Dams	231	673	29,079
Rest group	19	56	762
In total	***521***	***3688***	***67,001***

studies, involving upwards of 67,000 individual elicitations. An overview of the applications is presented in Table 13.1.

Weights for a performance-based combination of expert distributions are derived from experts' calibration and information scores, as measured on seed variables. Seed variables serve a threefold purpose:

(i) to quantify experts' performance as subjective probability assessors
(ii) to enable performance-optimized combinations of expert distributions
(iii) to evaluate and hopefully validate the combination of expert judgements.

The name "classical model" derives from an analogy between calibration measurement and classical statistical hypothesis testing. It contrasts with various Bayesian models in that it does not assume prior information.

The performance based weights use two quantitative measures of performance, **calibration** and **information**. Loosely, calibration measures the statistical likelihood that a set of experimental results correspond, in a statistical sense, with the expert's assessments. Information measures the degree to which a distribution is concentrated.

These measures can be implemented for both discrete and quantile elicitation formats. In the discrete format, experts are presented with uncertain events and perform their elicitation by assigning each event to one of several pre-defined probability bins, typically 10%, 20%,...,90%. In the quantile format, experts are presented an uncertain quantity, taking values in a continuous range, and they give pre-defined quantiles, or percentiles, of the subjective uncertainty distribution, typically 5%, 50% and 95%. The quantile format has distinct advantages over the discrete format, and all the studies reported below use this format. In five studies, the 25% and 75% quantiles were also elicited. To simplify the exposition, it is assumed that the 5%, 50% and 95% values were elicited.

Calibration. For each quantity, each expert divides the range into 4 inter-quantile intervals for which his/her probabilities are known, namely, $p_1 = 0.05$: less than or equal to the 5% value, $p_2 = 0.45$: greater than the 5% value and less than or equal to the 50% value, etc.

If N quantities are assessed, each expert may be regarded as a statistical hypothesis, namely that each realization falls in one of the four inter-quantile intervals with probability vector

$$p = (0.05,\ 0.45,\ 0.45,\ 0.05).$$

With the realizations $x_1, \ldots x_N$ of these quantities, the sample distribution of the expert's inter-quantile intervals may be formed as:

$$\begin{aligned}
s_1(e) &= \frac{\#\{i \,|x_i \le 5\%\,\text{quantile}\}}{N} \\
s_2(e) &= \frac{\#\{i \,|5\%\,\text{quantile} < x_i \le 50\%\,\text{quantile}\}}{N} \\
s_3(e) &= \frac{\#\{i \,|50\%\,\text{quantile} < x_i \le 95\%\,\text{quantile}\}}{N} \\
s_4(e) &= \frac{\#\{i \,|95\%\,\text{quantile} < x_i\}}{N} \\
s(e) &= (s_1, \ldots, s_4).
\end{aligned}$$

Note that the sample distribution depends on the expert e. If the realizations are indeed drawn independently from a distribution with quantiles as stated by the expert, then the quantity

$$2NI(s(e)|p) = 2N \sum\nolimits_{i=1,\ldots,4} s_i \ln\left(\frac{s_i}{p_i}\right) \tag{1}$$

is asymptotically distributed as a chi-square variable with 3 degrees of freedom. This is the likelihood ratio statistic, and $I(s|p)$ is the relative information of distribution s with respect to p. If we extract the leading term of the logarithm we obtain the familiar chi-square test statistic for goodness of fit. There are advantages in using the form in (1) (Cooke, 1991).

If, after a few realizations, the expert were to see that all realization fell outside his 90% central confidence intervals, he might conclude that these intervals were too narrow and might broaden them on subsequent assessments. This means that, for this expert, the uncertainty distributions are not independent, and he learns from the realizations. Expert learning is not a goal of an expert judgement study, and his joint distribution is not elicited. Rather, the decision-maker wants experts who do not need to learn from the elicitation. Hence the decision-maker scores expert e as the statistical likelihood of the hypothesis

H_e: *"the inter quantile interval containing the true value for each variable is drawn independently from probability vector p."*

A simple test for this hypothesis uses the test statistic (1), and the likelihood, or p-value, or calibration score of this hypothesis, is:

$$\text{Calibration score}(e) = p - \text{value} = \text{Prob}\left\{2NI(s(e)|p) \geq r|H_e\right\}$$

where r is the value of (1) based on the observed values $x_1, \ldots, x_N$. It is the probability under hypothesis H_e that a deviation at least as great as r should be observed on N realizations if H_e were true. Calibration scores are absolute and can be compared across studies. However, before doing so, it is appropriate to equalize the power of the different hypothesis tests by equalizing the effective number of realizations. To compare scores on two data sets with N and N' *realizations*, the minimum of N and N' in (1) is used, without changing the sample distribution s. In some cases involving multiple realizations of one and the same assessment, the effective number of seed variables is based on the number of assessments and not the number of realizations.

Although the calibration score uses the language of simple hypothesis testing, it must be emphasized that this does not reject expert-hypotheses; rather, this language is used to measure the degree to which the data supports the hypothesis that the expert's probabilities are accurate. Low scores, near zero, mean that it is unlikely that the expert's probabilities are correct.

Information. The second scoring variable is information. Loosely, the information in a distribution is the degree to which the distribution is concentrated. Information cannot be measured absolutely but only with respect to a background measure. Being concentrated, or "spread out", is measured relative to some other distribution. Commonly, the uniform and log-uniform background measures are used.

Measuring information requires associating a density with each quantile assessment of each expert. To do this, the unique density that complies with the experts' quantiles is used and is minimally informative with respect to the background measure. This density can easily be found with the method of Lagrange multipliers. For a uniform background measure, the density is constant between the assessed quantiles and is such that the total mass between the quantiles agrees with p. The background measure is not elicited from experts, as indeed it must be the same for all experts; instead it is chosen by the analyst.

The uniform and log-uniform background measures require an "intrinsic range" on which these measures are concentrated. The classical model implements the so-called k% overshoot rule: for each item, the smallest interval $I = [L, U]$ is considered, containing all the assessed quantiles of all experts, and the realization, if known. This interval is extended to

$$I^* = [L^*, U^*]; \qquad L^* = L - \frac{k(U-L)}{100}; \qquad U^* = U + \frac{k(U-L)}{100}.$$

The value of k is chosen by the analyst. A large value of k tends to make all experts look quite informative and tends to suppress the relative differences in information scores. The information score of expert e on assessments for uncertain quantities $1,\ldots,N$ is

$$\begin{aligned}\text{Information Score}(e) &= \text{Average Relative information wrt Background}\\ &= \left(\frac{1}{N}\right)\sum\nolimits_{i=1,\ldots,N} I(f_{e,i}\,|g_i)\end{aligned}$$

where g_i is the background density for variable i and $f_{e,i}$ is expert e's density for item i. This is proportional to the relative information of the expert's joint distribution given the background, under the assumption that the variables are independent. As with calibration, the assumption of independence here reflects a desideratum of the decision-maker and not an elicited feature of the expert's joint distribution. The information score does not depend on the realizations. An expert can give himself a high information score by choosing his quantiles very close together.

Evidently, the information score of e depends on the intrinsic range and on the assessments of the other experts. Hence, information scores cannot be compared across studies.

Of course, other measures of concentratedness could be contemplated. The above information score is chosen because it is

- familiar
- tail insensitive
- scale invariant
- slow.

The latter property means that relative information is a slow function; large changes in the expert assessments produce only modest changes in the information score. This contrasts with the likelihood function in the calibration score, which is a very fast function. This causes the product of calibration and information to be driven by the calibration score.

Decision-maker. A combination of expert assessments is called a "decision maker" (DM). All decision-makers discussed here are examples of linear pooling. For a discussion of pros and cons of the linear pool, see French (1985), Genest and Zidek (1986) and Cooke (1991). The classical model is essentially a method for deriving weights in a linear pool. "Good expertise" corresponds to good calibration (high statistical likelihood, high p-value) and high information. The aim is for weights which reward good expertise and which pass these virtues on to the decision-maker.

The reward aspect of weights is very important. The following optimization problem could be solved simply: find a set of weights such that the linear pool under these weights maximizes the product of calibration and information. Solving this problem on real data, it is found that the weights do not generally reflect the performance of the individual experts.

An expert's influence on the decision-maker should not appear haphazard, and experts should not be encouraged to play the system by tilting their assessments to achieve a desired outcome. Thus, a strictly scoring rule constraint must be imposed on the weighing scheme. This basically means that an expert achieves his/her maximal expected weight by, and only by, stating assessments in conformity with his/her true beliefs.

Consider the following score for expert e:

$$w_\alpha(e) = 1_\alpha(\text{calibration score}) \times \text{calibration score}(e) \times \text{information score}(e) \tag{2}$$

where $1_\alpha(x) = 0$ if $x < \alpha$ and $1_\alpha(x) = 1$ otherwise. Cooke (1991) shows that (2) is an asymptotically strictly proper scoring rule for average probabilities. This means the following: suppose an expert has given quantile assessments for a large number of variables, and subsequently learns that his/her judgements will be scored and combined according to the classical model. If the opportunity was then given to change the quantile values (e.g. the numbers 5%, 50% or 95%), in order to maximize the expected weight, the expert would choose values corresponding to his/her true beliefs. Note that this type of scoring rule scores a set of assessments on the basis of a set of realizations. Scoring rules for individual variables were found unsuitable for purposes of weighting (Cooke, 1991).

The scoring rule constraint requires the term 1_α (calibration score) but does not say what value of α should be chosen. Therefore, α is chosen here so as to maximize the combined score of the resulting decision-maker. Let $\mathrm{DM}_\alpha(i)$ be the result of linear pooling for item i with weights proportional to (2):

$$\mathrm{DM}_\alpha(i) = \frac{\sum_{e=1,\ldots,E} w_\alpha(e) f_{e,i}}{\sum_{e=1,\ldots,E} w_\alpha(e)}. \tag{3}$$

The *global weight* DM is DM_{α^*} where α^* maximizes

$$\text{calibration score}(\mathrm{DM}_\alpha) \times \text{information score}(\mathrm{DM}_\alpha). \tag{4}$$

This weight is termed "global" because the information score is based on all the assessed seed items.

A variation on this scheme allows a different set of weights to be used for each time. This is accomplished by using information scores for each item rather than the average information score:

$$w_\alpha(e, i) = 1_\alpha(\text{calibration score}) \times \text{calibration score}(e) \times I(f_{e,i} \,|g_i). \tag{5}$$

For each α the Item weight DM_α for item i is defined as

$$\mathrm{IDM}_\alpha(i) = \frac{\sum_{e=1,\ldots,E} w_\alpha(e, i) f_{e,i}}{\sum_{e=1,\ldots,E} w_\alpha(e, i)}. \tag{6}$$

The *item weight* DM is IDM_{α^*} where α^* maximizes

$$\text{calibration score}(\mathrm{IDM}_\alpha) \times \text{information score}(\mathrm{IDM}_\alpha). \tag{7}$$

Item weights are potentially more attractive, as they allow experts to up- or down-weight themselves for individual items, according to how much they feel they know about these items. "Knowing less" means choosing quantiles further apart and lowering the information score for those items. Of course, good performance of item weights requires that experts can perform this up-down weighting successfully. Anecdotal evidence suggests that item weights are an improvement over global weights, as the experts receive more training in probabilistic assessment. Both item and global weights can be pithily described as optimal weights under a strictly proper scoring rule constraint. In both global and item weights, calibration dominates over information, and information serves to modulate between more or less equally well-calibrated experts.

Since any combination of expert distributions yields assessments for the seed variables, any combination can be evaluated on the seed variables. In particular, the calibration and the information of any proposed decision-maker can be computed. It is hoped that the Performance Weighted

Decision-Maker (PWDM) would perform better than the result of simple averaging, the "Equal Weight Decision Maker" (EWDM). It is also hoped that the proposed DM is not worse than the best expert in the panel.

In the classical model, calibration and information are combined to yield an overall or combined score with the following properties:

1. Individual expert assessments, realizations and scores are published. This enables any reviewer to check the application of the method, in compliance with the principle of **accountability/scrutability.**
2. Performance is measured and hopefully validated, in compliance with the principle of **empirical control**. An expert's weight is determined by performance.
3. The score is a long run proper scoring rule for average probabilities, in compliance with the principle of **neutrality**.
4. Experts are treated equally, prior to the performance measurement, in compliance with the principle of **fairness**.

Expert names and qualifications are part of the published documentation of every expert judgement study in the data base; however, they are not associated with assessments in the open literature. The experts' reasoning is always recorded and sometimes published as expert rationales.

There is no mathematical theorem that either item weights or global weights outperform equal weighting or outperform the best expert. It is not difficult to construct artificial examples where this is not the case. Performance of these weighting schemes is a matter of experience. In practice, global weights are used, unless item weights perform markedly better. Of course there may be other ways of defining weights that perform better, and indeed there might be better performance measures. Good performance on one individual data set is not convincing. What is convincing is good performance on a large diverse data set, such as the TU Delft expert judgement data base. In practice, a method should be easy to apply, easy to explain, should do better than equal weighting and should never do something ridiculous (Goossens et al., 1996, 1998).

Applications of the classical model

Forty-five expert panels involving seed variables have been performed to date.[1] Because most of these studies were performed by, or in collaboration with, the TU Delft, it is possible to retrieve relevant details of these studies and to compare the achievement of performance-based and equal weight combination schemes.

The combined scores of EWDM, PWDM and Best Expert are compared pairwise in Figure 13.2. Figure 13.3 compares the calibration (p-values) and information scores of the EWDM, the PWDM and the best expert.

In 15 of 45 cases, the PWDM was the best expert, that is, one expert received weight one. In 27 cases, the combined score of the PWDM was strictly better than both the EWDM and the best expert. In one case, the EWDM performed best, and in two cases the best expert out-performed both equal weights and performance-based weights.

The EWDM is better calibrated than the best expert in 25 of the 45 cases, but in only 2 cases is more informative. In 18 cases, the combined score of

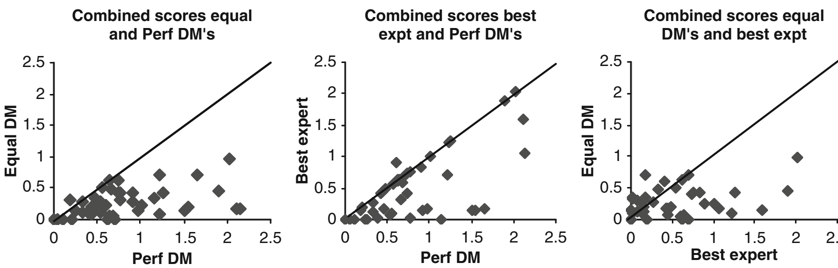


Figure 13.2 Combined scores of EWDM, PWDM and best expert

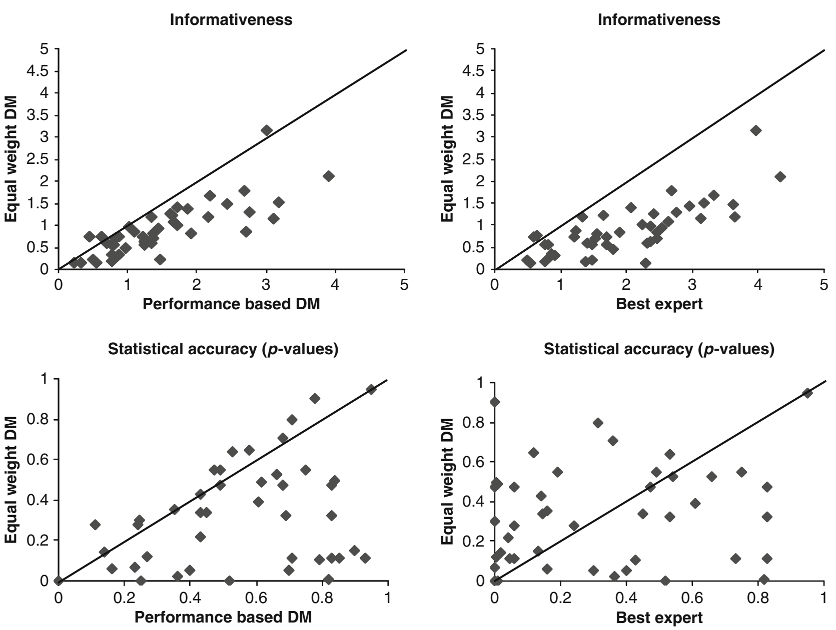


Figure 13.3 Calibration (p-values) and information scores of EWDM, PWDM and best expert

the EWDM is better than that of the best expert. In 12 of the 45 cases, the calibration of the best expert is less than or equal to 0.05; for the EWDM this happened in 7 cases (15%).

The motivation for performance-based weighting above equal weighting speaks for itself from this data. Sometimes the difference is marginal, but sometimes it is quite significant. Most often, the EWDM is slightly less well calibrated and significantly less informative, but sometimes the calibration of the EWDM is quite poor. Finally, it is noted that the experts have overwhelmingly supported the idea of performance measurement. This sometimes comes as a surprise to people from the social sciences but not for natural scientists. The essential point is that the performance measures are objective and fully transparent. It is impossible to tweak these measures for extra-scientific expediency. This does not preclude the discovery of better performance based schemes in the future. Some probes in this direction may be found in (Cooke et al., 2008a; Wisse et al., 2008).

Seed variables, variables of interest and robustness

A recurring question is the degree to which performance on seed variables predicts performance on the variables of interest. Forecasting techniques always do better on data used to initialize the models than on fresh data. Might that not be the case here as well? Obviously, there is need for expert judgement because the variables of interest cannot be observed, so this question is likely to be around for some time. Experts' information scores can be computed for the variables of interest and compared with the seed variables (see below). More difficult, is the question whether calibration differences in experts and DMs "persist" outside the set of seed variables. Questions related to this are as follows:

1. Are the differences in experts' calibration scores due to chance fluctuations?
2. Is an expert's ability to give informative and well-calibrated assessments persistent in time, dependent on training, seniority, or related to other psycho-social variables, etc.?

There has been much published and speculated about these questions, and the issue cannot be reviewed, let alone resolved here (see, however, Lin and Bier, 2008). If differences in experts' performance did not persist beyond the seed variables, then that would certainly cast a long shadow over performance-based combination. If, on the other hand, there are real and reasonably persistent differences in expert performance, then it is not implausible that a performance-based combination could systematically do "better than average".

Closely related is the question of robustness: to what extent would the results change if different experts or different seed variables had been used. This last question can be addressed, if not laid to rest, by removing seed variables and experts one at a time and re-computing the decision-maker. There follows one example where the variables of interest were later observed, and performance with respect to seed variables could be compared.

Real estate risk. In Qing (2002) the seed variables were prime office rent indices for large Dutch cities, published quarterly (variables 1 through 16). The variables of interest were rents of the actual properties managed by the investment firm. After 1 year, the realized rents were retrieved and compared with the predictions. The results for the equal and performance DM are shown below. Evidently, for both PWDM and EWDM, the performance on seed variables and variables of interest is quite similar. Note that the EWDM has larger 90% confidence bands (Figure 13.4).

Out-of-sample validation? In his review of Cooke and Goossens (2008), Clemen (2008) raised the important question: Does the performance of the Performance-Weighted DM (PWDM) persist beyond the set of seed variables? Clement believes that there is no significant difference between the PWDM and the EWDM outside the variables on which PWDM has been constructed.

As noted above, PWDM does use optimization to remove a degree of freedom in the definition of the classical model. In every study, robustness analysis is routinely performed by removing seed variables (and experts) one at a time and re-computing PWDM. It is not uncommon to see the calibration scores of PWDM fluctuate by a factor 2 or 3 on 10 seed variables.

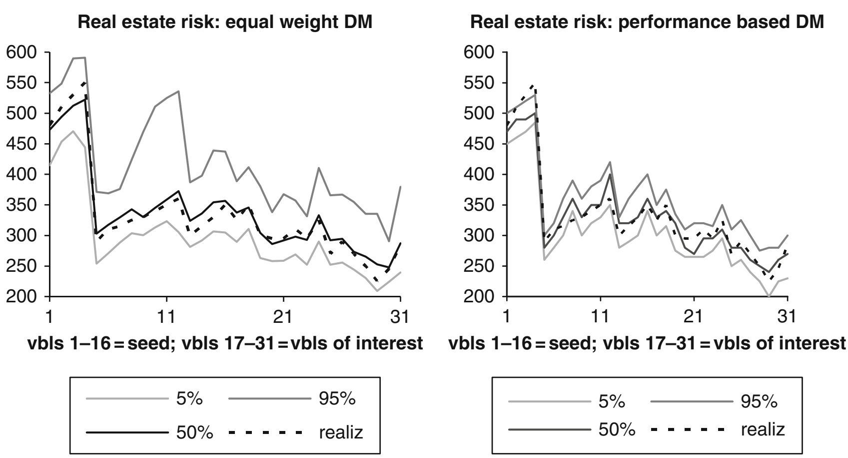


Figure 13.4 Performance versus equal weight combinations for real estate risk, seed variables and variables of interest

Out-of-sample validation involves basing PWDM on an initial set of seed variables, then using this PWDM on other variables, and comparing performance with EWDM on these other variables. This corresponds to the way PWDM is actually used. This can be done by splitting the set of seed variables into two halves, initializing the model on one half and comparing performance on the other half. Of course, this requires a relatively large number of seed variables. Thirteen studies with at least 16 seed variables are available. Dividing the seed variables in half gives two validation runs, using the first half to predict the second and conversely. Note that the variables on which the PWDM is initialized in these two runs are disjoint. The item weight PWDM could not be computed without writing a new code, so the choice of item versus global weights is denied the PWDM on this exercise.

There are 26 out-of-sample studies (two for each of the 13 studies). In 20 of the 26 studies, the out-of-sample PWDM outperforms EWDM. The probability of seeing 20 or more "successes" on 26 trials if PWDM were no better than EWDM is 0.0012.

Clemen reports results on 14 validation studies that are somewhat more pessimistic (9 "success" on 14 trials). His method involves removing seed variables singly, computing PWDM on the remaining seeds, and using this PWDM to predict the eliminated seed. On a study with 10 seed variables there are thus 10 different PWDMs. Each pair of the 10 DMs shares 8 common seeds. The criteria for selecting the 14 studies are not specified. It is difficult to see how all these factors would affect the results. Perhaps the following reasoning partially explains Clemen's less optimistic result: With a small number of seeds, removing one seed favours experts who assessed that seed badly and hurts experts who assessed that seed well, thus tilting the PWDM towards a bad assessment of that seed. This happens on every seed, thus cumulating the adverse effect on PWDM. This does not happen when one PWDM predicts the entire out-of-sample set of seeds. In any case, Clemen's method is not the same as picking one PWDM and comparing it to new observations with the EWDM.

Conclusions

Structured expert judgement has become an applicable tool in quantitative studies when input from data or experiments is lacking. The expert judgement data base provides a resource for evaluating the performance of various expert judgement combination schemes. It is clear that performance-based expert judgement models are statistically superior to simple averaging. This conclusion is based on extensive experience over a wide range of studies from diverse areas.

Modelling stakeholder preferences is less well articulated, but important ideas emerge from the field of discrete choice. The key issue for further progress is to develop tools for validating proposed models on the basis of observed preference data. Lack of external validation has plagued many of the multi-criteria approaches. It has been argued that probabilistic inversion techniques may suggest ways forward in this regard.

Note

1. These results are obtained with the EXCALIBUR software, available from http://delta.am.ewi.tudelft.nl/risk/. The windows version upgraded chi square and information computational routines, and this may cause differences with the older DOS version, particularly with regard to very low calibration scores.

References

Anderson S, de Palma A, and Thissen J. (1996). *Discrete Choice Theory of Product Differentiation*. MIT Press: Cambridge.

Budnitz R et al. (1998). Use of technical expert panels: applications to probabilistic seismic hazard analysis. *Risk Analysis* **18** (4): 463–469.

Clemen R. (2008). Comments. *Reliability Engineering and System Safety*, special issue on Expert Judgement **93** (5): 760–765.

Cooke R. (1991). *Experts in Uncertainty*. Oxford University Press: Oxford.

Cooke R (ed.) (2008). *Reliability Engineering and System Safety*, special issue on Expert Judgement **93** (5): 655–778.

Cooke R, El Saadany S, and Xinzheng HX. (2008). On the performance of social network and likelihood based expert weighting schemes. *Reliability Engineering and System Safety*, special issue on Expert Judgement **93** (5): 645–756.

Cooke R and Goossens L. (2000). *Procedures Guide for Structured Expert Judgment*. Project report EUR 18820EN, Nuclear science and technology, specific programme Nuclear fission safety (1994–1998). Report to: European Commission. Luxembourg, Euratom. Also in *Radiation Protection Dosimetry* **90** (3): 303–311.

Cooke RM and Goossens LHJ. (2008). TU Delft Expert Judgment Data Base, Special issue on expert judgement *Reliability Engineering and System Safety* **93** (5): 657–674. Available online 12 March 2007.

Cooke R, Mendel M, and Thijs W. (1988). Calibration and information in expert resolution. *Automatica* **24** (1): 87–94.

Cooke RM and Misiewicz J. (2007). Discrete choice with probabilistic Inversion: Application to energy policy choice and wiring failure. *Mathematical Methods in Reliability* July 1–4. University of Strathclyde: Glasgow.

Csiszar I. (1975). I-divergence geometry of probability distributions and minimization problems. *Annals of Probability* **3**: 146–158.

Deming W and Stephan F. (1944). On a least squares adjustment to sample frequency tables when the expected marginal totals are known. *Annals of Mathematical Statistics* **40** (11): 427–444.

Du C, Kurowicka D, Cooke RM. (2006). Techniques for generic probabilistic inversion, *Computational Statistics and Data Analysis* (50): 1164–1187.

French S. (1985). Group consensus probability distributions: a critical survey. In: Bernardo J, De Groot M, Lindley D, and Smith A (eds). *Bayesian Statistics*. Elsevier: North Holland, pp. 182–201.

Genest C and Zidek J. (1986). Combining probability distributions: a critique and an annotated bibliography. *Statistical Science* **1** (1): 114–148.

Goossens L, Cooke R, and Kraan B. (1996). *Evaluation of Weighting Schemes for Expert Judgment Studies*. Final report prepared under contract Grant No. Sub 94-FIS-040 for the Commission of the European Communities, Directorate General for Science, Research and Development XII-F-6. Delft University of Technology: Delft, Netherlands.

Goossens L, Cooke R, and Kraan B. (1998). Evaluation of weighting schemes for expert judgment studies. In: Mosleh A and Bari R (eds). *Proceedings PSAM4*. Springer: New York, pp. 1937–1942.

Goossens L, Cooke R, and van Steen J. (1989). Final Report to the Dutch Ministery of Housing, Physical Planning and Environment. *On The Use of Expert Judgment in Risk and Safety Studies* I – 5. Delft.

Goossens L and Harper F. (1998). Joint EC/USNRC expert judgement driven radiological protection uncertainty analysis. *Journal of Radiological Protection* **18** (4): 249–264.

Kurowicka D and Cooke R. (2006). *Uncertainty Analysis with High Dimensional Dependence*. Wiley: New York.

Lin S and Bier V. (2008). A Study of Expert Over-confidence. *Reliability Engineering and System Safety*, special issue on Expert Judgment **93** (5): 711–721.

Neslo R et al. (2008). Modeling stakeholder preferences with probabilistic inversion: Application to prioritizing marine ecosystem vulnerabilities. In: Linkov I, Ferguson E, and Magar V (eds). *Real Time and Deliberative Decision Making: Application to Risk Assessment for Non-chemical Stressors*. Springer: Amsterdam, pp. 248–271.

NUREG/CR-6372. (1997). Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts. *US Nuclear Regulatory Commission*.

O'Hagan A et al. (2006). *Uncertain Judgments, Eliciting experts' Probabilities*. Wiley, Chichester, West Sussex.

Qing X. (2002). *Risk Analysis for Real Estate Investment*. PhD thesis, Department of Architecture, TU Delft.

Train K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press: Cambridge, UK.

Winkler R et al. (1995). An assessment of the risk of chronic lung injury attributable to long-term ozon exposure. *Operations Research* **43** (1): 19–27.

Wisse B, Bedford T, and Quigley J. (2008). Expert judgement combination using moment methods. *Reliability Engineering and System Safety* **93** (5): 675–686.

Part IV

Analysing Information

14

Exploratory Quantitative Analysis of Emergent Problems with Scant Information

Tim Bedford

Introduction

This chapter compares and contrasts three methods for handling quantitative decision analysis when information is limited: Bayesian Robustness, Bayes Linear and Minimum Information. The way they utilize partial model specification from experts and their potential use as an exploratory tool is considered. The possibility of carrying out sensitivity analysis with these methods is explored and the recommendation made that this type of analysis is useful in extending the small world scope of such decision analyses to include various potential control mechanisms. The discussion is illustrated by simple examples.

Early stage decision-making is recognized as highly important to the course of a project and elsewhere (for simplicity, this chapter refers mainly to "projects", but this does not mean that the techniques discussed are only relevant to project risk management – in fact they are relevant to a wide range of emergent problems). The usual difficulty is that the information available at this early stage is very limited. There are various sources of uncertainty, such as the following:

- Lack of understanding about the major uncertainties and their interactions that could impact on the project
- Lack of understanding about the degree to which this problem is similar to previous problems
- Lack of understanding of the way in which future decisions will affect outcomes.

The first point raises the issue of assessing multiple uncertainties using procedures that capture the relevant dependencies. The second point acknowledges that if the analogies are to be made explicit, there is usually some knowledge that is relevant to specifying uncertainties. The final point underlines that the situation is usually a *dynamic* one, where decisions taken down the line might mitigate or exacerbate outcomes. The decision-maker should ideally take account of this dynamic aspect in selection of strategies.

This chapter considers the use of exploratory quantification to help "firm up" ideas around the above issues. It enters the realm of subjective probability, Bayesian methods and sensitivity and uncertainty analysis. Exploratory quantification has to be performed in a realistic way that is commensurate with the quality of data available. Hence the structure of the model has to reflect the main drivers influencing outputs. It should be robust to its quantitative data, and it should not be more costly to generate the model than is justified by the level of savings that it could make.

The basic modelling paradigm used here is Rational Decision Theory. Rational Decision Theory (Clemen, 1997; French, 1988; Morgan and Henrion, 1990) deconstructs a decision problem in terms of decision alternatives, uncertainties about outcomes in the real world and consequences (depending in a deterministic way on both choices made and the realized outcomes). Utility theory models a rational decision-maker's preferences for decisions with uncertain outcomes in terms of a utility function – a real-valued function defined on the set of possible outcomes – in which assessments of preferences on simple uncertain choices (usually called lotteries) can be extrapolated to more complex problems using consistency criteria.

Such decision models are "small world" models that can help guide behaviour. However, they require a number of specifications in order to operate at all. It is necessary to make assessments of which uncertain quantities are important in determining the consequences of decisions and to be able to quantify them in some way. It is also necessary to assess what type of consequences would affect an evaluation of the decision and to quantify the relative evaluation of those consequences (in the technical sense of the utility function referred to above).

This type of evaluation can be quite costly in practice, and hence there is a need to be pragmatic. Good (1983) refers to the notion of a Type II Bayesian as one who wants to stay within the paradigm of Rational Decision Theory but who takes into account the costs of doing a full analysis.[1] This is an attractive philosophy. When dealing with problems with scant information, it is useful to do some kind of "back of the envelope" calculation, as a fuller and more satisfying analysis would necessitate the collection of so much extra data that the decision point may well have been passed. This is certainly the case, for example, with many systems engineering projects, where vital decisions have to be taken early on before much of the information that could drive them is really available. The aim here, then, is to consider three different possible analytic approaches to modelling uncertainties where there is only broad-brush information. This information typically comes from experts and from analogy[2] data. Hence, in conjunction with the discussion of the methods, it will be shown how these methods can be applied to deal with expert and analogy data.

Expanding the small world model

The main theme of this chapter is uncertainty representation and modelling. However, to give a broader view of project problems, it is worth first considering the bigger picture. The use of decision-making tools requires the specification of decision alternatives, uncertainties and consequences, together with the specification of value judgements about consequences – typically through the use of costs, loss functions or utility functions (in order of increasing sophistication). However, it should be recognized that many of these aspects are determined by the context in which the decisions are taking place.

Decision alternatives are often determined by previous choices and consequences. For example, in engineering design projects, the availability of technologies that have previously been developed to a mature (or near-mature) level has a huge impact on the breadth of technical alternatives available. An example of this is with NASA's Project Apollo, where the US technological lead over the USSR in heavy-lift launchers was the key to providing a set of decision alternatives from which Kennedy could select a program goal. On the other hand, decisions taken at one stage of a program can narrow down the set of alternatives available later.

Value judgements, encoded in formal analyses in the utility function, are also highly contingent on previous experiences and elements of the program. This can be seen operating at a gross level in public life, where trade-offs between cost and safety are apparently related to the time after major accidents (Ale, 2005). Ale has noted the correlation over several centuries, between legal initiatives to reduce industrial risk, on the one hand, and major accidents on the other. Hence the existence of a feedback loop between disasters and the level of legal and administrative attention to mitigating risks is shown at a very broad level. As time since the disaster increases, attention wanders to other more urgent aspects, such as economic development. Rules and/or their implementation are relaxed until the next major event refocuses attention on safety.

Simplistically, this could be seen as a shifting trade-off between cost and safety, which depends on the distance from historical events. Those events serve as a feedback mechanism to prevent safety issues becoming negligible, but the very existence of the feedback mechanism shows that, as a society, we are not capable of truly making the choice commonly voiced in the aftermath of an accident, "to make sure this never happens again". While this example is a very large-scale one, such feedback loops occur at all sorts of places within large projects.

John Adams (1995) introduced the idea of the risk thermostat. Broadly speaking, at an individual level, risk-attenuating behaviour is seen as one element describing the way people adjust to the requirements placed on them. However, risk-reduction is not the main goal, and hence they will tend to adjust their behaviour to achieve other goals, as long as risk is perceived to be at an acceptable level. According to the risk-thermostat idea, the notion of what is perceived to be acceptable is influenced by what others think and by the way others behave. The risk thermostat is a very explicit feedback system, at least at a qualitative level.

More generally, the notion of control can be important in linking together instances of decision problems to the broader context of the program within which the decision problems play a part. Bedford et al. (2006) used this notion within the context of Systems Engineering to talk about the way that poor early decisions can be compensated through decisions available later in the systems engineering process. For example, lower than desired system reliability can be compensated by increased maintenance (at a cost) to deliver the desired overall level of system availability.

More generally, the context of a specific decision problem can be defined qualitatively through looking not only at the usual

- Decision alternatives
- Uncertainties
- Consequences

but also by assessing control mechanisms available later, which is outside the scope of the present decision problem. It is suggested that such control mechanisms should be incorporated into a conventional analysis through the use of sensitivity analysis. This will be discussed below.

The three methods presented below are all targeted at modelling the uncertainties within the decision problem, and thereby providing a mechanism to compute expected utility or other proxy quantities. The technical problem these methods seek to address is how to use limited information available from experts about uncertainties. In practice, information about these uncertainties can be obtained by asking for quantiles (e.g. the median of the distribution), moments (e.g. the mean and variance) or about other information which implicitly restricts the possible distributions. Bayesian theory is relevant to this issue, as this step corresponds to the establishment of a prior distribution for Bayesian analysis.

Robust Bayesian methods

One very relevant body of literature that has emerged and attempts to deal with the issues raised by scant knowledge in probabilistic modelling is that of Bayesian Robustness. This section draws heavily on Berger et al. (2000), who state, "Robust Bayesian analysis is concerned with the sensitivity of the results of a Bayesian analysis to the inputs for the analysis". Thus Robust Bayesian methods seek to address one of the perceived problems of Bayesian methods, namely, that the outcome depends on an apparently arbitrary choice of prior distribution. In addressing this problem, these methods also give support to the problem in which a decision-maker is not willing or able to fully specify a prior. The methods should not only help to assess the uncertainty in outputs driven by the decision-maker's partial specification but also give guidance as to how that uncertainty can be reduced, i.e. where the decision-maker could best concentrate on providing additional specification. Some limited help in this direction is possible.

Berger et al. identify three broad classes of technique within Robust Bayesian methods: informal, global and local techniques. Of these, the first two are discussed, being most relevant to this chapter.

Informal methods

These methods experiment with a number of different prior distributions that meet the criteria of the decision-maker. For example, if a number of quantiles have been specified for an uncertain variable, some prior distributions that fit those quantiles could be chosen. Then by calculating the output quantities (which might be an expected utility, or some other quantity), the extent of the difference made by the different priors can be assessed.

While an informal approach gives a reasonable first-pass assessment of sensitivity, it is certainly not systematic. It is not clear how the various prior distributions should be chosen nor how to judge what the effect is.

Global methods

A number of different global methods are possible, but are based on the idea of perturbing a "baseline" prior to another nearby prior.

The contamination method works by making a mixture of the default prior[3] with another distribution drawn from some reasonable class. It is critical here that only a "reasonable" class is used, because the quantities that are of interest – typically means and variances – in judging sensitivity can change by arbitrarily large amounts, even under the smallest perturbations of the baseline prior. Examples of reasonable classes are those that are unimodal and satisfy given quantile constraints.[4] Another approach uses a distribution band, where upper and lower distribution functions that bound the possible prior distribution are specified.

Finally, it is possible to simply specify a number of quantiles and look at the class of all distributions compatible with them. Upper and lower bound results for means are calculated quite easily, as they correspond to degenerate situations where the probabilities are concentrated on quantiles.

However, it is also possible to impose other constraints related to unimodality, which exclude the degenerate discrete cases and impose quite tight constraints on means.

Robustness feedback to decision-makers

The use of quantiles is very common in quantifying models, as these can be elicited fairly easily from decision-makers and experts. Because of this, the latter technique for modelling global robustness is rather interesting. It naturally raises the question of whether advice can then be given to decision-makers about how they could improve the robustness of their models – e.g. by specifying extra quantiles. Broadly, the idea here is to consider how the variation in the quantity of interest depends on the different intervals defined by the current set of quantiles. For example, Moreno et al. (1996) consider a variation on the contamination method described above, in which they allow separate contamination on each interval. This then allows them to compute the contribution to variability on each interval separately and to select the interval with the greatest contribution as the one that should be targeted to be split by eliciting another quantile from the expert/decision-maker.

Robustness to other model choices

The above discussion has concentrated on robustness with respect to the choice of prior distribution (or with respect to the partial specifications of the prior). However, a Bayesian model also consists of a likelihood function, enabling the decision-maker to change his/her prior through observations, and a loss function, enabling the decision-maker to make a choice of optimal parameters. Clearly the problem of robustness with respect to these choices is also important. Robustness with respect to loss function choice is not so difficult to consider, e.g. by studying one-parameter families of loss functions, and considering the changes. However, robustness with respect to likelihood choice is more complex. In both cases, an informal approach seems to be the most straightforward.

Bayes linear methods

A "simplified" representation of uncertainty is given by the Bayes linear approach propounded by Michael Goldstein and his co-workers (Goldstein and Wooff, 2007) and the references therein. This methodology takes a second order approach, in which uncertain quantities are modelled through their expectation values, variances and co-variances. In a sense this gives a "finite dimensional representation" of a belief system, because a finite number of quantities are only ever modelled, using a finite number of parameters, and conclusions can only be derived about those quantities. By contrast, in a conventional probability model, the probabilities are typically defined on some infinitely detailed space,[5] and the model can be used to derive conclusions about fine level details that are wholly artifacts of the representation.

Bayes linear uses the notion of expectation as the primitive concept from which other quantities can be derived. My expected value for an uncertain real-valued quantity X is the quantity $E(X)$ for which I am indifferent between the sure value of $E(X)$ and the random quantity X, i.e. $E(X)$ is the "fair price" for a lottery ticket whose outcome is the prize X. In contrast to the situation in probability, where $E(X)$ is defined through an integral involving the probability distribution for X, in Bayes linear, the expectation concept could be used to define probabilities, if that were deemed necessary for modelling purposes. To be precise, if one is interested in an event A then one can consider the characteristic function $a(x)= 1$ if $x \in A$, $a(x)= 0$ otherwise. It is easy to see that $E(a)$ is equal to the probability of A, and hence that probabilities can be derived from expected values. For the sake of this discussion, it can be assumed that there is a finite amount of exclusive and exhaustive events. To obtain the expectation value of a random quantity defined on these events from probabilities, requires that the probabilities on the full set of exclusive and exhaustive events have been specified, and enough information is hence provided to work out the expectation of any random quantity defined on these events. By contrast, if one starts with expectation as the primitive, it is possible just to work with a smaller set of functions.

Hence one of the main characteristics of the Bayes linear approach is that it allows a "partial prior specification" in cases where it would be too time-consuming or costly to develop a fully specified probabilistic model. The approach has a full axiomatic basis and therefore provides a well-founded alternative uncertainty representation to "full" probability analysis.

The second main characteristic of Bayes linear is that it allows for a simple form of adjustment of prior beliefs when making observations. This adjustment is analogous to Bayesian updating in conventional probability models but is carried out through a process of linear fitting. Suppose that the decision-maker is interested in a vector of quantities $\underline{X}$ and has specified $E(\underline{X})$ and $\mathrm{var}(\underline{X})$. Suppose also that the decision-maker is able to observe some vector of quantities $\underline{D}$, which will be used to improve the assessment of $\underline{X}$. The decision-maker specifies $E(\underline{D})$ and $\mathrm{var}(\underline{D})$, and in order to quantify the relationship between $\underline{X}$ and $\underline{D}$ must also specify $\mathrm{cov}(\underline{X}, \underline{D})$. Once these values are elicited and observations made, the decision-maker can adjust their prior assessments by linear fitting. The adjusted expectation of $\underline{X}$ given the observation of a collection of quantities $\underline{D}$ is

$$E_D(\underline{X})= E(\underline{X}) + \mathrm{cov}(\underline{X}, \underline{D})\mathrm{var}^{-1}(\underline{D})[\underline{D} - E(\underline{D})]$$

where $\mathrm{var}^{-1}(\underline{D})$ is the Moore-Penrose generalized inverse. The adjusted variance of $\underline{X}$ given $\underline{D}$ is $\mathrm{var}_D(\underline{X})= E(X - E_D(X))^2$.

The expression used here is derived from the idea of trying to form a linear estimate of each variable $\underline{X}$ from the observables $\underline{D}$. The best linear estimate is the one that minimizes the prior expected squared error

$$E\left[\left(\underline{X} - \underline{c_0} - \underline{c_1} \cdot \underline{D}\right)^2\right].$$

The minimization of this expression is straightforward and can be shown to be given by the linear expression of the variables $\underline{D}$ given above.

Link to Gaussian models and graphical methods

One connection between Bayes linear methods and full probabilistic methods is as follows: If the assumption is made that the variables have a joint normal distribution, then the adjusted expectations and variances are those that could be obtained by full Bayesian updating. Graphical models can be used to represent the structure of a Bayes linear model in a similar fashion to the way that graphical models are used to represent Bayesian nets, as illustrated in the example below.

Example

Here we give a simplified presentation of an example on reliability testing taken from Goldstein and Bedford (2007). Further examples on applications of Bayes Linear to reliability management are discussed in (Revie, 2008) and (Bedford et al., 2008). A sequence of observations of lifetimes $X_1, \ldots, X_n$ of a new product can be obtained from an in-service trial (at a cost); historical data $Y_1, \ldots, Y_m$ about a similar system can be inspected (at a cost); and data Z_i can be observed from a test rig (at a set-up cost, and cost per test). The model assumes that each component lifetime X_i is equal to the mean time to failure $\overline{X}$ plus a random term that is uncorrelated with everything else in the model and that each Y_i and Z_i is similarly related to their mean values. It can be assumed that the sequences of observations are related through the relationships between $\overline{X}, \overline{Y}$ and $\overline{Z}$. The notation $\overline{X}_n, \overline{Y}_n, \overline{Z}_n$ is used to denote the sample means (where it is understood that the number of samples on which the sample mean is based is possibly different for each variable). It can be shown that the sample means are sufficient for any inferences to be made about the underlying population means $\overline{X}, \overline{Y}$ and $\overline{Z}$.

Bayes linear analysis makes use of graphical models in the same way that Bayesian analysis uses Bayesian networks. The graphical model for the situation modelled here is shown in Figure 14.1.

The diagram illustrates the main relationships that have to be defined. Hence in this model, a key decision is to model the extent to which the old and the new systems are similar. The difference is modelled as $\overline{Y} = \overline{X} + R_1$ where only the mean and variance of R_1 has to be specified, as opposed to a full probability distribution in a probabilistic model. In order to model the

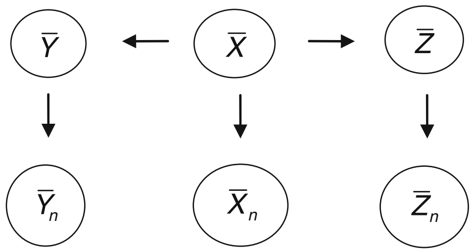


Figure 14.1 Bayes linear network

test rig, $\overline{Z} = \overline{X} + R_2$ is assessed, where R_2 has a small mean but a large standard deviation to capture the idea that the testing is probably accurate but that (through lack of understanding) the wrong thing may have been tested.

Decision-makers in this type of problem need to know how much testing or historical data must be collected to reduce the uncertainty in the new system mean time to failure. This type of problem can be easily explored by looking at how the adjusted variance for $\overline{X}$ changes as more data is received. Figure 14.2 shows the amount of reduction in the adjusted standard deviation for $\overline{X}$ achieved for various numbers of observations of X_i and Z_k without including the historical data. Figure 14.3 shows how that deviation changes by including the historical data. Hence, this type of plot can be used to work out the cost/benefit of acquiring information.

The value of this kind of analysis is that it shows that many observations of the test system, or indeed of the historical system, only reduce uncertainty about $\overline{X}$ up to a point. Eventually the uncertainty in how $\overline{Y}$ and $\overline{Z}$ are related to $\overline{X}$ dominates and makes extra data worthless.

This example reflects some of the known problems in the use of reliability data, where some databases have shown huge accuracy, but this is an

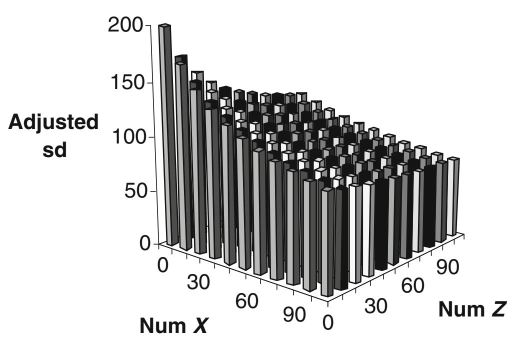


Figure 14.2 Adjusted standard deviation without historical data

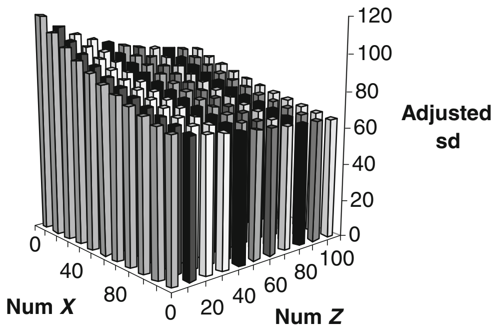


Figure 14.3 Adjusted standard deviation using historical data

accuracy that has been built up by extensive testing and therefore does not capture the true field reliability.

Maximum entropy and minimum information

The concept of entropy is one that crops up in many different areas of science. It will be shown here how the concept can be used to develop analysis tools on the basis of scant data or partial specification. Entropy, defined by Shannon and others in the area of information theory, has been used by many authors, notably Jaynes (2003), as a measurement of the degree to which a given probability distribution departs from being uniform. Its generalization, relative information, is a measure of the degree to which a given probability distribution departs from a specified background or reference distribution.

For a discrete probability distribution $\underline{p} = (p_1, \ldots, p_n)$ the entropy is defined to be $H(\underline{p}) = -\sum p_i \ln(p_i)$. This function has a number of appealing mathematical properties (Khinchin, 1957). First of all, it is non-negative and equals 0, if and only if exactly one of the p_i is one while the rest are zero. Secondly, the maximum value is taken when all the p_i are equal to $1/n$. Thirdly, the entropy is unchanged, if an impossible event is included, i.e. if $\underline{q} = (p_1, \ldots, p_n, 0)$ then $H(\underline{p}) = H(\underline{q})$. Fourthly, it varies continuously in the parameters. Fifthly, the entropy can be calculated quite simply for conditional probabilities: Suppose there is a partition into exclusive and exhaustive events $A_1, \ldots, A_n$, with associated probabilities $\underline{p} = (p_1, \ldots, p_n)$, and that there is another partition $B_1, \ldots, B_m$. Write $\underline{q} = (q_1, \ldots, q_{nm})$ for the probabilities over the partition refining these partitions, i.e. the partition into sets $A_i \cap B_j$. Then the entropy for the refined partition can be computed as an entropy from the coarse partition $A_1, \ldots, A_n$ plus an additional entropy gain on moving from the coarse partition to the finer partition. To be precise,

$H(\underline{q}) = H(\underline{p}) + E(H(\underline{q}_A))$, where $\underline{q}_A$ is the conditional distribution of $\underline{q}$ given that the event A from the partition $A_1, \ldots, A_n$ has occurred, and the expectation sign denotes taking the mean value over such events.

While these properties are nice mathematical properties, it is legitimate to question how arbitrary the definition of entropy given above really is. In fact, it can be shown that these properties characterize entropy up to a constant multiple. Hence, given these properties, the only arbitrary choice to make is in the base of the logarithms.

Entropy is a concept that appears in a number of different branches of mathematics, statistics and physics. It is a key concept in the mathematical theory of information (used to measure the amount of information in a signal and hence allowing an estimate of best possible signal compression from which the original signal can be reconstructed). It is also a key concept in thermodynamics. Because of its appearance in different disciplines, there has been a certain amount of parallel development. This also means that definitions may change very slightly between one area and another. Entropy has been used frequently within statistics and has been suggested as a tool to specify prior distributions. This view is most notably advanced by Jaynes (2003), who strongly advocated using the principle of maximizing entropy to select prior distributions.

Unfortunately, the definition of entropy given above is not appropriate for most applications. To try and write down the generalization of the formula given above to a continuous density reveals that it lacks invariance under coordinate transformations, i.e. the number would depend on the particular way the formula for the density had been written down. The way to avoid this is to consider the concept of relative information instead. This is also known in statistics as Kullback-Leibler divergence. Suppose there is some "background" distribution, e.g. the uniform distribution, and its density is written as q. Given another distribution with density p, then the relative information of p with respect to q is given by

$$I(p|q) = \int \frac{p(x)}{q(x)} \ln\left(\frac{p(x)}{q(x)}\right) . \mathrm{d}q(x).$$

(A warning: Jaynes considers the negative of this quantity, and in the literature there appears to be a little inconsistency in the precise name and sign of this quantity across the different disciplines that use the concept.) It is well known that it is easy to calculate explicit formulae for maximum entropy distributions given moment constraints (Jaynes, 2003), and an extensive formulation in terms of statistical mechanics with the link to Gibbs measures is found in Lanford (1973). In other words, suppose a distribution p is wanted that (in the sense of minimizing information) is most like q but has the additional property that the expectations of functions $f_1, \ldots, f_k$ are equal to specified values $\mu_1, \ldots, \mu_k$. There is a unique solution p for this optimization problem that can be written in the form $p(x) = q(x).\exp(\lambda_1 f_1(x) + \cdots + \lambda_k f_k(x))/Z$, where Z is the so-called partition function

$$Z = Z(\lambda_1, \ldots, \lambda_k) = \int \exp(\lambda_1 f_1(x) + \cdots + \lambda_k f_k(x)) \mathrm{d}q(x).$$

And the Lagrange multipliers $\lambda_1, \ldots, \lambda_k$ are determined by solving

$$\mu_i = -\frac{\partial \ln Z(\lambda_1, \ldots, \lambda_k)}{\partial \lambda_i}, \quad i = 1, \ldots, k.$$

For Jaynes, and indeed for many others, the maximization of entropy (or the minimization of information) is a clear solution to the problem of identifying a subjective (prior) distribution when there is limited prior information. One of the main difficulties remains the reliance on the "background" distribution, q. "This is the shortcoming from which the maximum entropy principle has suffered until now, and which must be cleared up before we can regard it as a full solution to the prior probability problem" (Jaynes, 2003).

Jaynes' solution is to return to the notion of non-informativeness and the principle of insufficient reason, but to give these a different form, by looking – in the specific application context – for invariance principles that could be applied to the background distribution, in order to fix it. These invariance principles are coded into the form of invariance under a group of transformations, such as changes to scale and location. It is well known that for some spaces there are natural invariant measures (e.g. the uniform measure is the unique continuous rotation invariant measure on the circle). However, it is also clear that there may not always be any obvious invariance principle to apply. One of the two major criticisms of the maximum entropy method is that it is not obvious how to choose this background measure: "Unfortunately, having to choose a base measure is almost as hard as choosing a prior so that this solution is rather circular" (Kass and Wasserman, 1998). This point is discussed below.[6]

In the author's own work, frequent use has been made of the minimum information principle in the context of building simple models for joint distributions (Bedford and Cooke, 2001, 2002; Bedford and Meeuwissen, 1997; Cooke, 1997; Kraan and Bedford, 2005). This work recognizes that in the practice of assessing subjective distributions, some aspects are easier to elicit than others. In particular, assessing marginal distributions, i.e. single variable uncertainty, is considerably easier than assessing multi-dimensional distributions. Hence, it might be possible to pin down most information about marginal distributions but only have limited information about how are uncertainties are linked together. The elicitation aspects of this will be discussed later. From a modelling point of view, however, this leads naturally to the situation where – taking just two variables, to make the presentation simpler – distributions for X and Y are known and the aim is to build a model for the *joint* distribution of X and Y. This type of problem is an important topic of current research, with much interest in financial modelling, engineering applications, etc. For the two variable problem, we can pose it as follows:

Given distributions F_X and F_Y, can information be elicited that will allow the modelling of the joint distribution F_{XY}?

This question is naturally reformulated in terms of a *copula*. A copula is a probability distribution on the unit square with uniform marginals. Every joint distribution has an associated copula, because each individual variable can be transformed through its distribution function to be uniform: since $F_X(X)$ and $F_Y(Y)$ are always uniform, the distribution $C(u, v) = F_{XY}\left(F_X^{-1}(u), F_Y^{-1}(v)\right)$ must be a copula. However, this process can be reversed, and by specifying C, F_X, F_Y the joint distribution F_{XY} can be defined through this formula (Sklar, 1959).

Relative information is now a nice way to proceed, because the usual objection – that there is no clear way of choosing a background measure – does not apply in this case. There is a very natural choice of background measure, namely, the independent copula. Use of this as background measure implies the aim to make the joint distribution as much like the independent distribution as possible, given the additional constraints imposed.

The major technical problem when constructing minimum information copulas relates to the constraint that the marginals be uniform. The usual approach to dealing with constraints, using Lagrange multipliers as outlined above, does not work, as there is an uncountably infinite number of constraints. Fortunately, it is possible to construct algorithms that converge rapidly. Bedford and Meeuwissen (1997) showed how to compute the minimum information copula with given Spearman rank correlation and computed analytic expressions (unfortunately not in closed form). The potential for this sort of technique can be illustrated quite simply with an example computed using only the power of Excel.

Example

In reliability theory applications, it is common to assume that component lifetimes have a constant failure rate (there is some practical justification for this parametric assumption, since early "burn in" failures may be removed by quality procedures, and late "wear out" failures may be irrelevant due to component replacement). However, many decision-making problems require knowledge about the joint behaviour of two (or more) components, i.e. to be able to specify a joint distribution with exponential marginals. There are various different approaches, but to illustrate the procedures suggested, a minimum information approach is applied in two different ways. To avoid dealing with a perfectly symmetric situation, it can be assumed that there are two components: one has failure rate 1, and the other failure rate 2.

Case 1

In contrast to Bedford and Meeuwissen, who looked at rank correlation (on the grounds that rank correlations can take any value between -1 and $+1$ and also that they could compute explicit expressions), product moment correlation is considered here or, equivalently, $E(XY)$. This quantity is considered more appropriate than rank correlation, as XY is a quantity that has a real interpretation (despite this, it is not an easy quantity for an expert to think about, which is one of the reasons for looking at a different quantity in Case 2). A single constraint is used on $E(XY)$ and, therefore, a single Lagrange multiplier. Figure 14.4 shows the range of possible values that $E(XY)$ can take as a function of that Lagrange multiplier. This type of figure can be used to support experts in making assessments of the quantity in question, by indicating the range of permissible values.

The selection of a specific value for $E(XY)$ then specifies the Lagrange multiplier and hence the functional form of the minimum information copula. Figure 14.5 illustrates two copulas. On the left there is a positively correlated density ($\lambda = 2$), while on the right a negatively correlated copula ($\lambda = -2$) is shown.

Case 2

As noted in Case 1, the elicitation of $E(XY)$ will not be without its problems. This is because the product of two lifetimes XY does not have a ready contextually meaningful interpretation. Better in this case is to consider a quantity such as the difference in the two lifetimes, $X - Y$. Better still is

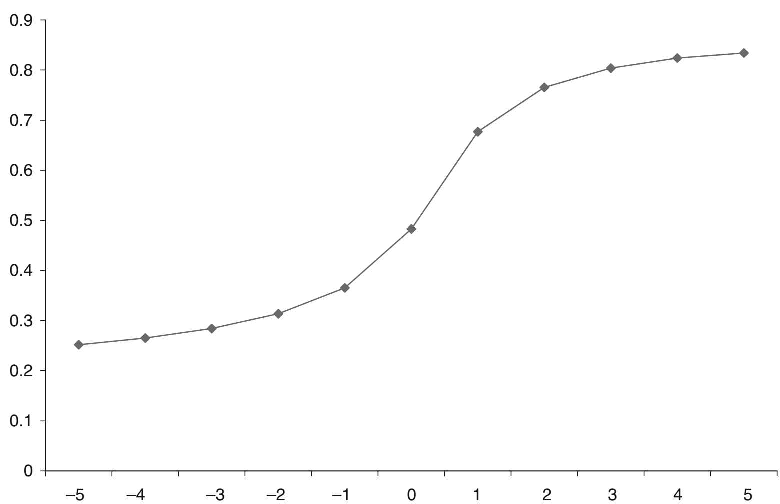


Figure 14.4 Range of values for $E(XY)$ as a function of the Lagrange multiplier

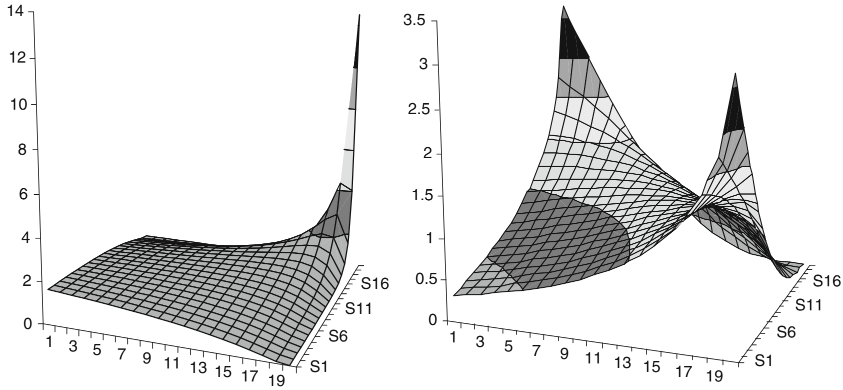


Figure 14.5 Minimum information distribution with given product-moment correlation (left $\lambda = 2$, right $\lambda = -2$)

to consider quantiles of this difference, as experts are generally better at assessing quantiles than at assessing expectations. Fortunately the minimum information solution can still be determined with this type of specification. Figure 14.6 shows the minimum information copula given that the expert asserts $P(X - Y < 0.3) = 0.3$ and $P(X - Y < 0.9) = 0.7$. Note that as soon as there is more than one constraint, the range of possible values for the second constraint is affected by the choice made for the first constraint. The constraint values cannot be chosen "independently" of each other.

The work of Bedford and Cooke (2002), and Kurowicka and Cooke (2002) on vines provides a useful way of extending minimum information copulas to minimum information joint distributions on multiple variables.

Expert judgement

Since the qualitative and quantitative assessments of models in early stage decision-making have to be driven by expert judgements, it is worth reflecting a little on the effectiveness of experts and on methods for eliciting that information. It is not the purpose of this chapter to go far into expert judgement issues and refer the reader elsewhere for more information (Bedford et al., 2006; Cooke, 1991; Meyer and Booker, 2001; Morgan and Henrion, 1990; O'Hagan et al., 2006; as well as Wright's chapter in this book).

The first point is certainly not new but is often forgotten. This is that judgement is present at all stages of the modelling process. Even the most data-driven frequentist statistician uses judgement to select appropriate models and appropriate tests. Judgement is an unavoidable element in establishing an outline modelling approach that is designed for a specific context.

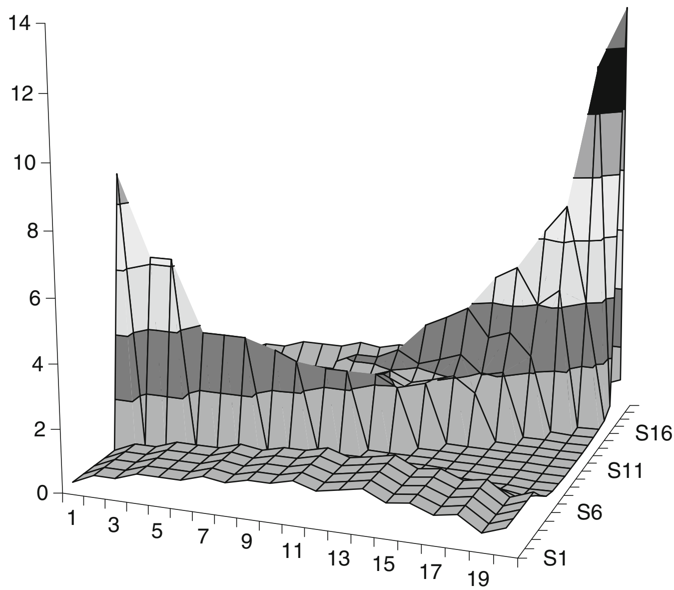


Figure 14.6 Minimum information copula with quantile constraints on lifetime differences

One good example of the way qualitative structuring can go hand in hand with quantitative modelling is the REMM approach to reliability design (Walls et al., 2006). Here design-engineer elicitation is used to explore areas of concern within a nascent design and to track the potential for excluding or reducing specific failure modes.

The qualitative structuring of models is of vital importance, particularly in early stage decision-making. The elicitation literature demonstrates that assessment of low quantiles is particularly difficult for experts. Therefore, it is necessary to tease out potentially significant low probability events at the qualitative stage, in order to make conditional quantitative assessments where appropriate. Morgan and Henrion (1990) discuss the benefits and partial experimental evidence of this approach. When multiple experts are involved, it is also important to compare assessments and reasoning for their assessments, just in case different mechanisms come to light.

Much of the statistical literature on constructing subjective distributions is concerned with modelling ignorance. However, it should be remembered that in almost no practical context is there true ignorance. The real problem is how to capture and make explicit that which is known.[7] As an example, in Systems Engineering design problems, there are almost no really new systems. Almost every "new" system is developed by scaling existing technologies, and those few technologies that are truly new have to go through a long process of maturation (which includes lots of types of testing) before being capable of real application. Therefore, for these systems, there is almost always some kind of experience with heritage or test systems that can be used to provide some relevant context. In this sense, it could be argued that for practical problems statistical definitions of ignorance do not help, and more effort should go into pinning down what is actually known.[8]

A bigger difficulty is how to utilize expert opinion effectively. Although expert opinion is important in developing the qualitative model structure, there are also different types of opinion that can be used at the quantitative stage. Much of the expert judgement literature concentrates on specific and detailed information, such as quantile assessment. However, there is potential to use more. Robust Bayesians are able to take account of information such as unimodality, while the minimum information method requires the specification of a background distribution. These are both examples of "broad-brush" information. In fact, it is surely an advantage rather than a disadvantage that the minimum information method takes account of this. It allows the splitting of the expert judgement process into two stages, one broad-brush and the other detailed. The broad-brush view is not detailed, but simply a statement that "in the absence of other information I would like the distribution to look broadly like this". The specific detail information adjusts the broad picture to include specific features.

Sensitivity analysis and model outputs

The need for sensitivity analysis within Decision Theory has been recognized for some time. Indeed, one of Morgan and Henrion's Ten Commandments for Good Policy Analysis (Morgan and Henrion, 1990) is to perform systematic sensitivity and uncertainty analysis.[9] They give an extensive discussion of sensitivity analysis methods, three of which are particularly relevant to this chapter.

The expected value of perfect information (EVPI) is the extra utility that can be gained (in expectation) by knowing an uncertain quantity exactly. This can be computed in a decision analysis by performing "what if" analyses to discover the best decision and corresponding utility, if a specific value for the uncertain quantity is known and then averaging over the different utility values. Perfect information should always give an improvement in the overall expected utility and represents the maximum possible benefit that could be gained from the acquisition of extra information about an uncertain variable.

The expected value of including uncertainty (EVIU) is a quantity introduced by Morgan and Henrion to consider the benefit of a "finer level" of uncertainty modelling. This fits well into the general theme of this chapter, where it is made clear that information is only available at a rough level of detail. EVIU is defined as the difference between an optimal decision ignoring uncertainty and the optimal expected utility decision. The optimal decision ignoring uncertainty is the one that is optimal after a given uncertain variable is replaced by its point estimate. Of course, the full distribution for the uncertain quantity is needed before the value of including it can be calculated. Because of this, Morgan and Henrion use it as an indicative tool, exploring how the form of loss/utility function affects the EVIU on different cases, and they compare it with EVPI as follows:

> The EVPI is the expected cost of being uncertain about x, whereas the EVIU is the additional expected cost of pretending you are not uncertain.

They show that for linear and quadratic loss functions, the EVIU is zero, while for bi-linear loss and catastrophic "plane catching" loss, the value can be positive. They suggest, however, that there are benefits to not ignoring uncertainty, even when EVIU is zero or near zero. One such as that further introspection may change the assessment of what the "best estimate" is, and this it may indicate where further model development must take place. It seems that this requires returning to the initial stages of analysis and reviewing the major influential factors to be included therein.

The final quantity discussed here is the expected value of perfect control. To calculate this for an uncertain quantity it can be treated as if it can be chosen, and then one can work out (using the maximum expected utility rule) what its optimal choice would be. The difference in overall utility between this problem and the original one is the value of perfect control. In some problem types, the expected value of perfect control can be used to indicate what variables need controls. Typically, these are factors that are outside the current scope of the decision-maker (e.g. they may relate to aspects of a design process that will happen downstream, such as the specification of a maintenance procedure, or the requirements on user training), but which are relevant to the wider decision-making that has to be supported.

Conclusions

The methods that have been discussed here are different quantitative modelling approaches that try to provide mechanisms by which quantitative models can be used, even when the input information available is limited. In the case of Bayes linear, model specification is made in terms of means, variances to specify uncertainties about individual quantities and covariances to specify the interactions of these quantities. Robust Bayesian analysis uses full probabilistic modelling to model joint behaviour of different quantities but has developed a range of methods to enable consideration of the sensitivity of model outputs to the level of detail provided for inputs.

Methods also exist to include some qualitative features of distributions in the assessment. Minimum information methods allow the provision of gross-level information through the specification of a background distribution, in combination with more detailed information of a level suitable for experts. While the method is usually presented in terms of moment specification, the example given in this chapter shows that it can be adapted to allow for quantile specification of observable quantities. Minimum information methods can be used to build up multivariate distributions, as in vines, but become computationally more complicated.

This chapter raised a number of areas of uncertainty arising in emerging projects:

- Lack of understanding about the major uncertainties and their interactions that could impact on the project
- Lack of understanding about the degree to which this problem is similar to previous problems
- Lack of understanding of the way in which future decisions will affect outcomes.

The methods discussed here allow the construction of exploratory models to address these issues.

Bayes linear provides a particularly simple way of building up models with multiple interacting uncertainties, though at the cost of providing fairly limited scope for the nature of those interactions. Minimum information methods provide more flexibility in modelling dependencies, although there are no clear rules for deciding how much detail should be included.

The Bayes linear example outlined gives a practical illustration of how that modelling approach can be used to explore similarity between current and previous problems. Again, the simplicity of that approach enables the capturing of dependencies. Minimum information models can be constructed along similar lines but without the same computational ease.

The three methods described here to do not directly address the point about future decisions. However, by providing models with a low number of parameters, they provide the necessary tools to explore the impact of gross future changes, which can then be approximately modelled by changes in mean, variance and other parameters.

There are now very sophisticated modelling and computational tools available to support decision-making. The challenge to modellers is to use these tools at an appropriate level of accuracy. It is worth remembering that the biggest uncertainty in numerical model outputs is often in the early significant figures, not in the later ones, and is due to incorrect structural assumptions. In other words, the tools and methods used have a tendency to direct the focus on deriving an unhelpful degree of unjustifiable accuracy.

The antidote to this is the use of simple, yet sophisticated, exploratory models that can take the limited information available from domain experts and provide a range of different sensitivity outputs.

Notes

1. As a simple example, a Type II Bayesian might really want to do full Bayesian updating on a particular variable of interest, but recognize that since he has a lot of data which he is willing to view as exchangeable, his Bayesian updating will give a very concentrated posterior distribution. Pragmatically, he might then decide to use a maximum likelihood estimator which will give a point estimate of the parameter and will be – except if he is very unlucky indeed – in his narrow Bayesian confidence interval anyway.
2. By analogy data, we mean data from systems that are similar but not exactly the same. Hence there may be systematic differences between the analogy systems and the systems under study – indeed this may be the point of developing the new systems. In Systems Engineering the expression "heritage data" is sometimes used in the same way.
3. That is, the prior that the analyst had in mind to model the information coming from the decision-maker/experts – often this is chosen on technical grounds to make calculations easier, for example, a so-called "conjugate" prior (O'Hagan, 1994). By taking a mixture we mean that we specify a small quantity $\varepsilon > 0$ and then use the density $(1 - \varepsilon)f_{\text{baseline}}(x)+\varepsilon f(x)$ where f_{baseline} and f are the mixing densities.
4. If f_{baseline} and f have the same quantiles, then the mixture $(1-\varepsilon)f_{\text{baseline}}(x)+\varepsilon f(x)$ has the same quantiles.
5. Technically, on a Borel sigma-algebra. It is worth remarking that the usual subjectivist constructions of probability only produce finitely additive probabilities, rather than the sigma-additive probabilities that mathematicians use. This is mathematically convenient but provides a lot of extra detail that is largely model construct.
6. Their other main criticism, is the inconsistency between Bayesian updating and maximum entropy constraints. However, if as we believe, the max entropy method is used to determine a prior, then there is no reason why it should be consistent with Bayesian updating. Selecting the prior is not something Bayes' Theorem can help us with. Similarly there is no good reason for trying to do updating on exchangeable data using the maximum entropy principle.
7. Perhaps this is one of the reasons why (Lindley, 2000) approvingly quoted a colleague (while acknowledging a little exaggeration) as saying, "There are no problems left in statistics except the assessment of probability."
8. Having said this, it should also be acknowledged that the ignorance *bias* is a significant problem in expert elicitation. Therefore ignorance is an elicitation problem, but it is not a statistical problem. Fox and Clemen (2005) offer an interesting discussion on the way in which the presentation of exclusive and exhaustive events during elicitation is biased by a tendency to rate every category as equally likely.
9. They define sensitivity analysis as the computation of the effect of changes in input variables or assumptions (including boundaries and model functional form) on the outputs. Uncertainty analysis is defined by them as the computation of the total uncertainty induced in the output by quantified uncertainty in the inputs and models, and the attributes of the relative importance of the inputs uncertainties in terms of their contributions.

References

Adams J. (1995). *Risk*. UCL Press: London.

Ale B. (2005). Living with risk: a management question. *Reliability Engineering and System Safety* **90**:196–205.

Bedford T and Cooke R. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence* **32** (1–4): 245–268.

Bedford T and Cooke R. (2002). Vines – A new graphical model for dependent random variables. *Annals of Statistics* **30** (4): 1031–1068.

Bedford T and Meeuwissen A. (1997). Minimally informative distributions with given rank correlation for use in uncertainty analysis. *Journal of Statistics and Computer Simulation* **57**:143–174.

Bedford T et al. (2006). Expert elicitation for reliable system design. *Statistical Science* **21** (4): 428–450.

Bedford T et al. (2008). *Applying Bayes Linear Methods to Support Reliability Procurement Decisions*. IEEE: RAMS08, Las Vegas, Nevada.

Berger J et al. (2000). Bayesian robustness. In: Ríos Insua D and Ruggeri F (eds). *Robust Bayesian Analysis*. vol. 152. Springer: New York.

Clemen R. (1997). *Making Hard Decisions: An Introduction to Decision Analysis*. Duxbury Press: Belmont, CA.

Cooke R. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press: Oxford, UK.

Cooke R. (1997). Markov and entropy properties of tree- and vine-dependent variables. *Proceedings of the ASA Section on Bayesian Statistical Science*. American Statistical Association. Alexandria: VA, 166–175.

Fox C and Clemen R. (2005). Subjective probability assessment in decision analysis: partition dependence and bias toward the ignorance prior. *Management Science* **51** (9): 1417–1432.

French S. (1988). *Decision Theory: An Introduction to the Mathematics of Rationality*. Ellis Horwood: Chichester.

Goldstein M and Bedford T. (2007). The Bayes linear approach to inference and decision-making for a reliability programme. *Reliability Engineering and System Safety* **92** (10): 1344–1352.

Goldstein M and Wooff D. (2007). *Bayes Linear Statistics: Theory and Methods*. John Wiley and Sons Ltd.: Chichester, UK.

Good I. (1983). *Good Thinking. The Foundations of Probability and its Applications*. University of Minnesota Press: Mineapolis.

Jaynes E. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press: Cambridge.

Kass R and Wasserman L. (1998). The selection of prior distributions by formal rules (vol. 91, p. 1343, 1996). *Journal of the American Statistical Association* **93** (441): 412.

Khinchin A. (1957). *Mathematical Foundations of Information Theory*. Dover Publications Inc.: New York.

Kraan B and Bedford T. (2005). Probabilistic inversion of expert judgments in the quantification of model uncertainty. *Management Science* **51** (6): 995–1006.

Kurowicka D and Cooke R. (2002). *The Vine Copula Method for Representing High Dimensional Dependent Distributions: Application to Continuous Belief Nets*. IEEE: San Diego, CA, USA.

Lanford O. (1973). Entropy and equilibrium states in classical statistical mechanics. *Statistical Mechanics and Mathematical Problems*. A. Lenard (ed.). Lecture Notes in Physics **20**. ISBN 978–3–540–06194–6. Springer-Verlag.

Lindley D. (2000). The philosophy of statistics. *Journal of the Royal Statistical Society Series D-the Statistician* **49**:293–319.

Meyer M and Booker J. (2001). *Eliciting and Analyzing Expert Judgment, A Practical Guide*. SIAM: Philadelphia.

Moreno E et al. (1996). Local robustness and influence for contamination classes of prior distributions Bayesian robustness. *IMS Lecture Notes Monograph Series* (Berger, J., Betro, B., Moreno, E., et al. IMS. **29**:137–154).

Morgan G and Henrion M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press: Cambridge, UK.

O'Hagan A et al. (2006). *Uncertain Judgements: Eliciting Expert Probabilities*. Wiley: Chichester, UK.

Revie M. (2008). *Evaluation of Bayes Linear Modelling to Support Reliability Assessment during Procurement Department of Management Science*. Unpublished PhD, Strathclyde University: Glasgow.

Sklar A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* **8**:229–231.

Walls L et al. (2006). Modeling to support reliability enhancement during product development with applications in the UK Aerospace Industry. *IEEE Transactions in Engineering Management* **53**:263–274.

15
Analyzing Information. Techniques and Analyses

Bjørn Andersen

This chapter presents and discusses a selection of analysis tools and techniques that can be used to make the most out of the information available in a project's early stages. The selection covers four types of tools, with increasing complexity and abilities: (a) pure data presentation aids, (b) methods for structuring information and data, (c) assessment or ranking tools and (d) proper analysis techniques. All of the approaches presented are well-established methods, either in the field of project management or in other fields or professions. Those that are rarer to find in a project setting should be especially helpful. The structure of the presentation divides the techniques into different phases with different analysis needs: (a) gaining a better understanding of the project's position and "reception" among relevant participants who will be affected by it, (b) developing a project concept idea further, (c) analyzing a project concept's strengths, weaknesses and other characteristics to better understand its chances of success and (d) comparing alternative project concepts and selecting one or more of these for inclusion in further work to refine the project concept.

Introduction

This book discusses the quandary every project faces: having to make a multitude of decisions, from when the earliest ideas appear, very often without much information on which to base these decisions. Several strategies can be pursued to increase the likelihood of making the right decisions, including obtaining more and/or better information, postponing decisions until information is available, involving more and better qualified people in the decision-making, etc. Another approach is to make the most out of the information available, and for this purpose there are many different tools and techniques at the project's disposal that have proven worth in clarifying issues or aiding the decision process. The purpose of this chapter is to present and discuss a selection of such analysis tools and techniques, always in the light of application in a project setting.

The selection presented spans a wide range of characteristics, although there is an established tradition in many fields of classifying everything as tools. These can range from a simple diagram to a fairly advanced analysis procedure. It is quite permissible to attach a common label to the entire spectrum, describing the full range as a *toolbox*.

The tools or techniques presented will typically be one of the following types:

1. **Pure data presentation aids**
   Charts or diagrams which portray data and information in different ways to help identify new meanings or patterns and thus produce a better basis for decision-making. An example is the simple spider chart.
2. **Methods for structuring information and data**
   Sorting or reshuffling the information to glean more insight from it, again for the purpose of aiding subsequent decision-making. SWOT analysis is a possible example of this group, although one could argue it also belongs among the analysis techniques.
3. **Assessment or ranking tools**
   Techniques that facilitate a more systematic evaluation and/or ranking of alternatives to choose from. Paired comparisons is a technique that falls within this group.
4. **Proper analysis techniques**
   A more extensive method of analyzing information and data that form the basis for pending decisions. Both stakeholder analysis and force field analysis are examples of more extensive analysis tools.

There are no doubt many ways to categorize analysis tools and techniques. The purpose here is to show that the approaches presented are of varying nature. The keen reader will undoubtedly have noticed that the four groups were presented in order of increasing complexity, as shown in Figure 15.1. Typically, increasing complexity means a larger effort is required when using the tool. The tools presented in this chapter are included in all four categories.

All tools presented in this chapter are well-known. Some are widespread enough to be familiar even to school pupils, others are much used in certain fields or professions, for example, the aforementioned force field analysis in change management situations. A number of them will already be familiar to project management professionals, as they are frequently applied/developed for project purposes. Some are much rarer to find in a project setting, but they can also work well.

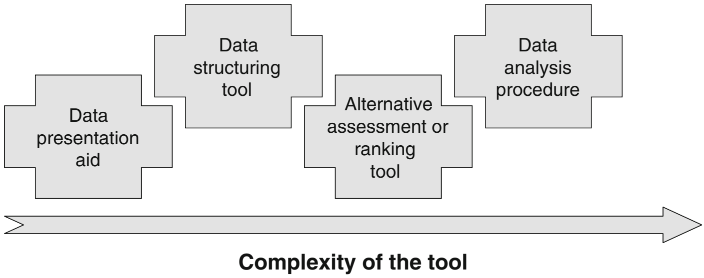


Figure 15.1 Spectrum of tools presented in the chapter

The tools and techniques and their application in a project

To facilitate the use of relevant tools for improving decision-making processes at different stages and situations in a project, it is necessary to explain where in a project each tool typically belongs. There are many different "lenses" that can be used to look at a project, the *time lens* being a favourite, which often produces a model of project phases.

Such phase models come in many different shapes and forms. A very basic approach divides a project into an early/front-end/planning phase and an execution phase. At the other extreme, a model can depict a two-digit number of stages, covering items like pre-engineering, detailed engineering, construction, hand-over, etc. The main purpose of such models is to give people a common language and a shared reference about the structure of the project. In this chapter, the purpose of the phase model is to show where the different tools are typically utilized throughout a project, as shown in Figure 15.2. Although this model deviates somewhat from other well-known models, it is more suitable for plotting the selection of tools.

The choice of which tools to include in this chapter was a difficult decision. Different schools and fields have developed an abundance of candidates along the full scale shown in Figure 15.1. There would be no problem at all filling an entire book with relevant tools, so narrowing down the field to fit inside one chapter is a challenge. Here, the final selection consists of eleven methods (any of which could be replaced by an alternative tool omitted here due to spatial confines). These are sorted to indicate the purpose of each tool in a project's early phase and in which phase of the project it is typically used – many of them can be used at several different stages. The tools are listed below, and Figure 15.2 illustrates at which stage of the front-end part of the project process they can be applied:

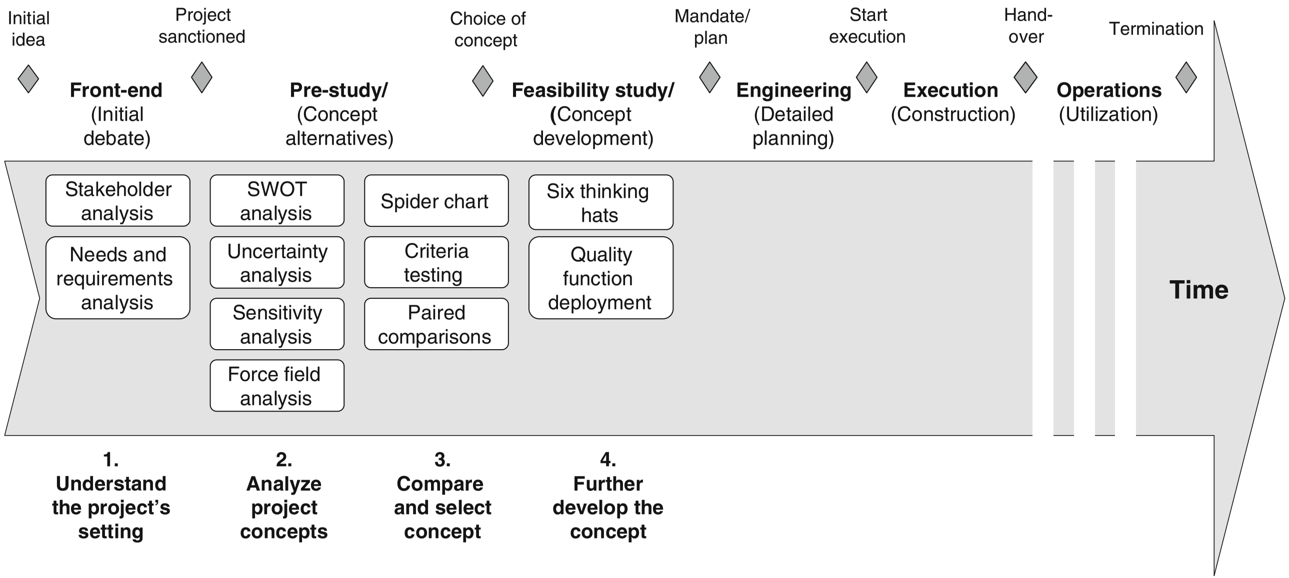


Figure 15.2 The project timeline with the different tools placed in logical phases of application

1. **Understand the project's setting**
   Gaining a better understanding of the project's position and "reception" among relevant participants who will be affected by it, predicting the effects of the project and different concepts for realizing the project's goals.

   - **Stakeholder analysis**
     Used to identify relevant organizations/individuals/entities that in some way influence or will be influenced by the project and gaining insight into their attitudes towards the project and likely behaviour.
   - **Needs and requirements analysis**
     Whose purpose is to understand in greater detail the position of the stakeholders with regard to what they expect to gain/receive from the project.

2. **Analyze project concepts**
   Analyzing a project concept's strengths, weaknesses and other characteristics to better understand its chances of success.

   - **SWOT analysis**
     Which, in this setting, can be used to understand which features of a concept are strong and which are weak, thus allowing further development of the concept to improve it.
   - **Uncertainty analysis**
     Used to identify uncertainties, both risks and opportunities that make an alternative project concept attractive.
   - **Sensitivity analysis**
     Whose main purpose is to demonstrate how costs, durations, benefits, etc. are sensitive to change, due to different external and internal factors.
   - **Force field analysis**
     A tool used to analyze the forces, both for and against, that implementation of a concept is likely to encounter, thus giving insight into difficulties that must be expected during project execution.

3. **Compare and select concepts**
   Comparing alternative project concepts and selecting one or more of these for inclusion in further work to refine the project concept.

   - **Spider chart**
     Used to portray information about the alternatives compared in a common diagram, to allow easier identification of gaps and differences.
   - **Criteria testing**
     A quantitative analysis of how different characteristics of the alternative project concepts influence the project goals.

- **Paired comparisons**
  An analysis tool that enables the selection of one alternative among a large number of candidates, requiring comparison of paired alternatives.

4. **Further develop the concept**
   Developing a project concept idea further, either to allow comparison with other concepts, or after the main concept has been chosen.

   - **Quality function deployment**
     Used to ensure that the requirements of the stakeholders are maintained in the concept development process and to analyze which aspects of the concept best address these requirements.
   - **Six thinking hats**
     Often termed lateral thinking, applied in the creative stages of the concept development process to ensure that ideas and concept specifications are viewed from several different angles.

Before presenting the individual tools, other dimensions characterizing the tools are worth examining. Some people prefer analysis techniques based on the use of quantitative data, others like to use more verbal information, some require more pre-existing data and information than others, etc. Understanding such requirements for each tool enables the selection of a suitable one for a specific situation. Several such dimensions could be utilized to describe the tools:

- Low vs. high requirements for volume of background data and information.
- Qualitative assessments vs. quantitative analysis.
- Background data and information in the form of estimates vs. facts.
- Analysis based on subjective assessments vs. objective calculations.
- Stochastic vs. deterministic analysis.
- Partial/local vs. holistic/total view taken when applying the tool.
- Use by individuals vs. best used by groups.
- Use of the tool by the project organization itself vs. requiring expert assistance.

A subjective analysis of the different tools' properties is summarized in Table 15.1. However, since all these tools can be applied in group assessments and internally by the project organization, the latter two dimensions are not included in the table but rather mentioned in the chapters below.

In the chapters below, each of the tools and techniques is presented in more detail. For each of them, the description contains a short summary

Table 15.1 Analysis of the tools along the dimensions chosen to characterize such tools.

Tools vs. dimension	**Volume of background information**	**Qualitative vs. quantitative**	**Estimates vs. facts**	**Subjective vs. objective assessments**	**Stochastic vs. deterministic**	**Local vs. total view**
Stakeholder analysis	Low	Qualitative	Estimates	Subjective	Deterministic	Total
Needs and requirements analysis	High	Both	Both	Subjective	Deterministic	Both
Quality function deployment	Medium	Both	Both	Both	Deterministic	Both
Six thinking hats	Low	Qualitative	Estimates	Subjective	Deterministic	Both
SWOT analysis	Medium	Qualitative	Both	Subjective	Deterministic	Both
Uncertainty analysis	Medium	Quantitative	Both	Both	Stochastic	Total
Sensitivity analysis	High	Quantitative	Both	Both	Stochastic	Both
Force field analysis	Medium	Qualitative	Estimates	Subjective	Deterministic	Both
Spider chart	Medium	Quantitative	Both	Both	Deterministic	Both
Criteria testing	Low	Both	Estimates	Subjective	Deterministic	Both
Paired comparisons	Medium	Both	Estimates	Subjective	Deterministic	Both

of what the tool is about, an explanation of its application in an early phase project setting, an outline of the steps involved in applying the tool and a discussion of the nature of the tool, as briefly depicted in the table above.

Understand the project's setting

Stakeholder analysis

Any project will in some way have an impact on its surroundings, otherwise there would not be much point in carrying out the project in the first place. Projects aim to achieve effects on different levels, but they all involve some changes or influences that will affect the environment of the project. The environment may refer to physical entities, but, more importantly, consists of local organizations and individuals. Every single project is "surrounded" by such entities that directly or indirectly participate in or influence the design, execution and effects of the project. These are commonly termed stakeholders, defined by PMI as:

> Individuals and organizations who are actively involved in the project, or whose interests may be positively or negatively affected as a result of project execution or successful project completion.
>
> (PMI, 2000)

Some typical stakeholders of a project are shown in Figure 15.3. Obviously, some of these are more central to the project than others, and some appear throughout the duration of the project, others only in certain phases.

Empirical findings have shown that stakeholders often play a significant role in the design and execution of projects. Analysis of a large number of projects has also demonstrated that stakeholders create as many difficulties for a project as they represent support for it. Surveys of projects that have experienced problems often show how, in hindsight, too little effort was invested in understanding the peripheries of the project and its stakeholders, and little consideration was given to the development of mitigating strategies to handle the stakeholders. Stakeholders are an important "unit of analysis" in a project's early phases, and stakeholder analysis a useful tool for any project.

Simply creating an awareness, within the project and its promoters, of the stakeholders, potential and actual, that it could expect to deal with can be an important eye-opener. Delving deeper into the matter and mapping the needs and expectations of the various stakeholders provides additional useful insight. Throughout the project, from its very first idea, through to completion, this knowledge can prove important:

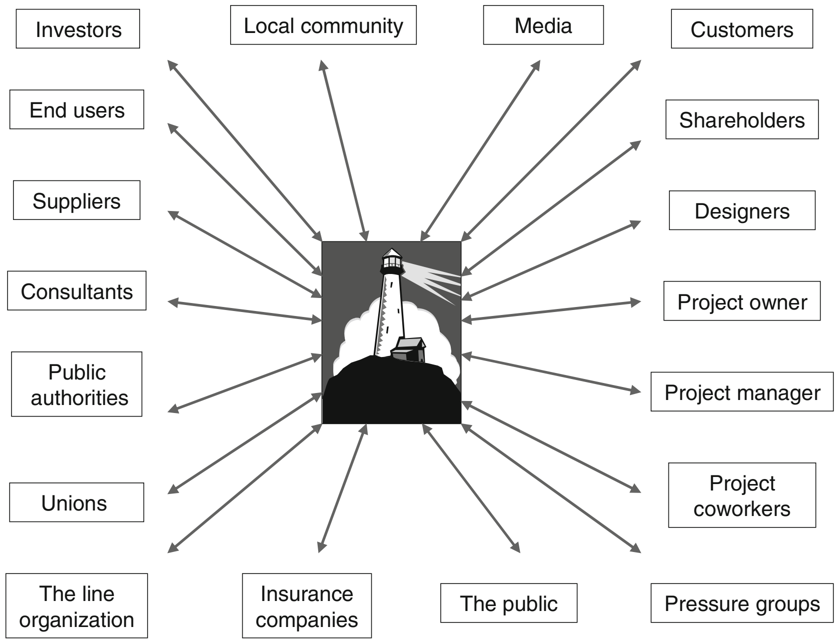


Figure 15.3 Project stakeholders

- Firstly, it allows the promoters to understand the climate for the project and its main goals. Should the stakeholder environment already prove overwhelmingly negative at this stage, the best solution might be to close the project right away or at least postpone it.
- If the decision is to initiate discussions about choice of concept and design, an understanding of the expectations held by the various stakeholders regarding the outcomes of the project is valuable. This knowledge can be used to directly influence the design of alternative concepts in order to maximize the likelihood of stakeholder acceptance.
- When evaluating different competing alternative concepts, stakeholder expectations can be used as evaluation criteria, at least as a subset of the overall criteria employed. A concept that scores well for many important stakeholders, and does not offend the expectations of others, should be a good choice.
- In the engineering phase, when work is aimed at developing detailed solutions for the chosen concept, the understanding of the stakeholder requirements can help decide on specific designs.

- Throughout the execution phase, numerous decisions can be eased by stakeholder analysis understanding, e.g. choice of contractors, remuneration policies, communication with the outside, etc.
- Conducting a stakeholder analysis is a fairly simple tool to complete. In practice, most stakeholder analyses follow steps along the following lines:

    1. Identify the project's existing and presumed future stakeholders, normally through some variant of brainstorming within the project team or its promoters.
    2. Classify the identified stakeholders according to some criteria to allow sorting them into more and less important ones. Typical criteria may be importance for the project, the ability to influence the project and the likelihood of the stakeholders supporting the project.
    3. Delve deeper into each stakeholder, or only the most critical ones if time and resources are limited, to gain a better understanding of their needs and expectations of the project and its outcomes (Andersen and Fagerhaug, 2002).

Depending on when in the project the stakeholder analysis is carried out and for what purpose, the understanding gleaned from steps 1–3 might be sufficient. In other cases, the analysis is continued by using the insight gained to evaluate alternative concepts and develop detailed design solutions, etc. There seems to be no better way of identifying stakeholders than by brainstorming, preferably involving participants from different parts of the project organization or its promoters. Such an exercise will usually result in a rather extensive list of stakeholders, some of whom will be less important to the project than others. To limit the extent of the analysis, and avoid wasting efforts on marginal stakeholders, it is helpful to determine which are the few stakeholders worth analyzing in more depth. Again, there are many ways to sort the stakeholders for this purpose, but a commonly used approach builds on the matrix in Figure 15.4.

These labels do not of course represent exact predictions of behaviour for each group but indicate from experience how they might act. Suitable strategies may be recommended for dealing with each group:

- Supportive: should be involved in relevant discussions and decisions.
- Marginal: should simply be monitored.
- Non-supportive: defensive strategy should be used to minimize the dependency on the stakeholder.
- Mixed blessing: best handled through cooperation.

The stakeholders who fall within the two left-hand fields will normally represent those whose acceptance of decisions and choices made within the

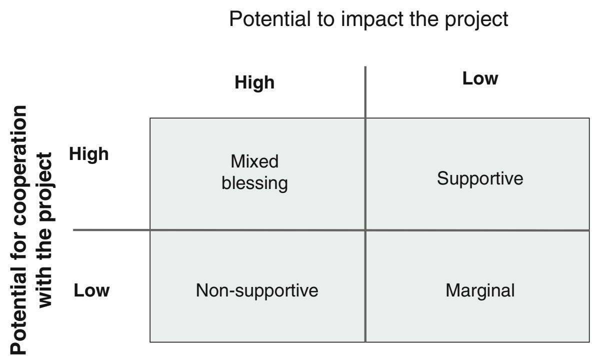


Figure 15.4 Matrix for classifying stakeholders

project is important to elicit. It is often sufficient to include only these in subsequent steps.

It should be apparent from the above that stakeholder analysis is exclusively a qualitative approach that relies on subjective assumptions and conjecture by the participants. As such, it might be easy to dismiss the analysis as random or haphazard. However, bear in mind that the power of such analysis lies in the fact that a group of people, with different fields of expertise and knowledge of areas and details of the project, pool their insights to make the best possible analysis. One person conducting a stakeholder analysis would very likely end up with rather different findings than when the analysis is performed by such a group. Also, being qualitative, the analysis does not require a high level of precision from the individual assessments made. The purpose of the analysis is to understand which stakeholders the project needs to accommodate and what they typically expect from the project. Experience shows that this is normally achieved without any problems.

Needs and requirements analysis

It is debatable where the stakeholder analysis ends, and the needs and requirements analysis starts; one could easily define stakeholder analysis as a complete assessment, encompassing an in-depth evaluation of stakeholder expectations. For practical purposes, it has become common to separate the two, as a stakeholder analysis can be highly worthwhile even when it only achieves a sorting of the different stakeholders.

There are also many ways to study expectations held by the stakeholders, from conducting discussions with individuals in a survey to extracting information from similar previous projects. The organization probably already possesses a fairly high level of knowledge about some stakeholders' requirements, for instance owners or customers. For others, it might be difficult to know exactly what they do or, perhaps, do not want. Expectations of pressure groups advocating equal opportunities, the media in general or public authorities, are probably much less clearly understood, and hidden agendas may even exist. Trying to map all of the probable and improbable expectations harboured by different stakeholders will undoubtedly reveal that these are plentiful, not always coherent but very wide-ranging.

The Kano model is a very useful diagram to bring some order to these expectations and differentiate between the important and less important ones (Kano et al., 1984). Figure 15.5 shows that it is essentially nothing more than an awareness-creating diagram showing that there are different types and levels of stakeholder requirements.

The straight diagonal line of the diagram portrays the clearly expressed requirements of the stakeholder. Generally, these are the only demands the stakeholder will make if asked about his or her desires. If the stakeholder is a major shareholder of the organization pursuing the project, he could, for instance, express requirements that the return on investment should be a minimum of 7.2%, that he be granted a seat on the project board, etc. In addition, there exists a set of requirements that are so basic that they are not even expressed, as indicated by the lower curve. For the shareholder, these could be that the organization does not go broke and lose its capital, that it does not get involved in criminal or other unethical activities that could

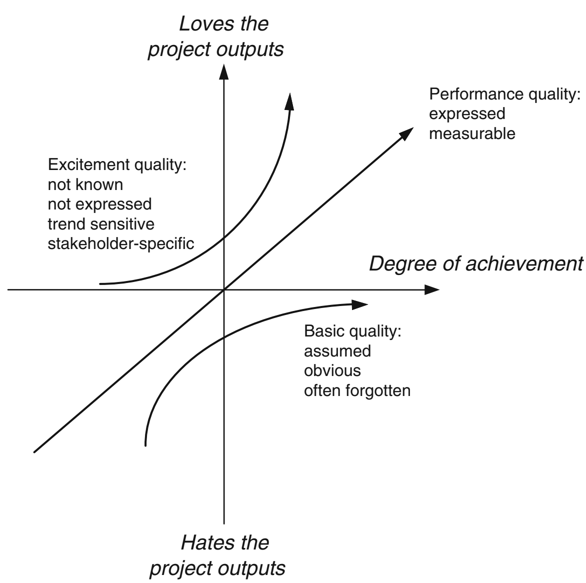


Figure 15.5 The Kano model and the three types of stakeholder requirements

harm the shareholders and that business is conducted according to general rules and customs, etc.

Together, these two requirement sets constitute a complete set of demands imposed by the stakeholder towards the organization. The satisfaction depends on how well both sets of requirements have been met. It will be of no help if the investment returns 10%, but the shareholder is arrested for his involvement with the company due to its engagement in criminal activities. In other words, satisfying expressed requirements cannot rectify shortcomings in the basic demands. On the other hand, satisfying every single one of the basic requirements will not lead to complete satisfaction, unless the expressed requirements have also been fulfilled. This will, at best, eliminate dissatisfaction. The danger is that the stakeholder takes it for granted that the organization is aware of the basic requirements, while this might not be the case. Such silent assumptions are one of the main focuses when clarifying requirements in the stakeholder analysis.

If these two sets of requirements are defined and satisfied, the foundation for satisfaction should be firmly established. To further enhance satisfaction, and even create delight for the stakeholder, we can look at the third set of requirements. "Requirements" is not really the correct word, as these conditions are not expressed by the stakeholder: often the stakeholder himself is not even aware of these needs. For the shareholder, this could include the organization making all arrangements for transportation to project board meetings, free access to the organization's products or services, a special website for the shareholders which presents updated information of interest to them, etc. If both the basic and expressed requirements have been satisfied, the fulfilment of such extra "requirements" can create true delight. These are often the little extras required to ensure loyalty and access to the best stakeholders. However, it should be noted that where such extra requirements are delivered on one or more occasions, they often become expressed or even basic requirements that must be fulfilled to avoid dissatisfaction.

In a project's early phases, this analysis is usually combined with the stakeholder analysis. The purpose is to understand better what criteria different stakeholders will use when considering their possible participation or sponsorship, as well as the outcomes delivered by the project. This information can be used for many purposes:

- The expectations identified can be employed as evaluation criteria when comparing alternative concepts and how they will be received by the different stakeholders, thus providing a decision basis for the choice of concept.
- The expectations can also be considered as requirements for the further development of the chosen concept and subsequent detailed design solutions.

- When making design decisions, the requirements can be used to determine which direction to follow to maximize stakeholder satisfaction.

Conducting the analysis really consists of nothing more than applying the Kano model systematically to all or the most important stakeholders. Information about the expectations of each can be based on the collective knowledge of the project team or, as mentioned earlier, the stakeholders can be asked directly. This will counter the obvious source of error that exists in speculating internally about what the stakeholders think of the project.

Although apparently based on graphs in a coordinate system, the Kano-based needs and requirements analysis is primarily a qualitative undertaking. However, some requirements can naturally be expressed in quantitative terms, although the figures are not used for any calculations. Depending on how insight into the stakeholder requirements is developed, the analysis can be based on both pure subjective estimates or objective statements from the stakeholders themselves. This analysis is a typical group exercise, the background information and data required being extensive. It should also be mentioned that it might be of limited use, as the needs and requirements identified may be less than conclusive, pointing in many different directions in terms of how the project should be designed.

Analyze a concept's strengths, weaknesses and other characteristics

SWOT analysis

SWOT analysis is probably one of the best known simple strategic analysis techniques. The name itself is an acronym representing the four analysis perspectives addressed by the tool: *strengths, weaknesses, opportunities* and *threats*. The analysis originated as a strategic planning aid but has been applied in many other contexts, including project management (Lewis, 1999). Identifying elements and factors internally and externally in the project and its periphery within these four perspectives, the purpose of the analysis is simply to create awareness of forces that will impact the project in the future. By understanding these forces and making them known throughout the project organization, better strategic decisions can be made, and the whole organization will be better prepared for future developments in the project.

In a project setting, the SWOT analysis can be used in much the same way as the stakeholder analysis, i.e. gaining a general understanding of the forces facing the project. However, such a general SWOT analysis can become insignificant, in that the assessments are so high-level that they give little specific insight. Rather than performing the analysis for the overall project, a more meaningful analysis can be obtained by identifying strengths, weaknesses, opportunities and threats for each alternative project concept contending for selection. This will allow the project team to identify certain threats facing the various contenders which, if eliminated or properly mitigated, can render them attractive alternative concepts. Similarly, certain strengths that contribute significantly to a concept's favourable ranking must be safeguarded if the concept is selected. The SWOT analysis on its own does not produce direct conclusions about which alternative should be chosen, but it helps to increase the insight into each alternative, and how it is most likely to develop in the future, given the forces it faces.

Conducting a SWOT analysis is technically very simple. It mostly entails brainstorming within the four perspectives, preferably in a group representing different disciplines and areas of the project organization. Although "technically simple", performing a good SWOT analysis can be difficult, especially in terms of managing to come up with those elements that are really relevant. This takes training. The following steps should be taken:

1. Compose a team to undertake the analysis, drawing on different types of competence within the project organization, possibly supplemented by external representatives.
2. Within each of the four analysis perspectives, and for each alternative project concept being analyzed, brainstorm any issues that seem relevant.
3. Compile the analysis results, using a simple table or matrix.
4. Discuss which of the issues identified is believed to have the strongest influence on each of the project concepts.

The SWOT analysis is quite a simple analysis tool, based primarily on qualitative information (although quantitative data can form a basis for identifying the different forces). It utilizes both estimates and facts to make subjective assessments within a group and can be applied both to the total project or locally for different concept alternatives.

Uncertainty analysis

Risk analysis/uncertainty analysis, or, more correctly, uncertainty management, is in itself an extensive field covering both specific analysis approaches, psychological issues about risk and opportunities, principles for continuous uncertainty monitoring, etc. Of these, the basic uncertainty analysis approach is of most interest in a project's early stages. This consists of a simple approach for identifying uncertainty elements, assessing the likelihood of each occurring, indicating the consequences should they occur, developing strategies for avoiding the occurrence of negative elements (risks) and increasing the chances that positive elements (opportunities) can be favourably exploited.

From this description, uncertainty analysis can seem similar to SWOT analysis, as both focus on threats and opportunities. On the other hand, uncertainty analysis goes much further than SWOT analysis in quantifying both likelihoods of occurrence and consequences. It is therefore compatible with the SWOT analysis and furnishes the findings from the SWOT analysis with more detail. In the early stages of a project, an uncertainty analysis can be used to improve the understanding of the uncertainty encumbering each alternative concept decision. As with the other tools in this category, this analysis will provide greater insight into the alternative concepts before a choice is made. However neither analysis is solely sufficient to provide a complete picture.

Using the basic logic outlined above for each alternative concept in the running, the following must be done:

1. Identify uncertainty elements, of both a positive and negative character.
2. For each uncertainty element, assess the likelihood that the uncertainty element will occur.
3. For each uncertainty element, assess the consequences for the project should the element materialize.
4. Summarize the identified uncertainty elements in an uncertainty matrix.
5. For the most important uncertainty elements, consider whether actions can be implemented to alter the likelihood of occurrence or the consequences of the uncertainty elements (Husby et al., 1999).

Uncertainty element refers to any event or development that might occur and impact the project. Examples can be unforeseen events or conditions such as change of government, bankruptcy of a contractor or the discovery of clay in the ground of an excavation site. Further, frame conditions can evolve along a continuous scale, e.g. lower interest rates, higher prices in the construction sector or increased demand for the product or service the project will supply.

To indicate the likelihood of an uncertainty element occurring and the consequences should it occur, a scale from 1–5 is normally used. The Value 1 indicates very low likelihood of occurrence and very low consequences, and Value 5 that the element probably will occur and that the whole project may be in danger, while the values in between indicate a continuous scale between these extremes. Having assigned these two assessments to each uncertainty element, the elements are plotted in an uncertainty matrix. Depending on the location in the matrix, the elements are ranked as shown in Figure 15.6. Elements located in the red area of the upper right part of the matrix should be given most weight when comparing concept alternatives.

If uncertainty analysis is used during the concept selection process, an added benefit is that the results from this analysis can be used during the later stages of the project. The findings will both influence cost estimation and will also indicate areas where management attention must be directed to avoid problems.

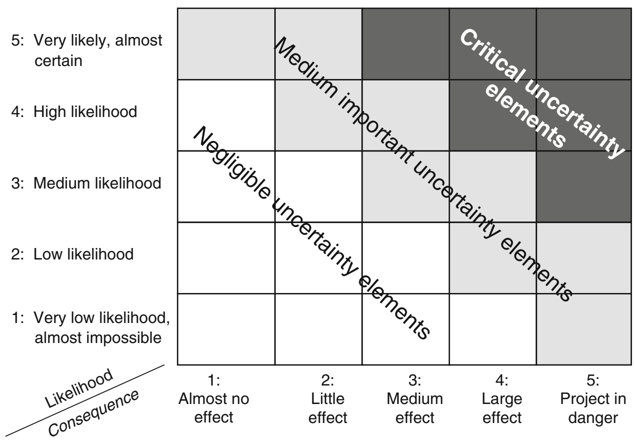


Figure 15.6 Principal uncertainty matrix

Uncertainty analysis builds on qualitative data but uses quantitative factors to undertake the analysis. To some extent, subjective facts are used to characterize the uncertainty elements, but normally a fair amount of subjective conjecture goes into the analysis. Uncertainty analysis is most powerful when performed by a group. Being a somewhat tricky analysis to undertake inside a tight group where the members know each other well, it can be beneficial to include an external facilitator.

Sensitivity analysis

When comparing different concept or design solution alternatives in general, or using one or more of the different tools presented here, a common approach is to assume that different figures are known and fixed and thus directly comparable. Unfortunately this is often too simple an assumption. Even with minor changes in the premises on which the figures were based, conditions like costs, durations, amounts, etc. can change significantly. Sensitivity analysis is a tool to assess how sensitive different estimates are regarding changes in the premises they build upon. In this respect, the tool is closely linked to uncertainty analysis, whose purpose is to identify factors in the project that seem likely to change or occur and thus influence the project.

For application in the early stages of a project, the sensitivity analysis can therefore be combined with an uncertainty analysis. The purpose will be to supplement the comparison of assumed fixed figures for different decision alternatives with an analysis of the sensitivity of these figures. In cases of big differences in sensitivity, this knowledge will provide further insight required to prioritize, e.g. where the most favourable concept is also the one with most volatility attached to key estimates. In such a case, the best approach could be to choose a less favourably ranked alternative, which is less sensitive to changes in conditions affecting the estimates.

When conducting a sensitivity analysis, there are at least two principally different approaches. The easiest way is to allow only one parameter to change at a time, but the richest insight comes from portraying changes of several parameters in the same analysis, a so-called multivariable analysis. The steps to perform such an analysis are:

1. Determine for which parameter of the project decision alternatives you want to analyze the sensitivity (the dependent variable), e.g. costs, amount of steel, time, emissions, etc. If more than one parameter is to be included, this must be done by repeating the analysis one or more times.
2. Decide which underlying factors (the independent variables) to include in the analysis, e.g. interest rates, currency exchange rates, construction standards, etc.
3. Calculate the changes in the dependent variable as a consequence of the changes in the independent variables.
4. Portray the results in a sensitivity diagram, where curves show how changes in the independent variables influence the dependent ones, as shown in Figure 15.7.
5. Having performed the analysis on several factors for several decision alternatives, the sensitivity of each alternative is compared to determine which seems most resilient to changes in conditions inside and outside the project.

Being a purely quantitative analysis tool and potentially involving a large number of calculations, the requirements for available background data can be large for a sensitivity analysis. As far as possible, the relationships between the dependent and independent variables should be established as precise, objective functions. Where this is not possible, subjective discretion must be used to determine these links. This is a tool that can be used by an individual, as opposed to a group, although underlying data must often be gathered from many different persons and sources.

Force field analysis

Force field analysis is based on the assumption that any situation is a result of forces for and against the current state of equilibrium. An increase or decrease in the strength of some forces will induce change, a fact which can be used positively. The analysis is simply an assessment of the forces working for and against a certain situation, for the purpose of understanding the challenges likely to be faced when implementing a certain concept.

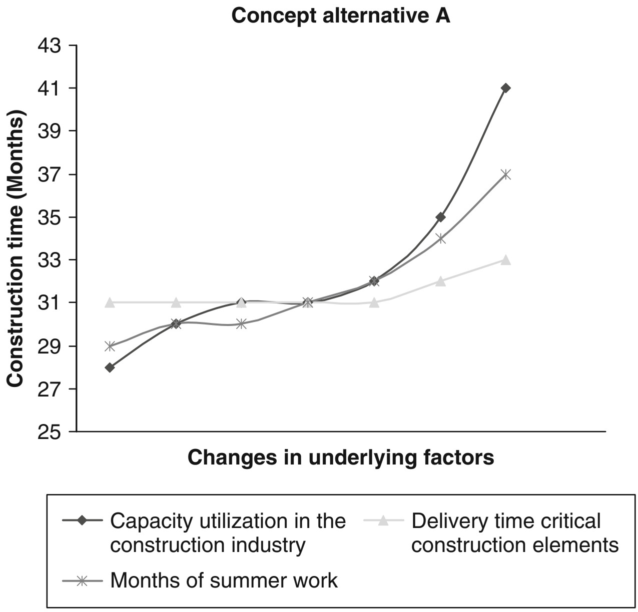


Figure 15.7 Sample sensitivity diagram

Originally developed for change management purposes, a force field analysis can be useful in a project's early phases. Where different alternative concepts are being considered, some of them quite different in nature, applying this tool can arm the project with important insight. Understanding how the different concepts are likely to be received and which forces will have to be countered to implement them, as well as which forces will support the implementation, provides one more piece of information to help facilitate the choice of concept.

The procedure for using force field analysis is:

1. Define the concept to be analyzed and the changes it will mean for stakeholders and the project periphery.
2. Brainstorm all possible forces inside and outside the project that could be expected to work for or against the implementation of the concept.
3. Assess the strength of each of the forces and place them in a force field diagram (a principal force field diagram is shown in Figure 15.8). The length of each arrow expresses the strength of the force it represents.
4. For each force, especially the stronger ones, consider actions that could increase the forces for the change and reduce those against it.

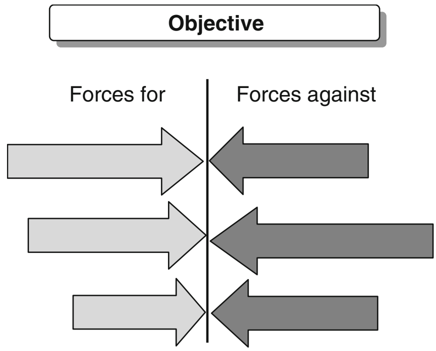


Figure 15.8 Force field diagram

Although the strength of the forces is quasi-quantitative, the analysis is primarily a qualitative one. The basis for the analysis is subjective assessment of the forces working for or against a concept, and the analysis should be carried out in a group setting.

Compare and select among alternative project concepts

Spider chart

Much of the information gathered and compiled about alternative concepts is either purely numerical or can be portrayed numerically. Comparing such data by the use of lists, tables or matrices is of course a relevant and useful approach. On the other hand, it is a well-known fact that graphical portrayal of data eases their interpretation and makes patterns stand out more clearly. The spider chart is an analysis tool offering additional capabilities for such graphical representation of comparison data. Figure 15.9 shows a principal spider chart. Each spoke in the chart represents one evaluation criterion for the project concepts, and each concept's "performance" for the criteria is indicated by the profile curves.

The usage in a project setting should be obvious. Two or more competing project concepts, or other decision alternatives, can be plotted in the chart and thus easily compared. This simple data portrayal tool can be used at any point in a project where alternatives must be compared and one selected.

To construct a spider chart, follow this procedure:

1. Determine how many evaluation criteria to include in the chart and assign one variable to each spoke.
2. Divide each spoke into logical segments by using a separate unit of measurement for each evaluation criterion. The further from the centre of the chart, the higher the performance.

3. Plot the performance data for each variable along the right spokes, using different colours or symbols to separate data points from those of different organizations.
4. Draw lines between the data points for each organization to generate performance profiles.
5. Examine the resulting chart to identify the concept that seems to score best overall according to the evaluation criteria.

The spider chart is a quantitative data presentation tool that relies on both estimates and facts. Where objective figures are unavailable, subjective assessments must be used.

Criteria testing

Criteria testing falls under the umbrella term of multiple goal analysis. Multiple goal analysis is not a singular technique but rather a common term for different approaches, the main purpose being to evaluate the impact of different decision alternatives on a set of several objectives. Such an analysis

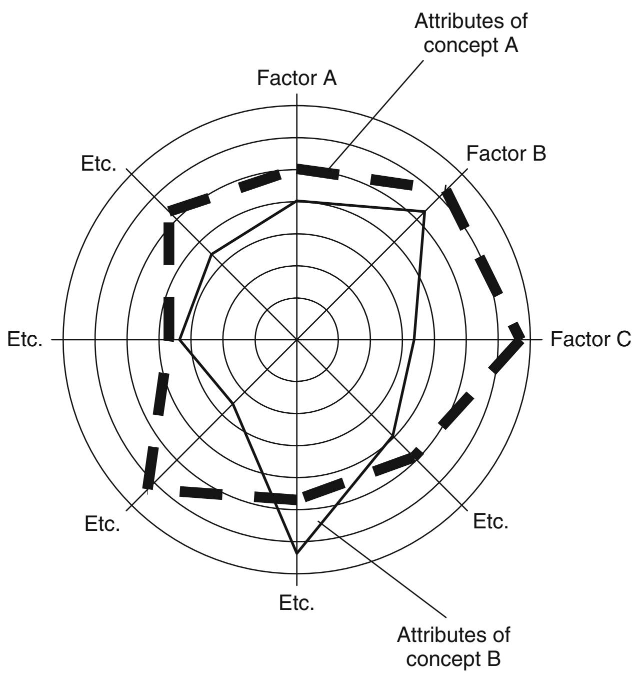


Figure 15.9 Spider chart

Project concepts	Objective	1	2	3	4	5	Total score
	Weight	3	1	1	3	2	
Concept 1		3	1	2	9	4	19
Concept 2		9	3	1	3	2	18
Concept 3		9	2	3	6	6	26
.							
.							
.							
.							
.							
Concept *n*		3	2	2	3	6	16

Figure 15.10 Sample criteria testing matrix

can be done in many different ways: graphical depictions using arrows of different thickness, qualitative influence assessments or numerical analysis based on weight and impact factors. Criteria testing builds on the latter principle, where the decision alternative that scores best in achieving the defined objectives can be identified through simple calculations. A simplified example of a matrix constructed during a criteria testing exercise is shown in Figure 15.10.

Criteria testing is thus another way to look at and compare competing concept alternatives, supplementing the insight gained by using a spider chart. The application of the spider chart in a project is useful when alternatives need to be compared to determine which one best serves the defined set of objectives.

The procedure for using criteria testing is as follows:

1. Place the set of objectives pursued by the project, typically three to five, in the upper field of the matrix. If desired, assign each of these a different weight factor that expresses relative importance. In Figure 15.10, weight factors from 1 to 3 are used, but other numeric values can be employed, e.g. a scale from 1 to 10.
2. Next, in the left-hand field of the matrix, place the competing concept/decision alternatives being considered.
3. For each decision alternative, assess its impact on each of the objectives. The example uses impact factors from 1 to 3, where 1 means low impact and 3 high impact. Again, a scale from 1 to 10 or a discontinuous set of factors, e.g. 1, 3 and 9, can be used. A low number means a low contribution to the objective.
4. Multiply the impact factor by the weight factor of the objective, and place the product in the correct matrix cell.
5. For each project concept, these products are summarized horizontally and the total sum placed in the right-hand column of the matrix. This numeric value indicates the collective impact of the concept on the

complete set of objectives. The higher the score, the better reason to believe that a concept will contribute to fulfilling the project goals.

The core calculation approach might be recognized from the quality function deployment. Criteria testing is useful in providing a comprehensive analysis of all the concepts' impact on all the objectives, without suffering too much from errors or inaccuracies in individual assessments. The tool is primarily a quantitative one, although the numbers used represent subjective assessments of influences between concepts and objectives. A word of warning about such quasi-quantitative techniques is that the numbers, once compiled, tend to be taken very much at face value and as absolute truth. Being the products of qualitative assessments, they are of course prone to some error and should be treated accordingly, more to indicate a direction than to dictate a conclusion.

Paired comparisons

A common problem for individuals or teams trying to choose one among a number of alternatives is that the sheer complexity of the choice can be debilitating. Relating to all the alternatives at once and trying to rank them can be extremely difficult, and the results might be determined by coincidence. Paired comparisons aim at prioritization and consensus reaching and do so through a sequence of paired comparisons. Single decisions are easier to make than selecting from a large number of possible solutions, which makes this a powerful approach.

In a project, this represents another technique for choosing among different concept/decision alternatives. These different selection tools can be thought of as lenses (see Section 2, "The Tools and Techniques…") Each looks at the alternatives in a different manner, and using two or more of them allows for a more qualified final decision.

The steps in paired comparisons are as follows:

1. Clearly identify the alternatives to be compared. The total number, denoted N, should be manageable, i.e. not more than eight.
2. Make a matrix with the alternatives, coded by letters, as row headings and the pairs as column headings, indicated by letters only, to save space. The number of pairs, P, is determined by the following formula: $p = [N \times (N - l)]/2$. A generic example of such a matrix is shown in Figure 15.11.
3. Column by column, each participant votes for one of the alternatives; the votes are logged in the matrix.
4. After participants have voted for all pairs, add the total number for each pair; this should equal the number of participants.
5. Add the number of votes given for each alternative to give the row totals. The highest-scoring alternative is the preferred one according to the group.

	A/B	A/C	A/D	A/H	B/C	B/D	B/H	C/D	C/H	D/H	Total
A	6	2	5	3							16
B	1				0	4	5				10
C			2			3		1		4	10
D		5			7			6	5		23
H				4			2		2	3	11
Number of votes	7	7	7	7	7	7	7	7	7	7	

Figure 15.11 Generic matrix for paired comparisons

As with criteria testing, paired comparisons use numbers to determine a conclusion, but the numbers are votes, which are of course only subjective assessments of individuals. The technique is best used by a group but can also be used by a single person.

Further develop the project concept

Quality function deployment

The Quality Function Deployment tool, QFD, was developed to represent a customer-oriented approach to product development. For this usage, it is a methodology for structuring customer needs, expectations and requirements and translating these into detailed product and process specifications. The principles can, however, also be used for a number of other purposes, including developing project concepts. Before showing how QFD can be applied this way, the basic theory behind the tool is presented. Firstly, the product development process consists of several sequential phases (Akao, 1990):

- Transforming customer requirements into a product concept
- Transforming the product concept into a product design
- Transforming the product design into a process design
- Transforming the process design into production documentation.

Each step of this process must adhere to the original customer requirements. The basic structure of QFD is a relational matrix at different stages of the process, as shown in Figure 15.12.

WHAT constitutes the goals of the analysis, which, in the first phase of product development, are the customer requirements and expectations. HOW expresses the means to reach these goals, the product development's first phase being technical product concepts. In the next phase, these will form the WHATs, and HOW will represent detailed design solutions for the product concept. If a weight factor for each element of WHAT is multiplied

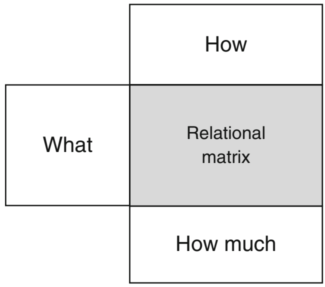


Figure 15.12 The basic structure of QFD

by a grade indicating how well each element of HOW contributes to satisfying the requirement, an indicator for the performance of each HOW element is generated, which is placed in the field HOW MUCH. This way, each phase of the product development process is linked together as a chain of relational matrices, to ensure that the voice of the customer is transmitted throughout the entire process (see Figure 15.13).

Additional information can be added, creating the chart popularly known as the "house of quality" (see Figure 15.14).

The QFD process is conducted by entering data into each room of the house of quality. WHAT represents the external requirements: in product development, the customer's product requirements. To each element in WHAT is attached a weight factor expressing the element's importance, which renders it possible to emphasize some requirements more strongly than others. After determining how the customer requirements can be fulfilled (HOW) the relational matrix linking WHAT and HOW is completed. To make the matrix as clear as possible, it is usually preferable to use as few

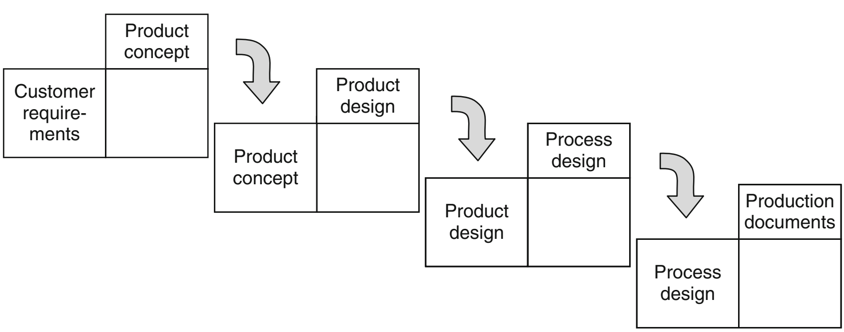


Figure 15.13 A chain of QFD charts

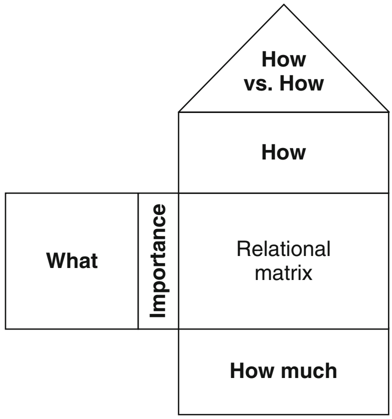


Figure 15.14 The house of quality

Relation	Symbol	Weight
Weak	△	1
Medium	○	3
Strong	◉	9

Figure 15.15 Symbols for the relational matrix

types of relations as possible. A set of commonly used symbols is shown in Figure 15.15. In the same manner, the roof of the house of quality forms a relational matrix to be used for investigating whether there are any relationships among the different elements of HOW. In this matrix, it is possible to indicate both positive and negative relations, i.e. factors that work together or that create trade-offs or conflicts. Some common symbols for this matrix are shown in Figure 15.16. For each HOW, the weight for the determined relationship to the individual element of WHAT is multiplied by the corresponding factor in importance of each requirement element. All products

Relationship	Symbol
Strong positive	◉
Weak positive	○
Weak negative	×
Strong negative	⋇

Figure 15.16 Symbols for the roof matrix

are summarized and placed in the lower field of the chart, HOW MUCH. Elements of HOW with a high score in this field should be preferred to others, if all elements cannot be incorporated into the product.

In a project setting, the usage of QFD is similar to its original purpose. Developing early concept alternatives, detailing the chosen one further and making detailed designs are all quite similar in nature to product development and can be seen as variants of the stages depicted in Figure 15.13. As such, the use is exactly the same: understanding how best to convert the stakeholder requirements into a project concept and detailed design that best satisfy the stakeholders.

The steps when applying QFD in a project are as follows:

1. Identify the stakeholder requirements (usually already done in the stakeholder and needs and requirements analysis) and place these in the WHAT field of the house of quality.
2. Assign importance factors to each of these requirements.
3. Identify the possible design elements that can be part of the project concept or detailed design solutions, and place these in the HOW field.
4. For each HOW element, assess to which extent it impacts each of the WHAT elements. Indicate this by assigning an impact factor, normally 1, 3 or 9.
5. Multiply the impact factor with the importance factor for the WHAT element, and write the result in the appropriate place in the relational matrix at the heart of the house of quality.
6. If using the roof of the house of quality, go through each cell in the matrix at the intersection of two WHAT elements. Analyze whether the two elements influence each other, and if so, place the appropriate influence symbol in the cell.
7. For each HOW element, summarize the scores down the column of the relational matrix and write the sum at the bottom in the ABSOLUTE IMPORTANCE row.
8. Calculate the percentage of the total scores accounted for by each HOW element, and write the number below the summarized scores in the RELATIVE IMPORTANCE row. The design elements with the highest score are best suited for satisfying the collective requirements of the stakeholders.

QFD is the first advanced data analysis tool to be presented in this chapter. It is both qualitative and quantitative in nature and uses both estimates and facts in the matrix. The background data are often based on objective information, but the individual assessments made in the relational matrix are highly subjective. The advantage of QFD is, however, that each assignment of weight or impact factor has relatively little influence on the overall outcome of the analysis, so any inaccuracies are not too dangerous. As with the preceding tools, QFD is also a typical group tool.

Six thinking hats

The purpose of the so-called six thinking hats is to actively encourage people to view a problem or situation and its solutions from several different perspectives. In this respect, it is useful in creating a mindset where teams avoid locking onto a specific design or solution too quickly. Taking distinctly different angles in a discussion is very efficient in encouraging ideas that would otherwise not have surfaced. Deliberately having one team member to play the devil's advocate ensures that suggested solutions are debated with regard to feasibility and possible flaws.

The technique achieves this by encouraging one to recognize what type of thinking one is using and to apply different types of thinking to the subject. The method was originally created by Edward de Bono (1985) when working to improve creativity and lateral thinking. He gave the six thinking hats the following colours and respective roles:

- The white hat: Cold, neutral and objective; the person "wearing it" should be systematic and careful in looking at the facts and figures.
- The red hat: Represents anger; the wearer should make sure she or he listens to intuition/gut feeling and their own emotions.
- The black hat: Pessimistic and negative; thinking with this hat on should focus on why an idea will fail.
- The yellow hat: Optimistic, sunny and positive, focusing on seeing ways ideas will work and trying to overcome obstacles.
- The green hat: Represents grass, fertility and growth, and the person underneath it should be creative and trying to cultivate new ideas.
- The blue hat: Connected with the sky, focusing on seeing things from a higher perspective.

Although some of the connotations used in describing the hats can be construed as negative, it is important to understand that each hat is equally important in ensuring fruitful discussions. This technique is a highly general and versatile approach that was certainly not developed specifically with projects in mind. However, it can serve several different purposes in a project setting:

- In the early phases, when several alternative concept ideas are being discussed and developed, or when the chosen concept is being considered in more detail, applying the six thinking hats can help create better concepts by stimulating creative thinking.
- In the detailed engineering stage, general design issues or specific engineering problems can be addressed using the technique, to ensure that the problem is viewed from all possible angles.

The general steps in using the six thinking hats are as follows:

1. Assign hats to the people in the discussion team, preferably one colour per person and make sure that everybody understands that when speaking during the session, they must clearly identify with the colour of their hat.
2. The team engages in a creative discussion about the problem or topic, with individuals contributing according to hat colour (other team members can of course also contribute at each step).
3. The facts about the problem are presented by the white hat.
4. The green hat presents ideas on how the problem could be solved.
5. The possible solutions are discussed, with the yellow hat focusing on benefits, the black hat on drawbacks.
6. The red hat works to elicit all team members' gut feelings about the solutions.
7. The blue hat summarizes the discussion and closes the meeting.

This technique is purely a qualitative approach that involves no analysis as such. It is essentially a structured way of discussing a topic, and can only be used within a group setting.

Conclusion

This chapter has presented a range of different tools and techniques that should be useful in a project setting where front-decisions need to be made but where the information to make fully qualified choices is lacking. It is my firm belief that employing various approaches to glean the most wisdom from the information you have available will lead to much better decisions than simply relying on gut feeling, coin tossing or other hasty ways of deciding something in order to proceed. Like all skills, mastering the skill of using such tools and techniques requires theoretical knowledge as well as experience. You can gain a theoretical understanding from this text, but do not expect to master the techniques right away. Keep practising, and I am sure you will end up with better projects and project decisions.

References

Akao Y (ed.) (1990). *Quality Function Deployment: Integrating Customer Requirements into Product Design*. Productivity Press: Cambridge, Massachusetts, USA.

Andersen B and Fagerhaug T. (2002). *Root Cause Analysis: Simplified Tools and Techniques*. ASQ Quality Press: Milwaukee, Wisconsin, USA.

De Bono E. (1985). *Six Thinking Hats*. Little, Brown, and Company: Boston, Massachusetts, USA.

Husby O et. al. (1999). *Usikkerhet som gevinst. Usikkerhetsstyring i prosjekt (Uncertainty as a Benefit. Uncertainty Management in Projects)*. Norwegian Center of Project Management, Trondheim, Norway.

Kano N, Seraku N, Takahashi F, and Tsuji S. (1984). Attractive quality and must-be quality. *The Journal of the Japanese Society for Quality Control* **14** (April, 2): 39–48.

Lewis JP. (1999). *The Project Manager's Desk Reference*. McGraw-Hill: New York, USA.

Project Management Institute. (2000). *A Guide to the Project Management Body of Knowledge*. PMI: Newtown Square, Pennsylvania, USA.

16
Parametric Analysis

Philip Pugh

Introduction

Provision of quantitative forecasts of cost is unavoidable, even at the very earliest stages of a project. Commonly such forecasts are excessively optimistic. However, this is not inevitable. Accurate forecasts of this kind are possible, provided suitable techniques are employed. Of particular importance is the fact that the forecaster concentrates upon total costs and avoids being drawn into excessive detail. To descend prematurely into detail is to base forecasts upon what is not yet known and can only be conjectured.

This chapter describes means for the construction of realistic forecasts via "parametric analysis", i.e. the statistical analysis of the total out-turn costs of past projects. Firstly the basic principles of such work are introduced. There follows discussions of the collection and normalisation of data. Next a variety of analytical techniques are set out. The chapter concludes with advice on their implementation based upon the author's extensive experience of the successful use of such methods.

Forecasts at the concept stages of projects

Notwithstanding the merits of qualitative analyses, it is never long before advocates and evaluators of major projects are forced into making quantitative forecasts. How much will it cost? How long will it take? How big will the revenues be? These are inescapable questions. No project will command support for long unless such questions can be answered promptly and convincingly. On the other hand, early publication of ill-founded forecasts has hamstrung the subsequent progress of many a project.

The need for early, yet accurate, forecasting is especially acute in the case of costs and revenues. Potential backers of commercial schemes will need to be assured of their financial viability. Governments will need to make provision within their budget plans for major public works. This must be done in fiscal terms. Quantitative forecasts at the concept stage of a project are therefore unavoidable.

For ease of exposition, the discussion that follows will concentrate upon the forecasting of costs. However, the techniques described are equally applicable to other aspects of a project. Revenues and time scales can be forecast in the same ways as costs.

In providing such early forecasts, the great danger to be avoided is that of being drawn into excessive detail. Too often, detail is thought erroneously to be synonymous with accuracy. The reverse is true, especially at the concept stage of a project. To descend prematurely into detail is to base forecasts upon what is not yet known and can only be conjectured. Resisting calls for detail in forecasting is not easy, but it must be done. Otherwise, a forecast is almost certain to be excessively optimistic.

Why this should be so is illustrated in Figure 16.1. Here an estimate of the future costs of a project is shown broken down into three components. (The relative proportions shown are representative of the author's experience of defence projects at the start of full development.)

First is the work that is known to be required and can be planned in detail. This is only part of the total cost but is all that can be forecast in detail. Next is an allowance for "known risks", i.e. recognised possibilities of extra work being required, some of which will eventuate. This is an important part of the total cost. However, it can be difficult to justify fully "line by line". Hence it is commonly underestimated. At the top is an allowance for "unknown risks", i.e. additional work that is impossible to specify in detail but which

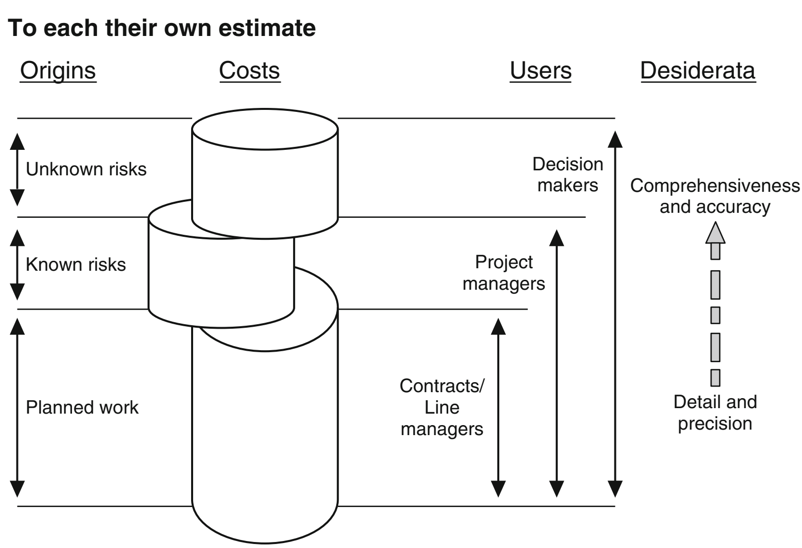


Figure 16.1 Components of a cost estimate

experience shows almost always to be necessary. This is also an important part of the total. Often "unknown risks" are more numerous than "known risks" (National Audit Office, 1996). But, by its nature, this element of costing is excluded from any detailed forecast built "bottom-up" by aggregating the costs of individual work elements.

Note that the requirements of different participants vary. Line managers and contracts staff need detail and precision but are concerned only with the planned work immediately ahead of them. On the other hand, those who decide whether a project should start (or continue) need first and foremost to know the total cost. Details of where that money will be spent are of secondary consideration. Particularly at the concept stage, their primary requirements are not for detail and precision but for comprehensiveness and accuracy.

Such purposes are ill-served by traditional "bottom-up" forecasting. For example, over recent times, actual costs for major UK defence projects have exceeded those forecast at the start of development by 40% on average (Pugh and Faddy, 2000). Experience in the USA is little different (USA Government Accountability Office, 2006). Moreover, defence compares favourably with many sectors of civil industry in this respect (Rand, 1986). Much better results can be achieved through parametric analysis, especially when this is adapted, as in recent developments, to the particular needs of forecasting the costs of projects in their earliest stages. Systematic bias towards underestimating future costs can thus be avoided.

Two further benefits of parametric analysis deserve mention:

First is the explicit treatment of uncertainty. Forecasts can, and should, include a quantification of the extent to which outcomes may vary from their mean, or most likely, value – usually expressed as the standard error of the estimate. This is valuable information for the decision-maker. A large uncertainty in cost or benefit can be tolerated, if the latter greatly exceeds the former. Conversely, a project for which the net benefit is small will be viable only if both cost and benefit can be known precisely.

Second is that the techniques of parametric analysis are simple in principle, if not always in application. Hence, the basis of a forecast can be explained readily to decision-makers and other interested but non-specialist parties. At the very least, some cogent and readily understood illustration of the reasonableness of a forecast can be offered. Forecasts are useful only in so far as they influence decision-making. It is always better to offer to decision-makers a rationale that they can make their own than to ask them for an act of faith in numbers quoted without understandable explanation. In what follows, the basic principles of parametric analysis are introduced. The collection and normalisation of data are then discussed. Next, a variety of analytical techniques are set out. The chapter concludes with advice on their implementation in producing a forecast.

Basic principles of parametric analysis

Everyday usage provides many examples of parametric analysis in its simplest form or, at least, of the starting point for such analysis. For example, it is common to speak of a "10-speed bicycle", a "2-litre car" or a "3-bedroom house". The speaker is understood to be referring not only to a particular set of gear wheels, the swept volume of an engine or merely to a special arrangement of internal walls. Rather, they are using a salient feature to convey an overall impression of the size and quality of a complex item. It is to be assumed that other attributes are consistent with that mentioned.

It is just so with other products or activities, whatever their scale. A myriad variables may be necessary to describe each exactly. However, in the working out of a rational design, these will usually be found to be highly interdependent. Hence, a few, if properly chosen, can stand surrogate for all of them.

Returning to everyday life, most listeners of descriptions of a bicycle, car or house as stated above will have little difficulty in recalling pertinent examples from their own experience. They will need only a little more information (e.g. the age of the vehicle or location of the house) in order to have a good idea of its likely price.

The same approach can even be applied to major projects. In this case, an exercise in data gathering may be necessary. A statistical analysis might be needed to link description and cost. However, the basic approach is the same. Firstly salient characteristics and costs of relevant past projects are assembled. Next a relationship between those characteristics and costs is sought. That relationship is then used to forecast the cost of a new project of similar kind.

Much depends upon a choice of "explanatory variables" to describe the past and the new projects. A correct choice of these is crucial in the development of a useful "Cost Estimation Relationship" (CER), i.e. the relationship linking explanatory variables and cost (or other outcome to be forecast). Explanatory variables should be capable of precise definition. They should be objective in nature. For concept costing, they should be known from the earliest stages of a project. A CER is of little use if the values to be included in it are unknown or are subject to much uncertainty.

Moreover, explanatory variables should make organisational and engineering sense, being clearly pertinent to cost (or to any other outcome being forecast). The resulting CER is less likely to depend upon some chance correlation between the chosen variables and those that truly matter. Such chance correlation may no longer pertain.

The number of explanatory variables to be used is touched upon later. It is sufficient to note that if the first has been chosen correctly then the addition of others in succession will bring diminishing returns. The use of pairs of variables that are well-correlated with one another should be particularly avoided. The second is largely implicit in the first, so its addition can add little to the description of the project and hence can be of little benefit to the explanatory power and utility of the CER.

A useful distinction can be drawn between what may be termed "performance" and "design" characteristics. "Performance" characteristics are those that specify what a project or product is to do. For a vehicle, these would include payload, speed and range. "Design" characteristics describe how that performance is to be achieved. For a vehicle these would include structure mass, engine power and fuel load. For projects at their earliest stages, the use of "performance" characteristics as explanatory variables is to be preferred. These are likely to be specified at the start, whereas "design" characteristics become known only as work proceeds. An exception to this rule can be a well-chosen measure of size. In some instances, e.g. the number of passengers to be carried by an airliner is itself a performance characteristic. In other instances, engineering factors link it firmly to an aspect of performance. For example, the displacement of an aircraft carrier largely determines the type and numbers of aircraft that it can operate.

A further advantage of using performance characteristics as explanatory variables is that it is here that measures valid over extended periods of time are most likely to be found. In radio systems, valves have been replaced by transistors, and these by integrated circuits, but numbers of subscribers, channels and data transmission rates have remained valid measures of performance. The number of past projects available for analysis is thus extended. This can be important when dealing with major projects, which may be undertaken only infrequently, with design technology changing rapidly.

Having chosen explanatory variables, one can proceed to the collection of data, the construction of a CER and the making of a forecast, as illustrated in Figure 16.2.

Figure 16.2 illustrates how, data permitting, it is also possible to derive "apportionment rules". These are similar to CER but describe how the total cost is likely to be apportioned between various systems (or suppliers). Thence, implications for industry, employment and the like can be assessed. Note, however, that these are secondary objectives. They are always subordinate to the primary task of forecasting the total cost (or time etc.). This process is often known as "top-down" estimating.

Collecting and normalising data

The analyst must be prepared to devote much effort to the collection of data. These then have to be normalised, i.e. put on a comparable basis. Often data collection and normalisation will take up the larger part of the time needed to prepare a forecast.

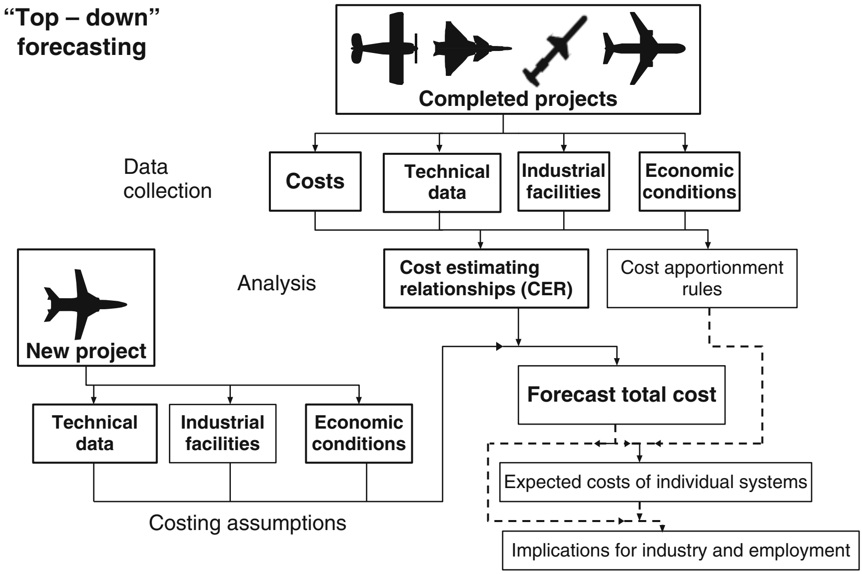


Figure 16.2 Forecasting of costs via parametric analysis

However, any analysis is only as good as the data upon which it is founded. Effort expended on assembling a sound data base is never wasted.

Too often it is claimed that relevant data are lacking, that technological change has made the outcomes of past projects irrelevant. The author's experience is to the contrary. Useful precedents are nearly always available, provided explanatory variables are chosen appropriately, as discussed earlier. Then date (e.g. of entry to service) can be used as a surrogate for technical advance. Generally, ample data on the costs of past projects can be gathered via a diligent search of trade journals and the daily press. For defence and for major public works, the reports of government audit departments are a valuable source of authoritative information. The essential tools here are patience and access to a good library.

However, cost data must not be collected indiscriminately. Data retained for analysis should meet the following criteria:

- It must be clear as to what each cost relates, e.g. to the procurement of an item or number of items all of essentially the same design and without any "extras" other than is usual in such purchases.
- Where there is a substantial non-recurring element of cost (e.g. where development precedes series production) recurring and non-recurring costs must be distinguished separately. Each will require its own analysis later.

- As far as is practicable, all costs should be out-turn costs, compiled and reported after completion of the project to which they relate. None of the data should be forecasts made early in the life of a project. Too many projects are permeated with what has been aptly called a "conspiracy of optimism". Such data are not to be relied upon.

Subject to the above, too great a precision should not be insisted upon. Approximate values can be admitted, provided their limitations are recognised. The object is to obtain as comprehensive a picture as possible. It is better to move forward with a wide view, even if some details are fuzzy, than to venture forth with tunnel vision. Likewise, care should be taken not to pre-empt later analysis by rejecting data at this stage simply because they are old or unexpected. On the contrary, such data can provide useful clues as to the form of CER that will be appropriate.

As indicated in Figure 16.2, each cost should be accompanied not only by the technical characteristics of the related project but also by some notes of its context. Of importance here are the industrial facilities employed and the economic conditions at the time. Where the project involves serial production of an item, the requisite context includes the number of items produced and the extent of any prior production using the same design. Knowledge of these production quantities is necessary for the first step in normalising each cost.

It is usual for the cost of making an item to fall, as the number of identical items made increases, a trend known as "learning". For a valid comparison, production costs have to be adjusted to correspond to the same production quantity. Accordingly, it has become customary to put unit production costs onto a "cumulative average" basis prior to their analysis. "Cumulative average" denotes an average taken across all units produced from the first up to some stated production quantity. Thus cumulative average N (written often as CuAv N) for unit production cost is the total cost of producing the first N items divided by N. The number N can be chosen arbitrarily, provided it is the same for all of the costs being normalised. However, it is usual to make it representative of typical production quantities of the type of item in question.

A common, and usually accurate, assumption is that every doubling of production quantity N reduces CuAv N to a fraction L of its former value. A value of $L = 0.90$ is typical of much production, but L can be as low as $L = 0.80$, when labour costs predominate. On the other hand, little reduction of unit cost with quantity is to be expected ($L \approx 1.00$), if the main source of costs is the purchase of goods and services "ready-made" from suppliers with many other outlets for their wares.

Suppose that I units were produced. These were followed immediately by the further production of a batch of F units at a cost T. It is required to calculate unit production cost at cumulative average N – denoted here as UPC (CuAv N). The requisite algorithm is as follows:

1. Estimate a value for "learner"(L) – see above
2. Evaluate $n = \text{Ln}(L)/\ \text{Ln}(0.5)$
3. Evaluate $E = N^n\ .\ \{(F + I)^{(1-n)} - I^{(1-n)}\}$
4. Compute UPC (CuAv N) $= T/E$.

Note that any units produced prior to a significant break in production should be ignored for the purposes of this calculation. If the batch of F units of cost T are the first to be produced (or the first following a break in production), then the algorithm as above remains valid, but with $I = 0$, the evaluation of E simplifies to $E = N^n.F^{(1-n)}$.

Note also that, having normalised all costs to CuAv N, any CER derived using costs thus normalised will yield forecast costs also at CuAv N. These may need correction to the actual quantity produced. The relevant formula is as below:

$$\text{Production cost} = \text{UPC(CuAv } N\text{)}.\ N^n\{(I + F)^{(1-n)} - I^{(1-n)}\}$$

(Cost of a batch of F units following on continuously from production of I units).

Whether involving serial production or not, costs will need correcting due to the effects of inflation, i.e. the changing value of money. It is usual to bring all costs to a common price date, assuming they will vary in proportion to some price index. A broad-based index of output prices is to be preferred for this purpose. Retail Price Indices (RPI) or Gross Domestic Product (GDP) deflators are frequent choices. Use of indices specific to an industrial sector is to be avoided since this involves an element of circular argument (that prices have risen because prices have risen). It will often be necessary for the purpose of normalisation to convert all costs into a common currency. This is frequently done prior to correcting costs due to inflation (of the common currency), using market exchange rates at the time each cost was incurred. As with any other use of market exchange rates, this brings the risk of distortion, due to the influences of short-term speculative pressures upon those rates. Current research into the calculation of rates that better reflect parity of purchasing power is of interest here. Pending such development, the best approach is to use not spot rates but annual (or longer) averages of exchange rates.

Size, specific costs and temporal trends

A simple and illuminating approach is to regard cost as the product of size and "specific cost", i.e. cost per unit size. Sometimes, this can be enough in itself as the basis for a forecast. It is always a useful first step. The measure of size needs to be chosen carefully. Mass is often a convenient measure of size, but specific cost should not be taken as being synonymous with cost per unit mass. "Size" can be any measure of scale to which cost may be reasonably proportional, at least to a first approximation. This is the basis of some widely used proprietary models. In essence, the user sets a value of specific cost by analogy with past projects. Multiplication of that by the size of the new project can yield a forecast of its cost. This is quick and convenient, but it often leaves the user with the problem of making (and justifying) their choice of specific cost unavoidably subjective.

Less open to objection on grounds of subjectivity is the use of trends in specific costs over time. Here, time usually refers to In-Service Date (ISD). Where there is a unique product (e.g. a major bridge or a railway line) ISD is the date upon which it came into use. Where multiple examples of a design are produced (e.g. aircraft or trains) ISD is the date upon which the first example of that design was delivered to any user in a state fit for routine service. A plot of specific cost against ISD will nearly always reveal a pattern of smooth change with time. This is usually one of exponential growth or (much more rarely) decay in specific cost. Figure 16.3 provides some examples of this. Note the diversity of products, the various measures of size and, especially, the continuity and persistence of the trends shown.

It was recommended earlier that costs be brought to a common set of economic conditions using economy-wide indices of output prices such as

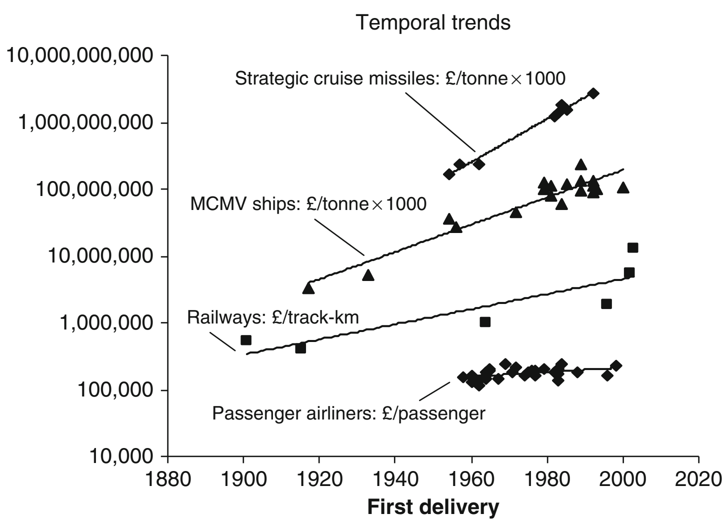


Figure 16.3 Examples of temporal trends in specific cost (all costs in £ at 2004 prices)

RPI or the GDP deflator. After correction for inflation, the costs of most goods offered for sale should show little change over time, if these are of unchanged quality. Trends in specific cost, such as those shown in Figure 16.3, reflect changes in quality as new designs supplant old in the market place. Differences in the rate of change of specific cost reflect differences in the reasons for bringing new designs to the market. In a few instances, demand is for a functionally unchanged product, but improvements in technology (beyond the usual run) enable it to be made more cheaply. Specific cost will then decline. In other instances, rising standards of living bring demands for, and make affordable, goods and services of better quality. Specific cost will then increase. In the civil field, such rates of increase tend typically to be at circa 2% per annum, matching growth in (real) incomes.

Defence goods are a special case, reflecting the peculiar nature of military competition. In the military sphere, there are no prizes for coming second. Cost may buy performance, but effectiveness derives not from performance per se but from advantage in performance over an adversary. Put simplistically, performance is how far your gun can fire. Effectiveness is how much further you can fire with your gun than they can fire with theirs. Hence, defence procurement is characterised by intense competition to acquire the "latest and best". Specific costs thus grow rapidly, rates of 4% or 5% per annum being typical for much defence equipment.

Plots such as those shown in Figure 16.3 can be used as a basis for forecasting. Regression analysis, with ISD as the explanatory variable and Ln(Specific cost) as the dependent variable, yields a formula for specific cost as a function of ISD. Insertion into this CER of the ISD of the new project under consideration gives the specific cost of the new project. Multiplication of this by the anticipated size of the new project provides a forecast of its cost. (See later for discussion of regression analysis and the evaluation of errors in such a process.)

The analyst needs to be alert to the danger of the size of the new project being understated, through over-optimism regarding its technical possibilities. Nevertheless, given care, such a forecast may well suffice for many purposes. Working thus can be especially useful at the very earliest stages of a project. It may then be necessary to work through a long list of possibilities in order to select a short list of the most promising for more detailed study. Even very approximate forecasts will be sufficient, while ease and speed in their preparation will be important. When more elaborate methods of forecasting are used, plots of specific cost against ISD remain most helpful. Forecasts prepared otherwise, or offered by others, can be placed on such a plot. It is obvious whether or not the forecast sits comfortably within the context provided by the costs of past similar projects. Such "context modelling" is a valuable guard against error. Moreover, being so readily understood, it is a persuasive form of presentation to decision-makers. There is no better way of demonstrating the accuracy of a forecast or of showing why one forecast is to be preferred over another very different one. This is of considerable merit given that, as noted earlier, forecasts are valuable only in so far as they aid the making of correct decisions.

Regression analysis

The role of regression analysis in the forecasting of costs (or times etc.) is to derive a CER linking the dependent variable (e.g. cost) with the explanatory variable(*s*), choice of which has been discussed earlier. In its usual form and with one explanatory variable, regression analysis establishes the best straight line to be drawn through a plot of dependent variable (y) against explanatory variable (x). The line found ($y = m.x + c$) is the "best" in the sense that it minimises the mean-square of the deviations of data from the line, that is, $\sum d^2$ is minimised where $d = y - (m.x + c)$.

This line of best fit or "trend line" passes through the centroid of the data. Hence, the process can be restated as being that of determining a value m such as to minimise $\sum d^2 = \sum(y' - m.x')^2$, where $y' = y -$ Mean(y) and $x' = x -$ Mean(x).

This can be readily extended to more than one explanatory variable. With explanatory variables $x_1, x_2, x_3 \ldots$ the process is that of finding values of the slopes $m_1, m_2, m_3 \ldots$ such as to minimise $\sum d^2 = \sum(y' - m_1.x'_1 - m_2.x'_2 - m_3.x'_3 \cdots)^2$. Values of dependent variable y corresponding to other sets of values of the explanatory variables can then be forecast using the relationship:

$$y = m_1 \cdot x_1 + m_2 \cdot x_2 + m_3 \cdot x_3 + \cdots + C$$

where

$$C = \text{Mean}(y) - m_1 \cdot \text{Mean}(x_1) - m_2 \cdot \text{Mean}(x_2) - m_3 \cdot \text{Mean}(x_3) \cdots$$

The fundamentals of such regression analysis are not special to forecasting costs. They are to be found in any good textbook on statistics, e.g. Moroney (1951 and reprinted frequently since). They are implemented in numerous statistical packages.[1] Accordingly, they are not repeated here. Rather, the focus here is on the application of regression analysis in cost forecasting. A number of points are worth mentioning. In these, the application of regression analysis to cost forecasting differs from the general use of this technique, even though the underlying mathematics is the same.

One of these points is that proliferation of explanatory variables is to be discouraged. This is for two reasons. Firstly, the number of data sets to be analysed is often modest. Adding explanatory variables leads quickly to excessive loss of degrees of freedom. Secondly, even with a large data set, such addition is unlikely to be fruitful. After all, the whole basis of parametric analysis is that a few variables can stand surrogate for many others. If the first few have been chosen properly, any that follow are largely implicit. Hence, they can be of little benefit to the explanatory power and utility of the CER.

In general, two performance or design characteristics serve to describe the project. ISD can be added. This can be a surrogate for technological change and/or the evolution of secondary characteristics, e.g. refinements in quality. At the concept stage of a project, this total of three explanatory variables should suffice for the forecasting of costs. If it does not, there will be a need to question the choice of explanatory variables. There may also be a need to question whether the data being analysed is truly for a homogeneous set of past projects of like kind.

Cost and performance or design variables are often found to be related in power-law fashion. Likewise, cost is commonly found to vary exponentially with ISD. Hence, a CER in the following form is frequently sought:

$$\mathrm{Ln(Cost)} = m_1 \cdot \mathrm{Ln}(V_1) + m_2 \cdot \mathrm{Ln}(V_2) + m_3 \cdot \mathrm{ISD} + C$$

where V_1 and V_2 are performance or design characteristics and m_1, m_2, m_3 and C are constants.

The slopes m_1, m_2 and m_3 are then found via solution of the following simultaneous equations:

$$\begin{aligned}
m_1 \cdot \sum x_1'^2 + m_2 \cdot \sum (x_1' \cdot x_2') + m_3 \cdot \sum (x_1' \cdot x_3') &= \sum (x_1' \cdot y') \\
m_1 \cdot \sum (x_2' \cdot x_1') + m_2 \cdot \sum x_2'^2 + m_3 \cdot \sum (x_2' \cdot x_3') &= \sum (x_2' \cdot y') \\
m_1 \cdot \sum (x_3'^2 \cdot x_1') + m_2 \cdot \sum (x_3' \cdot x_2') + m_3 \cdot \sum x_3'^2 &= \sum (x_3' \cdot y')
\end{aligned}$$

where

$$\begin{aligned}
&y = \mathrm{Ln(Cost)}, \quad x_1 = \mathrm{Ln}(V_1), \quad x_2 = \mathrm{Ln}(V_2), \quad x_3 = \mathrm{ISD} \\
&y' = y - \mathrm{Mean}(y), \quad x_1' = x_1 - \mathrm{Mean}(x_1), \quad x_2' = x_2 - \mathrm{Mean}(x_2) \\
&\text{and } x_3' = x_3 - \mathrm{Mean}(x_3).
\end{aligned}$$

The constant C is then obtained as

$$C = \mathrm{Mean}(y) - m_1 \cdot \mathrm{Mean}(x_1) - m_2 \cdot \mathrm{Mean}(x_2) - m_3 \cdot \mathrm{Mean}(x_3).$$

Note that this is the usual form of multi-variate regression analysis, except for the preliminary logarithmic transformations of cost and of variables V_1 and V_2. Note also that it is prudent to measure ISD from some arbitrary datum close to Mean(ISD). Otherwise, there may be a risk of errors due to rounding, the effects of ISD being derived as a small difference between two large numbers.

Having obtained a CER thus, its utility has to be evaluated. A convenient way of doing so is via a further regression analysis.[2] Ln(Cost) as predicted by the CER is evaluated for each of the past projects upon which the CER is founded. A further regression analysis (or "test correlation") is then performed. In this case, the dependent variable (y) is again the actual value of Ln(Cost) for each of the past projects. However, the (single) explanatory variable (x) is now the corresponding value of Ln(Cost) as obtained from the CER. For this "test correlation", the slope m should be found to be $m = 1$. The following test statistics should be noted:[3]

- The sample size (n) – the number of data pairs analysed
- The (product moment) correlation coefficient (r)
- The Adjusted Standard Error of Estimate (ASEE)
- The mean value of the dependent variable (Mean(y)) i.e. the mean value of the natural logarithms of the actual costs of the past projects analysed.

The significances of these test statistics are appreciated most readily in terms of the geometric interpretations shown in Figure 16.4.

Much textbook advice concerning interpretation of test statistics needs to be treated with caution. It is usually directed towards common applications, such as in epidemiology. Here the concern is to establish whether

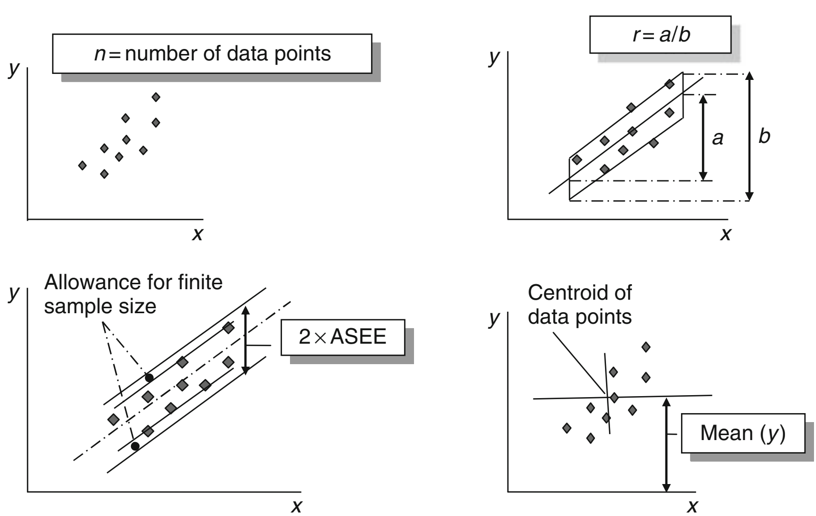


Figure 16.4 Geometric interpretation of test statistics

an association of y with x exists. There is great emphasis on the value of the correlation coefficient and upon the possibility of the value found having arisen by mere chance. This is misplaced in the context of forecasting. In cost forecasting, what is at issue is whether a relationship between the explanatory and dependent variables is useful as a means of forecasting the outcomes of individual future projects. Hence, emphasis is far more upon obtaining a suitably small value of ASEE, i.e. of a tight grouping of data about the trend line in the test correlation. (Note that in allowing for finite size of the data set analysed, the number of degrees of freedom lost in the analysis is that corresponding to the original derivation of the CER, e.g. for 3 explanatory variables the remaining degrees of freedom number $n - 4$.)

Two further points pertinent to cost forecasting deserve a mention here:

The first is that extrapolation beyond the range of costs of past projects is often unavoidable. A full evaluation of uncertainty in forecast values of cost is therefore essential. This must allow not only for scatter of data about the trend line but also for uncertainty in the derived slopes of the line and in location of the centroid of the data.

As usual for regression analysis, the uncertainty of a forecast is quantified as the root-mean-square deviation of possible out-turns from that forecast. This is denoted by $\mathrm{SE}(y_f)$ and, with test statistics for the "test correlation" denominated as above, it is given by the following equation:

$$\mathrm{SE}(y_f) = \surd \left\{ \mathrm{ASEE}^2 \left(1 + \frac{1}{n}\right) + \frac{\left(\frac{(1-r^2)}{r^2}\right)\left({y_f}'^2\right)}{(n-2)} \right\}$$

where y_f is the value forecast for the dependent variable, i.e. for the form of CER given earlier, it is the forecast value of Ln(Cost), and $y_f' = y_f - \mathrm{Mean}(y)$. Note that $\mathrm{SE}(y_f)$ varies with y_f, becoming larger, the greater the difference of that from $\mathrm{Mean}(y)$. Hence, it increases as the extent of extrapolation increases.

The second of these points special to cost forecasting is that if a CER is developed, as suggested earlier, following logarithmic transformation of cost, then the value forecast is not cost but Ln(Cost). Since Ln(Cost) is a stochastic variable, and exponentiation is a non-linear transformation, deriving the corresponding value of forecast cost is not just a simple matter of exponentiation. Rather, if the forecast of the (natural) logarithm of cost has a (mean) value, obtained by evaluating the CER of y_f and is subject to uncertainty $\mathrm{SE}(y_f)$, then the corresponding (mean) value of cost is given by the following expression:

$$\text{Mean forecast cost} = \left(1 + \frac{\mathrm{SE}(y_f)^2}{2}\right) \cdot \mathrm{Exp}(y_f).$$

The standard error of this forecast is then given by:

$$\text{Standard error of forecast cost} = \text{SE}(y_f) \times \text{Mean forecast cost.}$$

Bayesian techniques

There is no easy route to cost forecasting. There is no "silver bullet" – no set of rules to guarantee a sound estimate. Should more than one approach be possible, the analyst is wise to take each to a conclusion. Thereby, several forecasts will be obtained. Each should comprise a mean value, the average of all possible outcomes, and a standard error (SE), the root-mean-square deviation of possible outcomes around that mean.

It is then necessary to combine these estimates in order to obtain a single final forecast. Simple averaging is not appropriate, since this gives equal weight to every estimate, however precise or uncertain. On the other hand, neglect of all estimates except for the most precise cannot be optimal either. Even if the other estimates are less certain they do bring some information, which ought to be used. What is required is a weighted average in which estimates are weighted according to their degree of certainty. The optimum weighting is that which minimises the uncertainty of the final forecast.

Suppose we have two estimates, A and B. In Mean $\pm$ SE format, one estimate is Mean(A) $\pm$ SE(A) while the other is Mean(B) $\pm$ SE(B). We wish to form C, a weighted average of these, in which optimum weightings are given to A and to B. In other words, we wish to find the value of a parameter $\eta (0 \leq \eta \leq 1)$ such that $C = \eta . A + (1 - \eta) . B$ and SE(C) is minimised. If A and B are statistically independent then

$$\text{Mean}(C) = \eta \cdot \text{Mean}(A) + (1 - \eta) \cdot \text{Mean}(B)$$

and

$$\text{Var}(C) = SE(C)^2 = \eta^2 \cdot \text{SE}(A)^2 + (1 - \eta)^2 \cdot \text{SE}(B)^2.$$

Differentiating Var(C) with respect to η and setting the resulting first differential to zero gives the following condition for minimising SE(C):

$$\eta = \frac{\text{SE}(B)^2}{\{\text{SE}(A)^2 + \text{SE}(B)^2\}}.$$

Whence

$$\text{Mean}(C) = \frac{\{\text{Mean}(A) \cdot \text{SE}(B)^2 + \text{Mean}(B) \cdot \text{SE}(A)^2\}}{\{\text{SE}(A)^2 + \text{SE}(B)^2\}}$$

and

$$\text{SE}(C) = \frac{N}{D}$$

where

$$N = \sqrt{\{SE(A)^2 \cdot SE(B)^4 + SE(A)^4 \cdot SE(B)^2\}}$$
$$D = SE(A)^2 + SE(B)^2.$$

These formulae can be extended to include cases where A and B are not statistically independent but are correlated. However, this is a step too far for practical use. Evaluating the correlation coefficient of A and B is rarely possible with any accuracy. The practical approach has to be that of treating A and B as statistically independent whenever any correlation between them is likely to be weak. Conversely, where such correlation is likely to be strong, then A and B should be treated as if that correlation is perfect. In such cases, C should be put equal to whichever of A or B has the smaller SE (and the other neglected).

Other applications also exist for this Bayesian approach of taking a prior estimate (A) together with other information (estimate B) to form a refined posterior estimate (C). One such application is embodied in a suite of cost-forecasting models used with success in the defence field (Pugh, 2004). There a Bayesian approach is used to resolve the difficulty, noted earlier, of choosing between design and performance characteristics as explanatory variables. The architecture of these models is shown in Figure 16.5. Here,

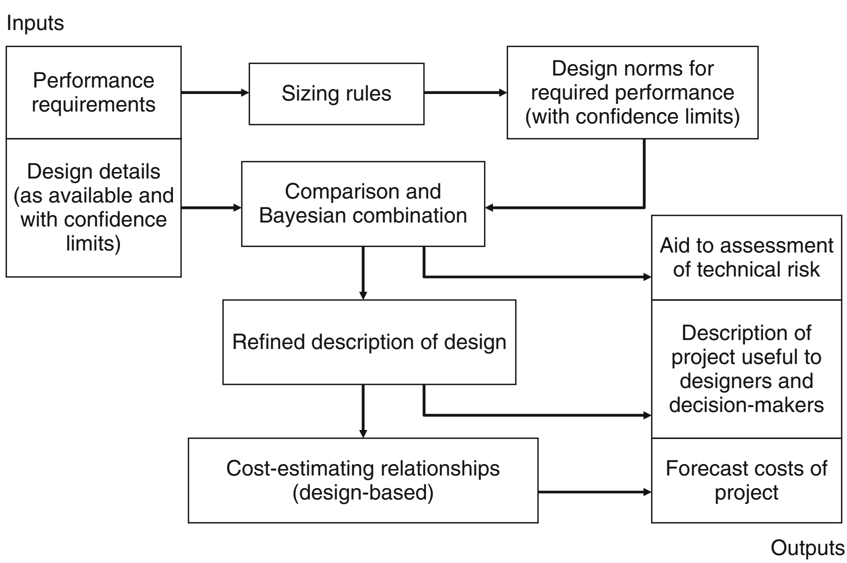


Figure 16.5 The architecture of a modern cost-forecasting model

the user inputs both performance requirements and design details. Both sets are input as they are known at the time of making the estimate, with the state of knowledge of each input being indicated in three-point form (minimum, most-likely and maximum).

The models include "sizing rules". These are relationships developed from past projects in the same way as CER but which relate performance requirements to the design characteristics used to achieve them. These sizing rules are applied to the performance requirements input by the user. This generates corresponding design norms – again each being accompanied by a measure of their uncertainty. These are then compared with values for the same design characteristics input by the user. Warnings are given of any discrepancies between the latter and the design norms as calculated via the sizing rules. Possible technical risks and/or excessive technical optimism are thus highlighted.

Each design norm and the corresponding input design characteristic are then combined in Bayesian fashion, as described above when combining estimates *A* and *B*. This yields a refined description of the design (a set of estimates *C* in the description above). This description is itself useful to designers and decision-makers. However, its primary use is as input to the (design-based) CER of the model from whence cost estimates are generated.

Consider now the evolution of a project. At its start, prior to any design or development work, estimates of design characteristics can only be imprecise. Through the Bayesian approach, the model will rely upon the design norms generated from the known performance requirements. However, as design and development get under way, more certain design information will become available. Automatically, forecasts will be based progressively more upon such data. When design and development are complete, design characteristics will be known exactly. The model will then rely upon these alone, having evolved seamlessly to this state from early reliance upon its inbuilt design norms. Thus, a single model is able to respond appropriately to all the circumstances encountered throughout a project.

Of particular value is the way in which such models guard against the influence upon their forecasts of the technical optimism which is habitual at the early stages of projects. The tendency to underestimate cost is thereby much reduced.

Practicalities of implementation

To summarise, parametric analysis, as applied to the forecasting of costs, times, etc. comprises a number of steps:

1. Collection of data concerning the out-turns of past projects
2. Identification of explanatory variables

3. Normalisation of the data so as to enable like-with-like comparisons
4. Analysis of the data to obtain CER (and other relationships as required), usually via regression analysis
5. Use of CER and any similar relationships to forecast the outcomes of future projects.

We conclude with a few notes on the practicalities of implementing such an approach.

A principal difficulty is where relevant data are sparse and may be subject to some uncertainty. This is especially likely for large projects. These are undertaken least frequently. Recent close analogies to the new project under consideration may therefore be hard to find. Analysts thus have to be prepared for much effort in data collection. They must spread their net widely. Data should not be rejected simply because they relate to projects completed some decades earlier. Nor should data be set aside merely because it is not absolutely precise. It may be that when all that can be done has been done, only a few, imprecise data are available. Even so, it is upon these that a forecast must then be built. Better a forecast that acknowledges a wide spread of possible outcomes than no forecast at all.

Where a variety of approaches suggest themselves, each should be taken to its conclusion in a forecast. These can then be combined in Bayesian fashion. Even a very approximate forecast can make a useful contribution, provided it is sensibly independent statistically of other forecasts. Any independent check upon the reasonableness of the forecast being offered has its value.

Above all, at the concept stage of a project the analyst must resist being lured, or pressured, into detailed discussion of the minutiae of design or organisation. Focus must remain on what really matters. This is to forecast the total cost (and overall time scale and benefit) of the project. Descent into detail is immaterial to deciding whether or not to go ahead. Moreover, it can be delusive, involving conjecture regarding aspects of the project that cannot really be known until at a much later stage. Such conjecture can easily lead to what has been described aptly as a "conspiracy of optimism". Premature concern for detail is the principal cause of the chronic tendency to underestimate the costs of major projects. During the preparation of a forecast, it may not be easy to resist including too much detail, as numerous estimates of individual work packages come in (i.e. to work "bottom-up"). All too often, such detail is thought by some to be indicative of a good estimate.[4] Nevertheless, such pressure must be resisted. Detail and accuracy are not synonymous. In fact, the reverse is true at the concept stage of a project, for the reasons discussed above. It is the simpler ("top-down") method, making far fewer assumptions and based only upon what is known reliably at the time, that actually yields the most accurate forecasts.

Moreover, when matters come to the presentation of results at the highest level of decision-making, there is a further reason for favouring simple parametric analyses: their very simplicity makes them easy to explain to non-specialists. Hence, at the highest level, their results are understood and accepted more readily. Placing a forecast within the context of trends in specific cost is, then, particularly useful. It can be indicative of the accuracy of a forecast, or in showing why one forecast is to be preferred over another very different one.

Of course, simple parametric methods have their limitations. Absolute precision is not to be expected and should not be attempted at the concept stage of a project. When much has yet to be decided, it is unwise to promise great precision regarding cost, etc. However, when sparseness and uncertainty of data do make for much uncertainty of outcome, then this will be spotted within the computations of SE of parametric analysis and should be reported. If, for lack of precedent (or otherwise), possible outcomes do vary widely, then that is important information in itself. It should have its place within decision-making. As is often the case in matters of logic and philosophy, the last word can be left to the ancient Greeks:

> It is the mark of an instructed mind to rest satisfied with the degree of precision which the subject admits, and not to seek exactness where only an approximation of the truth is possible.
>
> Aristotle (384–322 BC)

Notes

1. For example, Microsoft Excel performs regression analysis of the type discussed here very conveniently via its LINEST function.
2. This is not strictly necessary since the test statistics discussed later can be derived directly from the initial (multi-variate) regression analysis and are provided by some commercially available software packages, e.g. the LINEST function of Microsoft Excel. However, their significance and use can be understood best as the product of a "test correlation" as described here.
3. These may appear in different guises within commercially available software. For example, the LINEST function of Microsoft Excel does not give the correlation coefficient (r) but a "coefficient of determination" (r^2). The required correlation coefficient is the square root of this "coefficient of determination". Likewise, the same LINEST function of Microsoft Excel denotes ASEE as "sey" – terming this somewhat ambiguously as "the standard error for the y estimate". The two (ASEE and sey) are identical.
4. Such belief can be all the greater amongst advocates of a project when, as is usual, the "bottom-up" approach underestimates cost and, so, makes that appear to be more readily affordable. There are few reviewers of a forecast who are without a vested interest.

References

Moroney MJ. (1951). *Facts from Figures*. Penguin: Harmondsworth, UK (reprinted frequently since).

National Audit Office. (1996). *Ministry of Defence Initiatives to Manage Technical Risk on Defence Equipment Programmes*. HC361, Session 1995–1996, HMSO: London.

Pugh PG. (2004). Concept costing for defence projects: the problem and its solution. *Defence and Peace Economics* **15** (February, 1): 39–57.

Pugh PG and Faddy D. (2000). *Risk Analysis of Capital Investment Programmes*. Revue Actualités Perfectionnent, Gestion 2000, Bimestriel 1, Janvier/Févier 2000.

Rand Corporation. (1986). *Improving the Military Acquisition Process – Lessons from Rand Research*. R-3373-AF/RC.

USA Government Accountability Office. (2006). *Defence Acquisitions, Major Systems Continue to Experience Cost and Schedule Problems Under DoD Revised Policy*, GAO-06-368, April 2006.

Part V

Making Decisions

17
The Impact of New Information

Kjell J. Sunnevåg

Information will become available at different points of a concept's development. The precision of information with respect to cost estimates will increase as the project is developed, until all uncertainty is resolved at completion. However, major uncertain elements also relate to project benefits, and this is information that, for a large part, will only be available after project completion. It can nevertheless have an impact on the ranking of project alternatives upfront. This chapter basically investigates three questions: How much should we be willing to pay for new information? And how should we treat new information coming before <u>and</u> after project completion analytically in ranking project alternatives?

Introduction

What is information? One common definition is that it is the result of gathering, processing and organizing data in a way that adds to the knowledge of the receiver.[1] What is interesting here is that it implies that obtaining information requires some effort, i.e. that costs are involved. Obviously, some information is worth the cost and some is not, but how can one decide? Also, how should the knowledge provided by new information be used in the assessment of various project concepts?

We know that more information will become available at different points of a concept's development. We also know that the precision of information with respect to cost estimates will increase as the project is developed, until all uncertainty is resolved at completion. However, some uncertain elements also relate to project benefits, and this is information that, for a large part, will only be available *after* project completion. It can nevertheless have an impact on the ranking of project alternatives upfront.

Thus, some types of information may be available before crucial decisions are made, whereas other types of information will only be revealed in time, after the decision-making process (see Table 17.1 below).

Table 17.1 Main impact of different types of information.

	Information revealed *before* project concept is decided	**Information only revealed *after* project is put in use**
Internal uncertainty	Value of information Project value Real option value	NA
External uncertainty	Project value Real option value	Project value Real option value

This is the background to this chapter, which investigates three questions: How much should we be willing to pay for new information? And how should we treat new information coming before major decisions analytically in ranking project alternatives? And finally, what about information that will be revealed only after project completion, and how should this be treated analytically in ranking project alternatives? Bayesian updating is a keyword relevant to the first two questions, while real options offer a useful analytical framework related to the last question.

Strengthening the information basis

The issue of strengthening the information basis leads naturally to the questions, *how* and *to what extent*. Before attempting to answer these questions, the traditional approach to estimating relevant values should be examined. Very often, this is to assess a 50/50 estimate, supported by a low and a high estimate.

Uncertainty with regard to key variables essential to major decisions is obviously highest when the process of assessing project values commences but will gradually reduce as existing information is processed and new information is collected. Samset and Sunnevag (2008) refer to a Norwegian Official Report (NOU) which evaluates reasons for cost overruns on the Norwegian Continental Shelf (NOU, 1999: p. 11).[2] Here, when the oil companies present their plan for development and operations of an oil field to the Ministry of Petroleum and Energy for approval, i.e. at the end of the early phase, the total standard deviation for the project would have been reduced to 20%. For some components, however, the standard deviation will still be around 40%. As planning commences, this span will obviously be much higher, but as the project "matures", it will become more detailed, and more precise estimates will be possible. This is illustrated in Figure 17.1 below.

This process of gradual reduction of uncertainty, and the corresponding increase in estimate precision, can be attributed to several factors. Firstly, the project becomes more detailed. This enables more precise use of existing data

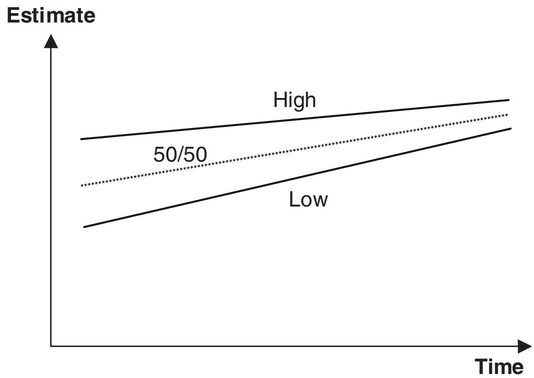


Figure 17.1 The gradual process of reducing uncertainty as project planning proceeds

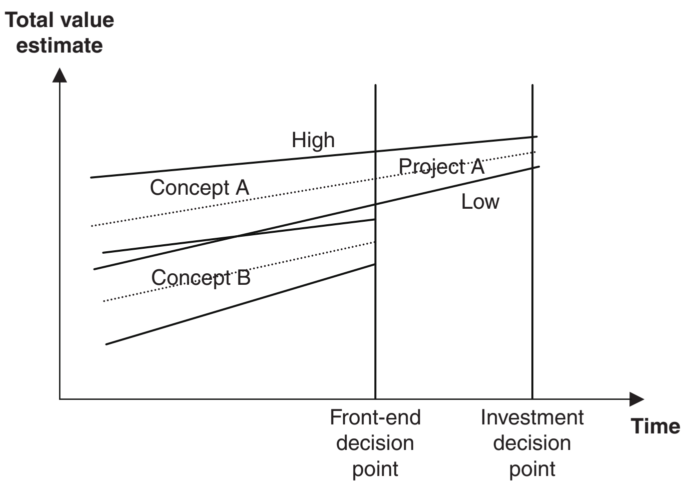


Figure 17.2 Challenges in making a choice between two concepts where estimates overlap

sources. Secondly, as time passes, more information relevant to the project will be revealed, and uncertainty accordingly diminishes. This will have a potential impact on prior estimates.

Returning to the particular challenges faced in relation to upfront decision-making, in Figure 17.2, a distinction is made between two decision points. The front-end phase is defined as the period leading up to a decision between at least one realistic investment alternative to the zero alternative (doing nothing). If a decision is taken to proceed with a project different from the zero alternative, this project will have to be detailed to such an extent that the final investment decision can be made, and building can commence. The particular challenges at the front-end decision point are illustrated in Figure 17.2.

Here the front-end decision point is placed at a point in time where it is relatively clear-cut which alternative realises the highest total net value, measured, as net present value (NPV). There is still room for surprises, but even more so, had the decision point been earlier on, where the high estimate for project B intersects the low estimate for project B.

With regard to value estimates, front-end decisions involving the choice between concepts may be more challenging (and more decisive) than subsequent decisions regarding the size of investment. One of the reasons is time, together with shortage of information. During the earliest phase, with scant information, and project alternatives that only exist as "rough sketches on the back of a fag packet", it is meaningless to use very sophisticated methods for assessment. To illustrate the type of rough, experience-based data that might be available, it could, for instance, be expressed as, "a road tunnel, with a certain width and length, in a specific type of rock, would typically cost in the range of $ to $$ per meter".

In addition, descriptions of the project alternatives may be imprecise. If each of the alternative concepts is only more or less vaguely defined, it adds to this problem.

In the total assessment, time will have the effect of increasing the span of possible outcomes, where these evolve more or less stochastically. This will particularly apply to input factors, whose value is determined by market forces. However, there will come a point when the decision-maker has to decide whether information is sufficient to make a choice between concept alternatives.

Additional problems arise when assessing the economic value of projects and project alternatives. Traditional techniques like NPV tend to discount future costs and benefits into today's values, using a discount rate that reflects project risk in order to assess the economic value of projects. It must be decided whether a project contributes to increased wealth and whether alternatives provide even greater wealth contribution. Obviously, this is a problem that involves uncertainty and the need to value decision-flexibility.

Simultaneous interrelations between the uncertainties in cash flow estimates, the risk-adjusted discount rate, the optimal strategy and the results obtained also provide analytical challenges. Moreover, it can be argued that where the risk of a project varies throughout its life, there is no single correct discount rate.

In the following, a distinction is made with respect to the circumstances under which new information becomes available, i.e. how and when uncertainty is revealed. Uncertainty may be external or internal to the project. With external uncertainty, new information will become available in time, independently of whether certain actions are taken. Price uncertainty is a good example of this. With internal uncertainty, new information is only revealed, if certain actions are carried out. In an oil field development context for instance, exploration drilling, drilling of appraisal wells and production testing are examples of actions that can reveal information, and alter ranking of project alternatives, thus creating option value, given the operating flexibility to make use of the information.

New information during concept assessment

There are two important messages at the outset. The first is that new information has no value unless it can potentially alter a decision. Thus, if it has been decided to build a new bridge, and nothing can change that decision, it would be a waste of resources to generate much information on costs. The focus should be to get more information on variables crucial to the decision.

The second point is that the marginal value principle should be used. This advocates that investment should be made in obtaining more information, as long as the marginal value of that information exceeds the marginal cost of getting it. These points will be explored further in the following sections.

Bayes' theorem (also known as Bayes' rule or Bayes' law) is a central result in probability theory that relates conditional probabilities. If A and B denote two events, $P(A|B)$ denotes the conditional probability of A occurring, given that B occurs. In Bayes' theorem an initial probability estimate is known as a prior probability. When Bayes' theorem is used to modify a prior probability in the light of new information, the result is known as a posterior probability. The process of revising initial probability estimates in the light of new information, using Bayesian techniques, is illustrated with practical examples in Goodwin and Wright (2004).

Assessing the expected value of imperfect information requires the decision-maker to judge how reliable the information will be, in order to obtain the conditional probabilities for the Bayes' theorem calculations. In some circumstances, this assessment can be made on the basis of statistical theory. In most cases, however, the assessment of the information's reliability will ultimately be based on the subjective judgement of the decision-maker.

Value of new information

It is possible to distinguish between perfect and imperfect information. Perfect information removes completely the uncertainty involved in a decision, whereas imperfect information has only a partial potential to reduce uncertainty in one or more of the underlying variables involved in a decision. To what extent uncertainty is reduced depends on how credible the new information is.

The **value of perfect information** can be illustrated using an example from road construction. It may be assumed that the construction of a highway involves the building of a tunnel. There are three basic steps to building a safe tunnel. The first is excavation: engineers dig through the earth or rock using different tools or techniques. The second step is support: any unstable ground around the construction workers must be supported while they are digging. The final step occurs when the tunnel is structurally sound and involves internal lining, the roadway and final installations such as the lights.

Before carving a tunnel, geologists and engineers investigate ground conditions by analyzing soil and rock samples and drilling test holes. However, these activities involve costs. Performing a poor job here can have a huge impact on costs and safety during construction and use.

The engineers have two different construction alternatives for the tunnel (Figure 17.3). The first one is the shortest. However, there is substantial uncertainty with regard to ground conditions for this alternative. The initial cost estimate is €500m., but if conditions turn out to be difficult, costs will increase by 50% to €750m. For simplicity, a 50/50 chance can be assumed for each outcome. Thus, the expected value of costs for the short alternative is $0.5 * 500 + 0.5 * 750 =$€625m.

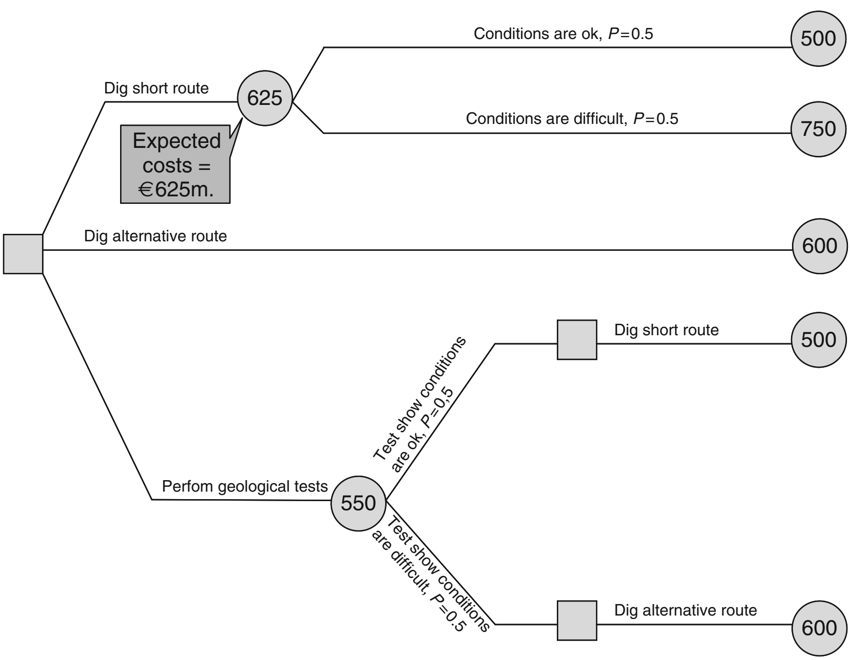


Figure 17.3 Decision tree for tunnel example

The alternative route is longer and costlier, but here the ground is well known, and geologists do not expect any surprises. The cost estimate for this alternative is €600m.

The geologists have been asked to test whether ground conditions are okay. If the test shows conditions are difficult, the alternative route will be chosen. If the test shows that ground conditions are okay, then the short and cheapest route will be chosen. Since there is a 50/50 chance for each outcome, it is also a 50% probability that the test will show that conditions are fine and vice versa. Thus, the expected outcome can be calculated if the geologists are asked to conduct the test. The expected costs, if the test is performed, is $0.5 * 500 + 0.5 * 600 =$€550m.

The difference between the expected value without the test, and the expected value if geologists and engineers investigate ground conditions is €625m. – €550m. = €75m.

Even if it is not yet known what the investigation will cost, it can be concluded that it should not exceed €75m. – the maximum amount that should be paid for perfect information in this case.

What if there is uncertainty related to the new information, i.e. the **information is imperfect**? Extending the example, the geological test is not entirely perfect and can thus result in a wrong conclusion. There are still two different possible outcomes, which are equally likely to occur. The test performed by geologists correctly shows that ground conditions are okay in 90% of the cases where ground conditions actually **are** okay. However, in 20% of the cases where ground conditions are actually difficult, the test will erroneously show that they were fine. Figure 17.4 below presents this information in a structured format.

Before proceeding, Bayes' rule should be applied. This is central to discussions on the value of new, but imperfect information, and provides a way to update the initial assessment or probabilities, when new information becomes available.

There was an initial distribution of outcomes equal to 50/50. This is the *a priori* distribution of outcome. A usual way of defining the phrases "*a priori*" and "*a posteriori*" is "from what comes before" and "from what comes later", respectively. Thus, the "a posteriori" distribution is the probability distribution after new information.

Assume that $P(B)$ assigns the probability that the test shows that ground conditions are okay. From the multiplication rule, the probability for B can be expressed as:

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2), \tag{1}$$

according to the total probability law. Here A_1 is the event that ground conditions are okay, whereas A_2 is the alternative that they are difficult. $P(A_1)$ is the "a priori" probability for the event that ground conditions are okay.

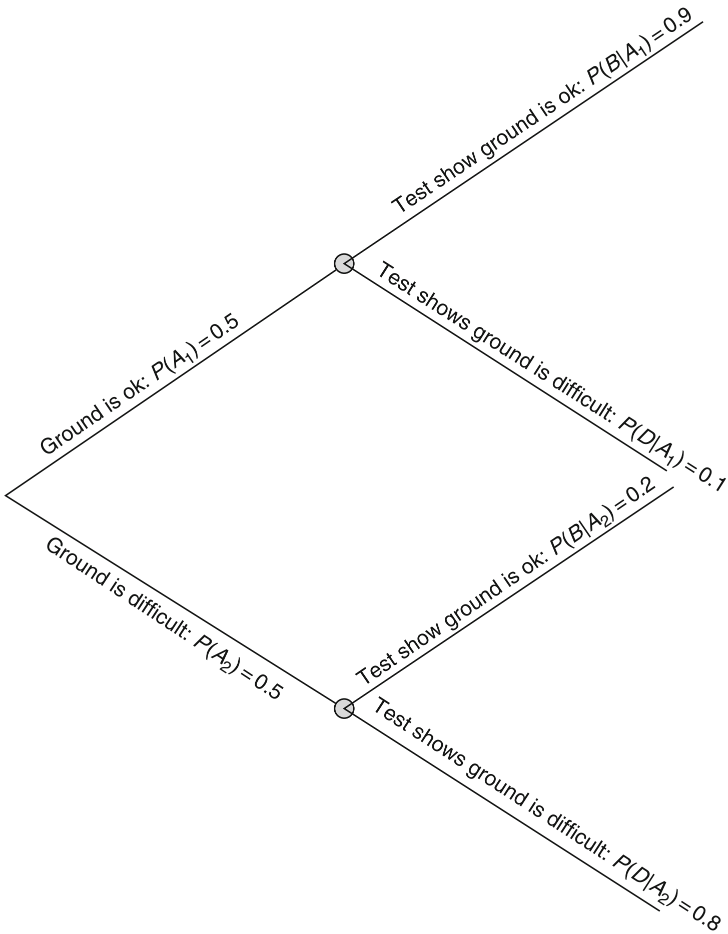


Figure 17.4 Probabilities for the test to the correct answer

The conditional probability for event A_i is:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)}. \tag{2}$$

When (1) is applied to (2), the outcome is:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}, \tag{3}$$

which is also known as Bayes' rule.

Thus, the probability that the test shows that ground conditions are okay is:

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) = 0.5 * 0.9 + 0.5 * 0.2 = 0.55,$$

i.e. a 55% probability that the test shows that ground conditions are okay. On the other hand, the probability that ground conditions are okay, if the test shows so, is:

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} = \frac{0.5 * 0.9}{0.55} = 0.82.$$

Consequently, new and uncertain information leads to a revision of the probability distribution, the extent of which depends on how credible the new information is.

To continue the example started in the previous section, the conditional probability must be known for the case that the test wrongly showed that ground conditions are difficult. D can be assigned to this event. First:

$$P(D) = P(A_1)P(D|A_1) + P(A_2)P(D|A_2) = 0.5 * 0.1 + 0.5 * 0.8 = 0.45.$$

The probability that ground conditions are okay, but the test erroneously shows they are difficult, is:

$$P(A_1|D) = \frac{P(A_1)P(D|A_1)}{P(A_1)P(D|A_1) + P(A_2)P(D|A_2)} = \frac{0.5 * 0.1}{0.45} = 0.1111.$$

The probabilities are now put into decision three, where a geological survey can be ordered, but the results of the survey have some uncertainty attached. At each chance node, the expected value is shown, and at each decision node, the value at that node is shown, given the optimal decision at that point (Figure 17.5).

We now see that the expected value if we do not perform the geological survey is €600m., whereas the expected value, given imperfect information is, €570m. Consequently, we should not be willing to pay more than €30m. for the survey.

The next step is to take a more general look at the impact of the credibility of the test on the "a posteriori" probabilities that the ground conditions are okay. Figure 17.6 presents the prior probabilities, and the posterior probabilities that the conditions are okay, for the case presented above. In addition, two alternative cases are presented to illustrate the importance of credibility. In the first, the test is right in 99% of the cases, that the ground conditions are okay, and only tells wrongly that they are ok in 2% in the cases, where

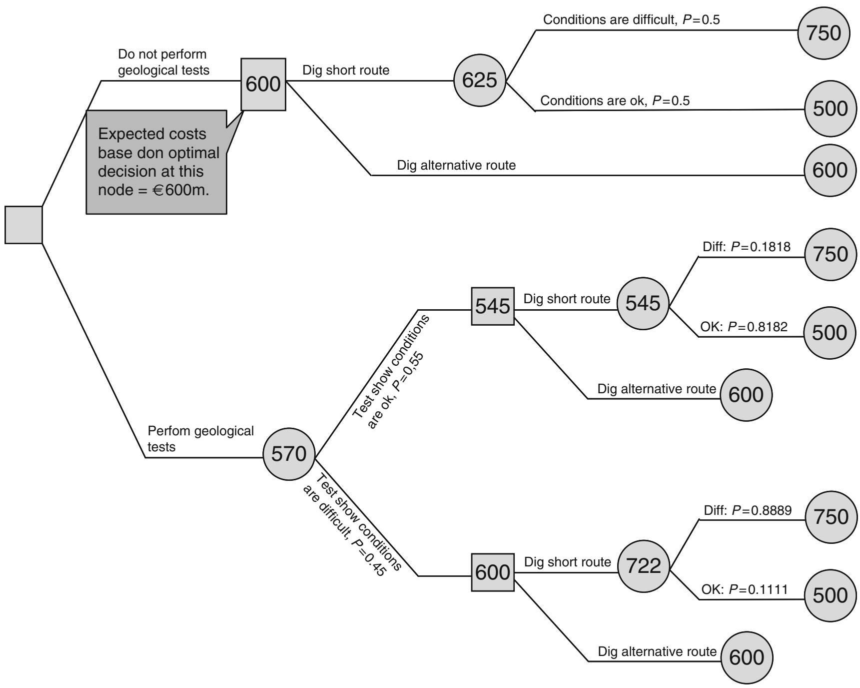


Figure 17.5 Decision tree for extended tunnel example

they are actually difficult. In the last case, the corresponding figures are 40 and 50%, respectively.

In the last case, the test shows very little: it is more confusing than enlightening. This picture can be presented more generally. In the graph shown in Figure 17.7, "a posteriori" distributions are seen as the function of four levels of credibility, varying from very low to very high.

Returning to the tunnel example, as the project planning proceeds, the geology of the different project alternatives will be revealed through geological examinations, and the requirements for the tunnel itself will be more precise, e.g. with respect to outfitting and safety requirements. However, until uncertainty is completely removed, there will always be some room for surprises. Their bigness is determined by the quality of existing and new qualitative and quantitative information sources. No less important is how this information used.

Knowing the conditional probabilities, the "a priori" probability distribution can be substituted by the "a posteriori" probability distribution. However, the calculation of posterior probabilities is a complicated procedure which needs a lot of information and requires intensive calculations. In

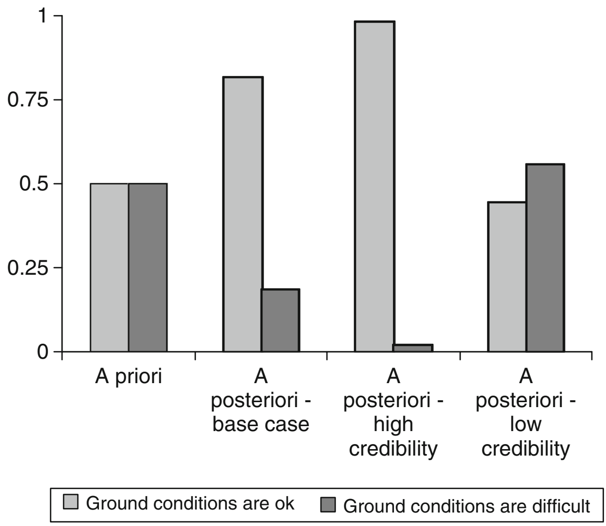


Figure 17.6 A priori and a posteriori probabilities

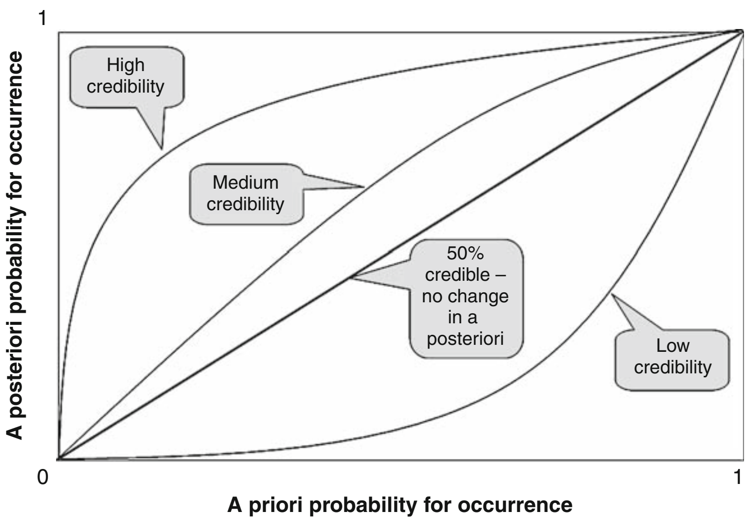


Figure 17.7 A posteriori distributions are seen as the function of four levels of credibility

practice, the decision-maker will have to allocate money and time to these activities, before it is possible to calculate the value of additional information. Also, it seems improbable that a decision-maker will be able to assess all likelihoods.

Rommelfanger (2003) points out that empirical surveys indicate that posterior probabilities are not applied when solving real decision problems: “Empirical opinion polls indicate that posterior probabilities are not applied in case of solving real decision problems”. There are several explanations for this. One explanation is that in many real-world situations, the decision-maker is quite simply unable to specify a priori probability distributions. In order to improve the situation, the decision-maker could look for additional information. Another explanation is that the Bayesian framework requires advanced theoretical skills.

With this in view, Rommelfanger suggests the use of *fuzzy decision theory* to improve upon the situation. The objective of the fuzzy decision methodology is to obtain a decision based on attaining a set of goals while observing (i.e. not violating) a simultaneous set of constraints.

Fuzzy decision theory is based on fuzzy logic and fuzzy set theory. Fuzzy logic allows intermediate values to be defined between conventional evaluations, such as true/false, yes/no, high/low, etc. One of the most useful aspects of fuzzy set theory is its ability to represent mathematically a class of decision problems known as “multiple objective decisions” (MODs). This class of problems often involves many vague and ambiguous (thus fuzzy) goals and constraints. The vagueness of expert judgement can be represented by probabilities here.

Fuzzy decision theory is particularly interesting as a decision support tool in the front-end phase. The reason is that fuzzy decision theory involves making decisions with imprecise information and measures. It deals with situations that might happen, as opposed to assuming situations that will happen. Uncertainty about probability is taken to be a form of (fuzzy) vagueness rather than a form of probability.

A lot of literature has emerged in this field, and the interested reader is referred to, for example, Wang et al. (2007).

New information after project completion

Investment like a tunnel, or building a bridge, is, to a large extent, irreversible or “sunk”. In general, investment expenditures are often sunk cost when they are firm or industry specific. Bad news for one firm in an industry will often be so for others as well. If market conditions change significantly, it may not be possible to reverse the investment decision or, at least, not without notable costs. The analytical challenges posed by specific, irreversible investments are not specific to the front-end phase, but one alternative has to be chosen, and very often this is a point of no return, at least where alternative concepts are concerned. If the alternatives differ significantly with respect to asset specificity, and flexibility to take account of changed conditions, the choice of a correct analytical approach may be crucial with respect to reaching the overall goal – to choose the project alternative that contributes the most to expected wealth-creation.

This section is concerned with information that will only be revealed after project completion. For a major road construction project, this could be the actual traffic on the road. There may have been some qualified estimates, but the actual use will only be revealed with time, after the road has been built and put into operation. Obviously, this information has an impact on the estimate of benefits, which, when compared with costs, will be crucial to the decision as to whether to build the road, and its designed capacity. Another example can be whether to expand the capacity for public transport in a city using light rail or buses. If demand conditions differ significantly from projections, buses offer a more flexible solution.

These two examples show that investment decisions share some important characteristics:

- investments are often partially or completely irreversible;
- there is uncertainty regarding future rewards from the investment;
- most investments have some flexibility with regard to timing;
- concept alternatives differ with respect to flexibility.

External uncertainty, flexibility, irreversibility and the possibility of delay are shared characteristics, all influencing the optimal decision.

The traditional method of assessing profitability is the NPV approach. Here, the investment is compared to the present value of future rewards from the project, having applied a risk-adjusted discount rate. If the net value is positive, the project will result in an increase in wealth for the shareholders (or the society, if this is a public construction project).

However, traditional methods of valuing revenue from investments can suffer from methodological difficulties. Traditional NPV analysis typically ignores flexibility. There is often some leeway about the timing of the final investment decision. It can be postponed to get more information about conditions important for the success or failure of the project. Furthermore, added value can, to varying degrees, be brought to the project through the management's ability to make operating decisions during the life of the project, i.e. to adjust the investment to existing market conditions as these change over time. Value can also be created through flexible development solutions. The degree of irreversibility can vary with the development concept. These issues come in addition to the challenge of finding the appropriate risk-discount rate for the project.

Real options approach

It is by now well known, that optimal investment rules can be obtained from methods that have been developed for pricing options in financial markets. Paradoxically, while uncertainty in the traditional NPV approach reduces value, through the increase in the risk adjusted discount rate, uncertainty in these valuation methods creates extra value to the flexible solutions. The extra value created through decision flexibility, or operational flexibility, is rarely quantified through traditional NPV approaches. However, experience has shown that the concept of *real options* provides a fruitful alternative valuation approach. But first, a few words on financial options.

A call (put) option is a right and not a duty to buy (sell) the underlying asset. For a financial option, the underlying asset may be company stocks. There is a difference here between American and European options, where the right in the former case applies up to a specified date, whereas the right in the latter case can be exercised at the expiration date.

The close connection between a call option on a stock and the real option on a given project (i.e. public construction project) is illustrated in Table 17.2 below.

The underlying asset in the case of a call option is the stock price; for an undeveloped project, it is the usage value or the revenues from the project. The last row in the table requires some special attention. The value and optimal exercise rule for a call option depends on the stock dividend rate. The holder of a stock call option does not receive dividends. The same applies to real options: the holder of a real option on a project does not receive the benefits from the project.

To deliver equilibrium, the expected net pay-off from the project (pay-outs plus capital gains) must compensate the owner for the opportunity cost of investing in the project. If authorities hold the option open to invest in the project, they will not receive the full benefit from the project. The higher the

Table 17.2 Comparison of a call option and a real option on developing a new project.

Call option	**Option on project**
Stock price	Revenues/usage value of project discounted for development lag
Exercise price	Investment costs (plus present value of operating costs)
Time to expiration	How long before decision is taken to develop project
Volatility of stock price	Uncertainty in revenues or usage value
Dividend on stock	Revenues/usage value lost if project is delayed, i.e. annual cost of delay

operating pay-out, the greater the opportunity cost of holding the option open rather than exercising it.

The corresponding question at the investment decision point is whether information that has an impact on the decision to invest or not will be revealed. Waiting can have a value, in either case. The proper methodological framework must be applied in order to assess project flexibility, i.e. to correctly reflect the ability to adjust to changing circumstances in project values.

As mentioned, investments, e.g. building a tunnel or a bridge, are, to a large extent, irreversible or "sunk". If market conditions change significantly, it may not be possible to reverse the decision to invest, at least not without notable costs. This kind of irreversibility is not obvious in the front-end phase, but, nevertheless, one alternative has to be chosen, and very often this is a point of no return: reverting to an alternative concept may not be possible. In making this decision at the front-end decision point, one has to ask whether there is a possibility that more information will be revealed with time, and in that case, will this information also have an impact on the ranking of alternatives?

The methodology is illustrated by a simple example. Since CO_2 emissions is a much-discussed topic at the moment, the project in question relates to technology for CO_2 abatement. The government decides on the development of new technology for carbon capture and storage. Authorities are concerned that the country has to fulfill its obligations with respect to emission reductions. The annual obligations can be fulfilled either by investing in new technology and cleaning measures (e.g. capture and storage), or they can be fulfilled by acquiring quotas in the market. The plant has capacity to solve the government's annual obligation relating to the Kyoto target in the period 2008–2012, i.e. for 5 years. Alternatively, the country's obligations will be exceeded by 10 million tons of CO_2 emissions. Consequently, the value of the project depends crucially on the costs of acquiring CO_2 quotas. These can be obtained and traded in the market. Furthermore, it can be assumed that all uncertainty relating to the project's costs, and the efficiency of the technology, has been resolved.

The first step in the project is to develop a full-scale capture and storage facility for CO_2 emissions, costing €1200m. The quota price today is €30/ton. Thus, if the government commits and invests today, the annual cash-flow (saved costs of acquiring quotas in the quota market) will be 10 million tons $*$€30/ton = €300m. During the first year, however, only half of this can be stored and captured, so only €150m. will be saved in year 0. If the option to invest is kept open, it can be assumed that the investment occurs at the beginning of the year, with production commencing immediately, so that the full production value can be captured.

The traditional way to look at this is to calculate the NPV, discounting future costs and the expected return using a risk-adjusted discount rate.

In this case, the project (the technology) will be worth

$$\text{NPV} = -€1200\text{m.} + \frac{10\text{m.tons} * €30/\text{ton}}{2} + \sum_{t=1}^{4} \frac{10\text{m.tons} * €30/\text{ton}}{(1+0.1)^t} = -€99\text{m.}$$

assuming that the technology can provide a (saved) cash-flow discounted for the remaining Kyoto period at the risk-adjusted discount rate of 10%. The saved costs consist of the value of future saved emissions costs. The project is shown to have a negative expected NPV and should be abandoned according to the traditional criterion, which is to only accept the project that has a positive NPV.

However, the timing option needs to be evaluated. The project can be initiated this year, or the following year, to see what happens with the quota prices. The option expires next year, so at that point it will be a "now or never" decision. This is not unrealistic in many investment cases, due, for instance, to logistical reasons or legal rights which expire. When the option expires, the project is evaluated using the traditional NPV rule based on the current price next year (which may be high or low).

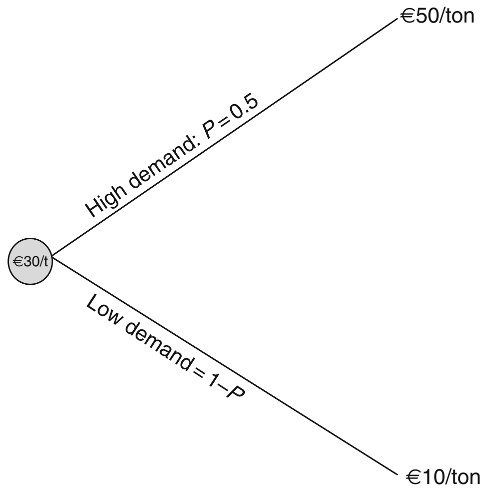
In this highly simplified example, there are only two possible outcomes next year: high demand for quotas and low demand for quotas.

If demand for quotas is high, the CO_2 quota price will rise to €50/ton, but if demand is low, the price will only be €10/ton CO_2. The probability for each outcome is equal, i.e. 50%. Thus, the expected price next year is €30/ton CO_2. The risk free interest rate is 5%.

The well-known economist, Robert C. Merton, expanded mathematical understanding of how to value financial options. In his work published in 1973, he used the term "Black-Scholes" options pricing model, by enhancing work that was published by Fischer Black and Myron Scholes. The fundamental insight of the Black-Scholes option pricing model is that the option is implicitly priced if the stock is traded. A common way of modelling price in this context has been to assume that prices move in the same way as stock markets, i.e. a random walk model.

Valuation of the project as a real option can now take place. To simplify, a binomial method is used for modelling outcomes. More sophisticated and realistic price development models are available. Hull (2003) for instance, presents some examples using Excel spreadsheets, where price is modelled more realistically for commodities markets, i.e. the price reverts around a mean. For this, he uses a trinomial mean reverting model for price development.

The project can be considered an option to benefit, if the quota price rises to the high level. But how much should this option cost? Obviously, the investment cost has its parallel in the option exercise price.



First, the return on the project must be calculated if demand is high:

$$\frac{\sum_{t=1}^{4} \frac{10\text{m.tons} * €50/\text{ton}}{(1+0.1)^t}}{€1200\text{m.}} - 1 = 0.32,$$

i.e. the return on the investment is 32%.

If demand is low, the cash flow and return is:

$$\frac{\sum_{t=1}^{4} \frac{10\text{m.tons} * €10/\text{ton}}{(1+0.1)^t}}{€1200\text{m.}} - 1 = -0.74.$$

Thus, in this case return is negative with −74%.

An essential concept in option valuation is risk-neutrality. In a risk-neutral world, the expected return must be equal to the risk-free interest rate, which in this case is 5%. Let p be the *risk neutral probability* for high demand. This probability can be found by solving:

$$\text{Expected return} = p * 32\% + (1-p) * -74\% = 5\%.$$

Solving this expression, the *risk neutral probability* for high demand is 74%.

To find the value of a call option on this project requires working backwards. If demand is high, the cash flow will be:

$$-€1200\text{m.} + \sum_{t=1}^{4} \frac{10\text{m.tons} * €50/\text{ton}}{(1+0.1)^t} = €384\text{m.}$$

Since buying the quotas in the market at the expected cost in this case would imply an expected cost of:

$$\sum_{t=1}^{4} \frac{10\text{m.tons} * €50/\text{ton}}{(1+0.1)^t} = €1585\text{m.}$$

it would be better to exercise the investment option in this case.

However, if demand is low, pay-off will be negative, if the option is exercised by paying the exercise price of €1200m., i.e.

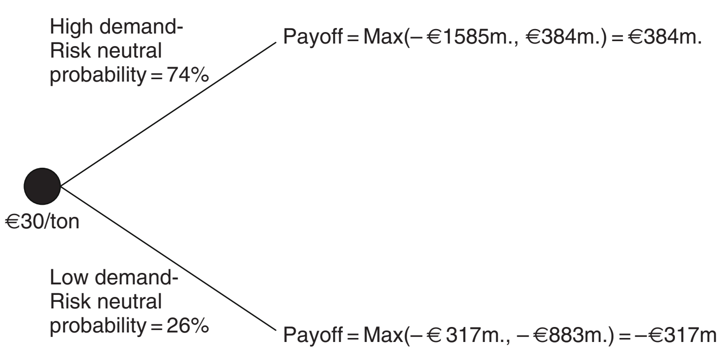
$$-€1200\text{m.} + \sum_{t=1}^{4} \frac{10\text{m.tons} * €10/\text{ton}}{(1+0.1)^t} = -€883\text{m.}$$

Evidently, investment will not be made in this case. But the quota obligation still has to be fulfilled. If demand and quota prices are low, it may still be cheaper to meet the obligation by buying quotas in the market than it would be to invest. A delay means buying quotas in the market to fulfill this year's obligations. If quota price become €10/ton, the cost of fulfilling the obligations will be:

$$\sum_{t=1}^{4} \frac{10\text{m.tons} * €10/\text{ton}}{(1+0.1)^t} = €317\text{m.}$$

Thus, in this case, it will be cheaper to buy quotas in the market. The figure below illustrates the pay-off in the two different outcomes:

The expected pay-off (using the risk neutral probabilities) in this risk neutral world has to be discounted to today's values using the risk-free interest



rate. The option to invest in this case amounts to:

$$\frac{€384\,\text{m.} * 74\% + (-€317\,\text{m.}) * 26\%}{1.05} = €195\text{m.}$$

However, delaying in investing still requires fulfilling this year's obligation, by buying quotas in the market:

$$\frac{10\text{m.tons} * €30/\text{ton}}{2} = €150\text{m.}$$

Thus, the real option is worth 195 – 150 = €45m., and this should be paid willingly in order to acquire the right to develop a full-scale CO_2 capture and storage facility next year. Since the value of the facility developed at current prices is −€99m., the value of keeping the option open, i.e. to wait and see, is higher than the value of investing today. But it is obvious that abandoning the project altogether since the NPV is negative is wrong.

Figure 17.8 presents option value and NPV as a function of investment cost. The two curves for option value present value before and after the first year quota obligation, which has to be fulfilled if the investment option is kept open. The NPV curve falls, as expected, to the right, since higher investment costs implies reduced NPV. The option value curve is initially increasing to the right until slightly above €600m. The reason is that under this investment cost level, the investment will be undertaken in both

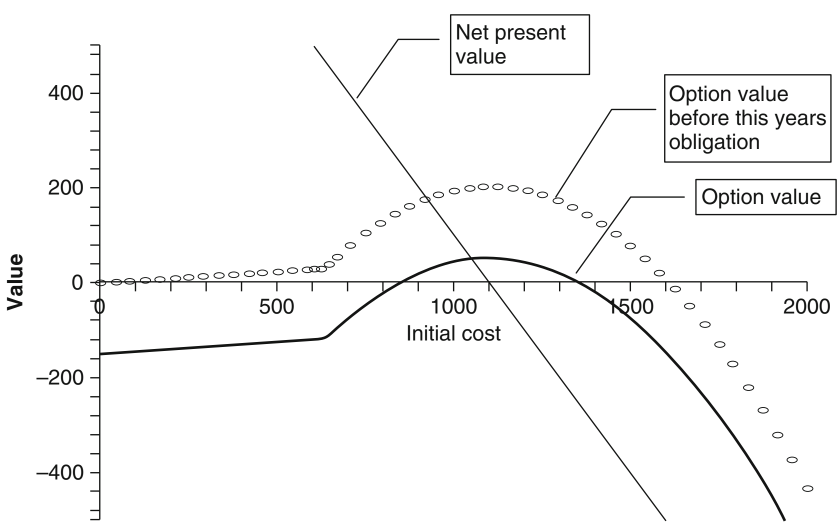


Figure 17.8 Option value and net present value as a function of investment cost

Table 17.3 Types of real options.

Real option	Description	Type of option
Option to wait	The option to postpone the project	American call option on project value
Expansion option	The opportunity to make further investment and increase the capacity of the project	American call option on the value of additional capacity
Contraction option	The option to reduce the scale of the project's operation	American put option on the value of lost capacity
Option to extend	The option to extend the project life	European call option on the asset's future value
Abandonment option	The option to close down the project	American put option on the project's value

the high and the low cost outcome. However, when the investment cost increases above this level, the option will not be exercised in the low price outcome but only in the high demand outcome, and the curved part commences. It will be profitable to exercise the option in the high cost outcome up to a certain point, where the investment cost becomes so high that it will be better to buy quotas in the market. At that point, the curve will fall linearly (and parallel to the NPV curve) to the right.

This is an example where information that will be revealed in the future will add value to the project, provided this information can be used in making a choice – in this case, whether or not to exercise the option to invest in the full-scale facility.

In reality, many different types of options can exist. Hull (2003) mentions five different types of real options, often found embedded in projects. These are described in Table 17.3 below.

Concluding remarks

According to the Oxford Dictionary of English, data are "facts and statistics collected for reference or analysis", whereas information is "facts provided or learned about something or someone". These terms are usually regarded as parts of a process leading from the collection of data to acquiring information and, finally, to obtaining knowledge, where knowledge, according to the same dictionary is "facts, information, and skills acquired through experience or education".

This process is particularly important – and challenging – in the early phase of development of project concepts and assessment, since information is limited and "rotten". Making the right choice between alternatives can make a huge difference in achieving the goals set out for the project.

Information will occur at different points in time during concept assessment. Some information may be obtained before crucial decisions are made, whereas other types of information will only be revealed with time, after major decisions have been made.

The purpose of this paper was to explore and illustrate some methodologies that can be used to take account of this so that the chance of making the right choice between concept alternatives improves.

The first issue explored was the willingness to pay for new information and how to use the obtained information to update prior beliefs with regard to the probability for different outcomes. Bayesian updating provides a valuable tool in this respect, when information is perfect and also when there is still uncertainty, or new information becomes available.

The real option approach provides a valuable tool for taking account of information that will be revealed only after the decision between concept alternatives is made. However, even if the real option approach eventually becomes established practice in evaluating project alternatives, the theoretical, empirical and conceptual challenges should not be underestimated. There is no doubt that the approach provides valuable insight into crucial aspects of the decision-making problem, particularly in the early phase of concept assessment. For instance, a more costly, but also a more flexible, concept may stand out as the preferred alternative when assessed using the correct analytical framework. The methodology illustrates the structure of the decision-making problem and provides insight into what may give extra value to different project alternatives.

Notes

1. See http://en.wikipedia.org/wiki/Information
2. Report from a Committee appointed by the Norwegian Ministry of Petroleum and Energy.

References

Goodwin P and Wright G. (2004). *Decision Analysis for Management Judgment.* West Sussex: Wiley.

Hull J. (2003). *Options, Futures and Other Derivatives*. Prentice Hall: New Jersey.

NOU (1999). Report from a Committee appointed by the Norwegian Ministry of Petroleum and Energy (in Norwegian), "Analyse av investeringsutviklingen på kontinentalsokkelen" (Analysis of the Development in Investments on the Norwegian Continental Shelf), Norwegian Official report: 11.

Rommelfanger H. (2003). Fuzzy decision theory. Intelligent ways for solving real-world problems and for solving information costs. In: Riccia G,

Dubois D, Kruse R, and Lenz H (eds). *Planning Based on Decision Theory*. Wien: Springer.
Samset K and Sunnevaag K. (2008). Making large decisions when little information exists. *Project Perspectives* **xxix**:58–64.
Wang P, Ruan D, and Kerre E. (2007). *Fuzzy Logic: A Spectrum of Theoretical and Practical Issues*. Berlin: Springer.

18

The Complexity of Decision-Making in Large Projects with Multiple Partners: Be Prepared to Change

Roger Miller and Brian Hobbs

The present chapter examines the reality of decision-making during the front-end of large projects that are particularly complex. The discussion is based on over ten years of empirical study of large projects in several industries and from differing perspectives. The chapter will first present the characteristics of complex projects and provide some examples. This is followed by a discussion of the major categories of decisions that must be made during the front-end of these projects. A presentation is then given of two very different approaches to the management of these projects in general and to decision-making in particular – rational choice and evolutionary shaping. The argument presented is that the complex projects discussed here require the latter approach, the governance rules and structures of which are presented in the final section of the chapter.

What makes projects very complex?

All large projects can be considered complex. They have the characteristics identified by Williams (1999) as sources of complexity, i.e. *structural complexity,* in that they are composed of many interrelated components, subsystems and technologies and also *uncertainty*. The sources of uncertainty include the technology, the political and social environment and the market, with many interactive effects among them. The types of projects discussed here have additional characteristics that further increase their complexity and the difficulty of managing them.

Three characteristics make these projects particularly difficult to manage. Firstly, they "push the envelope", i.e. they go beyond what is current practice either in the systems or markets they develop, or in the social arrangements they create. They are socially, technically or commercially innovative. Secondly, they all involve a division of labour among co-specialised organisations that form networks of participants. There is a dominant sponsor or sponsor group, but many of the relationships among participants are partnerships, in which both parties exert influence and participate in co-evolutionary processes. Thirdly, the environments in which these projects take place are dynamic: change is more the rule than the exception.

Three types of projects are employed here to illustrate the front-end of these very complex projects: (1) infrastructure projects developed, financed, delivered and operated by a consortia of private firms under concession agreements with governments; (2) projects to build state-of the-art industrial facilities; (3) and large IT projects in organisations whose primary business is not IT.

Pubic infrastructure built by private consortia

The first example is that of infrastructure projects developed, financed, delivered and operated by a consortia of private firms under concession agreements with governments. A significant number of the 60 large projects investigated on the International Research Programme on the Management of Large Engineering and Construction Projects during the 1990s were transportation or electricity production infrastructure projects (Miller and Lessard, 2001). These projects were social innovations, for they were carried out at a time when the role of private parties in infrastructure was being redefined. Many were early experiments in this type of arrangement for both the firms and the governments involved. The unfamiliarity and innovation produced social and political environments that were often unstable. Projects were often contested within the political area and institutional frameworks manipulated as a result. Although some governments now have reasonably well-established governance regimes for private-public partnerships in infrastructure, governments and firms are still experimenting in many countries, in various specific sectors of activity, rendering the results of these examples from the 1990s relevant in many contexts.

One of the most striking characteristics of these projects was the very long period of time required to develop the original idea into a viable project concept. On average, the front-end lasted 7 years. Decision-making during this front-end is best thought of as a non-linear and historical process, which goes through a series of episodes, during which the project concept is developed, tested and modified until it becomes viable. The project concept is much more than the technical solution. It also includes all the elements of the business case for the project, the organisations and organisational arrangements that will allow it to be executed and the institutional arrangements that will govern the project and provide it with legitimacy.

In the early stages, the need that the project is to satisfy is poorly defined. The process of identifying both the need and the concept that will meet this need is a search for a match, during which both the need and the concept are redefined several times. At the outset, it is not clear who all the participants and influencers will be. The participants must be identified, the roles defined and the relationships built up and translated into contracts through lengthy negotiations. The institutional framework may have to be modified or developed. More than 60% of the projects cited required significant changes to the laws and regulations governing projects. This is a political process, the results of which are very difficult to predict. The "project concept" derives from a combination of the network of participants, the business case for the project as seen from the points of view of each of the relevant participants and stakeholders, together with the proposed changes to the institutional framework. All parts are highly interdependent. As the project concept is moulded through a series of irregular episodes, a viable project concept is formed and anchored into the institutional framework.

These projects are particularly vulnerable to many types of risk. Demand for services provided by infrastructure investments is notoriously difficult to predict. Because of the way these types of projects are financed, they are very vulnerable to delays or shortfalls in market response. In addition to the risks associated with any large infrastructure project, they are vulnerable to a lack or loss of political and social acceptability, which sometimes leads to changes in the organisational network supporting the projects and even to changes in the institutional structure that was to anchor and legitimise them. The expression "legislative risk" has been coined to describe just such a situation. The projects are particularly vulnerable after they have reached a point of irreversible commitments, e.g. during the construction phase. Each successive project concept must be tested against the possible risks to the project's viability and modified to control the level of exposure. The emergence of risks or the awareness of potential risks has an important impact on the way the project concept is developed.

This type of project is exposed to both *known-unknowns* and *unknown-unknowns*. In the IMEC study cited above, the projects were hit by an average of four potentially catastrophic and unforeseen events. Examples included: bankruptcy of a partner, changes to the political power structure, rapid changes in the price of key resources and the Asian financial crisis of the late 1990s. The projects are exposed to these emergent unforeseen risks because of both the long period of time over which their development takes place and because of the volatile environments in which they operate. These emergent risks add to complexity and indeterminacy. In fact, the entire front-end process is largely indeterminate. Traditional conceptualisations of decision-making and decision-makers are poor representations of this reality.

Projects to build state-of the-art industrial facilities

A major investigation identified six archetypes of innovation or what the researchers describe as "games of innovation" (Miller and Olleros, 2007). Each "game" has a different structure and dynamics as it pursues different innovation strategies. A discussion of all six is beyond the scope of this chapter. One of the games was, however, particularly relevant and is used here as the second example of a type of extremely complex projects.

This "game of innovation" is centred on industrial corporations that build major facilities on a regular basis. Manufacturing and petrochemical companies are examples of industries where such firms and projects are found. These companies seek to build state-of-the art facilities, but recognise that their expertise is in operating the facilities, not in their design or construction. Large engineering projects are usually undertaken in mature markets and focus on the building of closed, integrated product architectures. Sponsors of projects rely on networks of leading suppliers to develop and install the latest technologies. The developments and innovations produced on these projects often have significant value beyond the specific project. The management of this intellectual property adds both incentives and complexity.

The entrepreneurial function in large engineering projects is thus shared among owner/sponsors, consultants, engineering firms and equipment suppliers. Projects are usually one-off ventures that introduce varying degrees of innovation into the productive system of the sponsor organisation and also into the economy in general. Large engineering projects innovate by pushing the envelope of existing technologies to new heights. Such projects focus on the design and implementation of mission-critical systems. Owner/sponsors are usually operators who have the expertise to evaluate, understand and even improve on choices. Projects produce tightly integrated closed systems intended to perform complex tasks, such as oil refining and manufacturing. Innovations emerge from interactions between operators and co-specialised consultants who understand the evolution of rapidly evolving infrastructure, such as production technologies.

New systems of operations incorporate the latest knowledge and technologies, thus breaking new grounds in terms of performance. Using know-how available on the market, engineers, architects and designers push the envelope a little bit further with each new capital investment project, be it a machine, a plant, or a power network. Information and Communication Technologies have been key drivers in recent years, not only for the powerful analytical insights they provide but also because they act as the backbone of these large systems.

The dominant levers for innovation in building state-of the-art industrial facilities are (1) the selection of experienced experts and suppliers, capable of envisioning bold new solutions to improve the performance of the whole system; (2) the building of governance structures and processes for collective problem-solving between the owner/sponsor and the experts and suppliers; (3) project management competencies, allowing both experts and serial innovators to accumulate know-how, abandon projects and incorporate worthy changes.

The ecosystem for building state-of-the-art industrial facilities is composed at the core by large clients with the need to build new systems or new facilities. Aiming to serve them are rings of diverse experts, from strategy consultants to architects, designers, engineers and IT specialists, and critical "Tier One" suppliers, who master a technology and are expected to share the "state-of-the-art" challenge of their clients.

Typically, the process starts with interactions at the senior management level, between large operators and strategy consultants to help define competitive issues, opportunities and risks. The goals pursued are usually to radically transform cost structures, improve the delivery of new products or understand emerging market dynamics. Owner/sponsors want to significantly improve productivity and coordination of material and information flows around the world.

Once the strategic issues have been decided upon, clients build a project team internally, or with consultants, to sketch out the architecture of the new system and outline specifications. They then invite systems engineers with accumulated expertise to imagine, design and articulate innovative solutions. In turn, specialised suppliers involved in implementation may do their creative work internally or engage in further generative interactions with their own contractors. New systems are thus integrated and assembled from closed, open or interoperable technologies and subsystems.

The dominant logic of innovation is thus a shared process of problem definition, solution design and implementation. Demand influences the supply of solutions, but in turn, solutions can reshape demand. The reputation of strategic advisers and the experience of systems engineering firms are core selection criteria. Owner/sponsors generally choose from a limited range of consultants and engineering suppliers nationwide or worldwide. In each sector, the same few names keep appearing on everyone's list of preferred partners. Novices find the going rough.

Innovations emerge from the sharing and shaping of ideas in generative debates. Consultants and systems engineers must stay significantly ahead of their clients to provide valuable advice. The means used by consultants to foster innovative competencies include working relationships with leading clients facing significant challenges; alliances with technology suppliers to understand expected new technologies; formalisation and codification of strategic or engineering methodologies; and accumulation of past experience in archives, knowledge management systems or expertise directories. Many strategy or engineering consultants fund research institutes to explore and build scenarios about the future evolution of the sectors that they serve.

Clients, especially large operators, capture value through the improved effectiveness and efficiency that stem from ramped-up projects. Consultants capture value by gaining reputation, experience and new knowledge. System integrators capture value by building partnerships with top-level strategy consultants and clients, involving both in learning about platform evolutions. Bankers or investors involved in the financing of projects capture value by better risk management.

Projects also emerge in successive shaping episodes that start with the client forming a hypothesis about the progress that can be achieved on the issues that need to be resolved and the effort required to develop strategies to bring about this resolution. Nested issues need to be addressed first, as well as the resources necessary to achieve progress towards closure. Each episode opens new options and closes old ones, until sponsors and partners achieve final lock-in, thus binding their commitments and losing a large amount of their freedom.

Projects are made economically viable, technically functional and socially acceptable by the client and its consultants, by progressing on solutions to deal with the following issues: (1) negotiating a project concept that truly creates value and can be progressively refined; (2) developing a business model that provides stability to the project and ensures that investments will be repaid and protected against opportunistic behaviours; (3) gaining and ensuring legitimacy is achieved through the consent of affected parties and approval by governments; (4) achieving shock-absorption capabilities by the building of governability devices into the project structure: crisis funding, cohesion, reserves, flexibility, generativity and modularity.

Large IT projects in organisations whose primary business in not IT

The third type of example is that of large IT projects sponsored by organisations for whom IT systems are mission-critical but whose primary business is not IT (Haggar and Miller, 2008). These organisations have often outsourced strategic analysis and design, as well as the provision of IT services to expert suppliers. Sponsors are network operators in financial services, telecommunications, distribution or messenger services, who want to streamline cost structures, improve delivery of new products and build e-commerce transaction systems. In building or improving these systems, owner/operators undertake significant risks and thus rely on professional experts to address them. Risks do not come so much from technological uncertainties, as from the scale of the projects, the possibility of cost overruns and the performance of the improved systems due to functional failures.

The context of these projects is similar in many ways to the previous examples. Sponsors are owners/operators who rely on consultants and suppliers to develop and implement new or improved systems. The operator's expertise is in accumulated business knowledge but not in the design and delivery of the systems that are at the heart of their business. Operators depend upon the expertise of consultants and suppliers, who invest heavily in maintaining their know-how at the leading edge. IT or engineering consultants, for instance, invest about 15% of sales in R&D and capabilities building; eighteen per cent of staff time is allocated to innovation (Miller and Olleros, 2007).

There are, however, significant differences. The system operators tend to have longer-term relationships with their consultants and suppliers, to whom they often outsource significant parts of the operating of the systems. The projects are thus somewhat less “one-offs”. The biggest difference is the importance of change. Change is present in all the complex projects discussed here, but the order of magnitude of the changes, and the stages at which they are introduced, are very different. Infrastructure projects, and projects to build new facilities, must freeze the design of the physical asset much earlier because the cost of change becomes prohibitive. This is referred to as the irreversibility of decision.

Design decisions on IT projects are much more easily changed. Compared to other technologies, the relative ease with which changes can be made creates the possibility to make significant alterations relatively late in the project life cycle. As long as the benefits of change outweigh the costs, change can create value. Changes may even lead to modification of the architecture.

The modular nature of the technology allows for multiple deliveries over an extended period of time. The logic of their management is very much one of programme management. Each project in the programme delivers a part of the whole system, but several modules can be put into operation as the programme unfolds. This allows the project participants to test parts of their systems in terms of functionality, operability and the response of both the operators and the market. This information is fed back to the programme team, and adjustments are made to the programme.

In all of the types of projects discussed here, there is a conscious search for opportunities, that leads to changes in the project concept as it is shaped over time. However, with large IT projects, the search for improvement opportunities is much more prevalent. If managed poorly, this can induce high levels of system instability and problems of configuration management. If managed well, it can create significant value.

The execution phases of the three types of projects described here are quite different, but the front-end phases are quite similar. The focus here is, of course, on the front-end. The front-ends of all of these projects are characterised by open-ended searches and iterative processes involving multiple partners.

Decisions in the front-end

Complex projects require that three important and interrelated sets of decisions be made in the front-end. First, a concept must be developed along with a business model. The concept must define the functionality and be technically coherent and socially acceptable. What is crucial here is that the concept must meet the needs of multiple stakeholders, both within the project network organisation and the project environment. It is much simpler to develop a concept when the needs of one sponsor are the only important concern. Even on projects with an owner/operator as the sponsor, the project must meet the multiple business objectives of internal business units and partners in order to create value for all parties.

Secondly, because the projects must meet the needs of several partners or stakeholders, the choice of members of the project coalition is interrelated with the development of the needs to be fulfilled and the functionalities to be delivered. This is a social process of building a network of interested parties. In the example of infrastructure projects developed by consortia, the lead sponsor may be a single organisation, but in many cases it is a group of co-specialised firms. In the other two examples, the owner/operator is usually the prime sponsor.

Thirdly, processes and structures must be put in place to ensure that the project is managed in a way that is adapted to the specific conditions encountered on the project. The following section presents alternative approaches to managing large projects. This is followed by a discussion of the governance rules and structures that are appropriate for very complex projects.

Approaches to the management of projects: rational choice or evolutionary shaping

Large projects are certainly difficult technical tasks, but they are primarily complex managerial and socio-political challenges. Achieving problem-solving and coordination of the interests and contributions of each powerful party is made possible through successive episodes of issue resolution. Simplifying to the extreme, two types of management approaches may be used to achieve effectiveness in coordination.

Rational choice

Project management theories view projects as ventures that can be planned and specified in advance (Cleland and Ireland, 2006). A business case is built, and experts are hired to design a solution, define specifications for work packages and select contractors using bidding processes. It is assumed that complications can be solved by engineering calculus, computations and better coordination. Uncertainty is viewed as manageable, through investments in information analyses to scope, mitigate and eventually minimise risks. Unexpected project changes are eventually fought in the legal system.

The assumptions that underlie rational decision-making frameworks are often out of line with the realities of the projects considered here. The first assumption is that the needs of the owner do not change once specified. The second assumption is that planning can proceed on the basis of the possibility to predict future states of nature and choose the best option; the future is probabilistic, and planning must be done early. The third assumption is that projects unfold over time to form a sequential unity. Each stage is a detailed elaboration of the previous one. The fourth assumption is that the passage of time is not really important, as conditions will not change appreciably. Therefore, it is not economically justifiable to build in the capacity to react to important emergent threats or opportunities.

The perspective of rational planning fosters the idea that changes are not only bad, but are signs of bad management. Studies of large infrastructure and information technology projects conducted over the last few years have led to the opposite conclusion. Changes can be good and rational if the benefits surpass the costs. Benefits can arise from the emergence of new technologies, new opportunities or market conditions. Should new conditions arise, changing may be a profitable option and sometimes a necessity.

While rational approaches are still valid for simple projects and construction phases, they are inadequate for complex projects facing dynamic futures. The link between value creation and the attitudes to change is particularly visible in IT projects, where changes are normal and procedures are designed to capture them. Technical opportunities and market evolutions suggest using methods and applications that were not originally planned. Executives accept that justifiable change is to be encouraged: changes are absorbed into projects as long as benefits are superior to expenditures involved.

The reality of the types of projects discussed above is out of line with the assumptions of the rational choice approach. Instead of being fixed, the needs of owners change: the typical case is that owners face financial difficulties and cut the dimensions of projects. It is well nigh impossible to assume that conditions will not change for the 7–10 years that planning, construction and ramp-up require. Projects go through successive episodes that are the result of both planning and responses to change. The passage of time opens up opportunities, triggers constraints and elicits moves by stakeholders that call for change. These realities do not make projects totally unmanageable, but they require a different approach, one which is more strategic.

Evolutionary project shaping

Rather than evaluating projects at the outset based on projections of the full sets of benefits and costs over their lifetime, many sponsors view them in evolutionary perspectives (Miller and Lessard, 2001). Sponsors act as champions, actively shaping projects in response to changing and unexpected conditions.

Projects are innovations in themselves, as they introduce new capacities in productive systems and have both positive and negative effects. They involve multiple actors under the leadership of major sponsors. The degrees of innovation that they introduce range from marginal improvements to highly disruptive innovations. The more innovative projects usually call for extensive shaping activities.

Sponsors start with initial concepts that have the possibility of becoming viable. They then embark on shaping efforts and debates to refine, reconfigure and eventually decide on a concept that will yield value while countering risks. The seeds of success or failure are thus planted early and nurtured as choices are made.

Planning has been portrayed as ineffectual (Mintzberg, 1994), but, paradoxically, clients are spending increasing amounts of resources and time on this activity as projects are shaped (Miller and Lessard, 2001). As issues arise and call for decisions, high levels of planning expenditures are necessary. Finally, concepts are formalised into viable configurations.

A variety of intertwined issues need to be resolved one by one by sponsors, alone or in cooperation with sponsorship partners, or co-specialised expert firms. Rather than selecting the optimal project, the essence of managing these very complex projects consists of shaping. Should institutions be inappropriate, a large part of the shaping problem is to create substitute coalitions and eventually help to create new institutionalising patterns. This typically involves "buying in" some stakeholders and "buying off" others. In some cases, the roles of stakeholders can be specified in advance. In many cases, however, it is not clear how to accommodate various interests, so the leading sponsor must exploit the front-end period to identify a mutual-gains trajectory. The shaping process combines deliberate actions with responses to emergent situations, as projects progress through time. Managers introduce real-time mitigation strategies to influence chaotic situations.

The shaping of projects takes place over many episodes, which are not a sequence of stages but episodes of progressive problem-solving and issue resolution. The path taken by each episode depends on previous ones but is to some degree autonomous and indeterminate. For each episode, sponsors form hypotheses about the extent of progress that can be achieved on the issues that need to be resolved. The costs of shaping are so high that sponsors first identify projects that stretch their capabilities but that, because of their complexity and risk, offer substantial benefits that could not be achieved with simpler, less risky undertakings. Planning and shaping efforts are expended to make projects economically viable, technically functional and socially acceptable.

Each episode starts with momentum-building, continues with reactions to opposing forces, and ends with closure. Momentum is built by imagining concepts, promoting legitimacy and designing a configuration such that partners, affected parties and governments accept what is proposed. *Meeting countering forces* and criticisms calls for realism to avoid the temptations of unreasonable commitments to satisfy expectations or scepticism to reject good opportunities. Each shaping episode ends with suggestions to abandon the whole project or to come to a temporary agreement on a conceptual configuration. The main shaping episodes are as follows:

i) *Initiation and exploration* The initiation episode lasts a year or two and closes when a credible party states openly that it is ready to allocate funds and start debates on the ways and means of shaping and financing the idea. Resources of a few million dollars are assigned for exploration of the initial project hypothesis. Engineering, financial and marketing feasibility studies are undertaken. Conceptual closure is achieved by the production of a series of documents or position papers sketching out the initial concept.
ii) *The development of a holistic proposal* is concerned with the building of an investment scenario and a business case for investors. Preparing the proposal is expensive. The result is a see-saw between positive-sum collaboration and zero-sum gaming. Eventually, the balance becomes acceptable.
iii) *Negotiations and issue resolution* Many issues cannot be resolved early. To confront emergent fears, clients often have to clean up polluted sites, organise referenda or build complementary facilities. They may also need to organise economic development initiatives to demonstrate their credible commitments. Furthermore, numerous issues skipped in the holistic proposal require solutions through negotiation and problem resolution. Facing social and environmental fears is a very expensive affair.
iv) *Closure on a committable package* Commitment on a final package takes place when all major issues have been resolved. In many projects, clients may have spent several hundred million dollars to develop solutions, gain consent, solve social and environmental issues and build agreements.

The central theme of this chapter is that evolutionary shaping describes both the reality of how complex projects unfold and the management approach that is best suited to them. The following sections address the issue of how to manage projects based on this approach by presenting a few of the governance rules that can help achieve this.

Governance rules

As they gain experience in projects, sponsors develop rules for action in the planning, designing and execution of projects. Some sponsors manage a portfolio of projects and thus have multiple opportunities to learn over the years what works and what is not associated with success. Sponsors who undertake projects on an irregular basis must rely on best practices, promoted as industry rules. The following are among the key rules for decision-making:

1. The primary rule is that sponsors must build infrastructures to operate their networks but lack the competencies to design and construct them. Smart sponsors recognise the expertise of the specialised firms on which they can rely. Only in cases where systems are proprietary and markets are mature will firms rely on internal capabilities for front-end decisions.
2. The value of information in front studies to build multi-dimensional project concepts is very high. A few million dollars invested to build and test scenarios can lead to the elimination of errors that could cause unnecessary but substantial expenditure. It is more important in the front-end to text scenarios than to gather detailed information on a particular scenario. Experienced executives can estimate the benefits, costs and risks associated with rough scenarios for a much lower cost than detailed analyses of single concepts.
3. The shaping of projects is a costly business. Standard non-innovative projects require 3–5% of the total cost to be spent in front-end analysis. Moderately complex projects will require 8–9% of total costs to be spent in front-decisions to face technical, market and socio-political risks. Complex opportunities that call for changes in institutional conditions and involve some degrees of innovation will require up to 35% of total cost to be invested in front-end planning (Miller and Lessard, 2001). Effective sponsors have learned that such expenses are often necessary.
4. Obstacles to adequate front-end investments in planning are numerous. Many executives or political leaders view such expenses as frivolous, because they involve immaterial actions such as negotiations, legal advice, community involvement, etc. Such investments may be regarded as ineffective. Quite often, unless institutional rules prevent them from doing so, political leaders will make announcements to suit electoral purposes, thus leading to the skipping of analyses.
5. Sponsors do not sit idle, waiting for probabilities to materialise; they judge risks, imagine ways to cope and work hard to shape outcomes. Strategising about risks may start analytically, but it requires managers to quickly become experts in organisational science, diplomacy, law and public affairs. Strong uncertainties and indeterminacy are thus reduced by the use of repertoires of strategies based on prior experience.

Building governability: facing up to changes

In order to deal with potential but unknowable future risks, sponsors attempt to infuse governability, i.e. the capacity of project participants to steer through unexpected turbulence when projects face changing conditions.

Turbulence is a "surprise" to decision-makers. The design of project structures can enable governability by triggering emergent responses as turbulence arises (Miller and Floricel, 2001). Building governability is second-order strategic thinking, in which sponsors examine each relationship and organisational device for its ability to enable appropriate responses, should turbulence arise. The following are examples of relationships and devices that sponsors deliberately use to build in governability:

1. Ownership arrangements and incentives that induce partners to respond in ways to protect their investments. For instance, when conditions change, strong ownership positions render it imperative to make the required trade-offs so as to ensure the survival of the project as a business venture.
2. The building of coalitions and the inclusion of parties that have an interest in taking actions to ensure the success of the project. For instance, lenders, such as pension funds or government agencies, will respond to calls for increased participation to cover increased costs. In this fashion, they will contribute to the robustness of the project.
3. The presence of slack, callable or deep-pocket financial resources to face needs as they arise. Owners with high stakes will respond to calls to refinance the project and modify the business model so as to ensure its viability.
4. Access to a rich array of possible strategic responses to emergent threats and opportunities. The project is likely to face unforeseeable threats and opportunities. It is, therefore, not possible to predict exactly what the best response will be. A project with a wider array of possible strategic responses is better equipped to face such emergent events. Owners and partners with extensive expertise and easy access to extended knowledge networks have more strategic options. Facing such emergent events may require the exercise of influence in the project context. Owners and partners with access to political resources have a greater number of strategic options, which may also be more powerful. When facing the unforeseen, the availability of a specific strategic option is not the most important consideration; it is the variety of possible strategic responses that is important.
5. Broad functional specifications that make it possible for owners or contractors to propose innovative solutions. For example, contractors in Public-Private Partnerships can develop novel approaches when they are provided with demand levels and quality targets, instead of constraining design choices.
6. Contracts with governments, clients, or investors that specify actions to be taken in the case of difficulties. For instance, power purchase agreements can specify in advance rate changes as a function of specific conditions. Another example is a "rendezvous clause" in the contract

specifying when, or under what conditions, a particular aspect of the project is to be re-negotiated.

7. Flexible project design and modularity that make it possible to reduce the scope and size of projects, should difficulties arise. For instance, power plants can be designed to burn many types of fuel, giving them the flexibility to adjust to and profit from variations in the cost of energy sources. Computer system projects are among the strongest users of this particular strategy. Large systems are divided into modular components that can be developed and implemented successively over time, producing value early in the project/programme and providing flexibility to adjust the scope as it progresses. If a project/programme organised in this way is terminated earlier than planned, the functionalities that have already been delivered can remain in operation. This provides additional flexibility. Projects in other sectors do not have the same possibility to modularise, but modularisation has been observed on many types of complex/large projects.

Conclusion

The examples and discussion in this chapter have attempted to show that the reality of complex projects calls for a shift in the way large projects are viewed and managed. Many leading firms have already made the shift. The approaches and the practices observed over the last decade reveal a renewed theory of project management in general and decision-making in particular. The key elements of these theories in practice can be summarised under three headings:

1. Decision-making is an interactive process among powerful and expert actors. Social actors participate in a process of mutual influence and negotiation. While some actors are much more powerful than others, the single "decision-maker" as such does not exist. The project sponsor is attempting to shape the social process and project concept together to create a viable project.
2. The decision-making process is historical; shaping episodes unfold over an extended period of time. The process is iterative, non-linear and at least partially indeterminate. The process is also historical in that (1) it is embedded in a social reality; (2) each episode builds on what was created in previous episodes but creates new options and pursues avenues that may not have been foreseen; (3) the process is largely irreversible.
3. Embedding change into the process is not evidence of bad management. On the contrary, because the process is partially indeterminate, and the project context is uncertain, emergent threats and opportunities are very likely. Complex projects need to be infused with governability, to be able to respond to threats, and to induce beneficial change.

References

Cleland D and Ireland L. (2006). *Project Management: Strategic Design and Implementation* (5th edition). McGraw Hill: New York.

Haggar K and Miller R. (2008). Managing IT projects, working paper. *Jarislowsky Chair in Project Management*. École Polytechnique: Montreal, Canada.

Miller R and Floricel S. (2001). The governability of projects. In: Miller R and Lessard D (eds). *The Strategic Management of Large Engineering Projects; Risks, Governance and Institutions*. MIT Press: Cambridge, Mass.

Miller R and Lessard D. (2001). *The Strategic Management of Large Engineering Projects; Risks, Governance and Institutions*. MIT Press: Cambridge Mass.

Miller R and Olleros X. (2007). The dynamics of games of innovation. *International Journal of Innovation Management* **11** (1): 37–64.

Mintzberg H. (1994). *The Rise and Fall of Strategic Planning*. The Free Press: New York.

Williams T. (1999). The need for new paradigms for complex projects. *International Journal of Project Management* **17** (5): 269–273.

19

Project Profitability from Society's Point of View

Kåre P. Hagen

This chapter deals with project profitability to society and how it is related to market-based commercial profitability. The first part discusses the sources of divergence between profitability to private companies and to society as a whole and how commercial assessments can be modified in order to find the social surplus. The second part deals with the role of information and risk for project assessments. It shows how the social cost of risk-bearing can be derived from the pricing of risk in private risk markets. The value of choosing flexible project concepts in order to adapt to new information about important risk factors is demonstrated in various contexts.

Introduction

Profitability depends on how efficiently scarce resources are utilised. The relevant measure of efficiency depends on the economic objective. In a private company, the objective is to maximise its contribution to the income and wealth of the owners in the short or long term. The economic profit in each period is given by the maximum distribution to the owners, without reducing the firm's market value. The value concept is forward-looking, in the sense that the market value of the company is the present value of the future cash flow that it is expected to generate for the owners. A positive present value means that the enterprise is yielding an average rate of the return on invested capital that is at least as large as the owners could have achieved by reinvesting the financial capital in the external financial market, assuming this to be the alternative investment opportunity. Hence, profitability is defined in relation to the external rate of return opportunity for the capital required to operate the company.

The rate of return in the alternative investment opportunity will therefore be a benchmark which the profitability of the company is measured against. Usually, the rate of return in the financial market is used as a hurdle rate for the internal profitability and may be modified in order to take into account the risk the owners have to bear. The alternative rate of return opportunity is an opportunity cost for investing capital in the company.

The concept of opportunity cost is central for the analysis of economic profitability. This is the case not only for capital, but also for all resources required for running an enterprise or for undertaking a particular project. Scarcity values are often called shadow prices and are related to opportunity costs in the sense that allocating a resource to a specific purpose entails an economic sacrifice as given by the opportunity cost. Hence, the opportunity cost is an implicit price that the value generated by the resource in a given enterprise has to match. For a private company, the economic sacrifice is what the firm has to pay for labour and other necessary factors of production.

When the difference between a company's sales revenue and opportunity costs is positive, it is run at an economic profit. The opportunity costs are not necessarily what the company has actually paid for its factor inputs but the values they might have generated in an external opportunity. This means that even if the difference between its revenues and payable costs is positive, it may not be profitable in the economic sense of being the most profitable way the owners could have used their scarce resources. It depends on the relevant opportunity cost. If relocating the company to another country leads to a higher capital return, the return on this external opportunity will be the relevant benchmark for its profitability. The company's book profit is based on actual costs, which can be different from opportunity costs. Hence, a firm may be unprofitable in an economic sense, even though it is run at a positive book profit.

In principle, the same sort of reasoning applies to the concept of profitability to society. The perspective may, however, be different, as there are more stakeholders with an interest in the social surplus. Society's stakeholders are those affected by the enterprise's operations in different roles and various arenas. In addition to the owners, the main stakeholders are consumers, workers and taxpayers. Hence, society's stakes in the company are broader than that of its owners. This difference in perspective has implications for how profitability to society relates to company profitability. Customers are important for the firm's profitability, as they contribute to its revenue. From society's perspective, customers contribute to the social profitability of the company through the social value derived from the consumption of goods it has produced. This implies that for domestic companies producing only for export, the business and social valuation of the firm's revenue coincide, since the export revenue accrues to the country's citizens, whereas the consumption takes place abroad.

From theory to analysis and decision

Assessing the social profitability of a given project requires an overview with respect to what the project will generate in terms of services, necessary input of production factors and the value of the services and cost of factor inputs. The valuation of output that is intended for domestic end users should be based on consumers' willingness to pay. The cost of factor inputs from domestic sources should be based on the factors' marginal value product in the most profitable alternative use. Consequently, the opportunity cost is positive if, and only if, the actual input has an alternative use in the short or longer term. Hence, the concept of social profits implies that the project will generate a positive social surplus if, and only if, it yields a larger contribution to the national value added than the input of resources would have generated in the best alternative use. Essentially the valuation problem boils down to whether market prices can be used for assessing social profitability. There are, however, several factors that complicate the connection between social opportunity costs and market prices:

Lacking markets. Market-determined values and profits as guidelines for social values and surpluses require that all scarce goods and resources in the economy are bought and sold in well-functioning markets, so that social utilities and costs of a project are fully internalised in private profitability assessments. The market system, is however, generally incomplete, in the sense that markets for some goods and inputs are lacking. One important reason for lacking markets is due to goods for which it is difficult to enforce exclusive property rights. Hence, the use or consumption of such goods cannot be individualised. If enforcement of exclusive rights to such goods is not possible, it will be difficult to collect the willingness to pay as revenue in the market. Prime examples are national defence and services related to law and order. Such services are called collective goods as they have to be consumed collectively. Another example is recreational areas to which everyone has free access. Beyond the confinement of national boundaries, there are globally collective goods, such as the ozone layer and world climate. For such globally collective goods, enforceable national property rights are neither possible nor desirable.

The absence of a market price reflecting the monetary value of users' valuation makes it difficult to assess the social value on the consumption side. The value assessment must therefore rely on other means. A common practice is to ask potential users about their willingness to pay for the service. This may be relevant for assessing the gains from, say, projects aimed at improving the environment or for assessing the environmental cost for those affected negatively by environmental degradation as a side effect.

There are also different reasons why market prices in cases where well-functioning markets exist do not necessarily provide good guidelines for assessing social profitability. The challenge is to examine how market prices should be corrected in order to reflect true willingness to pay or true social opportunity costs.

There are three main reasons why market prices do not correspond to socially optimal prices. These are external effects, natural monopoly and fiscally motivated taxes and levies.

External effects. External effects are economic consequences that the decisions of individual agents impose on other agents which (i) are not taken into account in the economic calculations governing the agent's decisions and (ii) do not go through markets. Point (i) implies that the individual market agent does not take into account the total costs or gains that can be attributed to his decisions. In such cases, private or commercial profitability does not reflect the true contribution to social profitability. In the presence of external effects, economic decisions based on market prices and commercial profits will not lead to socially optimal uses of resources. Point (ii) views external effects as a deficiency within the market system, since market prices do not capture all the welfare-related aspects of consumption and use of scarce resources in the economy. This may be due to the fact that there are scarce resources that are not traded through markets and consequently have no market prices. In this respect, the problem of assessing social profitability for projects having external effects is closely related to that caused by the lack of markets.

External effects can be negative or positive, and they can arise in transactions between consumers, producers or between producers and consumers. Positive external effects imply that decisions taken by an agent have positive effects for other agents, with the originating agent remaining uncompensated for such positive side effects through the market reward system. As such effects are not reflected in market-based revenues and profits, they will not be taken into account by commercially motivated market agents. Thus, actions inducing positive external effects will be carried out on a smaller scale compared with what is socially optimal. Conversely, negative external effects are detrimental effects on utility or productivity imposed on other agents, without those causing such external costs being held economically responsible in the market place. From society's point of view such actions are undertaken on too large a scale.

Negative external effects are where market prices exaggerate society's net willingness to pay if the consumption of an individual imposes negative effects on others, or where a company's costs undervalue total costs to society, because outputs or inputs have negative effects on other agents. Correspondingly, positive external effects are where market prices undervalue social benefits and a company's costs overvalue the true social costs. Examples of external effects are prolific. An important class of negative external effects is pollution. When a firm emits a pollutant into a river which causes negative effects on downstream users, these effects are costs to society which do not show up in the company's book accounts. These costs may be due to qualitative degradation of the river's value as a recreational resource, or to the fact that other companies, dependent on clean water from the river as an input, will be subjected to increased costs. The polluting firm will therefore undervalue the total costs in determining its optimal production level.

Natural monopoly. Natural monopoly is a term used to characterise activities where there are economies of scale in production, to the extent that average production costs decrease with increased production. The defining property implies that a single production unit will produce goods at lower average costs, compared to the situation where the production is split up and produced by two or more production units. For single items, this will be the case if variable production costs are small compared to fixed costs related to establishing and maintaining the production capacity. A larger scale production means that there are more units of the product on which to distribute the fixed costs so that the total cost per unit will decrease.

Enterprises that require large investment in infrastructure are often natural monopolies. In particular, this is usually the case for firms that are based on a physical network, e.g. airports, railways, electricity networks, water supply and sewage. The problem with natural monopolies is that, in the absence of price regulation, the monopoly power can be used to set the price above marginal and average cost, in order to capture a larger part of the consumer surplus. As this will be a pure redistribution from the buyers of the monopoly's services to the owners of the monopoly, this monopoly profit will not contribute to the value added in society. On the contrary, by raising the price above the marginal production cost, the sum of the consumer surplus and producer profit is reduced. The monopoly will therefore set a higher price and produce a lower volume than where the social value is maximised, by marginal cost pricing. Price equal to marginal cost will, however, entail a commercial deficit, as marginal cost is less than average cost. If the monopoly is required to recoup its total costs in the market, optimally regulated prices will maximise total consumer surplus, subject to the constraint that its total costs are covered.

If the natural monopoly exerts its power in supplying inputs to other domestic companies, the price of the inputs will be higher than the marginal social sacrifice shown in the marginal cost. This leads to repercussions in the form of production volumes that are too small and social efficiency losses in enterprises that are dependent on inputs from domestic natural monopolies. In such cases, the problem arises as to what price should be used in assessing the social profitability of projects that depend on inputs from domestic natural monopolies being incompletely regulated. There are two issues here. One is the market monopoly price. The other is the social sacrifice, as given by the actual marginal cost in production. If the monopoly were optimally regulated, it would be natural to use the regulated market price so that private and public companies dependent on the monopolised supply of the actual input could use the same price for the input in their profitability assessment. In this case the monopoly supply would be allocated efficiently among public and private firms, if the latter have no market power.

Taxes and levies. Because of taxes and levies, the price the producer receives for selling a good or a service is less than the price paid by the buyer. This difference constitutes the tax or levy imposed on the good. Conversely, for goods and services that are subsidised, the price which the producer receives is higher than the price the buyer pays. Assuming the good is subject to a unit tax equal to t levied on the producer, and that the tax exclusive price is p, the buyer will pay $p + t$, while the seller receives p. Assuming the market is perfectly competitive so that the producer keeps producing until the tax exclusive price is equal to marginal cost, while the consumers buy until the marginal willingness to pay is equal to the tax inclusive price, the marginal willingness to pay in consumption will be higher than the marginal production cost in the market equilibrium. Due to the tax, the profit maximising volume will be less than that maximising the total social surplus, which results in an efficiency loss. The efficiency loss is due to the fact that the private producer considers the unit tax as a cost, whereas, from a social point of view, it is a transfer from the firm to the government. The commercial profitability of increased production is therefore less than what it is to society so that the production volume will be too small from a social point of view.

In this case, the question will also arise as to what price should be used in a social profitability analysis – the tax inclusive or exclusive price. As an example, this problem may be considered in connection with calculating the cost to society for using domestic labour. If the wage per time unit is w and the wage tax is t, the worker will receive $w(1-t)$. If, for the sake of the argument, it is assumed that the labour supply is perfectly competitive, workers will supply labour until the marginal sacrifice in terms of reduced leisure (measured in monetary terms) is equal to the after tax wage. The company, on the other hand, will demand labour until the value of the marginal product of labour equals the wage before tax. Here the question also arises as to what wage should be established when examining the social profitability of the project. The same type of problem arises regarding the treatment of payroll taxes levied on the employer.

Since taxes and levies cause a wedge between marginal willingness to pay on the demand side and marginal costs on the supply side, financing projects by taxes will entail a social financing cost, due to the resulting efficiency loss. In Norway, the official estimate of average social tax cost is 20% per krone of taxes.[1] This is called the social cost of public funds, as it will be the social cost of financing a project by tax revenue. This means that when financed by taxes, the tax cost of a public project amounts to 20% of the net revenue effect on the overall public budget.[2] The alternative to tax financing is financing the project in the market through user charges if that is possible or by a profit margin on the services from the project. This, too, entails an efficiency loss, as the price faced by the users is higher than the actual marginal cost on the supply side. From an efficiency point of view splitting the financing between financing in the market and financing by taxes may in many cases minimise the total financing costs. The criterion for the optimum financing mix would, then, be that the marginal efficiency loss per krone is equal for both modes of finance.

The sequel discusses how to make the valuation principles operative and how they are to be implemented in practice. Also discussed are valuation questions relating to large projects, how to treat incomplete information and uncertainty about the relation between inputs and outputs and the desirability of retaining flexibility in carrying out projects.

Outlining the project's boundaries and identifying alternatives

Profitability analysis is a tool for supporting decisions as to uses of scarce resources and is relevant only to the extent that the required resources have alternative uses. If there are no alternatives to using the resources in the project, there is no problem of choice, as there are no degrees of freedom as to the uses of inputs. The alternative use of resources may be in an alternative project, which will then be a benchmark, or it may be the initial use of the resources – often referred to as the base alternative. The project under scrutiny and the base alternative are mutually exclusive projects competing for scarce resources. Most projects have a time dimension both as to output and costs. The project is then defined in terms of the time path for its services, resource requirements and resource costs over its lifetime. In such cases, the base alternative is simply the time path of the initial situation without the project. This is often referred to as the reference path, with which the time path of the project in question is to be compared.

Some projects have a local feature, in the sense that they affect resource uses and outputs in a given market or within a given sector. A typical example is the building of a local road. The service rendered by the project is the gain in time for local traffic, and the costs are dictated by the construction cost. This describes the total consequences of the project. Other projects are large, with nationwide repercussions across several sectors. In such cases, it can be difficult to distinguish between new activities that are generated by the project and relocation of existing activities in space and time. New activities induced by the project should be treated as an integral part of the project, whereas spatially relocated activities affect primarily the regional income distribution. However, if such induced activities produce goods or employ resources that are not allocated through the market, or where market prices are distorted, the value of such externalities must be added to the benefit or cost side of the project, depending on whether it is an external gain or cost. For example, if building a railway between two cities leads to a relocation of ancillary activities to an area with structural unemployment problems, the increased activities reduce the efficiency loss in the local labour market, which is a positive effect that should be added to the project's benefit side.

The relation between market prices and marginal willingness to pay and social opportunity costs

The value of goods produced for domestic private consumption is dictated by the consumers' willingness to pay for such goods. For a given consumer, the marginal willingness to pay will normally fall with increased quantity of consumption so that the average willingness to pay is higher than the marginal willingness. The theory of consumer behaviour assumes that the consumer demands quantities of a given good until the marginal willingness to pay for that good is equal to the market price. In equilibrium the market price will reflect the consumer's marginal willingness to pay. If the project has a negligible effect on the market price, the correct price to be used in a social profit assessment is the price faced by consumers in the market. This implies that for goods that are subject to fiscally motivated levies or taxes, the price should include taxes and levies, since the tax revenue is the share of consumers' willingness to pay what is collected by the government. On the contrary, if the excise taxes are motivated by negative external effects in consumption, the price should be the market price net of such levies, as they represent the external social cost of consuming such goods and should therefore be deducted from the individual willingness to pay. An example of such a tax is that on gasoline. For goods that are being exported, the correct price should be the net export price, assuming that the world market price is not affected by the project. This means that it should be net of export taxes or subsidies, as these are neither national income nor national costs.

Generally, the value of collective goods is dictated by the sum of individual willingness to pay by all potential users. When the consumption cannot be individualised and consumers are not excludable, there will not be any markets for such goods and hence no market prices that can convey information about individual willingness to pay. Willingness to pay and social values will then have to be assessed without any support from market prices. Value assessments should distinguish between direct and indirect methods. Direct methods are based on interviews with potential users about their willingness to pay for the supply of a specified set of collective goods. This value assessment is hypothetical, in the sense that the respondent is asked to express his/her willingness to pay for particular collective goods if they were to be supplied. This hypothetical nature of the valuation of the goods is the main weakness with the direct method, since the expressed value is not part of the actual budget constraint so that the respondent is not faced with a market trade-off between the public goods in question and private market goods.

Indirect methods utilise the fact that in many cases there are market goods that are perfect complements to collective goods. The market valuation of such perfect complements may then capture the valuation of the collective goods. Such market complements can be private goods or activities that are required in order to get access to collective goods. This might be the case for many locally collective goods, e.g. where access to a good housing environment can only be obtained by buying a house in the actual area. The valuation of the collective amenity will then be capitalised in the market value of houses located in this area.

Market prices of inputs and social costs

Here it may be useful to use the analogy of a company's assessment of the cost for inputs that are partly bought in the market and partly procured by deliveries within the company. For a country, the international market will be the analogy of the external market. For imports, the economic sacrifice for the country will be the import price net of duties and other levies. Levies on imports are a pure transfer from the importing firm to the government, and the country at large will neither benefit nor lose from such transfers.

If the importing company is a government agency supplying a monopoly service, the above reasoning would undoubtedly hold true. However, if it is a private enterprise, the import duty would be an expense for the company and would be regarded as part of the costs for the imported input. If the private company was competing with the public agency, the import duty would give the public agency a competitive advantage, due to the treatment of the import duty and not due to the public agency being more cost efficient than the private firm. The possibility that private companies might be crowded out by less efficient public enterprises, due to different treatment of import duties in their profitability calculations, would result in a social cost and should be taken into consideration when laying down rules for the treatment of import duties in public enterprises. A practical and simple way of doing this would be through a two-step procedure. First one should decide whether competing private alternatives to the public project exist. If the answer is in the affirmative, the public enterprise should use the import price including the import levy. If not, the import price net of the levy should be used.

A similar reasoning applies to the assessment of the social cost of using resources that could alternatively have been exported. The most important case for Norway in this respect is the social cost of using natural gas as input in the domestic production of electrical power. If this does not affect the price of Norwegian natural gas in the international market, the export price represents the economic sacrifice for the country. If using natural gas domestically for electric power generation causes negative external effects for Norwegian citizens, the monetary value of this externality should be included as part of the input costs. As to natural gas in particular, most of the pollution is due to emission of CO_2. In this case, the consequences will be global, so that they will only be borne by Norwegian citizens to a limited extent. If the only consideration is the welfare effects for Norwegian citizens, these global effects should not be included in the cost benefit analysis. However, Norway has taken upon itself to observe a maximum limit for her CO_2 emissions. To the extent that this limit is binding, the national emission costs for a power plant fuelled by natural gas will be the negative effects on the value added, as other production activities have to reduce their CO_2 emissions. This will then be the national emission cost. If there was a well-functioning domestic market for emission quotas, this negative external effect would be reflected in the quota price. This is yet another example of the fact that social profitability analyses are simpler in economies with well-developed markets.

The remaining problem lies in assessing the social cost of factor inputs from the sheltered sector, i.e. situations where the import opportunity is not available or feasible. In the present situation, this is the case for certain segments of the labour market. However, due to income taxation, the social cost of labour depends on whether it is seen from the supply or the demand side in the labour market. The economic sacrifice of using labour is its social opportunity cost. If the labour market is healthy and import of labour is ruled out, increased input of labour has to be covered either by increased supply, by reduced input of labour in other domestic activities or by a combination of the two. In the first case the sacrifice is in terms of reduced leisure and, in the latter case, in terms of reduced value added in the alternative employment. However, the presence of a wage tax drives a wedge between the marginal opportunity value of labour in the two alternative uses of the workers' time endowment. In a competitive labour market, workers will offer labour until the marginal value of leisure foregone is equal to the after-tax wage. Letting w note the pre-tax wage and assuming a wage tax of 50%, the after-tax wage is $0.5w$. This is then the marginal monetary shadow price of leisure. Employers will demand labour until the gross wage including the payroll tax is equal to the marginal value product of labour. Setting the payroll tax at 0.15, the marginal value of labour on the production side will be equal to $1.15w$ in equilibrium. Thus, if the opportunity cost on the production side is the relevant social cost of labour, projects that have a positive effect on labour supply will be more profitable from a social point of view, compared with projects that draw labour away from existing production activities.

However, if taxes were set optimally initially, and there is no involuntary unemployment, it would not be reasonable to consider increased employment as a gain in the social profitability of a project. If this was the case, the tax system could not have been optimally designed in the first place, and it would be more efficient to attain this positive effect on employment by a tax change. According to this reasoning, one should assess the social cost of using labour in a public project as if the project was threatening available labour in private employment. This means that one should use the gross wage including the payroll tax as the social cost. Competitive neutrality is thus achieved between private and public use of labour.

The required rate of return and the public discount rate

Many public projects are capital intensive. This is the case within communications, energy supply, water and sewerage systems. The cost of using capital will therefore be an important factor for the profitability of such projects. In principle, the cost of binding capital in a project should reflect its opportunity cost, which in this case is the social rate of return on capital in the best alternative investment opportunity. In a closed economy with the domestic capital market in equilibrium, capital for a specific investment can be obtained in three ways: by increased domestic saving, or by reduced capital investment in alternative production activities or even by a combination of the two. The cost of providing capital by increased saving is compensation in the form of the capital returns savers will require in order to be willing to defer consumption. If consumer savings are privately optimal, the required compensation is given by the real after-tax interest rate. Alternatively, capital can be provided by private companies reducing their investments. The social cost of this is the reduced value added from the private use of capital. If the taxation of companies is neutral with respect to their investments, the opportunity cost is given by the real rate of return before tax. However, given that the capital market is initially in equilibrium, the provision of capital, either by means of increased savings or reduced private investments, presupposes that the project must lead to a higher real interest rate. If the project is sufficiently large and capital is scarce, this may not be entirely unrealistic. However, today, with access to an international capital market with full capital mobility, the domestic interest rate level is tied to the interest rate level in the international capital market. In such a scenario, it is quite improbable that even large domestic projects will have any interest rate effect on the domestic use of capital. Hence, in an open economy, the interest rate in the international capital market will be the relevant opportunity cost for domestic use of capital. This will also be the case for public projects.

Accounting for risk in the benefit-cost analysis

Capital intensive projects have often a long duration. This means that their profitability may depend on distant future factors about which there is considerable uncertainty at the time when the decision on the realisation of the project has to be taken. The uncertainty may partly be due to factors and events that are beyond the control of the decision-maker. It might be future rainfall for a water power plant, the development of air traffic in the case of a new airport, etc. The uncertainty may also be due to factors that can be influenced by the decision-maker to some extent, either by the project's design, or by its realisation or by collection of information about important factors for its profitability that are uncertain at the time of the decision.

It is commonly assumed that when it comes to important economic decisions, most people have risk aversion. One implication of risk aversion is that, with the mathematical expectation of an uncertain result as reference point, the decision-maker assigns a larger weight on negative deviations than on comparable positive ones. If the result is measured in monetary terms, this means that the safety equivalent result is lower than the expected value, and the difference reflects the compensation one requires in order to be willing to bear the economic risk associated with the decision. Thus, risk-bearing involves a cost, and this should somehow be included in the cost-benefit analysis of the project. For risks beyond the control of the decision-maker, the question is how this cost should be assessed in quantitative terms and how it should be accounted for in the analysis. For risks that can be influenced by collecting relevant information prior to the start up of the project, or by choosing more flexible project concepts, the problem is to assess the cost and gains from collecting additional information or from choosing more flexible concepts that are more adjustable to new information. The first type of risk is exogenous, in the sense that the decision-maker has to take it as given, whereas the second type depends to some degree on how the project is designed and carried out.

Exogenous risk and project profitability

It is a basic insight from the portfolio approach (see e.g. Markowitz, 1959) to investment under uncertainty that the risk of the project cannot be regarded in isolation, rather it should be seen in the light of its contribution to the total economic risk that the decision-maker bears. This is given by the risk associated with the returns on the total investment portfolio. In other words, the risk of a given project depends on what portfolio it is a part of and what the decision-maker can do in order to hedge against such risk. An important distinction in this respect is that of unsystematic and systematic risk.

Unsystematic risk is sometimes referred to as specific risk. This kind of risk affects a very small number of projects in the portfolio. An example is an event that affects a specific project, such as a strike by the employees working on the project. Systematic risk, on the other hand, influences a large number of projects. A significant political event or an unexpected general rise in the wage level is an example of systematic risk.

The impact of unsystematic risk is reduced when the project is included in a larger portfolio, where negative and positive variations in the returns of the underlying assets tend to cancel each other out. The scope for reducing the risk on the total portfolio by investing in several projects depends, therefore, on how the returns on the individual projects are correlated. Where the returns of a specific project are perfectly negatively correlated with the returns on the total portfolio net of the project, investing in the project can eliminate the risk of the portfolio and will act as a perfect hedge. If the returns are negatively but not perfectly correlated, the portfolio risk will be reduced but not eliminated. Only where the returns are uncorrelated with those on the portfolio, is a project's contribution to the portfolio risk given by its "stand alone" risk. Projects with returns that are positively correlated with the portfolio returns will have an effective contribution to the portfolio risk, which is larger than the projects' own risk.

Systematic risk characterises projects for which the profitability depends on one or more uncertain common factors. Examples are wage levels, import prices and the international interest rate, etc. Due to such common uncertain cost factors the projects' returns will be positively correlated. Systematic risk cannot be reduced by diversification, i.e. spreading the investment over several projects. If there are efficient markets for diversifying unsystematic risk, the profitability assessment only has to take into account a project's contribution to systematic risk.

For the public sector investing in many different projects, the risk of the public investment portfolio will be given by its systematic risk. If the government evaluates its investments according to their social returns, the relevant risk for a project will be given by its effect on the variability of the total national value added. This in turn will be given by the national income for the social surplus that is created in markets. Thus, from a national perspective, it is likely that most unsystematic risk will be washed away in the national portfolio of investment projects.

The market price of risk and the required social rate of return

In principle, it makes no difference whether one discounts expected project surpluses with a risk-adjusted discount rate or discounts certainty equivalent surpluses with the riskless rate of interest, as the former can be deduced from the latter. Generally, it is preferable to relate social shadow prices to market prices, so that objective price information in the market can be used in the social profitability analysis, if such prices exist. In the stock market, the risk is implicitly priced in the form of a required risk-adjusted rate of return. When market information about the cost of bearing risk is given by risk-adjusted rates of return, a project's risk is more readily accounted for in the cost-benefit analysis by using a risk-adjusted discount rate, rather than a risk-adjusted assessment of the project's surpluses.

The risk of an investment in a given portfolio can be quantified by the co-variance between the investment's rate of return and that of the portfolio. Using the statistical variance of the portfolio returns as a risk measure, the ratio between this co-variance and the variance of the total portfolio rate of return will express the project risk as a share of the portfolio risk. This is called the investment's beta (β). If beta equals zero, this means that the investment's contribution to the portfolio risk is zero. In this case, the investment is, in fact, riskless, as the portfolio risk remains unchanged. The case $\beta = 1$ means that the investment has exactly the same risk profile as the total portfolio. Interpreting each investment as an equity share, and the total portfolio as all shares traded in the stock market – the so-called market portfolio – the beta value of a given share will show how the share's risky return co-varies with the return of the market portfolio. Denoting the risk-free rate of return in the market by r and the expected return on the market portfolio by $E(R)$, then $E(R){-}r$ is the expected excess return required by the market investors in order to be willing to hold the market portfolio. Multiplying this expected excess return by the share's beta value shows the excess return required by the market in order to hold the share in question. Denoting the market's required risk-adjusted return on this share by k, we get the investment criterion in terms of a risk-adjusted hurdle rate:[3]

$$k = r + \beta[E(R){-}r].$$

It should be noted that for $\beta = 0$ the required rate of return will be equal to the riskless interest rate, while for $\beta = 1$ it will be equal to $E(R)$, which is the required expected rate of return on the market portfolio.

This approach can be also be used for assessing the risk-adjusted required social rate of return on investment projects generally. The analogy with the market portfolio will now be the total investment portfolio in the country, the rate of return on which is given by the national income. With the basic assumption that the risk profile of the stock market portfolio is representative of the risk profile of the country's total investment portfolio, the stock market pricing of risk can be used as a basis for the social pricing of risk. The social hurdle rate of return can then be determined by finding the beta-value that represents the actual investment's risk profile.

An essential assumption for basing the social cost of risk-bearing on the pricing of risk in the stock market is that the risk profile of the rate of return on the market portfolio is approximately representative of the risk profile of the national income, reflecting the total returns on the national investment portfolio. Moreover, the risk aversion of those trading in listed assets must be representative of the risk aversion of those bearing the social risk associated with real investments. In practice, relying on stock market data implies that, for a given investment, one must find a listed risk copy in the stock market – i.e. one with the same risk profile – and then use the beta-value of this risk copy. This procedure presupposes that the hurdle rate of return of the market portfolio also reflects the social hurdle rate of return for investments with the same risk profile as that of the market portfolio.

However, there is an important difference between the rate of return in the stock market and the rate of return that is relevant for society. The stock market value and the derived market price of risk relate to the returns that accrue to the shareholders of the underlying companies. The social returns, however, also include the share of the returns that accrues to those who have provided loans to the company and, in addition, the share that accrues to the government as corporate taxes. In order to make the social returns on an investment comparable in terms of risk with the rate of returns priced in the stock market, one should use the risk of the total returns before corporate taxes in listed companies. The task will then be to find a listed company with a risk profile for its total returns being representative of the risk profile of the investment in question, which may be assumed to be a public investment. The market's hurdle rate for the total returns before taxes for this company will be a market-based hurdle rate for this public investment. As the share holders bear most of the economic risk in corporations, it seems plausible that the beta-value for the total returns of the company must be lower than that for the equity returns. Assuming that the company's debt is completely risk-free (absence of default risk), the beta-value of the returns on total capital will be $\alpha\beta_{E'}$ where α is the share of equity capital, and β_E is the beta-value for the returns to equity. However, it must also be noted that the excess return $E(R)-r$, which the market requires in order to hold the market portfolio, is derived from stock market data, and hence it is net of the corporate income tax and the personal taxes on the returns to shareholders.

The required excess return is usually referred to as the market's risk premium, and MP^t denotes this risk premium after personal and corporate income tax. Assuming that the corporate income tax rate is equal to the personal capital income tax rate,[4] $MP^t = E(R^t)-r(1-t)$, where the superscript denotes after-tax value, and t is this common tax rate. The hurdle rate of return to equity after tax, k_E^t, will then be $k_E^t = (1-t)r + \beta_E\, MP^t$. Disregarding that lenders[5] may also bear some of the business risk, the pre-tax risk-adjusted hurdle rate of return on total capital will then be given by

$$k_{\text{Tot}} = \frac{k_E^t}{1-t} = r + \frac{\alpha\beta_E}{1-t} MP^t.$$

Assuming a risk-free real rate of interest of 2.5%, an average share of equity in listed companies equal to 0.4, a risk premium after taxes in the stock market of 4.5%, and a tax rate of 28%, there is a risk-adjusted real hurdle rate of return on total capital of 5%. This hurdle rate applies to public investments and to investments undertaken by non-listed firms having a risk profile of the total returns corresponding to that of a representative listed company.

Project evaluation at an early stage

At an early stage in the planning and profitability assessment of a project, one will typically face a demand for a given type of service, e.g. a transportation service, and the aim is to find a project concept that satisfies this specific need in the most efficient way. This preliminary analysis may take the form of a cost-efficiency analysis, to find the project concept that satisfies the given need at the lowest cost. Such a cost-efficiency analysis may however be complicated by the fact that future factors on the utility and cost side of the project may vary in an unpredictable way from that anticipated in the preliminary analysis.

The role of the discount rate for the profitability analysis at the preliminary stage

It should be emphasised that the problem of assessing a project's profitability at an early stage is mainly due to the lack of information about the factors that are most essential for its profitability. This requires collection of further information and choice of a project concept that is sufficiently flexible to draw advantage from additional information. Compensating this lack of information by adjusting the discount rate for the resulting risk seems to be somewhat off the mark. On the other hand, the hurdle rate of return as given by the discount rate is a benchmark that has to be matched by the rate of return of a profitable project.

A given hurdle rate of return has the effect that potential projects are sorted into two classes: one class of projects that are profitable and another class of projects that do not generate the required rate of return. Having rather scarce information about the profitability of the projects in question may lead to a considerable risk in terms of sorting errors. There are two types of erroneous sorting. One accepts the project – or the project concept – when it should have been rejected. In analogy with statistical inference theory, this may be called a Type II error. The other type relates to projects that are rejected when they should have been accepted – a Type I error. Given that information is initially incomplete, both types of errors will occur. If the criterion for acceptance is that the expected present value of a profitable project must be non-negative, the relative frequency of the two types of error will depend on the discount rate reflecting the required rate of return. With a higher discount rate, fewer projects will pass the profitability hurdle, and there will be fewer Type II errors, while there will be more Type I errors. Conversely, a lower discount rate will lead to more projects being classified as profitable, and there will be relatively more Type II errors, and fewer of Type I. In this perspective, the size of the hurdle rate of return as the criterion for sorting will depend on what type of error is considered to be most serious. With specifically irreversible projects, it may be important to avoid Type II errors, i.e. to avoid accepting projects that are shown to be unprofitable, given more accurate information. This is an argument for a high hurdle rate in the preliminary screening between profitable and unprofitable project concepts.[6]

Generally, the less information available regarding its true profitability, the more risky the project will be. The risk term in the risk-adjusted discount rate depends on the project's systematic risk. At the early stage, the systematic risk can be influenced by the choice of the project design. Hence, all being equal, it will be important to find project concepts where profitability is less dependent on the general state of the economy (the business cycle). It may be argued that Norwegian citizens are overly exposed to oil-related risk. Hence, from the perspective of social risk, it may be sensible to choose project concepts for which the social profitability is less dependent on the price of oil. In this way, the risk and the risk-adjusted hurdle rate will depend on the project design and is, to some degree, endogenously determined.

Endogenous risk: collecting information and the scope for utilising new information in designing and timing of the project

Most investment projects are irreversible. In particular, this is the case with transport projects such as roads, tunnels and airports. This implies that the larger part of the investment cost can be recovered once the investment is undertaken. Secondly, important drivers for the profitability of the project may be uncertain at the time of investment. If part of the uncertainty is of the milestone type, it will be revealed with the passing of time. Thirdly, there is often a choice as to timing and to the way the project is carried out. It might, for example, be sensible to wait for the milestone risk to be resolved before taking irreversible decisions. Optimal timing will depend on the cost of waiting relative to the value of keeping the investment decision open (keeping the option open). The wait-and-see decision requires a method for calculating what the investment option is worth, i.e. the value of keeping open the option of not undertaking the project. It may also be possible to carry out the project in a stepwise manner, assuming additional information will be accrued during the project period.

Investing in a new airport may serve as an example where the most important risk factor is uncertainty about the future development of air traffic. In this case a stepwise realisation of the project may be sensible. This can take place by first building one runway and allowing the question of two runways to depend on the development of the traffic. A stepwise realisation will normally be more costly compared to realising the project in full scale from the outset. However, this additional cost should be considered as a sort of insurance premium against the unfavourable outcome of having permanently excess capacity in the case of an unfavourable development of air traffic. The profitability of a flexible solution with a stepwise construction schedule will depend on how much additional information will be gained from waiting. The degree of the investor's risk aversion is also important here.

Collecting information prior to the decision

Expected values of future project surpluses are based on available information at the time of the decision. Usually, it is an implicit assumption that the information collected prior to the decision has been optimised. If not, the project risk can be reduced by collecting additional information before the decision is taken. Generally, it will be profitable to collect project relevant information, as long as the expected value of additional information is greater than the additional cost. The value of new information is that it is likely to change the initially optimal decision in a positive way. Making this operational requires that it is possible to revise the probabilities for the relevant outcomes in the light of new information. However, it is important to note that for the additional information to have any positive value, the probability must be greater than zero for it to lead to a changed decision. If the optimal decision relative to the existing information about critical factors will not be affected by new information, the additional information has no economic value.

Assuming that the project in question is to build a tunnel through a hill from A to B, and the cost depends on how suitable the rock is for the building of tunnels, if the choice is between such a tunnel and an alternative connection without a tunnel, information about the nature of the rock might be important for the choice. In such a situation it might be profitable to spend resources on test drillings in order to reveal the actual conditions for building a tunnel. On the contrary, if there is no suitable alternative to the tunnel, and it has to be built anyway, this information is of no importance for the decision, even though it might be important for planning the project.

Consequently, additional information will have the greatest value in the planning phase of a project where the possibilities for adapting to the new information are highest. Once the project is being carried out and various irreversible decisions are taken, the possibilities for project changes are consequently reduced, and hence new information will be of less value. An irreversible project that, at the time of decision-making would have been deemed unprofitable in the light of ex-post information, might therefore be profitable to operate once it has been completed as the investment costs are not recoverable in any case.

The theory of real options as an approach to investment under uncertainty

A feasible option is to postpone the decision on accepting the project, pending additional information. If the project is profitable relative to initial information, postponement means that reaping the expected value created by the project is simply deferred. This creates a waiting cost if the decision-maker has a positive time preference. The reason for deferring the decision on carrying out the project must be to obtain additional information about the profitability of the project in time or that it will become less costly to collect such information at a later point. This would be particularly the case for milestone risk that is resolved at a given point in the future. However, it would not be a valid reason for postponing the project, if it were possible to get back to the initial situation without incurring any costs, i.e. to the situation prior to the decision to undertake the project, if the information collected at a later stage revealed that an unprofitable project had been chosen. On the other hand, if the project is irreversible, e.g. because it necessitates investment of the sunk cost type that cannot be recovered even if the project is closed down, carrying out the project immediately would entail a cost in the form of lost decision flexibility in a future situation with new information. Thus, postponing the decision to carry out the project gives an *option value* of maintaining the option of rejecting the project, if the updated information shows it to be unprofitable. This can happen if an unfavourable outcome of the random factor underlying the milestone risk is realised. If the option value is positive, a positive risk-adjusted expected present value of the project would not be a sufficient condition for the immediate realisation to be profitable. The risk-adjusted present value has to be larger than the option value of deferring the decision on realisation, which is foregone by immediate realisation. Hence, the option value will be an opportunity cost for the decision on immediate realisation.

Summing up, there are three sources of uncertainty in an investment analysis. One is the uncertainty about future project surpluses, another is uncertainty about the investment cost and a third is uncertainty about the opportunity cost of capital that is being locked up in the investment.

Uncertainty about future project surpluses. Assume that either a given project can be implemented immediately or the decision can be postponed for 1 year. Let A_0 denote realising the project immediately and A_1 in 1 year. The investment costs are NOK 540 m. The annual net surpluses depend on the realisation of two possible, mutually exclusive random outcomes. One outcome gives an infinite stream of annual net cash flows of NOK 45 m. and the other infinite annual cash flows of NOK 15 m. The two outcomes are equally probable and the discount rate (hurdle rate of return) is 5%. The decision-maker is maximising net expected present value. We let $E[N(A_i)]$ denote expected net present value of project A_i, $i = 0, 1$. Hence

$$E\,[N(A_0)] = \frac{0.5 \cdot 15 + 0.5 \cdot 45}{0.05} - 540 = 60^7.$$

Realising the project immediately gives an expected present value of 60 m., and based on the conventional present value criterion, it is profitable. If the decision is postponed for 1 year the present value with the unfavourable outcome will be $(15/0.05) - 540 = -$ NOK 240 m. and the project would not be undertaken. If the favourable outcome materialises, the present value will be $(45/0.05) - 540 =$ NOK 360 m. Hence, by postponing the project it will only be undertaken in the favourable state. The expected present value assessed at the initial date will then be $E\,[N(A_1)] = 0.5\,\frac{45/0.05-540}{1.05} = 171$ m. which is twice as large as the present value of deciding immediately.

The reason for this is that by postponing the decision, one is preventing the outcome entailing a loss since the uncertainty is resolved after one period. The difference between the present value of waiting and that of deciding at once is the option value of maintaining the decision flexibility until the true state is disclosed. In this example, the option value is NOK 111 m. If the option value is treated as an opportunity cost for the decision on immediate realisation, this alternative is clearly not profitable. We may also note that realising the project immediately is the best decision, if the choice is between now or never, as the project seen from today yields an expected rate of return that is larger than the required rate. However, if the alternative is to wait for another year before deciding, realisation today is not the optimal one.

It is, however, conceivable that postponing the realisation of the project for 1 year may lead to higher investment costs. One might then examine the maximal increase in the investment cost without violating the profitability of the wait-and-see alternative. Denoting this critical investment cost by $I_{\max}$, it is given by

$$60 = (0.5)\,\frac{45/0.05 - I_{\max}}{1.05}$$

which yields $I_{\max}$ = NOK 786 m. That means that as long as the increase in the investment costs is less than 246 m., the additional cost may be seen as a profitable investment in decision flexibility as given by the choice of implementing now or in 1 year. Similarly, a stepwise implementation of the project instead of implementing the project in full scale may be possible. This would also give an option value as the capacity can be adjusted to new information about the need and demand for the services of the project. One could then calculate how large an increase in cost that is tolerable with stepwise implementation without making implementation in full scale at once the most preferred alternative.

The concept of real option value in concept assessment is explored further in Sunbeams' chapter, "The Impact of New Information", in which closer attention is paid to the latter topics.

Uncertainty about the investment cost. Uncertainty about investment cost is common to capital intensive projects with a long construction period. This is often the case within the energy sector, with large hydroelectric projects, and for power plants fuelled by natural gas, where the investment requirements associated with developing new technologies for handling CO_2 are uncertain. The above example examines this problem and is modified by assuming that future project surpluses are certain and equal to NOK[8] 45 m. per year, the cost of investing the first year is known with certainty, while investment cost the following year is uncertain. Assuming that immediate implementation requires an investment of NOK 800 m., this yields a present value of NOK 100 m. In 1 year, the investment cost is assumed to be NOK 1200 m. with probability equal to 0.5, and NOK 400 m. with probability 0.5, which means an expected investment cost of 800 m. This is the same as in the alternative of immediate implementation. The project should only be realised if the favourable outcome occurs. Expected present value will then be $0.5 \cdot$ NOK 500 m., and discounting the expected value yields $250/1.05 =$ NOK 238 m. so that the wait-and-see alternative is clearly better than realising the project immediately.

Uncertainty about the required rate of return (the discount rate). As a last example, it is assumed that investment costs and future project surpluses are known with certainty, while the future discount rate varies in an unpredictable way. The discount rate uncertainty will have two effects. Firstly, it will have a favourable effect on the present value. To illustrate this, it can be assumed that the time horizon of the project is infinite and that the uncertain discount rate can take two values – 5% or 15% with equal probabilities. The expected discount rate will then be 10%. Assuming that the project yields a net surplus of 1 Norwegian krone per period for the infinite future, the present value of the future income series evaluated at the discount rate of 10% is 10 NOK. However, the expected present value when the discount rate might be either 5% or 15%, with equal probabilities, is given by $0.5 \cdot 1/0.05 + 0.5 \cdot 1/0.15 = 13.33 > 10$. Hence, an uncertain discount rate renders the project more attractive, if maximum expected present value is the investment criterion.[9]

Secondly, even though the discount rate uncertainty is favourable for the expected present value, it might still be profitable to postpone the decision to implement the project until this uncertainty is resolved. This is due to the fact that the wait-and-see option enables the decision-maker to avoid the loss outcome given by a high realised value of the discount rate. Hence, the uncertainty about the discount rate will have the same effect on the decision to postpone as uncertainty about future project surpluses. This is shown by modifying the example above. It is now assumed that the project surpluses are known and equal to NOK 45 m. per year and that the investment cost is NOK 540 m. The project will be in operation 1 year after the investment decision, and the discount rate will be 5% or 15% with equal probabilities. The project values are then $45/0.05 = 900$ and $45/0.15 = 300$ with equal odds. Hence, the expected net present value of the project is NOK 600 m. – NOK 540 m. = NOK 60 m. so that the project is profitable according to the expected present value criterion. If the discount rate had been equal to its expected value with certainty, the net present value would have been $450 - 540 = -$ NOK 90 m., and hence the project would have been unprofitable. If the decision to invest is postponed by 1 year, i.e. until the discount rate is known, the project would not be undertaken if the discount rate turns out to be 15%. If the investment decision is postponed, the expected present value will be

$$0.5\left(\frac{45/0.05-540}{1.05}\right)=171$$

which is considerably larger than the expected value of taking the decision immediately at the beginning of the project.

To sum up, increased variability in the discount rate so that the expected rate remains unchanged ("mean preserving spread") will increase the present value of the project. Furthermore, it will be more profitable to postpone the decision as to whether or not to invest in the project. The reason for this is that the alternative of waiting enables the decision-maker to avoid loss states with high future rate of return requirements; the probability of ending up in loss states increases with the variability of the discount rate.

Concluding remarks

Social profitability analysis of projects in a market economy will be based on market prices, to the extent that these are readily available. If the market system is complete and well-functioning, this will also be conducive to a socially optimal choice of projects. However, in such an ideal economy, there is less need for social re-examination of private profitability calculations, as these two profitability concepts coincide. Hence, the raison d'être for a separate social cost benefit analysis to examine the social profitability of projects is that markets are lacking or market prices are distorted so that they do not reflect true social values and costs. The first part of this chapter discussed how to assess social values and costs without the support of market prices and how to correct prices when they are misleading because of external effects or distortions.

The second part discussed the problems raised by incomplete information for the choice of project concept. If project investments are irreversible, and the expediency of a given project depends on uncertain factors, where the uncertainty is likely to be resolved at a later date, this might call for keeping the option open as to the choice of concept, by postponing the decision until the relevant information is revealed. Thus the project analysis would take place in two stages. In the first stage the project concept is chosen, and in the second stage the profitability of the chosen project is calculated. In practice the decisions at these two stages will be connected. In two-stage planning one usually starts with the last stage, which in the present case means calculating the profitability of the various project concepts under proposal. Going back to the first stage, one then chooses the project concept and the timing of its realisation, by comparing net social project values with their time-dependent option values.

Notes

1. See Finansdepartementet: NOU 1997: 27 and Finansdepartementet 2005.
2. If the net effects are positive, the tax cost will be negative too. This might be the case with building a kindergarten that could result in an increase in the female labour supply and an increased income tax revenue exceeding the project costs.
3. The theory behind this formula expressing the hurdle rate of return for investment in shares with risky returns is the so-called capital market pricing model. See Mossin (1966, 1969) for a discussion of the underlying assumptions and the derivation of the investment criterion.
4. This is the case in the Norwegian tax system.
5. We assume that the interests on loans accrue to domestic lenders so that it must be considered a part of the national income.
6. It may be argued that the relative frequency of the two types of decision errors depends on the way the screening of projects is organised. A hierarchical structure with several decision gates will minimise the prevalence of Type II errors, whereas a flatter structure with delegated decisions will minimise Type I errors (Sah and Stiglitz, 1986).
7. We have here used the fact that the present value of an infinite series of cash payments with a discount rate r is given by $NV = \frac{1}{1+r} + \frac{1}{(1+r)^2} + \ldots + \frac{1}{(1+r)^n} + \ldots = \frac{1}{r}$.
8. 1 Norwegian krone = 1 NOK.
9. Technically, this follows from the so-called Jensens' inequality stating that the expected value of a convex function of a random variable is larger than the function value evaluated at the expected value of the variable. The result follows then from the fact that the present value function V/r is a convex function of r.

References

Finansdepartementet. (1997). Prinsipper for lønnsomhetsvurderinger i offentlig sektor. NOU:27.

Finansdepartementet. (2005). Veileder i samfunnsøkonomiske analyser.

Markowitz Harry M. (1959). *Portfolio Selection: Efficient Diversification of Investments*. Blackwell/John Wiley & Sons: Oxford/New Jersey.

Mossin J. (1966). Equilibrium in a capital asset market. *Econometrica* **34**:768–783.

Mossin J. (1969). Security pricing and investment criteria in competitive markets. *American Economic Review* **59**:749–756.

Sah R and Stiglitz J. (1986). The architecture of economic systems: hierarchies and polyarchies. *American Economic Review* **76**:716–727.

Index