Alistair Munro

Series Editor Ian J. Bateman

The Economics of Non-Market Goods and Resources 12

Bounded Rationality and Public Policy

A Perspective from Behavioural Economics

Springer

Bounded Rationality and Public Policy

THE ECONOMICS OF NON-MARKET GOODS AND RESOURCES

VOLUME 12

Series Editor: Dr. Ian J. Bateman

Dr. Ian J. Bateman is Professor of Environmental Economics at the School of Environmental Sciences, University of East Anglia (UEA) and directs the research theme Innovation in Decision Support (Tools and Methods) within the Programme on Environmental Decision Making (PEDM) at the Centre for Social and Economic Research on the Global Environment (CSERGE), UEA. The PEDM is funded by the UK Economic and Social Research Council. Professor Bateman is also a member of the Centre for the Economic and Behavioural Analysis of Risk and Decision (CEBARD) at UEA and Executive Editor of Environmental and Resource Economics, an international journal published in cooperation with the European Association of Environmental and Resource Economists. (EAERE).

Aims and Scope

The volumes which comprise *The Economics of Non-Market Goods and Resources* series have been specially commissioned to bring a new perspective to the greatest economic challenge facing society in the 21st Century; the successful incorporation of non-market goods within economic decision making. Only by addressing the complexity of the underlying issues raised by such a task can society hope to redirect global economies onto paths of sustainable development. To this end the series combines and contrasts perspectives from environmental, ecological and resource economics and contains a variety of volumes which will appeal to students, researchers, and decision makers at a range of expertise levels. The series will initially address two themes, the first examining the ways in which economists assess the value of non-market goods, the second looking at approaches to the sustainable use and management of such goods. These will be supplemented with further texts examining the fundamental theoretical and applied problems raised by public good decision making.

For further information about the series and how to order, please visit our Website
http://www.springer.com/series/5919

Bounded Rationality and Public Policy

A Perspective from Behavioural Economics

Alistair Munro
National Graduate Institute for Policy Studies,
Tokyo, Japan

and

Department of Economics, Royal Holloway, University of London,
Egham, Surrey, UK

Springer

Alistair Munro
National Graduate Institute for Policy Studies
7-22-1 Roppongi, Minato-ku
Tokyo, 106-8677 Japan

And

Department of Economics
Royal Holloway, University of London
Egham, Surrey, TW20 0EX, UK

ISSN 1571-487X
ISBN 978-1-4020-9472-9 e-ISBN 978-1-4020-9473-6
DOI 10.1007/978-1-4020-9473-6
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2008940653

© Springer Science + Business Media B.V. 2009
No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Cover Image: © 2008 JupiterImage Corporation
Cover design: Integra Software Services Pvt. Ltd.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

For my daughter, Hana, whose rationality becomes less bounded with each day.

Preface

This book is about bounded rationality and public policy. It is written from the perspective of someone trained in public economics who has encountered the enormous literature on experiments in decision-making and wonders what implications it has for the normative aspects of public policy. Though there are a few new results or models, to a large degree the book is synthetic in tone, bringing together disparate literatures and seeking some accommodation between them.

It has had a long genesis. It began with a draft of a few chapters in 2000, but has expanded in scope and size as the literature on behavioural economics has grown. At some point I realised that the geometric growth of behavioural research and the arithmetic growth of my writing were inconsistent with an ambition to be exhaustive. As such therefore I have concentrated on particular areas of behavioural economics and bounded rationality. The resulting book is laid out as follows:

Chapter 1 provides an overview of the rest of the book, goes through some basic definitions and identifies themes.

Chapter 2 is devoted to a survey of some of the evidence on anomalies. There are several excellent summaries of this literature including those by Colin Camerer (1995) and by Chris Starmer (2000). My aim is not to be systematic therefore, but to show that the evidence is substantial and that it points to predictable deviations from rational behaviour in many arenas of economic activity. The emphasis in this chapter is on anomalies found in static choice situations. Discussion of dynamic choice anomalies, learning and those associated with information processing are found later in the book.

Chapter 3 is a sequel and companion to Chapter 2 in the material it contains: I discuss evidence and theories about learning. The key issue here is whether repeated choice opportunities and the framing created by the marketplace lead to the elimination of anomalies. Some evidence on information processing is therefore also in this chapter.

While Chapter 3 considers the impact of markets on anomalies, Chapter 4 examines how deviations from consistency affect the organisation of markets. The theme in this chapter is that bounded rationality can say something about market behaviour and in some situations claims about the efficiency of markets can be made.

Chapter 5 is the welfare chapter. I tackle the tricky and possibly intractable issue of the normative implications of the data discussed in Chapter 3, looking at alternatives to standard welfare economics.

Chapter 6 is a summary chapter on public policy and bounded rationality. A central purpose is to clarify the relationship between bounded rationality on the one hand and traditional notions of market failure on the other hand. I also consider the direct implications of bounded rationality for the role of the state.

In Chapter 7 I review some of the existing literature on merit wants. The purpose is draw out its major conclusions and to show its limitations as a theory of policy. In particular I argue that its main weakness is in the general absence of articulated models of why preferences deviate from individual welfare.

Chapter 8 provides a discussion of the neglected role of agents. The fact that someone else knows what is best for another individual does not mean that they will be motivated to reveal that knowledge or, if given the power to act, that they will make choices which raise the welfare of their principal. While government or a group of experts may hold better information about welfare in some cases, in many other cases it is close friends or family who know best. In Chapter 8 I focus on the second case, which has been largely neglected in the literature on agency.

The main focus of Chapter 9 is tax policy when the policy maker has only limited information on individual welfare rankings. In what sense can we still promote efficiency as a desirable economic goal? Are some tax changes still likely to be desirable?

While Chapter 9 is devoted to manipulations of the budget set, Chapter 10 focuses on manipulations of the frame as an instrument of policy. I also discuss some intertemporal anomalies here in the context of savings policy, an area that has seen much interest from behavioural economists.

Chapter 11 concludes the major arguments of the book with a practical application of some of the arguments from Chapter 5 applied to the problem of non-market valuation. Much debate has occurred in recent years on the future of stated preference methods in the face of evidence on their unreliability. The conclusion drawn here is that decisions on the usefulness of contingent valuation and its allied methods for valuation should be comparative – i.e. based on an evaluation of the costs and benefits of alternative methods of making decisions. So, we compare stated preference methods to its alternatives and rivals in a framework drawn from decision theory.

In terms of routes through the book, readers familiar with the literature on anomalies may wish to skip Chapters 2–4 and proceed to Chapter 5. Readers particularly interested in non-market valuation may find it useful to proceed via Chapters 2–5 and then directly to Chapter 11. Tax policy is to the fore in Chapters 7, 9 and 10.

As with any research enterprise I have picked up intellectual debts along the way. Mick Common and Nick Hanley at the University of Stirling provided my first exposure to environmental economics and contingent valuation. The experimental group at University of East Anglia, notably Robin Cubitt, Chris Starmer and Bob Sugden deserve special thanks for introducing me to the startling notion that theories of economic behaviour were testable in the laboratory. Last but not least, Ian

Bateman has been a close collaborator for many years and from him I have learnt a great deal about environmental valuation in particular and research in general.

Egham, Surrey, UK *Alistair Munro*

Contents

1 Introduction 1
1.1 Merit Wants 3
1.1.1 The Link Between Bounded Rationality and Merit Wants 3
1.1.2 Other-Regarding Behaviour and Expressive Preferences . . 4
1.2 Definitions 5
1.2.1 Frames 6
1.2.2 Frames and Welfare 8
1.2.3 Uncertainty and Information Processing 10
1.2.4 Merit Goods 11
1.3 Main Points 13

2 Anomalies 17
2.1 Introduction 17
2.2 The Reliability of the Evidence 18
2.3 Reference Dependent Preferences 22
2.3.1 Organising the Evidence 32
2.3.2 Prospect Theory – Cumulative and Riskless 33
2.4 Mental Accounting 39
2.5 Preference Reversal 44
2.6 Conclusions 47

3 Information, Learning and Markets 49
3.1 Introduction 49
3.2 Probabilities, Information Processing and Bayes' Theorem 50
3.2.1 Representativeness and Availability Biases 50
3.2.2 Adjustment and Anchoring 51
3.2.3 Confirmation Biases 54
3.3 Ambiguity Aversion and the Ellsberg Paradox 56
3.3.1 Summing Up Bayes' Theorem 58
3.4 How Markets Might Lead to Individually Rational Behaviour 58

3.4.1 Summing Up Market Learning 64
3.5 Learning 65
3.5.1 Melioration 69
3.6 Conclusions 73

4 Markets and Reference Dependent Preferences 75
4.1 Introduction 75
4.2 Partial Competitive Equilibrium 76
4.3 Imperfect Competition 80
4.4 General Equilibrium 83
4.5 Reference Dependent Preferences 85
4.5.1 Equilibrium and General Equilibrium 87
4.5.2 Discussion 90
4.6 Conclusions 91

5 Welfare 93
5.1 Introduction 93
5.2 Frames and Paretian Welfare Economics 95
5.3 Responses to the Evidence 100
5.3.1 Switch Theories 100
5.3.2 Allow Individual Welfare to be Frame Dependent 101
5.3.3 No Ordering 102
5.4 Optimal or Golden Frames 103
5.4.1 Markets and Referenda as Optimal Frames 105
5.5 Information Aggregation 108
5.5.1 Extra-Preference Information – Relaxing WEF 111
5.5.2 Bayesian 112
5.6 Paternalism 114
5.6.1 The Justification of Paternalistic Intervention 114
5.6.2 Paternalism: The Contractarian View 121
5.6.3 Summing Up 122
5.6.4 Ranking Frames Using Consent and Other Arguments from Paternalism 123
5.6.5 Hypothetical Rational Consent 127
5.6.6 Contractarian Consent 128
5.7 Conclusions 130

6 Public Policy and Bounded Rationality 133
6.1 Introduction 133
6.1.1 Bounded Rationality: Identification and Valued Added . . . 133
6.2 Market Failure 138
6.2.1 Asymmetric Information 139
6.2.2 Near-Optimality 139
6.2.3 The Boundaries of the State 142

6.3 Voting for the Nanny State: Regulation of Markets by Self Aware Consumers . . . 148
6.4 Political Economy – An Illustration . . . 151
6.4.1 Efficiency and Bias . . . 152
6.4.2 Term Limits . . . 155
6.4.3 Candidacy . . . 156
6.5 Conclusions: Possible Worlds . . . 159
6.6 Appendix . . . 160

7 Standard Fiscal Policy and Merit Wants . . . 163
7.1 Introduction . . . 163
7.2 Individualistic Welfare Functions, Private Failure . . . 164
7.2.1 Individualistic Welfare Functions, Public Failure . . . 167
7.2.2 Two Spurious Categories of Merit Wants . . . 168
7.2.3 Non-individualistic Welfare Functions . . . 169
7.3 Objective Functions . . . 170
7.3.1 Risk and Uncertainty . . . 172
7.4 Optimal Policies . . . 172
7.4.1 Wrong Information Models . . . 179
7.4.2 Limits of the Literature . . . 179
7.5 The Value of Information . . . 180
7.5.1 The Value of Misinformation . . . 183
7.6 Conclusion . . . 185

8 Agency and Dependency . . . 187
8.1 Introduction . . . 187
8.2 Dependency Relationships: Children . . . 189
8.2.1 The Child's Development of Economic Ideas . . . 190
8.3 The Role of the State . . . 191
8.3.1 Relief from Duty . . . 192
8.3.2 Information Asymmetry . . . 193
8.4 Regulating Care with Asymmetric Information . . . 195
8.4.1 Compulsory State Provision . . . 197
8.4.2 Minimum Standards and Public Provision . . . 197
8.4.3 Voluntary State Provision of Care . . . 198
8.4.4 Mixed Policies . . . 199
8.4.5 Detection and Enforcement . . . 200
8.5 The Role of the State in Regulating Adulthood . . . 202
8.5.1 Welfare Functions with Children . . . 203
8.5.2 Optimal Policy . . . 204
8.6 Conclusions . . . 206

9 Tax Policy . . . 207
9.1 Introduction . . . 207

9.2 Monotonicity 207
9.3 Technical Efficiency – Marketed Goods 208
9.3.1 Technical Efficiency – Non-Marketed Goods 211
9.3.2 Cost Minimization 211
9.3.3 Summary 213
9.4 Tax Reform Issues and Loss Aversion 213
9.4.1 The Problems Created by Reference Dependent Preferences 213
9.4.2 Extensions to Consumer Theory 216
9.4.3 Implications of the Theory for Welfare Economics 218
9.4.4 Implications of the Theory for Tax Reform Models 220
9.4.5 Optimal Taxation 221
9.4.6 Final Thoughts 223
9.5 Conclusions 224

10 Framing Matters: Non-Standard Fiscal Policy 227
10.1 Introduction 227
10.2 Creating Mental Accounts 228
10.2.1 Theory 229
10.3 Intertemporal Issues: Pensions and Savings Policy 236
10.3.1 Intertemporal Choice and Self-Control 237
10.4 Regulating the Framing of Savings and Pensions Choices 242
10.4.1 Planning Ahead 245
10.5 The Framing of Tax Policies 249
10.5.1 Tax Evasion 251
10.5.2 Labour Supply and Tax Framing 252
10.6 Conclusion 256

11 Stated Preference and Non-Market Decisions 259
11.1 Introduction 259
11.2 A Formal Approach to Optimal Decisions 263
11.2.1 Part-Whole and Scope Effects 265
11.2.2 Elicitation Effects 267
11.2.3 Responding to Anomalies 270
11.3 Formal Comparison 275
11.3.1 Example 1: WTA Versus WTP 278
11.3.2 Example 2: Elicitation Methods 281
11.4 Summing Up 282

References 285

Index 313

List of Tables

2.1 Empirical Evidence on Reference Dependent Preferences 23
3.1 Representativeness Heuristics 51
3.2 Availability Heuristics 52
3.3 Payoffs in a Learning Environment 70
3.4 Learning with Regret 71
3.5 Bayesian Learning versus Reinforcement 71
4.1 Equilibrium Outcomes with a Decoy Brand 82
4.2 Reference Dependent Preferences 84
6.1 Political Economy Models and Bounded Rationality 135
7.1 Types in the Racionero Model 175
8.1 The Game of Guilt 192
9.1 LA and NS 218
10.1 Non-Hicksian Policies 228
10.2 Framing Tax Systems 249
10.3 Framing Income-Leisure Choices 254
10.4 Framing WTC Incentives 255
11.1 The Optimal Frame – NOAA Panel Recommendations 273
11.2 Evidence on the Coefficient of Population Variation, B 279
11.3 Critical Values of n for Equivalence of CV and ICJ 279
11.4 Mean WTP Values and Critical Jury Sizes for Different Acceptable Sets 282

List of Figures

1.1 Mapping the Territory of Merit Wants 4
1.2 Decoy Effects and Beans 8
2.1 WTP, WTA and Substitution 28
2.2 Reference Points 29
2.3 Organising the Evidence 32
2.4 The Value Function in Prospect Theory 35
2.5 The Probability Weighting Function 36
2.6 Mental Accounting and the Timing of Consumption 42
3.1 Willingness to Pay and Social Security Numbers 64
3.2 Bayesian Error Rates in an Experiment 72
4.1 Experienced Demand 77
4.2 Equilibrium Cycles 78
4.3 Evolution of Prices 79
4.4 Markets with Decoys 82
5.1 Framing and Welfare Economics 98
5.2 Intransitivity of Welfare Indifference 110
5.3 Information and Aggregation 111
5.4 A Hierarchy of Preference Information 120
5.5 Optimal Endowment with One-Sided Learning 130
6.1 Bounded Rationality and Pareto Optimality 140
6.2 Upper Hemi-Continuity of E and No Market Failure 142
6.3 Learning and the State 146
6.4 The Impact of Learning on the Boundaries of the State 147
7.1 Merit Wants and Merit People 167
7.2 Merit Wants and Incentive Compatibility 176
7.3 Incentive Compatibility and Merit Wants II 178
7.4 The Value of (Mis)Information 181
8.1 Regulation and Altruism 188
8.2 Minimum State Provision 198
8.3 Regulation of Carers 201
9.1 Ex Post Feasibility of Tax Reforms 215
9.2 Axioms on Loss Aversion 217
9.3 Optimal Taxation and Reference Dependent Preferences 222

10.1 Prospect Theory and the Winter Fuel Payment 231
10.2 The Effect of a Mental Account on Choice 233
10.3 The Effect of Labelling on Heating 234
10.4 The Value of Procrastination 241
10.5 Choices in Thaler and Benartzi (2004) 244
10.6 Factors Affecting Planning Ahead 246
10.7 Income-Leisure Trade-off 253
11.1 Decision Methods 261
11.2 Optimal Estimators 264
11.3 Part-Whole Effects in Pizza Valuation 266
11.4 The Effect of a Starting Point on Valuations 270
11.5 WTP for Teabags 270
11.6 A Comparison of Three Methods 278
11.7 Optimal Decisions for $s = 0.4$ and $n = 25$ 281

Chapter 1
Introduction

A government is considering changes to the nation's railway system. New technology will improve safety, but at a cost of £2 bn to be paid through higher fares or taxes. Two reputable companies are commissioned to find out if the benefit justifies the cost. The first discovers that aggregate willingness to pay is only £1.5 bn and therefore recommends that the safety improvements should not be implemented. The second company reports that citizens would require minimum compensation of £3 bn in order to forego the safety changes and therefore concludes in favour of the improvements. What should the government do?

An individual knows that there is a ten percent chance that, in the later stages of life, she will become senile and incapable of managing of her own affairs. Were that to occur she knows she would want a close friend or relative to make decisions on her behalf. While she trusts her family and most of her friends, looking around her she sees many carers who are incompetent or insensitive to the needs of their charges. Some are downright cruel. What regulatory policy would she advocate for government?

Fertilizer runoff into local streams and lakes means that the social cost of an arable crop lie above its private cost. One proposal is to tax the fertilizer in order to bring social and private marginal costs into line. However, a civil servant has noted that a fraction of the population who buy the crop seem to exhibit inconsistent preferences between all sorts of choices. The question is therefore whether the uncosted pollution actually represents a market failure, a distortion that should be remedied. Doesn't such a classification presuppose rationality?

The starting place for this book is Paretian welfare economics , which has formed the foundation of most economists' thinking about public policy issues for over half a century. The examples given above are typical public policy dilemmas yet they lie outside the scope of a Paretian welfare economics that is founded on three important principles: One is that all individuals have complete and consistent preferences that are in accordance with individual welfare. The second is the monotonicity property – social welfare is increasing in individual welfare. The third element is the exclusivity property, denoted 'neutrality' in Sen (1970) which holds that social welfare is exclusively determined by individual welfare. This book relaxes the first of the principles. It does so because there is abundant evidence that individuals do not

A. Munro, *Bounded Rationality and Public Policy*, The Economics of Non-Market Goods and Resources 12, DOI 10.1007/978-1-4020-9473-6_1,
© Springer Science+Business Media B.V. 2009

always have well-defined and consistent preferences and that, even when they do have preferences, they do not always choose what is best for their own welfare.

The ingredient added is bounded rationality which has many definitions. Some of them are discussed in the next section, but for the most part I identify bounded rationality with behaviour that is not wholly explicable in terms in terms of the satisfaction of complete and consistent preferences. To understand the relevance of bounded rationality for public policy, the book draws on two literatures, both of which are expanding at an explosive pace. The first of these deals with experiments on individual decision-making. Much of the research in this area crosses the boundaries between economics and psychology and has thrown up a serious and sustained challenge to the notion that individuals always have well-defined and consistent preferences. But deviations from the precepts of rationality appear not to be random. Instead behaviour is predictable, at least to a degree, on the basis of the type of decision problem faced by individuals and by context. The significance of these experimental results is widely contested, particularly with regard to their relevance to 'economic' environments such as real world markets. Nevertheless there is substantial enough supporting evidence from real world phenomena to suggest that a maintained hypothesis of unboundedly rational individuals is wrong.

The second literature is one more closely connected to public policy: it concerns valuation, particularly the methods used in environmental economics, but also the similar and complementary techniques used in sub-disciplines such as health and transport economics where cost benefit analysis is extensively deployed. Over the last forty years economists and other social scientists have grappled with the problem of how to elicit preferences for non-market goods. Most commonly used are stated preference techniques such as contingent valuation. Despite prolonged and careful refinement of the techniques involved, valuation methods continue to throw up large volumes of evidence about preferences that are almost impossible to reconcile with standard models of the consumer. As with experiments on decision-making, much of this evidence is controversial, but what is notable is its consistency with that gathered in experiments in decision-making.

As the opening example suggests, for project appraisers and others offering policy advice on costs and benefits the failure to produce reliable and context free figures for valuation raises serious questions. Should economists attempt to refine their methods further in the hope of obtaining consistent figures for values? Instead, should they cease doing cost-benefit analysis at least for some types of goods and projects? Should they offer a range of values or should they adopt one particular method for eliciting values and stick to it?

The results of investigations into decision-making have given rise to a new area of the subject, behavioural economics, which attempts to explain economic phenomena in terms of empirically rigorous models of individual behaviour. It has had some success in explaining the anomalies thrown up in markets, particularly in financial environments (see Shefrin and Thaler, 1988 and Shefrin, 1999 for example), where the approach has been labelled behavioural finance. The successes of behavioural finance suggest the benefits of a parallel approach (behavioural public finance (Krishna and Slemrod 2003) as it were) within public economics wherein some of the issues to be tackled include:

- How is it possible to make judgements about individual welfare when the information we have about preferences is often inconsistent?
- How can we predict individual and aggregate behaviour when standard models of the consumer are inaccurate?
- To what extent are policy makers best modelled as boundedly rational actors?

It is the task of this book to provide some initial answers to these questions.

1.1 Merit Wants

Before we had behavioural economics and bounded rationality there were merit wants. We might know that it is better to save for a rainy day, but still splash out on an impulse buy that will be the source of regrets tomorrow; we might opt for a complex medical procedure without any clear idea of whether it will cure us; we might smoke, even though we would rather not. Each of these is an example of a merit want or merit good – a situation where there is a difference between what we want or choose and what is good for us. Merit goods matter because their existence represents a potential limit to the ability of the marketplace to serve human welfare; a source of market failure. They also matter in a more practical sense, because goods often labelled meritorious, such as healthcare, pensions and education, represent a substantial component of public expenditure (and therefore GNP) in most countries. Healthcare, pensions, and education together make up over a third of UK government expenditure for instance, dwarfing the amounts spent on public goods such as defence. Moreover, other areas of expenditure (such as the personal social services and housing) also involve a substantial component of what is often viewed as merit good spending. Thus an understanding of merit goods is central to an understanding of non-marketed goods. Of course many goods may have multiple market failures. Education for instance can have external benefits, while arguably government provision of its finance is Pareto improving because of failures in human capital markets. And of course intervention may be driven by the rent seeking motives outlined by public choice theorists. Nevertheless there are many areas of government activity, particularly with regard to the 'nanny state' where other market failures struggle to provide an explanation as sensible as that offered by the merit wants story.

1.1.1 The Link Between Bounded Rationality and Merit Wants

As mentioned above, my working definition of bounded rationality is the failure to have complete and consistent preferences. Though there are other definitions, this is the simplest and probably the one most widely-used in economics. Meanwhile, merit wants are goods where individual preference or choice fails to accord to with individual welfare. In economic theory, the weak individual welfare relationship is usually taken to be complete and consistent, so someone who is boundedly rational is also therefore meritworthy, in the sense that there must be at least some comparisons between bundles of goods where individual preference rankings do not match

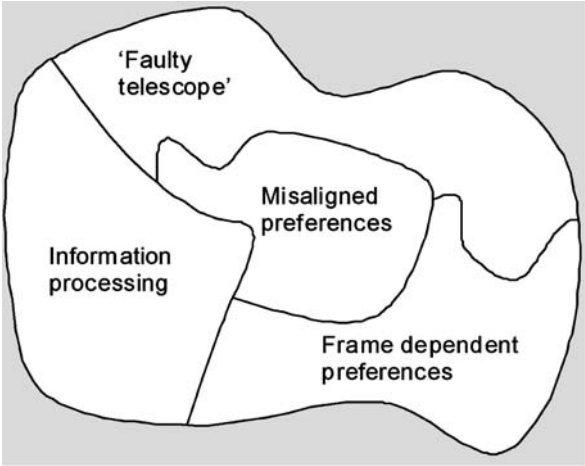


Fig. 1.1 Mapping the Territory of Merit Wants

with the individual welfare ranking. Logically, bounded rationality is therefore a special case of merit wants, but the stress in the book will be on the empirical link between the two concepts. One of the central problems with the merit wants literature is that it rarely attempts to explain the reasons why choice or preference does not accord with welfare or why supplying information might not remedy the consequences of information deficiencies. Nor does it provide much in the way of hard evidence that individual's choices are indeed sometimes sub-optimal. The data gathered on information processing and decision-making through economic experiments can provide meat on the otherwise bare bones of the theory of merit wants.

To change metaphors, Fig. 1.1 provides a cartographic guide to the links between behavioural economics and merit wants . By using the insights of the former we can divide up merit wants into different regions.[1] Failures of information processing provide one reason why individuals do not choose what is best for themselves; a faulty perspective on the future is its neighbouring territory etc. However, the region labelled 'misaligned preferences' in the map marks an unexplored core of the merit wants concept and leaves open the possibility that behavioural economics can only take us so far in explaining why individuals might prefer options that do not maximize their own welfare.

1.1.2 Other-Regarding Behaviour and Expressive Preferences

The textbook model of homo economicus is often criticized on two grounds: its failure to recognise the irrationalities of human behaviour and the lack of acknowledgement that we may care about things other than our own consumption. Altruism

[1] These explanations of behaviour may overlap. An individual who fails to use Bayes' Theorem to update information optimally, may also be prone to framing effects.

and other aspects of other-regarding behaviour are undoubtedly important, but this book largely ignores them on the grounds that no axioms of weak rationality are broken by such preferences. There are some occasions though, when I shall bring in some discussion of altruism. The first place where altruism is to the fore is in the discussion of the delegation of responsibilities for choice. Typically, when individuals are incapable of making rational choices, the power to make those choices is handed by the state first to members of the immediate family or, on occasion, to close friends. Little sense can be made of such a policy unless we incorporate some aspect of other-regarding behaviour into the analysis.

The second situation where other regarding behaviour may be important for this book is in the discussion of experimental results, some of which appear to indicate irrationality on the part of subjects, unless some note of other-regarding behaviour is taken. Recall the well-known ultimatum game, in which individual A must make an offer to individual B over how to split £1. Individual B can only accept or refuse the offer. If B accepts then the £1 is split according to A's proposal, but otherwise both players receive nothing. In experiments, it is common to observe that a significant proportion of 'B' subjects reject divisions of 90 pence to A and 10 pence to B or similar. Now most people prefer 10 pence to nothing, and when given the choice between the two would, in other circumstances, choose the 10 pence. Apparently therefore many individuals have inconsistent preferences. However, it is more likely that the choice between 10 pence and nothing is different depending on the context. When it is proposed by another individual who will then take away 90 pence it invokes feelings of injustice and encourages a spiteful response to the miserly offer. Most of the experimental results which will be used in this book arise from individual decision tasks, not in an interactive or game theoretic context. Hence other-regarding motives are unlikely to be significant in explanations of the results. But at least some of the stranger evidence, particularly in the case of contingent valuation exercises, can be explained, at least in part, by other-regarding motives such as the desire to preserve a wilderness for future generations.

1.2 Definitions

Broadly speaking rationality has two connotations.[2] The weaker notion is that of consistency, either in preference or in choice (and hence in revealed preference).

[2] Rationality can be defined on the basis of preference or on the basis of choice. In many policy contexts, the evidence on individual attributes that one might use to judge or test rationality can be a mix of revealed preference (i.e. choices) and stated preferences. For instance, the fact that a family has chosen to visit lake or wetland may reveal something about their willingness to pay for environmental quality, but a contingent valuation survey might also produce data on their preferences over environmental improvements. Moreover in theory choice and preference may differ, suggesting the necessity for some fairly clumsy definitions of rationality that cover both choice and preference together with the link between the two. For reasons of simplicity, I shall avoid that route, although in one or two places in the book I come back to the difference.

The wider notion of rationality – perhaps more common outside of economics – concerns the fitness of means to ends and also to those ends themselves. For instance Elster, 1986, proposes the notion of 'broad rationality' 'that allows a scrutiny of the substantive nature of the desires and beliefs involved in actions'(p. 15). McFadden (1999, p. 14) compares rationality 'in the broad meaning of sensible, planned, and consistent,' to the narrow version usually employed in economic analysis that equates to consistency. Within broad theories the rationality of the sun worshipper who holidays on the chilly beaches of Britain each year may be questioned, as is the behaviour of the drug user or the anorexic. But these choices cannot be interpreted as irrational in weaker notions of rationality, provided behaviour is consistent. Theories of merit wants therefore require notions of broad rationality if they are to allow the possibility that consistent choice is irrational. In Fig. 1.1 this corresponds to the region marked 'misaligned preference', where choices are consistent, but do not meet some broader or stronger notion of rationality.

1.2.1 Frames

Before getting into formal definitions of rationality, we need to have some preliminary discussion of frames. But before that I begin by defining some basic notation and concepts. Let there be H households and n goods and let X be the set of *social states*. A social state is an m vector with m $\geq nH$; which includes within it at least a list of consumption vectors for each individual, but it may contain other elements.[3] Let $\preceq_i$ be the weak preference relationship for individual $i = 1, \ldots, H$. Strict preference ($\prec_i$) and indifference ($\sim_i$) are defined from $\preceq_i$ in the usual manner.

Completeness (A1). For all x, y∈ X eitherx $\preceq_i$ *y or y* $\preceq_i$ *x.*
Transitivity (A2). For all, x, y and z∈ X,x $\preceq_i$ *y and y* $\preceq_i$ *z* $\rightarrow$ *x* $\preceq_i$ *z .*
If both A1 and A2 hold then $\preceq_i$ is an ordering.

Now consider the following famous example, reported in Tversky and Kahneman, 1981:

> Imagine that the US is preparing for the outbreak of an unusual Asian disease , which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences are as follow.

The subsequent part of the task differs between the two different sub-samples, I and II. Subjects in group I face a choice of options:

- If program A is adopted, 200 people will be saved.
- If programme B is adopted there is a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved.

[3] This is simply to allow the possibility of non-preference information in welfare evaluation.

Subjects in group II face the following choices:

- If program C is adopted, 400 people will die.
- If programme D is adopted there is a 1/3 probability that nobody will die and a 2/3 probability that 600 people will die.

When faced with these choices, 72 percent of subjects in group I opted for option A; 22 percent of subjects in group II preferred C to D. How does this square with A1 and A2? There are three possibilities. One is to take the view that there are four options in total in the examples and that $A \neq C$ and $B \neq D$. If this is the case A1 and A2 are not violated, but at the cost of trivialising their content. A second option is to accept that $A = C$ and that $B = D$ and conclude that for at least 50% of individuals that $A \sim_i B$. Of course this may be the case. But it is always possible to manipulate A say, so that fewer lives are lost and establish strict preference in the two examples. That leaves the third option that, at least for some individuals, $A \prec_i B$ and $D \prec_i C$, and so A1 fails.

Tversky and Kahneman's Asian disease problem is an example of a *framing effect*: the way a decision problem is presented to subjects affects the choices in a manner that suggests preferences differ between the two contexts. The concept of framing is central to this book. Having said that, what constitutes a frame is elusive. Tversky and Kahneman (1999) simply state 'the same option, however, may be framed or described in different ways' (p. 4).[4] In other words, frames differ, but the underlying decision problem remains the same. With some cases we can easily accept that the effect of a change in context can alter choices in a manner consistent with notions of rationality. For instance, if subjects in group II receive the information in Japanese then for a monolingual English speaker it is reasonable to conclude that the two contexts provide different information to the speakers. On the other hand, anyone with basic arithmetical skills can appreciate the logical equivalence of the two frames used by Tversky and Kahneman so it is difficult to sustain an argument that there is really is different information embodied in the two variants.

The notion of a frame is a wide one. The choice set itself may affect preferences and an interesting example of such is provided by Doyle et al. (1999), who investigated purchases of baked beans in a supermarket (see Fig. 1.2). For one week, the team monitored sales of two brands of beans, both sold in the same large sized tin. One brand, X, accounted for just 19% of sales despite being cheaper than the leading brand. The researchers then introduced a third good, namely small tins of X sold at the same price as the large can. Sales figures from the following week showed that, while (unsurprisingly) no purchases of the decoy were made, market share of X had increased significantly ($p = 0.034$) to 33% of sales, with the market share of the leading brand falling from 81 to 67%. The introduction of the small tin of X is an

[4] Earlier they write 'we use the term 'decision frame' to refer to the decision-maker's conception of the acts, outcomes and contingencies associated with a *particular* choice', Tversky and Kahneman, 1981, p. 453 (my emphasis).

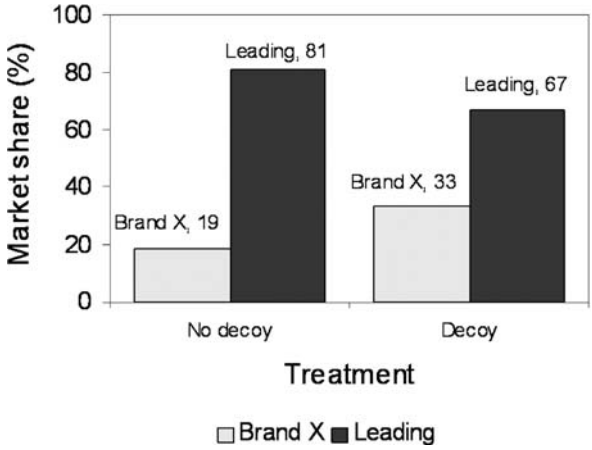


Fig. 1.2 Decoy Effects and Beans

example of a *decoy* effect. The decoy changes the way that the choice is framed and as a result shifts preferences for at least some customers.

More generally, we can think of there being a set of frames F, in each of which there is a preference between x and y. Typical members of the set are f, f′ and g and the preferences of an individual may depend on which element of F is employed. I use $\preceq_{fi}$ to indicate a preference relationship for individual i in the context of frame i. Different theories of economic behaviour are then alternative models of the relationship between f and $\preceq_{fi}$. In the traditional Hicksian model of the consumer for instance, $\preceq_{fi}$ is independent of f. Meanwhile in theories of status quo or endowment effects the current consumption bundle is a determinant of preferences.

1.2.2 Frames and Welfare

Having defined frames, the next step is to define some standard notions of preference and individual welfare in a world where, at least potentially, framing matters.

Axiom 1.1 (Preference (PF)) . $\forall i \in H, f \in F, \preceq_{fi}$ *is an ordering.*

Axiom 1.2 (Individual frame neutrality (IFN)) $\forall f, g \in F, x, y \in X, x \preceq_{fi} y \rightarrow x \preceq_{gi} y$.

These axioms are behavioural postulates. They state that all individuals have complete, transitive and reflexive preferences and that these preferences are not affected by the frame - in which case we can write $\preceq_i$ to denote the weak preference relationship.

Weak rationality is then identified with the existence of coherent preferences or choice.

Definition 1.1 Weak Rationality (WR). An individual is weakly rational if s/he satisfies PF and IFN

Recall that there are some important notions of rationality that go beyond simple consistency, often labelled broad or strong. Strong rationality we define by introducing welfare, specifically the notion that preferences and welfare yield the same ranking. I use the subscript w to denote the welfare relationship.

Axiom 1.3 (Welfare Equivalence (WEF)) . $\forall x, y \in X, f \in F, i \in H, \left(x \preceq_{fi} y\right) \longleftrightarrow \left(x \preceq_{fwi} y\right)$.

Definition 1.2 Strong Rationality (SR). An individual is strongly rational if s/he is weakly rational and also satisfies WEF.

This definition does not say much about what constitutes individual welfare, only that it entails a consistent and complete ranking of all options. In fact, one can think of a large number of criteria which are potentially informative as to individual welfare . Information might be derived directly from stated or revealed preferences, from individual empathy or from scientific information about, for example, the minimum number of calories necessary for survival. For the moment it need not concern us how welfare is constructed or the particular criteria employed. What is notable is that framing effects, where individuals prefer x to y in one context, but prefer y to x in other contexts violate the weak rationality postulate and therefore strong rationality as well.

What then is bounded rationality? The definitions of rationality given above suggest that there may be two kinds of definitions of bounded rationality as well, one involving the negation of SR and one involving the negation of WR. Following Simon (1955) there is another approach in which it is usual to view bounded rationality as something more than the absence of rationality. Instead the intention of the individual is maximizing but outcomes are bounded by the rationality of the processes. Although this is a fairly common approach to the concept of bounded rationality it is not universal. In his survey on bounded rationality and individual decision-making, Camerer (1998) identifies bounded rationality with the absence of weak rationality and makes no cross-reference to Simon. John Conlisk 's 1995 survey on the subject uses the evidence on non-rationality from experiments as evidence against unbounded rationality. Meanwhile, Robert Sugden (1991) identifies only two categories: rationality and irrationality . One reason why so many authors sidestep Simon is that there are problems with the intentional version of bounded rationality. Undoubtedly, there are cases where evidence of purposeful behaviour is at hand, but in many situations, intentions may not be independently observable to an outside observer. Conversely, behaviour or expressions of preference that defy any kind of interpretation of purpose or motive are extremely rare and possibly nonexistent. Because of this difficulty in defining purposeful behaviour or expressions of preference, I shall define bounded rationality in keeping with the literature on anomalies:

Definition 1.3 An individual has bounded rationality (BR) if not WR.

1.2.3 Uncertainty and Information Processing

To be a concept useful for economists, rationality has to extend to worlds characterised by risk and uncertainty. The definitions offered so far therefore need some further refinement, particularly with regard to the case of information processing. As they stand, the definitions are applicable to the world of uncertainty, provided that the sets of choices, S, X and so on are interpreted as sets of acts in the sense of Savage (1954). Arguably, though what constitutes rationality differs between situations of uncertainty and certainty. Basic principles of probability and information processing are also profoundly connected with rational behaviour in a world of incomplete information. As a result, for situations of risk it seems entirely reasonable to include the requirement that a rational actor should apply the logic of (subjective) probability and that this includes the use of Bayes' theorem to update beliefs in the face of new information**.**

An additional feature, the sure thing principle (STP) is also sometimes listed as a feature of rationality under risk (Samuelson1952). In Savage's 1954 framework in which STP is defined, there is an exhaustive list of possible states of the world, each of which is mutually exclusive. An event is a collection of possible states, with the set of events defined as a partition on the set of states. Individuals face choices between acts. For each act-event pair there is a consequence. The preference relationship is defined on the set of acts.

Let E be an event and let −E be its complement. Suppose that an individual is faced with a choice between two acts, X and Y. Outcomes are summarised in the table below:

	E	−E
X	g1	g3
Y	g2	g4

STP states that if g1 is not preferred to g2 and g3 is not preferred to g4, then X is not preferred to Y. If g3 is indifferent to g4 then STP still holds, so if it is then allowable to replace a prospect by something to which it is indifferent we can replace g4 in Y by g3 and infer that X is not preferred to Y if g1 is not preferred to g2. Alternatively, we could replace g3 by g4 in X and derive the same conclusion. It follows that, granted these intermediate steps, STP implies that preferences between prospects are determined only by states of the world in which outcomes differ. This property is known as independence.

The Allais paradox and later experiments on common ratio and common consequence effects cast extreme doubt on the empirical validity of the independence axiom, but it is its normative acceptability that is the issue here. The independence axiom is not implied by A1 and A2 or their equivalent for the world of uncertainty; it is a separate assumption. It is though, a principle of rational choice? Paul Samuelson, in a widely quoted statement argued that:

> Within the stochastic realm, independence has a legitimacy that it does not have in the non-stochastic realm. Why? Because either heads or tails must come up: if one comes

> up, the other cannot; so there is no reason why the choice between g1 and g2 should be "contaminated" by the choice between g3 and g4 (1952, p. 672–3).

Samuelson is asking us to deny the rationality of, for instance, feelings of regret or disappointment (Loomes and Sugden1982). Other justifications have been offered (e.g. Hammond 1988). The key point is that independence or its sibling, STP may be reasonable principles of behaviour under risk, but their status as foundational principles of rationality is questionable. Neither Samuelson nor Savage nor later writers on the subject ultimately offer a compelling reason why an individual should choose according to independence. However, in fact, I shall not make much of the issue of independence in what follows. When choosing anomalies to illustrate how widely actual behaviour departs from the predictions of rational choice theory, I will not use the common ratio and common consequence effects as motivating examples, despite the robust evidence for their existence. But in the context of risk I shall not suppose that the sure thing principle is a necessary features of rational behaviour.[5]

1.2.4 Merit Goods

Having defined rationality and bounded rationality we now need to give some attention to merit wants. It is useful to begin by noting that, although it is common to talk in terms of merit wants or merit goods, it is not generally possible to isolate goods in this manner. Suppose, for example, that we think that healthcare is under-weighted (in some sense) by individuals in their decisions, then, an equivalent formulation would be say that all other goods, taken together, are over-weighted compared to healthcare. So it is more exact to define an individual as meritworthy. In most formal models of merit wants individuals have transitive and complete preferences. It is just that these preferences are contrary to welfare. On the other hand, if we take the colloquial definition of merit wants – goods where individuals do not choose according to welfare – then this must include cases where choice does not satisfy WR. Hence I shall offer two variations on the definition of meritworthy to cover the possibilities:

Definition 1.4 (i) An individual i is meritworthy (M) if not SR. (ii) M^- if M but WR.

Which of these is the better definition? Some of the starkest examples of merit wants involve either information processing or irresolute choice, with addiction providing a concrete example of the latter. Suppose an individual prefers to take drugs

---

[5] That is not to say that regret or similar violations of independence does not play a role in public policy. Being blamed for failure is a quick way to end many a politician's career. And blame as it appears in the media often appears capricious: decision-makers are judged according to hindsight, not in acknowledgement of the risk and uncertainties of the moment when the original decision was taken. In these circumstance, regret or blame aversion would be a prudent strategy on the part of an individual who values the longevity of his or her political career.

now, but subsequently to give them up, thereby obtaining the pleasure to be had from illicit substances without suffering the misery and humiliation of addiction. Let C(...) denote a choice function and suppose x is total abstinence, y is try and then give-up and z is try and become addicted. If choices are irresolute, it is possible that $C(\{x, y\}) = \{y\}$, $C(\{x, z\}) = \{x\}$ and $C(\{x, y, z\}) = \{z\}$. Rationalised through a preference ordering we end up with a rejection of WR – either because preferences are dependent on the frame (the choice set in this case) or because of frame independent but intransitive preferences. Similarly, in information processing examples, individuals may respond to new information in a manner which breaks with notions of weak rationality. Consequently it seems reasonable to view M as the default definition of meritworthy, rather than the narrower definition represented by M^-, even though the latter version dominates formal models of merit wants (e.g. Sandmo [1983]).

Although I argued that the concept of merit wants is misleading, at least in some circumstances it does have the advantage of psychological intuition. One context in which it is possible to formalise this intuition arises when there is some degree of separability between the goods where preferences and choice accord with welfare and goods where preferences, choice and welfare diverge. Suppose we re-write a typical element of X as $z = (x, y)$ where x and y have n and m elements respectively. Let $X(y)$ be the feasible values of x, given y and write $x \preceq_i^y x'$ to mean that $z \preceq_i z'$ where $z = (x, y)$ and $z' = (x', y)$. Similarly write $y \preceq_i^x y'$ to mean $z \preceq_i z'$ where $z = (x, y)$ and z' is now (x, y'). The strict relationships, frame dependent relationships and the welfare relationships are defined in a similar manner.

Definition 1.5 Merit goods. The elements of y are merit goods and the elements of x are non-merit goods if,

1. $\forall\ f \in F, i \in H, \preceq_{fi}, \preceq_{wi}$ are orderings.
2. $\forall\ y, i$ and $\forall x, x' \in X(y), x \preceq_{fi}^y x' \rightarrow x \preceq_{wi}^y x'$
3. $\exists\ i, y, y', x, f$ such that $y \preceq_{fi}^x y'$ but $y' \prec_{wi}^x y$.

The first part of the assumption states that for any given y, preferences over x bundles satisfy the principles of weak rationality. The second part states that these conditional preferences are in accord with the welfare ordering. The final part states that at least for some values of x the conditional preferences between two y bundles depart from the welfare ordering.

It is worth stating this definition in full because one commonly encountered response to evidence of anomalies is the claim that bounded rationality is confined to specific goods, Cherry et al. (2003), particularly those that are not bought and sold frequently in the market place, Binmore (1999). The argument suffers from a number of empirical deficiencies, some of which are discussed in Chapters 2 and 3. Putting these problems aside for the moment, to be sustainable the argument requires that goods can be separated in some way such that behaviour is rational (in some sense) for at least some of the goods. The definition of non-merit goods goes some way towards splitting goods into two kinds in a reasonable manner, but it is still not strong enough to guarantee behaviour which is usually seen as a basic requirement

of sensible behaviour in the context of revealed preference (McLennen (1990)) . Specifically, for the choice function C *Contraction* is the property that if x $\in S \subseteq S^*$ $\subseteq$ X and $x \in C(S^*)$ then x $\in$ C(S). The following example shows that preferences may satisfy the definition of merit goods, but still violate contraction.

Example 1.1 Suppose an individual faces choices drawn from the set $X = \{A, B, C, D\}$. where $A = (x, y)$, $B = (x, y')$, $C = (x', y)$ and , $D = (x', y')$. There are two relevant frames. In frame 1, the preference ranking is, $B \prec D \prec A \prec C$. In frame 2, the ranking is $A \prec C \prec B \prec D$. The properties set out in the definition of merit goods only limit preferences between A and C and between B and D and these sets of frame-dependent preferences satisfy the restrictions. Now suppose (as in the baked bean example) that the choice set affects the frame evoked. In particular suppose that frame 1 is evoked by the choice set $\{$A, B, D$\}$ while frame 2 is evoked by choice set $\{$A, B, C$\}$. Given the pattern of preferences, we get the following choices: $A = C(\{A, D, B\})$ and $B = C(\{A, B, C\})$. However, no pattern of choices from the set of two element choice sets is consistent with the contraction property as we can quickly see by comparing the implications of $B = C(\{A, B, C\})$ and $A = C(\{A, D, B\})$. From the first of these we get $C(\{B, A\}) = B$, whereas the second implies $C(\{B, A\}) = A$.

So, for any given value of the merit good (y or y′) preferences between bundles involving only changes in the non-merit good do satisfy the requirements of consistency and completeness. However, for comparisons involving changes in the merit good, consistency can be elusive and behaviour in the market can be contradictory. As a result, neat separation of the world into areas affected and unaffected by framing is practically impossible.

1.3 Main Points

Standard welfare economics, together with its associated taxonomy of market failures, gives a comprehensive theory of action for the state. It is tempting to try to offer something equally complete for the world of bounded rationality. I am not sure that there is a single coherent message to be obtained in the absence of complete rationality, precisely because there are so many different kinds of bounded rationality. Nevertheless, it may be useful to set out some recurring themes.

1. The first message which bears repeating is simply the importance of boundedly rational behaviour within economic contexts. Bounded rationality behaviour is not necessarily confined to non-market goods or to unfamiliar or rarely bought commodities. Much of the behaviour found in the laboratory is also exhibited in field data.
2. Potentially, the evidence which points to bounded rationality can be interpreted in two ways: *measurement* and *incoherence*. The measurement interpretation of the problem is that, with the tools available to social scientists, it seems that we are unable to elicit reliable indicators of individual preferences for a wide

variety of goods. The incoherence interpretation of the problem is that individuals really do not have complete, coherent and context-free preferences. The fact that the data found in hypothetical choices (i.e. stated preference) and in real choices shows evidence of the same kinds of departures from rationality points to incoherence as at least part of the story. However, given that much of the evidence on preference is obtained from stated preference studies, measurement problems should not be understated.

3. Theories of economic behaviour are essentially theories about the properties of the demand function, a fairly general formulation for which is $x(p, m, f)$, where x is a vector of demands, p is a vector of prices (which may be virtual prices for rationed commodities such as public goods), m is income (again which may be virtual if there are constraints on consumption). The last term, f, is the frame in which the demand function is elicited. Theories of economic behaviour, such as the Hicksian model , place restrictions on this function. For instance, $x(ap, am) = x(p, m)$, the property of homogeneity of degree zero when a is a positive scalar. Another important feature of the Hicksian model is that of *frame neutrality* , that is $x(p, m, f) = x(p, m, f')$ $\forall$ p, m, f and f'. Behavioural models of the consumer relax this property of frame neutrality, thereby expanding the set of potential public policies and making the impact of traditional economic policies such as taxes more complicated, in the sense of being dependent on the manner in which their implementation is framed. Nevertheless within a given frame, the predicted effect of changes in traditional policy instruments can be standard.
4. Inconsistent behaviour and expressions of preference that conflict with the precepts of the rational choice pose major problems for welfare economics. Some commentators have viewed the issue as virtually intractable. For instance, McClure's criticisms of both Musgrave and Head have been repeated many times over the years (see for example the interchange in the ANU volume (Brennan and Walsh (1990))). His position, as stated by Head , is that the 'merit wants concept is completely irreconcilable with a normative policy framework based on consumer preferences or individual values.' (p. 218). Instead issues of irrationality, for instance, 'should be relegated to a limbo of questions for which no normative conclusions can be offered.' (Brennan and Walsh, 1990, p. 219). Undoubtedly there are problems, but as we shall see in subsequent chapters the situation is not quite so bleak. Specifically it is possible to make some headway if we are prepared to use non-preference information[6] and accept that a complete ordering of options may not be possible. In other words the entire citadel of welfare economics does not have to be completely abandoned, but a managed and partial retreat may be in order.

[6] In some formulations of Paretian economics, the third property of neutrality is defined in terms of preferences rather than individual welfare. From that perspective, I am suggesting a relaxation of the third property to allow non-preference information when preference information is equivocal.

5. My approach to welfare is pragmatic. Acts of choice, statements of preference and intuitions about what is good are viewed as sources of information that are not perfectly reliable. Greater weight is placed on evidence when it is supported by other data.
6. In general the rational and the boundedly rational are not easily separable. Instead there is a continuum of cases: individuals can be more or less capable of making their own choices; commodities can pose greater or lesser challenges to evaluation by the imperfectly intelligent agent. One of the ways that crude models of merit wants have been found wanting in the past arises out of their all-or-nothing approach to consumer sovereignty . In fact there are degrees of intervention into personal decision-making: a government can choose what a person can consume or it can influence the decision (e.g. through taxes) or it can restrict, by banning certain options. Thus, for policy purposes, a government does not have to know what is best, or have a complete ranking of alternatives; benefits to the individual may also flow if the state only knows what is bad or worst.
7. The fact that an individual does not know what is best for her or himself does not imply that any other individual or group of individuals knows any better. We can call this the *knowledge hurdle*. Were the gap between preference and welfare to be purely random, then there would be little to be said about grounds for intervention in individual decision-making. However, the evidence from the biases and heuristics literature in cognitive psychology (e.g. Kahneman et al., 1982), suggests that biases in processing information do tend to be systematic. Government policies are often constrained to be fairly uniform across the population, so a knowledge of the direction of cognitive biases may still not provide an opportunity for welfare-enhancing intervention, if individual welfare rankings are so varied across the population that uniformity would impose severe welfare costs on some or most individuals. Where systematic bias is combined with a small variance in welfare rankings across the population (and this is known to be the case), then the knowledge hurdle may be overcome.
8. More generally though, that agents are boundedly rational does not mean that it is better for the state to make choices on their behalf or to ignore possibly imperfect information about their preferences. Market-based allocation methods may be superior because markets give incentives and opportunities for individuals to learn that are much attenuated when the state makes choices on their behalf. Consequently, bounded rationality is not a 'market failure' in the traditional sense. Meanwhile, if the equilibrium of the economy is sufficient continuous in the underlying preference and technology parameters, then market failures such as externalities or public goods will still be causes of inefficiency when there is some bounded rationality present. In other words bounded rationality does not destroy the traditional concepts of market failure.
9. Understanding the psychology of information processing gives insight into why, in a context of imperfect information , simply passing on information to ill-informed individuals may not lead to them making the choices best for their own ends. In other words, the optimal solution to merit want problems may not

be information provision. Furthermore, because policy typically takes place in a second-best environment, information provision has effects on incentive compatibility constraints as well as direct effects on beliefs about optimal choices. Optimal government policy may therefore not involve the production of truthful information – to mislead may be better if it relaxes constraints on other policy tools.

10. To the extent that individuals are aware of their own cognitive limitations, they may be willing to play restrictions on their future choice sets. In a similar manner, individuals may vote for institutions and policies that limit or at least guide their future choices.
11. When there is great uncertainty about what is best for an individual, then apparently inferior methods of project evaluation might outperform theoretically superior methods. For instance in cost benefit analysis, if revealed preference methods only elicit a proportion of total value their greater reliability (i.e. lower variance) may make them preferable to methods such as contingent valuation which elicit total benefits, but which are notoriously prone to framing effects. Similarly, Citizens' Juries may outperform standard approaches to non-market valuation if the use of Juries can eliminate biases in decision-making.
12. Finally, there are a great many questions about public policy in a world of boundedly rational agents that remain to be raised, let alone resolved. Some of the most important questions are about calibration: how important quantitatively are deviations from rationality in spheres relevant for public policy? To what degree, for instance, is tax evasion affected by the way tax enforcement policy is framed? Related concerns include the degree to which Goodhart's Law[7]applies to the exploitation of framing effects for public policy purposes. These kinds of questions are raised throughout the book.

[7] Goodhart's Law states that, 'Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes.' Schwert (2002) documents the extent to which financial market anomalies, once revealed, are diminished and in some cases eliminated. Cherry et al., (2003) is an example of an experiment where intensive education and feedback to participants reduces a decision-making bias.

Chapter 2
Anomalies

2.1 Introduction

In economics, the term 'anomaly' usually refers to behaviour which does not conform to the predictions of rational choice models . The argument of the introductory chapter was that anomalies are sufficiently common within economic behaviour and that there is sufficient predictability in decision-making biases to provide useful guidance for policy makers. The purpose of this chapter is to provide instances. Now, the gap between the predictions of economic theory and actual behaviour has been extensively documented elsewhere. Camerer (1998), McFadden (1999) and Starmer (2000) provide recent summaries and there are whole books devoted to the subject, such as the volume edited by Kahneman and Tversky (2000). Consequently, I shall be selective in the anomalies I discuss, concentrating on the most familiar and on some of the objections which might be raised against their existence. In line with the objectives of the book, the focus will be on features which have significance for public policy.

Anomalies come in three categories. The first and most straightforward category is that of *explicit* anomalies, where observed behaviour is at variance with one or more of the stated axioms of the model. For instance, if an individual strictly prefers option A to B, chooses B out of B and C, but then strictly prefers C to A, then he or she breaks the transitivity axiom.

In the second category of *implicit* anomalies , it is what is tacitly assumed by the theory of rational choice that is contradicted by the evidence. For instance, when the axioms of consumer theory are stated, there is rarely if ever a statement to the effect that preferences should be invariant to the description of the choices, provided the descriptions are sufficiently transparent. It is an example of what Kenneth Arrow (1982), describes as 'a fundamental element of rationality, so elementary that we hardly notice it' (p. 6). Nevertheless description invariance is typically assumed in the actual use of the theory. Typical of this kind of anomaly are framing effects, as in the Kahneman and Tversky's Asian disease example utilised in Chapter 1. Another instance is that of preference reversal , where it is the elicitation mechanisms (valuation versus choice) that produce different rankings of the bundles under comparison.

The third category, which I shall call extreme event anomalies includes all cases where behaviour does not unequivocally contradict a rational choice theory, but

A. Munro, *Bounded Rationality and Public Policy*, The Economics of Non-Market Goods and Resources 12, DOI 10.1007/978-1-4020-9473-6_2,

© Springer Science+Business Media B.V. 2009

where the parameters of the theory would have to be extreme in order to accommodate the observations – and possibly, these parameters would imply behaviour inconsistent with that routinely observed elsewhere. For instance, Kahneman and Tversky (1979), report examples of the reflection effect , the tendency of subjects to choose the safer lottery out of a pair when all outcomes are non-negative, and to choose the risky option, when the signs on the outcomes are reversed, so that all potential gains become losses. This phenomenon is explicable in terms of EUT, but it requires the assumption that all the individuals who exhibit the reflection effect happen to be at the wealth levels where their preferences turn from risk loving to risk aversion.

In this chapter and the next I concentrate on well-known examples of anomalies which cut across the types. They have been chosen to satisfy three constraints: they are well-established through laboratory and field evidence, each has some significance for public policy, either normatively or in its positive aspect. Finally, they range over three key aspects of human decision making: preferences, budgets and information processing. The first anomaly is that of reference dependent preferences – a category that includes status quo bias or the endowment effect, and which has been the subject of much investigation, particularly by environmental economists troubled by the large measured gap between willingness to pay and willingness to accept measures of value. The second anomaly is that of mental accounting - the apparent tendency of individuals to partition spending and income in ways which contradict the standard predictions of consumer theory. The third anomaly is preference reversal - the dependence of the ranking of different options on the instrument used to elicit preferences. The final group of anomalies concerns information processing in decisions involving risk and uncertainty. I discuss this last group in Chapter 3.

2.2 The Reliability of the Evidence

Much of the clearest evidence on anomalies is drawn from experimental economics, which tends to provide sharp tests of rationality assumption in a manner which is rarely possible with non-laboratory data. There is an ongoing debate within economics about the significance of experiments. My position is this. Most modern economic theories of rational decision-making are universal in tone. That is, they set no limits to the domain of their applicability. Consequently, rejection of the theories in laboratory experiments represents rejection of the universality of the theories. One counter response is then to accept the evidence of laboratory experiments, but to withdraw the claim of universality, adopting a position in the spirit of Marshall, 1890, who drew a distinction between areas of society, such as the marketplace, where economic models of rational behaviour would be obeyed and other domains where non-rational behaviour might hold sway.[1] This Marshallian

[1] Marshall (1890) wrote that 'Economics is a study of mankind in the ordinary business of life; it examines that part of individual and social action which is most closely connected with the attainment and with the use of the material requisites of wellbeing.' p. 1.

viewpoint is defensible, provided its proponents have a model of why rational choice models might apply in one context, but not in another and provided there actually is no evidence that anomalies occur widely and are not confined to the laboratory. And as we saw in Chapter 1 it also requires preferences that are sufficiently separable to allow the rational to be separable from the currently irrational. Actually though, the evidence that anomalies are not confined to the laboratory is large and growing. Several anomalies, such as the winner's curse in auctions have their origin in market behaviour. Meanwhile Camerer (1998, 2000), provides surveys of the field evidence which supports prospect theory rather than EUT, while Shefrin (1999) is a book length study of boundedly rational behaviour in financial markets. Later on in this chapter, and in Chapters 3, 10 and 11, I shall present some of this evidence which challenges the presumption that anomalies are epiphenomena, confined to particular areas of economic activity.

A variant on the Marshallian objective argues that the failure of laboratory-based experiments to match the predictions of standard theory is not necessarily the unfamiliarity of the environment, but more the limited opportunities for learning and the lack of incentives. This is the basis of Charles Plott's discovered preference hypothesis (Plott, 1996) – that through experimentation, feedback and repetition, individuals gradually learn their true preferences, provided the environment remains stable. Let us take the issue of incentives first.

It is a common criticism of experiments is that they flood the subjects with unusual decision problems in an unfamiliar environment and offer only limited chances for learning. For economists, perhaps the most pertinent criticism concerns the apparently small incentives to choose optimally which most subjects face. Although often, the implicit payment per hour is relatively high in experiments, the difference in expected value between the choices that individuals face can be very low. Often this is deliberate, since for some experimental tests, it is desirable to have a fairly even split of subjects choosing each option. When more is at stake, perhaps individuals come closer to optimization.

To judge whether individuals are optimizing we have to know their true preferences. In many experiments this is not possible, so all that can be done is to increase the payoffs and see if the results are affected. Camerer and Hogarath (1999) pursues this route, but finds little evidence that increases in payoffs eliminates anomalies. Another alternative is to replicate the experiment right down to the original payoffs, in a country where incomes are low. Kachelmeier and Shehata (1991), used Chinese business students in Beijing and compared the results of their lottery-valuation experiments there, with results using Canadian students. Both samples showed similar patterns of behaviour, particularly with regard to the over-weighting of small probabilities. More direct evidence is available from a paper by Gneezy and Rustichini (2000), the title of which ('Pay Enough or Don't Pay at All'), sums up their results. Two experiments are reported. In one the task consists of fifty IQ-test questions, with paid subjects rewarded according to the number of correct answers. In the other task subjects were house to house volunteers, asking for charitable contributions. Payments in this case were linked to the promises of donations elicited by the volunteers. By splitting their sample of Israeli students into three and offering

one part no incentives for the task in hand, while the individuals in the other two sub-samples are offered a small monetary and a large incentive respectively Gneezy and Rustichini were able to test the hypothesis that higher incentives leads to higher performance. They find that a larger incentive yields better performance on the part of their subjects, but that subjects paid nothing perform as well as subjects paid the highest amount.

Camerer and Hogarth (1999), contains a review of seventy-four published studies in which incentives varied between subjects or between tasks within the same experiment. The experiments cover a variety of economic and psychological areas including physical tasks, choices under uncertainty and judgement. They point to 23 studies in which incentives improve mean performance (when a performance standard was clear) or reduced anomalies. In a further nine studies, raised incentives reduced performance, while in 28 studies no changes in mean performance were recorded. In 16 cases changes in incentives affected behaviour, but no clear performance standard was in evidence, while in eight studies the effects of incentives was confounded with other changes in the design. Overall the authors conclude with the 'provocative conjecture' that,

> There is no replicated study in which a theory of rational choice was rejected at low stakes in favor (sic) of a well-specified behavioral alternative, and accepted at high stakes. p. 26.

Underlying the idea that higher incentives should yield behaviour closer to the predictions of rational choice theory is the intuitive concept of the decision-maker as economist, who in making a choice or declaring a value, must devote costly time and effort to the problem (Conlisk, 1996; Smith and Walker, 1993). The basic weakness of this model is that, in many circumstances, it fails to offer an explanation of why choices should deviate from the predictions of rational choice in a non-random fashion. Of course ad hoc or ex-post arguments can often be found. There are several theoretical models that now show why resistance to trade (i.e. an endowment effect) may be advantageous in trading environments. But the features that propel such a conclusion are not part of the core model and do not help explain other anomalies.

Having explained why laboratory evidence deserves to be taken seriously by economists, let me now point to some of its obvious limitations. Economic experiments designed to test rational choice models are entitled to assume as part of their null hypothesis, that full or complete rationality extends to understanding the instructions and their implications, provided those instructions are clear. Once one accepts bounded rationality, experimental designs, predicated on the assumption that individuals are fully rational may tell us little about how boundedly rational individuals behave. In short therefore, economics experiments designed for the null of full rationality may be very good at telling us whether the assumptions of rationality are true, but quite poor at distinguishing between alternative models of bounded rationality.

A related issue is that one other, important limitation of many experiments is their qualitative nature. In many public policy contexts quantitative information about the degree to which individuals do not maximize is required. Of course, some quantitative information about the size of anomalies within samples can be produced by

experiments. Typically, for instance, individuals are willing to pay around 10–20% of expected value to avoid the ambiguous option in the Ellsberg two colour experiment (e.g. Bernasconi and Loomes , 1992). On the other hand it is not clear how this number would generalise to real world choices and samples other than the typical groups of psychology or economics students recruited for the laboratory.

The quantitative impact of deviations from the predictions of rational choice may be insignificant for several reasons. First the deviations themselves might be small in absolute terms. Secondly, they may be swamped by other kinds of factors. For instance, in the market place, endowment effects may look similar to habituation or to transaction costs. That said, the 'real-world' evidence we have on some anomalies suggests that they may be quantitatively important. The equity premium puzzle (Mehra and Prescott, 1985) is a puzzle because of the large and sustained gap in returns to equities and returns to bonds over the course of the twentieth century – about 8% per annum for the USA. Hardie et al. (1993), in their study of supermarket price elasticities found much larger price elasticities for price rises, compared to price falls. Coefficients of loss aversion of about 2.4 were required to explain their results for orange juice.

Third, the bounded rationality of some laboratory agents may be irrelevant to the item we are interested in. In a bond market driven by arbitrage for instance, it will be professional traders who ensure that the price of a short-run bond does not deviate from the price of a long run bond in a manner that provides opportunities for riskless profit.

Finally, perhaps individuals are not fully rational. Perhaps, though they are just rational enough, with heuristics and decision-making rules of thumb which are adapted to their environment. This is then intriguing thought behind theories of ecological rationality (Goldstein and Gigerenzer, 2002), which posit that since information processing is a costly activity, it is optimal to economise on its use, thus producing decision rules which maximize the expected net value of choices given the distribution of contexts in which decisions will have to be made. Clearly there may be some sense to this argument, but as we shall see below, much of the evidence from markets and experiments with repeated choice suggests some persistence in anomalies (to put it mildly), which suggests that adaptation is a slow process.

Some parts of this debate about the value of experimental evidence apply to all of economics; some parts though have a special resonance for public policy. After all, many of the reasons (strangeness of the environment, weakness of the incentives, absence of opportunities to learn etc.) why one might reject laboratory evidence are equally relevant to many public policy contexts, particularly those concerned with non-market goods. As Downs (1957), pointed out, when decisions are made collectively individuals have only low incentives to become informed about the value of different options. Many of the decisions to be made about public policies concern unfamiliar or new options and some, but by no means all, are rarely repeated. Caplan (2001), takes this further in his discussion of 'rational irrationality' in a public policy context. He argues that not only do voters typically have weak incentives to become informed, they also have poor reason to eliminate biases. So, for some kinds of public policy issues, we might expect to see behaviour similar to that found in

the laboratory. Arguably, in some other contexts, such as changes in taxation on common consumer goods, behaviour will be better modelled using more standard economic models.

2.3 Reference Dependent Preferences

In a classic experiment conducted by Jack Knetsch at the University of Victoria, 76 students were endowed with coffee mugs while another 87 students were each given a 400 g bar of Swiss chocolate (Knetsch, 1989). All subjects were offered the chance to exchange their endowment for the other good in the experiment. Out of the students endowed with mugs, 89% elected to keep their mugs. Meanwhile, 90% chose to hang on to their endowment of chocolate, contradicting economic theory which predicts that, provided the sub-samples were chosen randomly, the proportion preferring the chocolate to the mug should have been the same in each group. Subjects in this experiment provide evidence of the endowment effect (Thaler, 1980), or status quo bias (Samuelson and Zeckhauser, 1988) the tendency to prefer one's existing bundle of goods over alternatives. The endowment effect is an example of a wider class of phenomena, whereby preferences are conditional on the reference point from which an individual compares alternatives or expresses valuation. Typically the reference point is either the individual's endowment of goods or a bundle representing their routine consumption, but as we shall see below it can be other vantage points. Such preferences violate one of the implicit axioms of Hicksian consumer theory, namely that preferences are independent of the vantage point from which they are elicited.

Table 2.1 summarises many of the studies which have supplied the evidence on reference dependent preferences. It draws on Traub (1999), as well as Camerer (1995). Much of it comes from environmental economics, where in particular, the discrepancy between willingness to accept compensation (WTA) and willingness to pay (WTP) has been documented extensively within the contingent valuation literature. Horowitz and McConnell (2002), for instance report data from 201 studies, mostly from contingent valuation. In many cases, the ratio of mean or median WTA to its equivalent WTP value is much greater than one, often four or five (e.g. Bishop and Heberlein, 1979; Rowe et al., 1980; Viscusi et al., 1987).

If viewed separately from the experimental evidence, the data thrown up by contingent valuation might be classified as an extreme event anomaly, for in theory, as Hanemann (1991), shows, a large gap between WTA and WTP is compatible with the axioms of consumer theory, provided that substitution and income effects are large enough. Figure 2.1 illustrates the argument. The left hand panel shows WTP and WTA for an amount of good, x. The indifference curves, U^0, and U^1 indicate that the good is a perfect substitute for money (i.e. all other goods). In this case, WTP = WTA. In the right hand panel, x is not a perfect substitute with other goods and preferences are not quasi-linear, so there is a non-zero income elasticity of demand for x. In the case depicted, x is a normal good and hence WTA > WTP.

Table 2.1 Empirical Evidence on Reference Dependent Preferences

Study	Commodity	Method	Results (mean WTA/mean WTP Except Where Stated)
Contingent valuation and contingent choice: hypothetical goods			
Hammack and Brown (1974)	Hunting permits (waterfowl)	Open-ended	4.2
Sinclair (1976)	Preservation of fishing spot	Open-ended	2.9
Bishop and Heberlein (1979)	Hunting permits (Geese)	Open-ended	4.8
Banford et al. (1979)	Preservation of pier	Iterative bidding	2.8
Rowe et al. (1980)	Visibility reduction	Iterative bidding	11.8
Brookshire et al. (1980)	Hunting permits (Elk)	Iterative bidding	3.2
Bishop and Heberlein (1986)	Hunting permits (Deer)	Open-ended	20.8
Viscusi et al. (1987)	Health risks	Open-ended	2.5–2.7
Hartman et al. (1991)	Power supply reliability	Open-ended	4.1
Hartman et al. (1991)	Power supply reliability	Choice	60% of consumers with high reliability and rates opted for status quo; 58.3% with low reliability/low rates opted for status quo.
Samuelson and Zeckhauser, 1988	Car versus road safety	Preferences between alternative % allocation of funds given different statements about status quo	Status quo =50:50: 76% prefer status quo. Status quo = 60:40: 39% prefer 50:50.
McDaniels, 1992	Car safety	Open-ended	3.6
Dubourg et al., 1994	Car safety	Iterative bidding	3.1
Pinto Prades, 1997	Health states	Person trade-off for changes in health	Differences in rankings when options presented as gains compared to when presented as loss versus gain.
Bateman et al. (2000)	Traffic reduction measures	Dichotomous choice	>6, other results in line with reference dependent preference theory predictions.
Laboratory experiments, real goods			
Knetsch and Sinden (1984)	Lottery ticket, prize = US $50	Split sample – half endowed with lottery	29/38 endowed with lottery unwilling to accept $2 to give it up; 19/38 unwilling to pay $2 to acquire lottery.

Table 2.1 (continued)

Study	Commodity	Method	Results (mean WTA/mean WTP Except Where Stated)
Coursey et al. (1987)	SOA (bitter substance) solution	Open-ended	>3
		Iterative bidding	>3
		Vickrey auction, 1st trial	>3
		Vickrey auction, last trial	≈1.5
Brookshire and Coursey (1987)	Tree density in a park	Open-ended	61.1
		Smith auction, hypothetical	56
		Smith auction, 1st trial	3.9
		Smith auction, last trial	2.4
Knetsch (1989)	Chocolate bars	BDM	2
	Mugs and candy	Option, given endowment of one, to switch to other	Given mug, 89% prefer mug to candy; Given candy, 10% prefer mug to candy.
Kahneman et al. (1990)	Tokens	Induced value	11 out of 11 expected trades executed at expected price.
Knetsch (1992)	Coffee mugs	Trading game	2.2; across four trials, on average only 2.25 out of 11 expected trades made.
Boyce et al. (1992)	Mugs and pens	Option, given endowment of one to switch to other and gain 5 cents.	Given mug: 88% prefer mug to pen plus 5 cents; Given pen: 10% prefer mug plus 5 cents to pen.
Tietz (1992)	Norfolk pine	BDM	From 1.7 to 2.4
	Fruit	Vickrey auction	1.5
Adamowicz et al. (1993)	Movie tickets	Open ended	1.9
Adamowicz et al. (1993)	Ice hockey tickets	Open-ended with substitutes	1.7 (with substitutes); 1.9 (no substitutes).
Shogren et al. (1994)	Candy bars	Vickrey auction, repeated trials.	0.9 (for trial #5); rapid convergence of WTA & WTP.
	Food poisoning risk	Vickrey auction repeated trials	4 (for trials 17–20); no convergence.
Eisenberger and Weber (1995)	Lotteries	BDM	1.5
Bateman et al. (1997a)	Belgian chocolates	BDM; controls for substitution effects	64.5% subjects prefer chocolate option when endowed with money; 92.5% prefer chocolate option when endowed with chocolate.

Table 2.1 (continued)

Study	Commodity	Method	Results (mean WTA/mean WTP Except Where Stated)
	Coca-Cola	BDM; controls for substitution effects	15.8% subjects prefer Coke option when endowed with money; 60% prefer Coke option when endowed with money.
Myagkov and Plott (1997)	Lotteries involving losses	Market for gambles	Evidence of risk seeking in losses, consistency between market behaviour and hypothetical survey.
Bateman et al. (1997b)	Vouchers for meal at a pizza restaurant	BDM; compensation for income effects	2.27
	Vouchers for dessert & coffee at a pizza restaurant	BDM; compensation for income effects	1.97
Morrison (1997)	Mugs	BDM; compensation for income effects	2.2
	Chocolate	BDM; compensation for income effects	1.1
			0.9 (for trial #5); rapid convergence of WTA & WTP.
Rhodes (1997)	Supermarket goods	Vickrey auction; controls for substitution effects	>1; little evidence that substitutability drives loss aversion.
Herne (1998)	Chocolate and chewing gum	Choice between two options given endowment of a third.	Option 1.95 times more likely to be chosen if it dominates endowment.
Bateman et al. (2005)	Belgian chocolates	KKT procedure with controls for substitution effects	>2. Evidence for loss aversion in both money and goods.
Loomes et al. (2002)	Lotteries	Repeated median price auction	WTA/WTP ratio diminishes through repetition, but values sensitive to framing effects.
Plott and Zeiler (2005)	Travel mugs	BDM after training. All subjects endowed with mug.	0.9
Biel et al. (2006)	Charitable Donation	Dichotomous choice; Income controls	23 out of 51 subjects 'endowed' with donation kept it; 9 out of 48 subjects not endowed contributed.
Field experiments and market data			
Samuelson and Zeckhauser (1988)	Annual health plan enrollments	Choice by existing staff versus new staff	43.1% of existing staff opted for status quo versus 22.7% of new staff.

Table 2.1 (continued)

Study	Commodity	Method	Results (mean WTA/mean WTP Except Where Stated)
Johnson and Hershey (1993)	Car Insurance	Choice of right to sue or not, with variations in default option	Given extended rights, 73% opted for extended rights. Given restricted rights, 20% opted for extended rights.
Hardie et al. (1993)	Supermarket goods (scanner data)	Compare demand responses to price rises and cuts	Demand elasticity for price rise > elasticity for price fall.
Benartzi and Thaler (1995)	Investment decisions: bonds versus equities	Compare investment choices to EUT predictions	Level of risk aversion implied by bond-equity split incompatible with that implied by other risky decisions; compatible with prospect theory.
Camerer et al. (1997)	New York cab drivers' labour supply	Responses to demand	Labour supply lower on higher demand days – evidence of daily income targeting.
Bowman et al. (1999)	Teachers' consumption	Responses of consumption to bad income news	Consumption does not fall, supporting endowment effect.
List (2003)	Sportscards	148 sportscard traders randomly endowed with cards and invited to swap	6.80% of inexperienced non-dealers willing to swap; 46.7% of experienced non-dealers willing to swap; 45% (approx.) of dealers willing to swap.
	Disney lapel pins (badges)	80 pin traders randomly endowed with pins and invited to swap	40% of experienced traders willing to swap; 25% of inexperienced traders willing to swap 15.8%.
List (2003)	Sportscards	120 dealers and non-dealers asked wtp or wta for sportscards	For non-dealers, ratio = 5.58. For dealers ratio of mean WTA/mean WTP = 1.3. Evidence of risk seeking in losses, consistency between market behaviour and hypothetical survey.

Table 2.1 (continued)

Study	Commodity	Method	Results (mean WTA/mean WTP Except Where Stated)
Bell and Lattin (2000)	Supermarket goods (scanner data)	Compare demand responses to price rises and cuts, allowing for consumer heterogeneity	Significant evidence of loss aversion for detergents, hot dogs, tissues, paper towels, margarine and soft drinks; insignificant though positive effects for bacon, butter, crackers, sugar and ice cream.
Simonsohn and Loewenstein (2006)	Housing demand in USA	Uses Panel Study of Income Dynamics to relate price paid by movers in new city to price paid in old city	Moves from more expensive cities pay more for same housing services compared to movers from cheaper cities.
Munro and De Souza (2007)	Quest skills and artefacts	Use online roleplaying environment to buy and sell virtual commodities.	New players show large endowment effects. Experienced users show no effect, while highly experienced users over-trade.

Notes

Vickrey auction (second price auction .) In buying (selling) the good is allocated to the highest (lowest) bidder at a price equal to that bid by the second highest (lowest) bidder.

Median price auction As above but with the price determined by the median bid.

Smith auction If sum of bids < cost of providing the public good, money is returned. If sum of bids ≥ cost, then the good is provided. If sum of bids > cost, then contributions are scaled back in proportion so that the scaled sum equals the cost of provision.

BDM = Becker de Groot Marschak mechanism. In buying (selling) individuals receive (sell) the good if a randomly generated offer price is lower (higher) than their state price. The price paid or recieved is the offer price.

KKT = Kahneman, Knetsch and Thaler mechanism. In buying (selling) individuals see a sequence of choices offering them a change to buy (sell) the good at different prices. They must indicate which offers are acceptable and which are not. One element of the sequence is picked at random and the individual's choice executed.

Haneman proves that the gap between WTA and WTP is increasing in η/σ, where η is the income elasticity of demand, and σ is the Hicks-Allen elasticity of substitution between x and other goods. In theory, therefore, the gap between observed values for WTA and WTP could be due to income and substitution effects. One reason why this might be difficult to accept lies in the fact that WTA for x is actually WTP for x given the original income, plus WTA. Let y^0 be that original income,

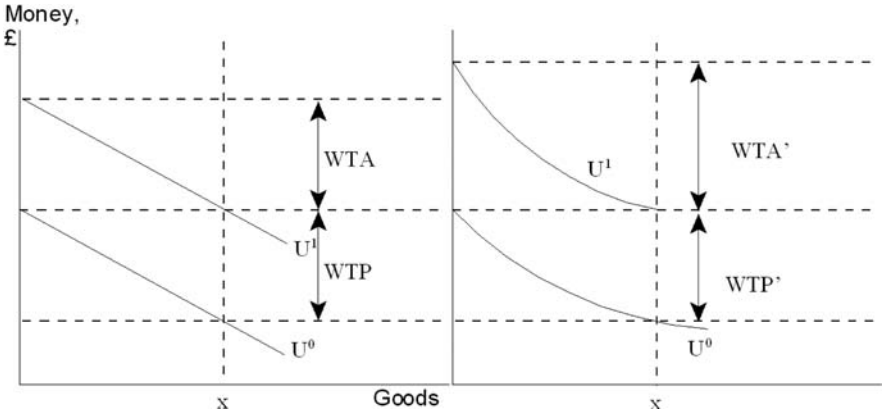


Fig. 2.1 WTP, WTA and Substitution

then WTA $=$ WTP($y^0 +$ WTA). It follows that dWTP/dy is approximately equal to (WTA $-$ WTP)/WTA. In the case where WTA is four or five times WTP , this means that the dWTP$/dy$ is approximately 0.75–0.8, or in other words, for every \$1 extra of income, consumers are willing to pay around three-quarters of that amount extra for the good in question. There is no evidence around that WTP is so sensitive to income.

In addition to the problems of interpretation posed by the presence of income and substitution effects, recall that contingent valuation is an exercise in hypothetical valuation, typically conducted for public goods. Not surprisingly therefore, the validity of evidence from contingent valuation has often been questioned (e.g. Diamond and Hausman, 1994), while practitioners have played down the WTA versus WTP disparity, pointing to the unfamiliarity to consumers of the former welfare measure as a reason to doubt its reliability. As the second section of Table 2.1 shows, though, extensive evidence which supports the existence of reference dependent preferences has also come from incentive-compatible laboratory experiments.

Boyce et al. (1992), Tietz (1992) and Adamowicz et al. (1993), all report gaps of 50–150% between WTA and WTP for a variety of goods bought and sold for real in the laboratory. Meanwhile, in Bateman et al. (1997b), subjects were asked to value vouchers for parts of a meal at pizza restaurant in Norwich , UK, using the incentive compatible Becker deGroot Marschak (BDM) mechanism to elicit preferences. For the components of a pizza meal the ratio of mean WTA to mean WTP was always in the range 2.5–4.

A potential weakness of the designs used for the experiments in the previous paragraph, is that, as with the contingent valuation exercises described above, the results are explicable in terms of unobserved income and substitution effects. This problem is shared with the design of Shogren et al. (1994), who attempted to see if the WTA-WTP disparity could be related to the degree of substitutability of

the goods in question. They chose two goods: candy, which could be expected to have a large number of close substitutes and one, health, which is widely recognised as having few if any substitutes. A notable feature of Shogren et al. (1994) was the repeated nature of the task which subjects faced. In repeated Vickrey auctions WTP and WTA for candy quickly converged, but the equivalent measures for health stayed well apart even after twenty rounds of the experiment. The authors put the difference down to the differences in substitutability, but as Gwendolyn Morrison (1998), points out the large WTA/WTP ratio for the food poisoning risk could have been due to reference point effects, since these are potentially confounded with income and substitution effects. She attempted to discriminate between the two explanations using an experiment in which subjects were given their WTA for giving up chocolate or a mug and then WTP values were extracted using a BDM. The ratio WTA/WTP remained above 1 even after several rounds of repetition. However, unlike the single task experiment the procedure used was not incentive compatible.

Moreover, as the original Knetch experiment showed it is possible to control for income and substitution effects. Figure 2.2 illustrates the tasks faced by subjects in an experiment reported in Bateman et al. (1997a). Subjects were endowed with a mixture of goods (Coke or Belgian chocolates) and money. Let Δ be the horizontal distance between a and b. Using a BDM, subjects whose endowment point was a were asked to state maximum willingness to pay (WTP) to obtain Δ. Subjects endowed with d were asked their willingness to accept compensation (WTA) to give up Δ. Subjects endowed with c were asked to state the minimum amount of money the would accept to forego a gain of Δ (equivalent gain or EG) and subjects endowed with b were asked to state a maximum amount of money which they would be willing to lose in order to avoid losing Δ (equivalent loss or EL). Each form of valuation reveals something about the individuals' preferences between a and d, but according to Hicksian theory, preferences are independent of reference points and hence the proportions of individuals whose values indicated preferences for a over d, should be invariant to the endowment point. Table 2.1 also summarises the results of the experiment. There is a clear pattern: as the endowment point moves rightwards

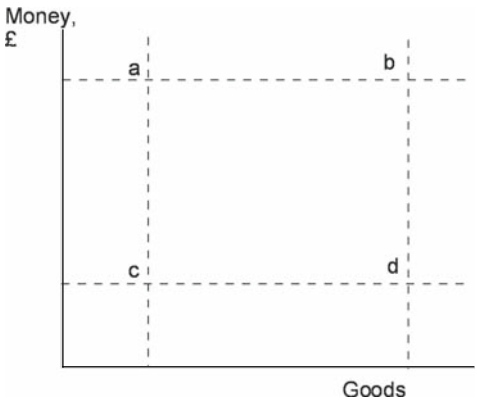


Fig. 2.2 Reference Points

or down in Fig. 2.2, the proportion of subjects preferring d to a increases. With the exception of the move from a to b for the case of chocolate, all such moves (a to b, a to c, c to d and b to d) produce a statistically significant increase in the proportion of subjects implicitly preferring d to a. Similar results are reported in follow-up experiments conducted by Rhodes (1997), and Bateman et al. (2005). In the former, Rhodes adapted the design to test the hypothesis that reference point effects were caused by, or at least correlated with, the degree of substitutability. The hypothesis was not supported by the evidence. The latter experiment was designed to test competing hypotheses proposed by Daniel Kahneman and by Bateman, Munro, Starmer and Sugden to explain the results of the Bateman et al. (1997a) paper. The results of the follow-up experiment were unable to distinguish unequivocally between the competing explanations, but in the parts of the experiment which replicated the 1997 framework the original results were clearly supported. In addition, whether subjects were asked to express valuation in goods (e.g. how many chocolates to acquire £1) or in money, similar patterns of reference point effects arose.

So, EG valuations have been found to lie between WTP and WTA whether or not the EG measure is elicited directly as a valuation (Bateman et al., 1997a) or via choice (Kahneman et al., 1990; Knetsch, 1989; Loewenstein and Adler, 1995). Similarly, EL valuations also lie between WTP and WTA (the EL effect) (see Bateman et al., 1997a).

Reference point effects have not been confined to the laboratory; similar phenomena have been found elsewhere though usually in cases where the reference point is viewed as customary consumption rather than the endowment point. Samuelson and Zeckhauser (1988) study the annual choice of medical and pension plans made by Harvard University employees. Individual employees face a free choice from a menu of plans, but year on year, choices remain exceptionally stable, even in the face of large changes in the rates of return on pension plans and the continual introduction of new options. This is interpreted by the authors as evidence of a status quo bias, where, unlike in most of the experiments presented above, the status quo is defined as customary consumption rather than the current endowment. In a similar vein, Johnson and Hershey (1993), find that simple changes in the default option (opting in versus opting out) yield large changes in the behaviour of U.S. citizens in their choice of insurance plan. Bowman et al. (1999), collect data on consumption by U.S. teachers and find that it is much more sticky downwards in the face of negative income shocks, compared to positive changes in consumption following good news on future income. Similar evidence of loss aversion has been repeatedly found in supermarket scanner data (Hardie et al., 1993, for instance), where the elasticity of demand for price rises is significantly lower than that for price falls. Meanwhile, Simonsohn and Loewenstein (2006), uses data from the U.S. Panel Study of Income Dynamics to examine the price paid by individuals who move city. Using information on 800 individuals he finds that renters have a 'previous cost of renting' elasticity of approximately 0.14. He rejects the explanation of a selection effect by noting that subsequent within city moves lead to changes in consumption in the opposite direction of any initial change. That is, an individual who moves from a more expensive city, for example, will tend to reduce their expenditure on housing initially, but then subsequently increase it.

In financial markets, Benartzi and Thaler (1995) explain the persistence of some stock market anomalies as evidence of endowment effects. More evidence is provided by Shefrin and Statman (1985) and by Ferris et al. (1988) . US tax rates for capital gains and losses diminish as the length of time for which an asset is held increases. An investor should therefore liquidate losses rapidly (thereby getting a higher tax rebate) and hold gains (thereby facing a lower tax rate). This implies a negative relationship between the length of time a traded asset has been held and whether or not the owner lost money on it. In fact the ratio of losers to gains traded is stable over time and this too provides evidence for an endowment effect.[2]

An enterprising sequence of field experiments has been conducted by List (2003), using memorabilia traders and collectors in the USA. In one experiment 74, mostly male participants in a sportscard trading show were randomly endowed with one of two types of potentially valuable baseball souvenirs. As with the mugs and chocolate experiments described above, in the absence of endowment effects, 50% of subjects should prefer the alternative to the good with which they were endowed. In fact only around 1/3 of subjects opted to trade. List divided his sample into 'professional dealers', 'experienced' non-professional traders and 'non-experienced' non-professional traders, with the latter defined as non-professionals who traded, on average, less than 6 items of memorabilia per month. Only in the last group was the endowment effect found to be statistically significant. Similar results occurred in a mostly female Disney commemorative lapel badge (or pin) marketplace, where subjects were randomly endowed with one of two types of Mickey Mouse badge and invited to trade. As in the first example the proportion of non-experienced traders willing to trade was significantly lower than the proportion of experienced traders. A follow-up experiment which involved eliciting willingness to pay and willingness to accept values for basketball cards using a random nth price auction again pointed to a marked difference in the behaviour of professional dealers and inexperienced amateurs on the one hand and the inexperienced on the other. List interprets his results both as support for reference dependent preference theories and as a challenge: 'within certain groups of subjects that had intense trading experience, the endowment effect essentially disappears' (p. 21). The wider significance of this point is unclear, but it is a result replicated in another context by Munro and Ferreira de Souza (2007) using the inhabitants of a virtual reality online role-playing game. In List's experiment, subjects were unusually experienced in trading compared to the vast majority of individuals in most markets – 'experienced' traders were defined as those who made 84 or more trades per year, a very high number of transactions compared to those conducted for most goods by most individuals. Secondly, the subjects were self-selected in that all were recruited through memorabilia trading shows and, as Kahneman et al. (1990), p. 1328 point out, this is exactly the group of subjects

[2] This is often called the *disposition* effect Shefrin and Statman, 1985. The underlying idea is that the endowment point does not immediately adjust to the nominal financial gain or loss. As a result further changes in the value of a stock that has lost money take place in the loss domain, where individuals are risk seeking. Hence there is a tendency to gamble on keeping the stock. Compare this to a nominal gain, where further changes are evaluated in the positive domain where individuals are risk averse and therefore more likely to liquidate the asset.

where the endowment effect might be attenuated, because goods may be 'purchased for resale rather than for utilisation'. Despite this, subjects who traded less than 7 items per month showed evidence of significant and persistent endowment effects.

2.3.1 Organising the Evidence

The considerable evidence on reference dependent preferences can be organised using Fig. 2.3. Consider two bundles, x and y, composed of two goods, such that neither bundle dominates the other. Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$, where $y_1 > x_1$ and $x_2 > y_2$. Let r and s be two other bundles which will be reference points in what follows. Write $x \preceq_r y$ to mean that y is weakly preferred to x as viewed from reference point r. Define indifference and strict preference in the usual manner.

The wide discrepancies between WTA and WTP listed above suggest that a movement in the reference point from x to y shifts preferences towards y. Most of the evidence in Table 2.1 is of this kind, but as we also saw, in some cases information from reference points other than x and y is available. For instance, the results of Bateman et al. (1997a) indicate that preferences are shifted towards y both if the reference point changes from x to $r^5 = (x_1, y_2)$ and if it changes from r^5 to y. Similarly, preferences are shifted towards y both if the reference point changes from x to $r^7 = (x_2, y_1)$ and if it changes from r^7 to y.

The experiments in the table also provides information about other changes in the reference point. Consider reference points r^4 and r^6, by $r_1^4 = x_1, x_2 > r_2^4 > y_2$, $y_1 > r_1^6 > x_1$, and $r_2^6 = y_2$. Tversky and Kahneman (1991, pp. 1044–1045) and Herne (1998) report that changes in the reference point from r^4 to r^6 shift preferences towards y.

Some limited information is also available on shifts of reference point between r^1 and r^{10}, defined by $x_1 > r_1^1$, $r_2^1 > x_2$, $r_1^{10} > y_1$, and $y_2 > r_2^{10}$. Tversky and Kahneman (1991, p. 1045) find that changes in the reference point from r^1 to r^{10} shift preferences towards y. Following Tversky and Kahneman (2003); Munro and Sugden (2003) label this the advantages/disadvantages effect. We can also compare

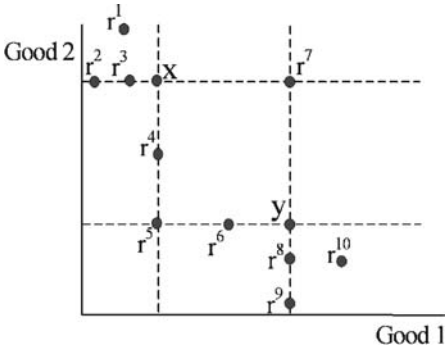


Fig. 2.3 Organising the Evidence

r^3 and r^8, defined by $x_1 > r_1^3$, $r_2^3 = x_2$, $r_1^8 = y_1$, and $y_2 > r_2^8$. The point r^3 is dominated by x but not y, while the reverse is true for r^8. Herne (1998), concludes that changes in the reference point from r^3 to r^8 shift preferences towards y.

Finally, compare r^2 and r^3, defined by $x^1 > r_1^3 > r_1^2$ and $r_2^3 = r_2^2 = x_2$. Herne (1998) finds that changes in the reference point from r^2 to r^3 shift preferences towards y. Since the labelling of the goods as '1' and '2' is arbitrary, an equivalent statement of this effect is the following: changes in the reference point from r^9 to r^8 shift preferences towards x.

An important aspect of all the experimental evidence is the rapidity with which reference points and hence preferences adjust. In the Bateman et al. (1997a) experiment, for instance, subjects were simply told about their endowments, but this was enough to alter preferences. In the Knetsch and Sinden (1984) experiment, subjects tended to prefer the status quo after only a few minutes of possession of their mug or chocolate endowment. This speed of adjustment suggests a form of myopia since the subject endowed with a mug tends to prefer it over a chocolate bar, despite that fact that were she to swap goods, she would most likely swap her preferences as well. Lowenstein and Adler (1995), offer some specific evidence on myopia in an experiment where subjects failed to anticipate changes in preferences resulting from changes in endowment.

Although the static relationship between reference points and preferences is fairly well established, the dynamic relationship between feedback, incentives and reference point effects is to date under-researched. Some studies listed above (e.g. List, 2003; Coursey et al., 1987; Shogren et al., 1994) investigate persistence without coming to a definitive conclusion, but the available evidence suggests a. that reference point effects diminish over time for some goods; b. where reference point effects weaken, WTA tends to towards WTP rather than vice versa. The clearest evidence on this issue is provided by the innovative field experiments of John List , discussed above, where high volume traders showed little evidence of a WTA-WTP disparity. Meanwhile Plott and Zeiler (2005), show that with a sufficiently intensive programme of subject training reference point effects can be eliminated in the laboratory. On the other hand, the fact that reference point effects have been apparently found in supermarket data (e.g. Hardie et al., 1993) and in financial markets suggests that for many individuals they are quite hard to eradicate through the normal repetition of everyday life.

2.3.2 Prospect Theory – Cumulative and Riskless

Having considered the evidence on reference point effects, I now turn to efforts to model them. Although there are not many alternative theories in this area, they cover the three categories of theories which are typically advanced to explain anomalies. At one extreme are models which suggest that the behaviour is rational or something very close to it. For instance, individuals may be have well-behaved preferences, but may take some time to learn about the incentives offered by unfamiliar trading environments. At the other extreme are models based more heavily on psychological

assumptions and which may lack a complete set of rigorous axioms. In between are models which replace one or more of the standard assumptions of rational choice theory with axioms based on alternative psychological intuitions.

Undoubtedly the leading contender as an alternative model of behavioural decision-making is prospect theory, due to Kahneman and Tversky (1979). Camerer (2000), for instance argues that prospect theory should replace EUT as the workhorse of modelling behaviour in the face of risk. In fact there are several different versions of this theory: the original model, 1979; cumulative prospect theory, Tversky and Kahneman (1992); and a reference dependent model, Tversky and Kahneman (1991). The first two models focus on choices made under uncertainty while the last considers preferences between riskless choices, but shares many of the same psychological underpinnings. Since cumulative prospect theory (CPT) is an up-dated version of the original theory of choice under uncertainty I shall concentrate on it, together with its close relative, the riskless choice model of reference dependent preferences.

As with its predecessor, CPT consists of two stages: editing or framing and valuation. In the editing stage, subjects understanding of risky choices is simplified by, for example the elimination of common elements and by the mental depiction of outcomes in terms of gains and losses compared to the status quo. This latter feature is fundamental to all versions of the theory and to its riskless sibling and marks a striking departure from standard models of rational choice in which it is total consumption and total wealth that are the carriers of utility.

Let S be the set of states of nature, mutually exclusive and exhaustive, with typical state s_i, i = 1,. . .,n. X is the set of (monetary) consequences and includes zero as an element. A prospect is an n vector x, which assigns a consequence for each possible state of nature. The prospect which gives zero in every state of the world is the neutral prospect.

In CPT, the utility function is,

$$V(x) = \sum_{i=1}^{i=n} \pi_i v(x_i) \tag{2.1}$$

where the π_is are the decision weights and v is the valuation function. Let $v'(x_i)$ be the first derivative of v and let $v''(x_i)$ be the second derivative. Tversky and Kahneman make four assumptions about the form of v(.):

V1 $v''(x_i)$ and $v'(x_i)$ exist and are continuous except at 0; v(0) = 0.
V2 $v''(x_i) < 0,\ x_i > 0$
V3 $v''(x_i) > 0\ , x_i > 0$
V4 $-v'(-x_i) > v'(x_i), x_i > 0.$

The first assumption is purely technical. The second assumption implies that individuals are risk seeking in losses, while V3 means that they are risk averse in gains. V4 states that the slope of the value function is steeper in its negative domain and implies that there is a kink in it at 0. Tversky and Kahneman give some psychological underpinnings to assumptions V2–V4. V4 comes from loss aversion – the

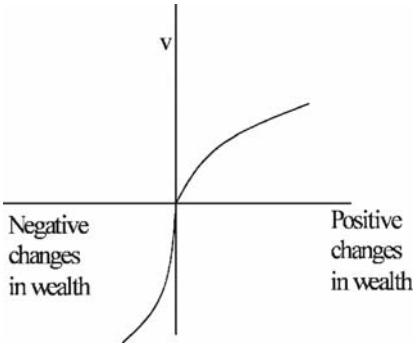


Fig. 2.4 The Value Function in Prospect Theory

tendency for individuals to weight losses more highly compared to gains of the same absolute magnitude. V2 and V3 they attribute to diminishing sensitivity , meaning that individuals become less sensitive to changes in outcomes as the distance from the status quo increases. Figure 2.4 summarises the properties of the value function, showing this pattern of slopes and the kink at the status quo point, all of which give rise to a form typically described as an 'S' shape.

The decision weights are different for positive and negative outcomes, so to derive them, we need first to re-order each prospect. Let y be a permutation of the elements of x, such that $i > j$ implies $y_i \leq y_j$. Let A_i be the sub-set of S in which y_i is the outcome and let y_J be the largest element in y which is negative. Consider first the positive elements of y. For these,

$$\begin{aligned} \pi_n &= w^+(A_n) \\ \pi_i &= w^+\left(\bigcup_{k=i}^{k=n} A_k\right) - w^+\left(\bigcup_{k=i+1}^{k=n} A_k\right) \qquad i = J+1\ldots n \end{aligned} \tag{2.2}$$

where w^+ is a function defined on the power set of S, such that

$$\begin{aligned} &w^+(\varnothing) = 0 \\ &w^+(S) = 1 \\ &w^+(A) < w^+(B) \forall A, B \in S \text{ such that } A \subseteq B. \end{aligned}$$

The decision weights for the negative elements of y are defined similarly. That is,

$$\begin{aligned} \pi_1 &= w^-(A_1) \\ \pi_i &= w^-\left(\bigcup_{k=1}^{k=i} A_k\right) - w^-\left(\bigcup_{k=i-1}^{k=i} A_k\right) \qquad i = 1\ldots J \end{aligned} \tag{2.3}$$

where the w^- function has the same properties as its w^+ counterpart.

As defined the functions are appropriate for both uncertainty (where probabilities are not necessarily known) and risk (where probabilities are known). For the case of risk, let p_i be the probability of state of the world i (with the events still placed in ascending order of payoff), the weighting function for positive outcomes is just $\pi_i = w^+(p_1 + \ldots p_i) - w^+(p_1 + \ldots p_{i-1})\, i = J + 1 \ldots n - 1$ and $\pi_n = w^+(p_n)$. Similarly the weighting functions for negative outcomes is $\pi_i = w^-(p_1 + \ldots p_i) - w^-(p_1 + \ldots p_{i-1})\, i = 2 \ldots J$ and $\pi_1 = w^-(p_1)$.

One interpretation of the weights is that they represent the distorted probabilities as perceived by individuals. If there is no distortion then $w^-(a) = w^+(a) = a$. However, Kahneman and Tversky (1979), argue against this interpretation, stating that the weighting functions are decision weights and hence 'should not be interpreted as measures of degree or belief'.

In their original formulation they suggest the shape for the weighting function shown in Fig. 2.5, where probability is measured along the horizontal axis and the weight is shown on the vertical axis. As can be seen, the suggested function overweights low probabilities, and underweights higher probabilities. Meanwhile, $w(0) = 0$ and $w(1) = 1$, but near to the extreme points there are discontinuities or areas where the function is not specifically defined. In the case of probabilities close to zero, for example, Tversky and Kahneman suggest that sometimes such probabilities are overweighted, but in some cases, the editing function removes very small probability events from consideration altogether.

Subsequently, specific functional forms for w have been suggested and estimated. For instance, Tversky and Kahneman (1992) propose the following one parameter form:

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \tag{2.4}$$

with $\gamma = 0.61$ in the gain domain and $\gamma = 0.69$ for losses. Meanwhile Prelec (1998), provides an axiomatic justification for the two parameter exponential form $w(p) = exp(-\beta(-ln(p))\alpha)$. All these suggested functions share the property of

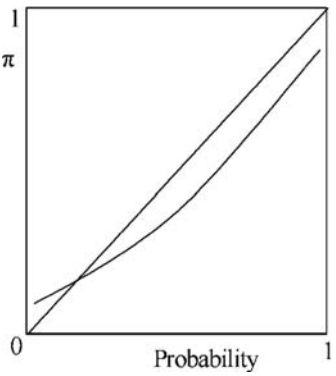


Fig. 2.5 The Probability Weighting Function

overweighting small probabilities and underweighting larger values. They also yield fixed points for the mapping at $p = 0$ and $p = 1$, but unlike the Kahneman and Tversky's original sketch there are no discontinuities as probabilities approach the extremes.

When suitable functional forms are chosen for the v and w functions, CPT can account for many of the most commonly encountered anomalies. The WTA/WTP discrepancy for instance and other manifestations of the status quo effect arise as a consequence of loss aversion, V4. Meanwhile the common ratio and common consequence effects (Starmer, 2000) are explicable in terms of the shape of the probability weighting function. Nevertheless, prospect theory is not without its critics and some properties of the model are not robust to changes in the way preferences are elicited.

This is most apparent with the property of risk loving in the loss domain. When gathering evidence about attitudes to risk, Kahneman and Tversky typically offered choices between pairs of gambles one of which was certain. For instance, a certain loss versus a gamble with some chance of a smaller loss and some chance of a greater loss. In such context, they found risk loving in losses. However, once both gambles involve the possibility of loss or when gambles involve mixtures of gains and losses, risk loving in the risk domain is no longer universal. Schneider and Lopes (1986), for instance find risk loving only when one prospect is a sure thing.[3]

Prospect theory and CPT are theories of risky choice, but most of the experiments and field trials reported above involve riskless choices. For this context, Tversky and Kahneman (1991), propose a related theory of reference dependent preferences which I will now briefly summarise.

Consider an individual's preferences over all possible bundles of n goods (with $n \geq 2$), represented by the set $\mathbb{R}^n_+$. Typical elements in this set will be represented by r, s, x, y, z, where $r = (r_1, \ldots, r_n)$ and so on. A preference relation is a binary relation on $\mathbb{R}^n_+$; $y \preceq x$ is read as 'x is weakly preferred to y'; the relations of strict preference ($\prec$) and of indifference ($\sim$) are defined from in the usual way. In Hicksian consumer theory, an individual's preferences are described by a single preference relation. In line with the approach of Chapter 1, we use the concept of a frame-dependent preference structure. Here the set of frames, $F = \mathbb{R}^n_+$ so that the frame dependent preference structure is defined as a function from $\mathbb{R}^n_+$ to the set of all preference relations; to each reference point $r \in \mathbb{R}^n_+$, a preference structure assigns a reference-dependent preference relation $\preceq_r$ (where again I drop the i subscript for

[3] Levy and Levy (2003), ask their subjects to choose between pairs of gambles where one of the options dominates the other, according to their interpretation of prospect theory. For example, $H = (-\$3,000, 1/4; 0, 1/4; \$4,500, 1/2)$ is dominated by $G = (-\$3,000, 1/4; \$3,000, 3/4)$ and in turn H dominates $F = (-\$1,500, 1/2; \$4,500, 1/2)$. In fact though, only 23% of the sample of 180 opted for G in a choice between it and F, leading Levy and Levy to conclude that, 'we strongly reject the S-shaped value function hypothesis' p. 27. However, Wakker (2003) shows that Levy and Levy misinterpret their data by omitting the possibility of probability weighting and in fact the results are consistent with prospect theory.

individuals for simplicity). The relation $\preceq_r$ describes the individual's preferences over consumption bundles when her reference point is r (or as I shall sometimes say, her preferences viewed from r). A preference structure is reference-independent if $\preceq_r$ is identical with $\preceq_s$ for all r, s. Preferences in this special case correspond with those of Hicksian consumer theory .

In their model of riskless choice, Tversky and Kahneman assume that, for any reference point, preferences are complete, transitive and satisfy monotonicity (increasingness). That is:

Axiom 2.1 (C1 Completeness) *For all r: $\preceq_r$ is complete (i.e. for all x, y: $x \preceq_r y$ or $y \preceq_r x$).*

Axiom 2.2 (C2 Transitivity) *For all r: $\preceq_r$ is transitive.*

Axiom 2.3 (C3 Increasingness) *For all r, x, y: if $x > y$ then $x \prec_r y$.*

Then Kahneman and Tversky place restrictions on how reference-dependent preferences alter with changes in the reference point. Say that a change in reference point from r to s strictly favours y relative to x if (i) y $\sim_r$x implies y $\preceq_s$x and (ii) y $\prec_r$x implies y $\prec_s$x. Similarly say the move from r to s weakly favours y relative to x if y $\preceq_r$ x implies y $\preceq_s$ x. Consider any good i $\in \{1, ..., n\}$, any bundles x, y such that y_i¿x_i, for some i and any reference points r, s such that $s_i > r_i$ and $r_j = s_j$ for all j $\neq$ i. Tversky and Kahneman assume:

Axiom 2.4 (K4 Loss Aversion) *If $y_i < s_i$ and $x_i = r_i$, then the move from r to s strictly favours y relative to x.*

Axiom 2.5 (K5 Diminishing Sensitivity I) *If $x_i < s_i$, then the move from r to s weakly favours y relative to x.*

Axiom 2.6 (K6 Diminishing Sensitivity II) *If $r_i < y_i$, then the move from r to s weakly favours x relative to y.*

The first of these assumptions represents loss aversion in a riskless world. It means that if a bundle y is preferred to x when view from x, then it will certainly be viewed as superior when y itself is the reference point. The second and third assumptions are the riskless counterparts of assumptions V2 and V3 in prospect theory. The intuition here is that, when the reference points are relatively close to x and y the perceived differences between the bundles being compared are magnified compared to a situation where r and s and distant.

Now the theory of riskless choice has certain drawbacks which limits its acceptability as a model of economic behaviour. First it allows non-convexity in preferences. Secondly, if the reference point is taken to be current consumption, it allows cycles of preferences in which B is preferred to A from A, C is preferred to B from B, but then A is preferred to C from C. Thirdly, there is no assumption of continuity in reference points, so that small changes in the vantage point can have drastic implications for the preference map. In addition the additive form favoured by Tversky and Kahneman implies extreme Slutsky substitution effects: either individuals make no change in consumption of good i when faced with a compensated rise in its price

or they cease to consume the good altogether. Finally, assumptions K2 and K3 are largely unnecessary in order to explain the empirical evidence which provides the backdrop to the model.

Alternative models of reference dependent preferences have been proposed. To a significant degree they are more 'economic' in tone. For instance, Sileo (1995) has a model composed of two elements. The first element axiomatises the idea that individuals may be unsure of their preferences. The second part introduces a decision rule which is extremely conservative: individuals switch away from their current consumption bundle only when they are certain that they truly prefer the alternative. This assumption generates a status quo bias, but is silent on the class of results discussed above where individuals choose between options which do not include the reference point. Meanwhile Carmichael and MacLeod (2006) formalises the notion that, in general, overstating WTA and understating WTP may be adaptive for humans used to bargaining situations. With inertia in behaviour, such strategies can carry over to contexts where they are no longer optimal. Masatlioglu and Ok (2005), propose a model where when there is a status quo there is a cost of moving from the status quo that is deducted from the utility of alternatives (see also Sagi, 2006).[4] Munro and Sugden (1997, 2003), have a model of reference dependent preferences which avoids psychological axioms in favour of an assumption of acyclicity. Munro (1997), offers another alternative which restricts the set of reference point changes which can alter preferences, essentially by dropping K2 and K3. These last two models are presented in Chapters 4 and 10 respectively, so for now I shall leave them to one side.

2.4 Mental Accounting

Mental accounting is a term used by Richard Thaler (1981, 1985), to describe models and rules that individuals use to organise their finances. It consists of three types of rules: those concerning the framing and perception of consumption and expenditure; those concerning the assignment of spending to particular accounts and finally rules about how often accounts are audited and balances settled. Because of this, mental accounting might better be viewed as a cluster of interacting anomalies and heuristics rather than a single effect, but their shared implication is that all the money in the budget is no longer fungible. My reason for including it here is primarily due to the importance that budgets and other constraints play in real-world decisions. Many common anomalies such as preference reversal or reference point effects are elicited in the context of experiments in which issues of budgets are not to the fore,

[4] Choice theories that model status quo bias as a fixed psychological switching cost tend to suffer from two empirical weaknesses. First for any strictly positive value of the switching cost some bundles that dominate the status quo would be rejected. Secondly, curves showing bundles indifferent to the status quo are discontinuous around the status quo – so that there are no alternatives local to the status quo that are strictly preferred.

so it is important to realise that anomalies can occur even when individuals do face the incentives created by budget constraints.

An instance of mental accounting is provided by the following example:

> Imagine that you about to purchase a jacket (calculator) for $125 ($15). The salesman informs you that the item you are about to buy is on sale for $120 ($10) at the other branch of the store, located 20 minutes drive away. Would you make the trip to the other store? (Tversky and Kahneman, 1981, p. 457).

When Tversky and Kahneman (1981), set this question they found that most respondents would travel to save $5 off the $15 calculator but not for the price reduction on the jacket. Such a decision is consistent with a model of accounting which relates losses and gains to a reference point closely connected to the choices on offer; it is not consistent with global budgeting.

In their discussion of this and other experiments, Tversky and Kahneman (1981), set out three levels at which accounts may be framed: minimal accounts, in which options are compared without reference to other choices or contexts for choice, topical accounts in which some context is introduced and comprehensive accounts where a wider set of considerations come into play. Partly the level is evoked by the size of the expenditure. So, for instance, in the calculator/jacket example they theorise that in the case of the calculator, the relevant account is minimal - subjects think only about whether it they are willing to drive fifteen minutes to save $5, but for the jacket example, the higher baseline price becomes part of the frame, evoking a topical account in which it is not worth the effort to save $5 on a $125 item.

One of the important lessons of the mental accounting literature is that sunk costs sometimes are not sunk. Instead, with capital goods bought for consumption purposes, individuals act as if the payment for the good is entered as a loss in a new mental account. Subsequent enjoyment of the good then produces a stream of benefits which are set against the initial loss in the same account. The good is then psychologically sunk when the account breaks even. If the good is lost or sold before that point, then its surrender involves a loss. Past the break even point, however, there is no significant feeling of loss. An example of this is provided by Okada (2001), who offered a hypothetical choice to her subjects, half of whom were given the option to trade in their older camera for $80 in order to buy a new and superior camera for $200. The other subjects were offered the $200 camera with an $80 discount. In Hicksian theory, the second option dominates the first, since the subjects end up with two cameras, but in fact subjects were more likely to accept the trade-in deal, compared to the straight discount. The difference in sub-groups was most marked for individuals who reported bad experiences with their current camera. Okada's explanation for this and other similar results is that individuals do not integrate the gains and the losses into a single account.

Field evidence on sunk costs is provided by, for example, work by Arkes and Blumer (1985) and by Gourville and Soman (1998) who show that the effect seems to fade after a while, in a manner termed 'payment depreciation' by the latter pair. They gathered data on health club use from a company which levies its membership fees twice a year. Use of the facility was highest in the months immediately

following payment of fees, tailing off until the next billing point, when it jumped up once more. Arkes and Blumer, meanwhile, conducted a field experiment amongst US purchasers of a season ticket for a college theatre company: one sub-sample received a small discount of 13%, one received no discount, while the other obtained a discount of 47%. Attendance at plays during the first half of the season was highest amongst play lovers who had paid full price for their tickets and significantly lower for the groups who received a discount, but in the second half of the annual season, there was no statistical difference between the three groups.

Thaler (1980), offers the example of 'Mr. H [who] mows his own lawn. His neighbour's son would mow it for $8. He wouldn't mow his neighbour's same-sized lawn for $20' (p. 43). If the time it takes to mow a lawn is valued by Mr. H at v, then his refusal to employ his neighbour's son (leaving aside issues of quality control) suggest that $v < \$8$, but then his refusal to mow his neighbour's lawn points to $v > \$20$. One way to explain this phenomenon is to think of the two lawns as different goods. Perhaps mowing a neighbour's lawn for $20 would mean a loss of dignity for Mr. H. On the other hand many similar examples suggest another explanation: that, despite the predictions of standard theory, out of pocket expenses seem to carry more psychological weight than opportunity costs.

Thaler, 1985, uses this observation to argue that the net utility of a transaction is best understood as a combination of two effects. Consider the following question posed to two sub-samples:

> You are lying on the beach on a hot day. All you have to drink is ice water. For the last hour you have been thinking about how much you would enjoy a nice cold bottle of your favourite brand of beer. A companion gets up to go and make a phone call and offers to bring back a beer from the only nearby place where beer is sold (*a fancy resort hotel*/a small run-down grocery store). He says that the beer might be expensive and so asks how much you are willing to pay for the beer. He says that he will buy the beer if it costs as much or less than the price you state. But if it costs more than the price you state he will not buy it. You trust your friend and there is no possibility of bargaining with (*bartender*/store owner). What price do you tell him? (Thaler, 1985, p. 203).

The only difference between groups was that one group saw the italicised portion of the question while the other group saw the alternative words. The mean reservation price for beer in the first case was $2.65; for the second case it was $1.50. Thaler explains this difference in terms of acquisition and transaction utilities. The former is more standard: it is the gap between the value of an item and the price paid; the latter is a reference-price effect – utility is increasing in the difference between the reference price for the good and the actual price paid. When this gap is large, the consumer feels they are getting a good deal; as it narrows so the benefits of the deal disappear. The key assumption here is that the reference price is conditional on the type of outlet: higher in the case of the luxury hotel; lower for the local corner shop.

As is clear from the previous example, much of the mental accounting literature is compatible with the predictions of prospect theory, especially if we are prepared to assume that individuals are able to choose their mental accounts strategically. To see this, consider two transactions, x and y and a consumer's choice of whether to enter them as a single, combined transaction or place them into separate accounts.

Let $v(x + y)$ be the value when x and y are integrated into one mental account and let $v(x) + v(y)$ be the value when they are treated as separate accounts. Recall that losses are weighted more heavily than gains and that the value function is concave for gain and convex for losses. Then we have the following properties:

1. $v(x) + v(y) > v(x + y)$ when both x and y are positive
2. $v(x) + v(y) < v(x + y)$ when both x and y are negative
3. $v(x)+v(y) > v(x+y)$ when x is a large gain and y is a sufficiently small loss and
4. $v(x) + v(y) < v(x + y)$ when x is a small gain and y is a sufficiently large loss.

In this framework, to maximize total value subjects should separate gains, integrate losses, separate out small gains from losses and integrate small losses into large gains. However, one of the weaknesses of the assumption that consumers have full control over the choice of accounts is in its implication for the treatment of gains. As Thaler (1985), points out, if a gain can be broken down into smaller and smaller gains utility can be still higher and higher. Specifically, property 1 implies that $nv(x/n) > v(x)$ for $n > 1$, implying that the optimal treatment of a gain is to break it down into infinitesimal parts. Thus it seems more reasonable to suppose that consumers do not have full control over their ability to manipulate accounts or that there are costs associated with mental accounting frameworks. There is as yet no formal theory of how such a cost structure might be constructed, though the lessons from the literature discussed above suggest that it is easier to integrate two items of income or expenditure when the goods are similar or when the two events are close in time.[5]

This leads to one of the key features of the mental accounts: their dynamic nature which means that individuals may have preferences about the timing of expenditures which are incompatible with standard models of consumer behaviour. Consider Fig. 2.6 which shows a pattern of consumption and its associated expenditure over time for one individual. For instance, the commodity might be a football or theatre season ticket and the expenditure could represent regular transfers from a bank account in payment. The solid bars indicate two instances of consumption (above the line) and one act of payment (below the line). In line with much of the literature on mental accounting, Prelec and Loewenstein (1998), argue that for many individuals there are simultaneous 'costs of consumption' and 'benefits of payment'. Costs of

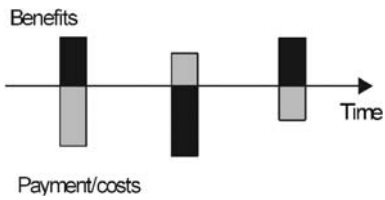


Fig. 2.6 Mental Accounting and the Timing of Consumption

[5] Further thoughts on the limitations of prospect theory for mental accounting are offered in Chapter 10.

consumption may involve guilt or feelings of regret that a good has been used up, while the benefits of payment can include relief at no longer having to worry about a debt. In the figure, they are shown in grey; they represent hedonic echoes of the real transactions that in part cancel out the pain or pleasure of the original experience. By manipulating the timing of consumption and payment, the consumer may be able to affect the size of the offsetting terms. To gain insight into how this occurs, Prelec and Loewenstein faced 89 individuals with a ranking task involving the timing of payment and use of a hypothetical time share, where the individual had purchased 3 weeks of use of a resort apartment to be enjoyed over a period of four years at a total price of $3,000. Some options involved pre-payment, some involved use of the apartment in the early or years, some involved a matching of the timing of payments to the year of use etc. To elicit discount rates, they also asked their subjects to rank options involving only benefits or only payments.

Individuals with a positive rate of time preference – which characterised most subjects in their sample – should prefer consumption and payment patterns with the payments end-loaded. In fact, a majority of subjects did not rank such options highly. The largest group favoured either matching year of use to year of payment, but with payment preceding use or preferred options with some pre-payment and then paying the rest of the fee in installments prior to use.

Prelec and Loewenstein argue that this is indicative of a high cost of consumption when a good has not been paid for. But they also conclude that there is a higher benefit of payment when it is close to the period of consumption. Now individuals with this kind of mental accounting system would be extremely averse to incurring debts, but as the authors acknowledge high levels of credit card spending and debt are a feature of many households. They suggest that this is due to hyberbolic discounting which in turn leads to procrastination .

This brings us to the final element of mental accounting: the control of the flow of savings and consumption over time. Thaler (1981), suggests that mental accounts represent one of the ways in which the controlling or planning part of the self can limit the temptation to consume rather than save. Labelling separate accounts as savings or capital accounts create psychological barriers to using money for day to day expenditure. Moreover, if payments into savings plans are automatically made, then they are not coded as a loss from current consumption, but instead are more likely to be integrated with the future gains from invested income.

Outside the laboratory, wider evidence consistent with mental accounting has come from a number of studies tracking consumption expenditure between pay days. Huffman and Barenstein (2004), use Family Expenditure Survey (FES) data to show that average consumption expenditure falls by about 12% between paydays for UK households. Stephens (2003, 2004) examines Consumer Expenditure Survey (CES) and FES to document a similar cycle of consumption for US social security recipients and UK consumers respectively. Meanwhile, after analysing energy intake data for food stamp recipients in the USA, Shapiro (2004), finds a declining path of consumption. In a similar line of research, Baker et al. (2005), use CES data to show that US households consume differentially out of dividend payments and stock appreciation. The point about this data is that it contradicts the standard

theories of consumption smoothing in the face of anticipated, but intermittent receipts of income. One potential hypothesis is that households face credit constraints. While this is possible for families on food stamps or claiming social security, it is less likely to be a compelling explanation for the wide range of households in the UK FES samples.

Summing up, the term mental accounting covers two types of behaviours: strategic and contextual. In the latter case, particular frames of reference are evoked by specific choice situations (e.g. Tversky and Kahneman's example of saving $5 by going to another store). In the former case mental accounts are used as commitment devices to control time inconsistency and as simplifying accounting procedures to limit the mental costs of decision-making. The difference between the two types of behaviour lies in the degree of control the individual has over the framing and design of mental accounts. With the current state of research into the subject it is not particularly clear how much control individuals have over mental accounts. Obviously though this is a crucial issue for public policy. If mental accounts can be evoked and controlled easily by external factors and agents, then effective government policy might extend to the design of mental accounts. Even if mental accounts are internally generated for strategic reasons, then the design of tax and savings policy might still be improved by taking into consideration the typical format of such accounts. These possibilities are taken up in Chapter 10.

2.5 Preference Reversal

Preference reversal arises when the relative evaluation of two (or more) items is dependent on the means by which value is elicited. In the classic task that gives rise to the phenomenon, there are two items: subjects are asked to choose between the two of them. They are also asked to place a monetary value on each item separately (typically this is a minimum selling price). When an individual chooses one item but places a higher valuation on the other item, they exemplify the phenomenon of preference reversal.

What makes preference reversal an anomaly rather than evidence of some randomness in behaviour is the evidence that at least for one class of items, the overwhelming majority of reversals are in one direction. There is a systematic tendency for individual decisions under risk to display preference reversal making it one of the most important and robust anomalies in observed decision-making. In the typical laboratory experiment (see for instance, Lichtenstein and Slovic (1971), Lindman (1971) and Grether and Plott (1979)), the items are two monetary lotteries, of which one (the 'P-bet') offers a relatively large chance of a relatively small prize (otherwise nothing) while the other item (the '$-bet') offers a smaller chance of a significantly bigger reward. Valuations are minimum selling prices, usually obtained via a Becker-De Groot-Marschak mechanism. The most common in these circumstances (and therefore often called a 'standard reversal'), is a tendency for experimental participants to choose the P-bet over the $-bet in the choice, but to place a bigger monetary value on the $-bet. For instance in the Lichtenstein and Slovic (1971),

experiment 56.2% of those who chose the P-bet then valued the $-bet more highly, while only 11.2% of those who opted for the $-bet valued the P-bet more highly (Seidl, 2002).[6]

In a much cited paper, Grether and Plott (1979), sought to examine the robustness of the earlier work by psychologists, acknowledging in their opening paragraph, 'this paper reports the results of a series of experiments designed to discredit the psychologists' work as applied to economics. Grether and Plott (1979, p. 623). Subjects in the Grether and Plott (1979) experiments therefore faced significant incentives and a random lottery design employed to reduce income effects; in different treatments the language of instruction was altered to test the theory that preference reversal resulted from misplaced bargaining behaviour[7] and the possibility of registering indifference was allowed. The authors find results similar to those previously obtained by psychologists: in the group that faced incentives, out of 273 choices there were 69 preference reversals in which the P bet was chosen and the $-bet valued more highly and 22 reversals of the opposite kind. The authors summarise thus:

> Taken at face value the data are simply inconsistent with preference theory and have broad implications about research priorities within economics. The inconsistency . . . suggests that no optimisation principles of any sort lie behind even the simplest of human choices . . . (Grether and Plott, 1979, p. 623).

Though numerous explanations for the preference reversal phenomenon have been put forward, for present purposes, it is helpful to separate them into two groups.

One class suggests that it arises from preferences that break the axioms of expected utility theory. For example, it could stem from intransitivity (as for instance, the regret theory explanation of Loomes and Sugden 1983; Loomes et al., 1989) or from violation of the independence of irrelevance of alternatives, interacting either with the Becker-De Groot-Marschak mechanism (Karni and Safra, 1987) or with the random lottery incentive system used in many of the relevant experiments (Holt, 1986), or from failure of the reduction principle for compound lotteries, interacting with the Becker-De Groot-Marschak mechanism (Segal, 1988). A feature of explanations in this group is that they do not deny the existence of coherent preferences; instead they relax some of the axioms of expected utility theory. For instance, a typical feature of most experiments that produce preference reversal is the use of a random lottery device to decide which of the tasks faced by the subjects will be played out for real. If the independence axiom of expected utility does not hold, then it is not necessarily incentive compatible for subjects to opt for P in the experiment when P is valued more highly than $ and vice versa.[8]

---

[6] Most examples of preference reversal involve monetary gambles of this kind, but it has been found elsewhere e.g. in the valuation of environmental improvements – see Irwin et al. (1993).

[7] In selling it is often optimal to overstate a reservation price. Grether and Plott (1979) argue that overstating might be more likely to occur with the $ bet, since the high probability attached to winning in the P bet limited the ambiguity of its value.

[8] Suppose there are only 3 tasks: choice and the two valuation exercises. Let $\overline{P}$ be a compound lottery with P as the payoff if the reservation price is less than the random selling price and the

A second set of theories, mainly proposed by psychologists, argues that individuals respond to tasks using heuristics that are sensitive to the exact nature of the task. On this view, there is no presumption that coherent, frame-independent, preferences exist at all. Rather, preferences are constructed in response to the tasks faced. Different methods for eliciting preferences, such as choice methods or valuation methods, may produce different sets of preferences. Specific psychological theories include the three known as prominence, compatibility and task goal hypotheses, respectively. To understand them, it is helpful to think of the items in the classic preference reversal experiment as having more than one attribute, such as probability or payoff, and of all of these attributes as being relevant to the decision in a given type of choice or valuation task. The prominence hypothesis presumes that there is one attribute that stands out and that it figures more heavily in choice tasks than in tasks, such as valuation, which involve matching (see Tversky et al., 1990; Slovic, 1995). The compatibility hypothesis is that the subjective weight attached to an attribute is greater the more compatible that attribute is with the task in question (see Tversky et al., 1990; Tversky et al., 1990; Slovic, 1995; Slovic et al., 1990). The third variant, the *task goal* hypothesis (Fischer et al., 1999), shares with the prominence hypothesis the notion that there is one prominent attribute. It adds to this by suggesting that the weight given to this attribute depends upon the goal of the task, as perceived by the subject. The theory then supposes that the prominent attribute is weighted more heavily in tasks whose perceived goal is to differentiate between items (e.g. choice) than in tasks whose perceived goal is to equate them (e.g. valuation).

To see that all three theories are consistent with the classic evidence of preference reversal, it is simplest to confine attention to gambles which give a particular prize with some probability and zero otherwise. Such gambles have two attributes: money-prize and win-probability. The classic pattern of preference reversal between such lotteries can be explained by the prominence hypothesis provided that win-probability is the prominent attribute. According to this view, choices are more affected by win-probability than are valuations creating a tendency for the P-bet to be chosen even when the \$-bet is valued more highly than it. Similar reasoning makes the task-goal hypothesis consistent with the classical form of preference reversal. Moreover the evidence can also be explained by the compatibility hypothesis, if monetary prizes are more compatible with monetary valuation tasks than with choice tasks. On this view, subjects pay more attention to the size of the monetary prizes when putting monetary values on gambles than they do when choosing between them. This induces a tendency to value the \$-bet more highly even when the P-bet would be chosen in a straight choice.

---

random selling price if the converse is true. Define $\overline{\$}$ in a similar fashion. With a random lottery, in choosing between P and \$ the subject actually faces a choice between the compound lotteries $A = (1/3, \overline{P}; 1/3, \overline{\$}; 1/3, P)$ and $B = (1/3, \overline{P}; 1/3, \overline{\$}; 1/3, \$)$. When Independence holds if P is preferred to \$, then A will be preferred to B, and vice versa, but if Independence does not hold then for some preferences an individual may prefer P over \$ (and value it more highly), but choose B over A.

In the ordinal payoff scheme, first proposed by Cox and Epstein (1989), and Tversky et al. (1990), each subject faces three tasks for each pair of bets: choice between P and $; valuation of P and valuation of $. Unlike standard BDM schemes, at the end of the experiment either *Choice* or *Valuation* is selected at random for the player. If *Choice* is selected at random, then the subject receives his or her chosen lottery and plays it out. This is identical to the standard random lottery. If *Valuation* is selected then the valuations of the two bets are compared and the subject receives and plays the one with the higher valuation. It follows that the 'valuation' tasks are in fact choices. As such, as long as an individual has context free preferences (whether or not they obey EUT), she or he should rank the two bets consistently. Cubitt et al. (2004), use the scheme and find the standard pattern of preference reversals, thereby rejecting the first category of explanations for preference reversal. In a second treatment, they ask subjects to 'value' the bets using the probability of winning £10 as the valuation device. According to prominence theories of preference reversal, the standard results should still obtain, whereas with the compatibility thesis, the standard pattern of reversals should disappear to be replaced by a predominance of reversals in the opposite direction. Since the task of participants is always to choose, the task-goal hypothesis suggests no reversals. The results from this second treatment are inconclusive: there are many reversals, with no particular pattern. Thus, individually, no psychological theory is supported, but there is still clear evidence that, for many subjects task context affects choices.

Preference reversal is a general term and as such it has been applied to different contexts. For instance, in List (2002), subjects are asked to value bundles of sports cards at a collector's fair. The bundles contain cards in excellent condition, but in some cases damaged, but still potentially valuable, cards are added to the collection. In other words the larger bundle dominates the smaller bundle. In valuation the larger bundles are often valued significantly below the smaller bundles, whereas in choice or when the bundles are valued side by side, the larger bundle is always chosen or valued more highly. This kind of preference reversal ('less is more', List calls it) can also be found in earlier work by Hsee et al. (1999), using hypothetical choices and students as subjects. As with the more common kind of preference reversal therefore, the lesson is that the form by which preferences are elicited can affect relative expressions of value even in market settings.

2.6 Conclusions

The main conclusions of this chapter are threefold. First, behaviour which departs from the precepts of the rational choice model is widespread. These anomalies are not confined to the laboratory or to non-experts or to situations with low incentives. Secondly, anomalies are not confined to preferences, but are also present in information processing. Finally, the deviations from the rational choice models are, at least to some extent, predictable – not in the sense that there is a model of behaviour which yields accurate predictions of decision-making in all circumstances,

but in the sense that there is a reliable link between the framing of decisions and the preferences elicited.

In the evidence presented in the preceding sections, I have drawn on a mixture of evidence: real-world data, experiments with and without incentives as well as field experiments. The anomalies found in that data are consistent in the sense that similar designs or context produce similar deviations from the Hicksian model. The WTA-WTP disparity is an instance of this. However, it is far less clear whether the anomalies are consistent with one another, in the sense that it is possible to put together a single model of decision making which adequately explains all of the results. Within economics, the most widely quoted alternatives to the Hicksian model are based upon cumulative prospect theory, as developed from the original prospect theory proposed by Kahneman and Tversky (1979). Reference point effects are consistent with this theory. The Ellsberg paradox which will encounter in the next chapter is also compatible with CPT provided decision weights can be modified by the degree of uncertainty about the probability of an event. CPT though has some deficiencies, notably the way it allows cycles in choice and in its less than clear predictions about editing effects. Nevertheless, in the chapters which follow we shall sometimes encounter the use of prospect theory as a relatively simple alternative to the Hicksian model when considering the implications of public policy for behaviour and vice versa.

Potentially, results of the kind summarised above have implications for both normative and positive aspects of public economics. On the positive side, they suggest that, compared to the set of assumptions which comprise the weak rational choice model, something richer is required to predict and explain adequately the behaviour of citizens and their rulers. On the normative side they raise two kinds of issues. First, how do we measure individual welfare, when indicators of it such as stated and revealed preference are prone to biases? Secondly, how should goods be allocated given these biases?

Before we face these questions in Chapter 5, I wish to consider two issues of special concern to economists. Most, but by no means all of the evidence discussed in this chapter comes from one shot decisions in laboratory settings. Mostly, though, economists are concerned with decisions made repeatedly in market setting. There are a number of theoretical arguments why this combination of repetition and incentive might produce quite different kinds of behaviour. Possibly, anomalies are eradicated by experience or possibly the behaviour of markets might not be altered by the existence of anomalies. So, in the next chapter I consider learning and then in Chapter 4, I examine the possible impacts of reference dependent preferences on markets.

Chapter 3
Information, Learning and Markets

3.1 Introduction

This chapter takes up some of the unfinished business of Chapter 2. In particular I consider the influence of markets and learning on anomalies. We consider the following three hypotheses:

1. Individuals incorporate information according to Bayes' theorem.
2. Through learning, demand converges to some underlying pattern of preferences given any frame. In other words through repeated choice $x(p, m, f) \rightarrow x(p, m)$ for any frame f
3. Through repeated choice $x(p, m, f^*) \rightarrow x(p, m)$, when f^* is a market frame .

Hypothesis three is an important special case of hypothesis two. Meanwhile hypothesis 2 is more likely to be true if hypothesis one is also true, because it means that information is processed consistently.

My primary reason for taking up these issues is one prominent view within economics that repeated exposure to the incentives offered by markets will eliminate anomalies (see Binmore, 1999, for instance). Moreover, within non-market valuation advice on the optimal frames for preference elicitation often recommends an elicitation mechanism that mimics the psychology if not the incentives of the market place (e.g. Sugden, 2003a). Partly this advice is based on the idea that the familiarity of a market frame will evoke behaviour which is less prone to anomalies.

Anomalies in information processing have special relevance in the context of merit wants or policy towards bounded rationality. One commonly-made criticism of merit want models is that rather than subsidizing consumption of merit goods or taxing de-merit goods, such as tobacco, governments should limit their liberty-constraining actions to delivering accurate information about the costs and benefits of different choices.[1] If, though, the processing of information is sub-optimal then simply providing it may not bring about the change in behaviour desired.

[1] In fact information provision may not be optimal even in the context of the rational consumer (see Chapter 7 for an example).

A. Munro, *Bounded Rationality and Public Policy*, The Economics of Non-Market Goods and Resources 12, DOI 10.1007/978-1-4020-9473-6_3,

© Springer Science+Business Media B.V. 2009

There are two main conclusions to be had from the chapter. On the one hand, there are good reasons to be sceptical about the power of markets to eliminate anomalies. Those reasons are both theoretical and empirical. On the other hand, it is clear that behaviour does change with repetition and it typically does so in a way which makes it more consistent.

3.2 Probabilities, Information Processing and Bayes' Theorem

The previous chapter dealt with some recorded examples of preferences which stray from the rational choice model, but rationality is often taken to imply more than a set of complete and consistent preferences. Consider two events A and B, which may linked in the sense that observation of A may be informative of the likelihood of B occurring and vice versa. Let P indicate probability with $P(B \mid A)$ meaning, in the standard notation, the probability of B given A and let $P(A \cap B)$ be the probability of A and B Bayes' Theorem is then,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

Bayes' Theorem is usually taken to be the rule for a rational person faced with the task of incorporating new information into their decisions. As with the axioms of utility theory, Bayes' theorem implicitly makes predictions about what should not affect beliefs as well as predictions about factors that should alter probabilities.

Tests of Bayes' theorem are probably not as common as tests of EUT, but nevertheless a considerable body of evidence has built up that individuals are not Bayesians. Tversky and Kahneman's 1982 volume collected together much of the psychological literature on tests of information processing much of it devoted to tests of Bayes' theorem. The general conclusion of this literature was that individuals were not rational information processors, but instead tend to rely on heuristics or simple information processing rules of thumb, the sub-optimality of which (or bias) could be demonstrated using simple experiments. Some typical heuristics and their biases are outlined below.

3.2.1 Representativeness and Availability Biases

Kahneman et al. (1982), argue that one broad class of anomalies can largely be explained by the heuristic of representativeness, meaning that, when asked what is the probability that A belongs to class X rather than Y individuals focus on the extent to which A is typical of the class X rather than class Y. As a result, subjects ignore sample sizes and prior information. For instance, average heights for American men are below six feet tall. As a result, the probability that 60% of a sample of men exceed six foot declines as the sample size increases. However, Tversky and Kahnemann reported that estimates were the same for samples of 10, 100, 100.

Table 3.1 Representativeness Heuristics

Bias	Example
Insensitivity to sample size	Subjects are told that in one sample 4 out of 5 balls taken from an urn were red. In another sample of the urn, 12 out of 20 balls turned out to be red. Subjects are asked which item of evidence provided more support for the hypothesis that the urn contains 2/3 red balls rather than the opposite? Most subjects opt for the for first sample and underestimate the posterior odds that the converse is true.
Insensitivity to Prior information	Subjects are given information about 100 people, and asked to state a probability that each person was lawyer or engineer. Reported probabilities are insensitive to being told that 70% of the 'sample' were actually lawyers or 70% were actually engineers. When information on the subjects is neutral – meaning that it did not conform to stereotypes about particular occupational groups, subjects tend to stick to a probability of $\frac{1}{2}$ that a person belonged to each category even when told population percentages which differed from 50%. Bayes Theorem in this context predicts that, for neutral information, subjects give a figure of 70%.
Insensitivity to predictability	Predictions are insensitive to reliability. Subjects are asked to score a student teacher's lessons on the basis of a few paragraphs of description, then asked to predict the score of each student in five years' time. Scores were identical or as extreme. In other words, subjects were apparently rejecting a model where some of the student teacher's performance was due to chance and hence that there would be some possibility of regression to the mean.
Misconceptions of Regression	Flight instructors who praise good landings and criticize bad landings or football coaches who praise performances in a win and criticize players for losing discover that, on average bad is followed by better and good by worse and thereby conclude that praise is misguided.

In another of their experiments subjects were informed about two maternity hospitals, one of which typically has 45 births per day, the other of which has 15. Subjects were asked to predict which hospital recorded more days where greater than 60% of the babies born were boys. Of course it is much more likely that the hospital with smaller numbers of births per day should have the greater sample variability and therefore the higher number of days where the proportion of boys exceeds 60%, but 21% went for each of the two hospitals while 53% stated that the number of days should be about equal.

Table 3.1 gives further examples of the effects of representativeness, drawn from Kahneman et al. (1982) and Nisbett and Ross (1980). Instances of another type of general heuristic at work, availability , are outlined in Table 3.2. Here, probabilities are partly assessed on the ease by which examples are brought to mind.

3.2.2 Adjustment and Anchoring

A third class of biases centre around the heuristic of anchoring and adjustment. In this group of effects, an individual's response is correlated with some initial

Table 3.2 Availability Heuristics

Bias	Example
Retrievability of instances	Subjects listen to list of well-known personalities then they are asked to state whether the proportion of men or women is higher. A different list is given to different groups; with some a higher proportion of more famous women is included, in the others, the reverse is true, but the proportions of each sex overall are the same. Subjects tend to judge proportions according to which sex has the more famous personalities mentioned.
Effectiveness of a search set	Are there more words where r is the first letter, or more where r is the third? Subjects tend to state that r first is more common (even though it is not). The same pattern is found for abstract words (love) over the concrete (door).
Imaginability	How many committees of k members can be formed out of ten people? Subjects imagine partitions and so construct a monotonically decreasing plot of number against k (whereas it is symmetric, bell-shaped). e.g. 70 committees of 2 guessed; 20 committees of 8, whereas it is 45 in both cases.
Illusory correlation	Subjects are shown information on mental patients consisting of a diagnosis plus a drawing of a person by the patient. Afterwards they are asked to judge how often diagnoses (e.g. paranoia, suspiciousness) accompanied certain features of the drawings. Subjects overestimated frequency of 'natural associates', even when there was a negative correlation; and missed out on actual associations present.

information available to them. When, according to rational choice theories, responses should be uncorrelated with the information, anchoring represents an anomaly. Tversky and Kahneman (1974), give the example of subjects asked to estimate the percentage of African countries in the UN. Prior to each question subjects watched a 'wheel of fortune' determine a number between 0 and 100. The answers to the question were correlated with the number produced by the wheel e.g. 25 was median estimate if number was 45, 45 if number was 65. Payment for accuracy had no effect.

According to Tversky and Kahneman anchoring is also the source of bias in a class of anomalies where individuals evaluate conjunctive and disjunctive events. For instance, consider the following example in which balls are to be drawn from a bag containing red and white balls. Subjects are asked to estimate probabilities for the following cases: the probability that any one ball will be red given equal numbers of red and white in the bag; the probability that 7 balls in a row will be red, given that 90% of the contents are red or that at least one ball will be red out of seven tries with 10% of contents being red? The first case is simple; the second involves conjunctive events and third case involves disjunctive events. The actual probabilities are 0.5, 0.48 and 0.52, but according to Tversky and Kahneman, people tend to overestimate conjunctive event probabilities and underestimate disjunctive. They argue that this is an example of anchoring, with the anchor provided by the simple event.

The first kind of anchoring effect has been widely replicated in economics, particularly in the kind of stated preference methods used widely in marketing,

healthcare evaluation and environmental economics. Hurd et al. (1997), found such an effect on a significant scale within the Asset and Health Dynamics of the Oldest Old (AHEAD) survey – a panel which provides information on elderly Americans and where, to reduce problems of non-response, individuals who answer 'don't know' to consumption questions are faced with a series of bracketed alternatives. Estimates derived by the methods used in the survey were sensitive to the anchors presented to the elderly subjects, to the extent that some figures for mean consumption could vary by as much as a factor of two, depending on the values used for brackets.

In the dichotomous choice methodology used widely in environmental valuation, subjects are asked whether they are willing to pay x, where x is varied over the sample. In some instances, follow up questions, based on the responses to the first question, are used to improve the efficiency of the estimates. For example, a subject who states that she or he is willing to pay £10, may then be asked if they are willing to pay £20. Conversely, if they are unwilling to pay the first sum, they may be offered a lower figure. It is a familiar feature of contingent valuation that the estimates for demand curves differ systematically, depending on whether follow up questions are asked or not. If anchoring occurs then saying 'yes' to £y is more likely if a subject has previously said 'no' to £x, where $x > y$. Similarly, saying 'no' to £z is more likely if the subject has previously said 'yes' to £x where $z > x$. Compared to a design with no follow-up, the net effect of anchoring therefore depends on whether follow-up questions are only asked of subjects who reject the first option, only offered to individuals who accept the first option or offered whether subjects reject or accept. In the first case, anchoring will raise estimates of willingness to pay; in the second case, willingness to pay figures will fall and in the third case the impact is ambiguous.

Using data from a survey of willingness to pay for management improvements for a specific lake in the USA, Herriges and Shogren (1996), compare anchoring effects for residents and visiting recreationists. They find a negative though insignificant effect for local residents and a positive and significant effect for recreationists. In fact for the latter group, the median willingness to pay estimate based on the follow up questions is over 50% higher than the figure based on the original question. They argue that the difference in significance between the two groups is due to the greater knowledge of residents about the properties of the lake. The role of expertise or ignorance in anchoring effects has been stressed by Mussweiler and Strack (2000) , who in a study involving German university students concluded that 'more pronounced anchoring effects occur if judges have very minimal knowledge about the target.'

The possibility of anchoring has been used to explain the difference in estimates of willingness to pay obtained via open ended and dichotomous choice, where it is proposed (see Boyle et al., 1998 for instance) that the bid value in dichotomous choice acts as the anchor. Frykblom and Shogren (2000), test this hypothesis using a laboratory experiment involving a Vickrey auction and a dichotomous choice task to elicit values for an atlas of Swedish environmental geography. They cannot reject the null hypothesis of no difference between the two distributions of elicited values.

As they hypothesise, one possible reason for this inconclusiveness is that the impact of anchoring on measures of the distribution of values is not unambiguous. In theory, the anchor of x may pull up bids which would be lower than x in the absence of the anchor and at the same time the anchor may pull down bids which would be higher than x in its absence. As a result, unless the anchor is at one extreme of the range of possible values, the sign and size of its impact on the mean are not predictable. However, the impact on the median can be predicted. Consider two groups, one of which is exposed to an anchor, x and a control group from which values are elicited without an anchor. Let M^x be the median in the group which is exposed to the anchor and let M^c be the control group median. If $x > M^c$ then anchoring implies $M^x > M^c$. Conversely if $x < M^c$ then $M^x < M^c$.

3.2.3 Confirmation Biases

Confirmation biases refers to a well-documented tendency of individuals to process or seek new information in a manner which supports their existing beliefs. Lord et al. (1979), for instance write that,

> The biased assimilation processes underlying this effect may include a propensity to remember the strengths of confirming evidence, but the weaknesses of disconfirming evidence, to judge confirming evidence as relevant and reliable, but disconfirming evidence as irrelevant and unreliable, and to accept confirming evidence at face value while scrutinising disconfirming evidence hypercritically. (p. 2099).

For the purposes of summarising this evidence we can split confirmation bias into a passive part – a tendency to misinterpret given evidence – and an active part: the tendency to seek out new information which supports the original hypothesis. Both aspects of the phenomenon have been the subject of investigation. A typical example of the passive bias is provided by Lord et al. (1979), who subjected 48 student subjects to the same set of information on the deterrence effects of capital punishment. The students, who had been selected on the basis of their prior answers to a questionnaire on capital punishment were split equally between proponents and opponents of the death penalty. Though both sub-groups received the same information, proponents concluded that most of it supported their own position on the death penalty and, after reading it, had increased confidence in their own beliefs. Simultaneously, opponents of capital punishment on average concluded that the new information supported their position and emerged with increased confidence in it. Other studies, such as Darley and Gross (1983) which examined perceptions of academic ability and Plous (1991), which focussed on nuclear power report similar conclusions.

In the active version of confirmation bias (usually called positive confirmation bias), individuals seek out only information which could support a maintained hypothesis rather than evidence which might reject it. For the Wason test for instance, subjects are presented with four cards. They are told that there are numbers on one side and letters on the other and then they are invited to turn over those cards which

test the hypothesis that (for instance) 'all cards with an X on one side, have a 1 on the other side.' In tests of this kind, most subjects correctly turn over the card where they can see an X. However, they typically also turn over cards where they see a 1, but fail to turn over cards showing 2 (or another number). In doing so they show evidence of the tendency to search for evidence which confirms a hypothesis rather than that which refutes it.

Some of the evidence discussed so far comes from psychological experiments of the kind that many economists would challenge because of the lack of monetary incentives . However, where economists have devised experiments that test Bayes' theorem using monetary incentives, information processing is rarely revealed to be optimal. For instance, Jones and Sugden (2001), present a version of the Wason test, adapted so as to provide monetary incentives for subjects to turn over the correct cards. Student subjects were paid for correct guesses, but each card turned over came at a 'price' set such that it was always optimal to turn over a card which represented a true test of the hypothesis under consideration. Results were in line with previous experiments conducted without monetary incentives: even after considerable experience, subjects turned over 'confirming' cards rather than 'rejecting' cards even when it was not in their financial interest to do so.

Experimental market-based tests of Bayes' theorem include work by Grether (1980), who finds evidence of a representative heuristic and Camerer (1987) who uses a double auction mechanism to investigate equilibrium price formation in an asset market where the two states of the world, X and Y, are determined by a random draw. Dividend payments from the assets are state dependent. Subjects do not know the state of the world, just the prior probability that it is X (0.6) and some ancillary information in the form of three draws, with replacement, of balls from a cage. Subjects are told that the X cage contains 2 red balls and one black, while the mix of colours is reversed in the Y cage. Camerer compares the predictions of a Bayesian model with heuristics including exact representativeness – a rule of thumb whereby subjects predict that the state is X, if the 3 drawn balls match the pattern of the X cage and believe the state is Y if the mix is two black balls and one red, but otherwise act according to Bayes' theorem. As with Grether, Camerer's results are mixed. Over time, asset prices move to values statistically insignificantly different from those which would be expected with risk neutral, Bayesian subjects. However, for the specific cases when the three balls matched the pattern from one or other of the cages, the null hypothesis of Bayesian subjects was rejected in favour of exact representativeness.

The evidence on failures of Bayes' Theorem is sometimes linked with the debate about the perception of probabilities. Much of the evidence from experiments on choice under uncertainty (see Starmer, 2000) suggests that small probabilities are over-weighted in choices. The fact that they are over-weighted does not imply that they are mis-perceived, but there is some direct evidence in favour of the hypothesis. In a famous experiment, Fischoff et al. (1982), asked two samples of American subjects, one drawn from undergraduates and one from the League of Women Voters, to report estimated probabilities of death from a variety of causes, common and uncommon, after giving the subjects some baseline information about

either the probability of dying in a car accident or from accidental electrocution. There was a wide variation in estimates, but a consistent pattern in the sense that subjects tended to overestimate the likelihood of death from rarer causes, but underestimate the possibility of dying from more common causes. Later analysis by Benjamin and Dougan (1997), and subsequently by Hakes and Viscusi (1997), cast doubt on the significance of the results, pointing out that mortality predictions may, quite rationally, be linked to personal circumstances such as age or gender and indeed that such factors (particularly those that are age related) account for much of the deviation of the perceived probabilities from their true values. As in other cases, we might also question the wider significance of such misperception, since it is not clear what economic incentives the subjects had to acquire more accurate information.

3.3 Ambiguity Aversion and the Ellsberg Paradox

John Maynard Keynes (1921), asked whether, 'if two probabilities are equal in degree, ought we, in choosing our course of action, to prefer that one which is based on a greater body of knowledge?' (p. 313). According to EUT, the answer is 'no' since compound lotteries (e.g. caused by uncertainty over the exact values of probabilities) are reducible to their simple equivalents. Ellsberg's 1961 examples challenge the behavioural relevance of this answer and suggest instead that individuals show aversion to choices involving ill-defined or unknown probabilities. In his two colour example, subjects are asked to bet on whether a ball pulled from an urn will be red or black, when these are the only two colours possible. A correct guess yields a monetary prize; failure yields nothing. In one urn there are fifty balls of each colour and the subject is told this; in the other urn the subject is told only that there are 100 balls in total and that both colours are present. For both urns, most individuals tend to be indifferent between betting on red and black, but most also prefer to bet on the urn where the proportions of each colour are known, rather than on the urn of uncertain composition. This pattern violates EUT . To see this, let p be the perceived probability of the ball being black for a subject faced with the urn where the true proportions are unknown. If p is below 0.5, then the individual should prefer to bet on either of the colours from the other urn, but also means that the probability of the ball being red is greater than 0.5, in which case out of all options available the subject should prefer to bet on a red ball being pulled from the urn of unknown composition. Conversely, if p is above 0.5, then the subject should opt for betting on black ball from the same urn.

In the 'three colour' example an urn contains balls of three colours. The proportion of white balls is revealed to be $1/3$, but the proportion of red and black balls is not stated. For task one subjects must choose between winning when a white ball is chosen and winning if a red ball is chosen. According to EUT, betting on white (which most people do) implies that the subjective probability of 'red' is below 1/3 and hence that the subjective probability that a black ball will be chosen is above $1/3$. Hence in task two, when subjects must choose between winning on white or

black or winning on red or black, an EUT maximizing subject who has chosen 'red' in the first task, should plump for 'white or black' in the second task. But most individuals opt for 'red or black' – the option that has a well-defined probability.

Early tests of the Ellsberg paradox and related phenomena include experiments conducted by Becker and Brownson (1964), who told subjects the range of numbers of red balls in an urn. They found subjects willing to sacrifice around 60% of expected value in order to avoid ambiguity.

The absence of well-defined probabilities is often defined as ambiguity and hence ambiguity aversion is used as the term to denote an individual's reluctance to choose bets when exact probabilities are unknown. In their survey of the topic, Camerer and Weber (1992) distinguish between situations where the probability distribution of probabilities is known (second-order probability) and cases where the distribution of probabilities is itself unknown. Whether this distinction matters in behavioural terms is unclear – Keren and Gerritsen (1999), for instance find no statistically significant differences between the behaviour of two samples of subjects, one of whom were given second order information about probabilities in an Ellsberg-style experiment and one of which were not. However it is clear that second order probabilities do elicit Ellsberg-style anomalies. In an earlier study, Yates and Zukowski (1976), used a BDM mechanism to elicit selling prices for three types of bets: on an 'known urn' containing equal proportions of blue and red chips; on an 'uniform urn' where the exact number of red chips was not specified, but where subjects were told that it was equally likely that the urn contained 0 to 10 chips (out of 10 in total) and for an 'unknown urn' where no information on the content was available. Subjects placed a higher value on the uniform urn compared to the unknown urn, but placed a still higher value on the known urn.

The psychological well-springs of ambiguity aversion are as yet poorly mapped. One symbol of this lack of understanding is the fact that the term ambiguity aversion often refers both to the phenomenon exhibited in the Ellsberg paradox and to one proposed explanation of it. Some alternative explanations have been offered. Frisch and Baron (1988) suggest that one possible rationale for the phenomenon is mistrust of the experimenter on the part of the subjects. If participants believe that the experimenter will manipulate the draw so as to bias it against the event chosen by the subject, then aversion to ambiguity would be quite rational. Alternatives to this include Heath and Tversky's 1991, competence theory. In several experiments, US college students were offered choices between betting on a draw from an urn where the proportion of balls was known and betting on uncertain events, such as American football matches and presidential elections where the subjects felt some expertise. Perhaps not surprisingly, subjects pre-selected for their knowledge of sports tended to prefer the bet on football matches to the urn, while subjects with a political interest opted for the presidential election. Since many of the subjects preferred the ambiguous bet to the risk one, Heath and Tversky argue that this shows preferences for betting are driven by feelings of competence rather than ambiguity aversion. However, results of this kind do not cast much light on the 'pure' Ellsberg experiments, where it is hard to see how competence plays a role. Moreover the results could simply be interpreted in terms of beliefs. Fox and Tversky (1995),

offer an explanation of what drives feelings of competence in terms of comparative ignorance. In one experiment, Stanford students were asked to state willingness to pay to bet on draws from one of two bags of balls. For one bag ('risky'), the split of red and black balls was known with certainty; for the other bag ('ambiguous') subjects were only told that red and black balls were in the bag and that the total number of balls was 50. One sub-group were asked to state separate WTP s for draws from both bags. Other sub-groups were only presented with one of the scenarios. Fox and Tversky found no significant difference in WTP figures for the two sub-groups who faced only one task. However, subjects who faced both tasks placed a much higher value on the risky bet, compared to the ambiguous bet (the mean WTP figures for the other two sub-groups lay in between). In other experiments subjects were asked to choose between a sure payoff and betting on an event where knowledge was likely to be vague, such as whether the inflation rate in Holland was above or below 3% in the previous 12 months. Psychology subjects who were told that the survey was also being presented to economics postgraduates and to professional business forecasters were much more likely to opt for the sure payoff – something that Fox and Tversky take as providing support for their notion that individual attitudes to choices involving ambiguity are influenced by comparisons with other options and with other individuals.

3.3.1 Summing Up Bayes' Theorem

The evidence of this section shows that departures from the predictions of Bayes' Theorem are common and robust. As a result we might anticipate that the ability of individuals to learn effectively will be limited. However, some environments important to economics, such as the marketplace might provide the right conditions for learning and the incorporation into choice of relevant information. It is this issue we explore in the next section.

3.4 How Markets Might Lead to Individually Rational Behaviour

Markets offer individuals incentives to choose carefully. They also offer opportunities to observe the choices made by other individuals and to learn from their own experience. Finally, in some contexts, agents who do not optimize may be eliminated from the market place or become infinitesimally small in their economic activities compared to agents who do optimise. Consequently, markets may lead to individually rational behaviour; the question is will they do so. Economists have attacked this question in a number of ways, through theory, observation and experiment. Some of this data on observation and experiment has already been discussed. The broad thrust of the literature is that anomalies found in the laboratory are also found outside of it. In this section I shall begin by reviewing some of the other experimental evidence before examining theoretical reasons why markets might produce

greater rationality in behaviour. As with the rest of the book, the primary focus of this section is on individuals and not firms.

The evidence for convergence to rational behaviour within markets is mixed. There are three major sources on the matter: field evidence from actual markets; repetition of purchase and selling of real goods in laboratory experiments and finally market experiments themselves. I shall take them in reverse order.

One of the largest apparent pieces of evidence that markets yield rationality is provided by the market experiments pioneered by Vernon Smith and his collaborators (e.g. Smith, 1962). In these experiments subjects are typically invited to buy and sell 'goods'. Buyer are offered a reward from the experimenter for each unit of the good they acquire. This reward may depend on the number of units they buy. Sellers meanwhile, face a cost of producing the good, meaning that, from any revenue they acquire from selling a unit of the good, they must deduct some amount set by the experimenter, which then determines their net revenue. For both sides of the market, therefore, the design of the experiment creates incentives to behave as a profit maximizer, it induces preferences. The costs set by the experimenter create a supply schedule, both for individuals and for the market as a whole (provided that players treat prices parametrically); meanwhile a demand schedule is also created for the buyers by the prices offered to them. If the reward paid to the buyers declines with the number of units they purchase, then the demand schedule will be downward sloping. Similarly, if the costs of the sellers rise with each unit they offer for sale, then the supply schedule will be upward sloping. Provided the demand and supply schedules intersect, there will be a price which represents a competitive equilibrium and a corresponding level of trade which will maximize the gains from trade, defined as the sum of the net financial gains of all market participants.

In market games of this kind, prices are set in a number of ways. In the double auction, sellers may at any time post a price at which they are willing to sell a unit, while buyers may post prices at which they are willing to buy. At any given time, the highest selling price and the lowest buying price are 'live'. When one of the buyers accepts the live selling price or one of the sellers accepts the live buying price, then the market for that unit closes and the two relevant players exchange the unit for the agreed sum. The game continues until there are no more units to be sold or the selling price remains stuck below the buying price for some reasonable and publicly known length of time. Other forms of trading rule include the posted bid system where one side of the market (typically the selling side) get to post take it or leave it bids and the buyers must whether or not to accept.

In general with market games of this kind, after a few periods of learning and with enough participants on both sides of the market, there is rapid convergence to the competitive equilibrium price and nearly all the gains from trade are exhausted. This tendency holds true even when the reservation prices of the participants are private knowledge. Typically, the results are strongest – in the sense that prices conform most closely the predictions of the competitive equilibrium model – when the double auction is used, even when the number of sellers is small but still greater than one. For instance in a survey of six double auction experiments Holt (1995),

finds the median efficiency to be 96.5%. When the posted offer mechanism is used, prices may remain above the competitive equilibrium level and trade is below the efficient level. However, it is still often the case that behaviour is compatible with profit maximisation.

Although this is a large and robust body of results, its implications are unclear for the status of the debate about whether market experience leads to rationality. It is apparent that the subjects do learn to choose actions which raise the surplus generated in the market. But the key point is that the preferences are induced by the experimenters, so the results do not indicate that subjects would learn to act according to their own preferences in goods and asset markets. It is also the case that very often the high degree of market efficiency hides some important variation between the participants, who on closer inspection appear to be less than rational.

Because preferences are induced in these experiments, it is possible to define efficiency in an objective manner, external to the participants. It is this feature that enables Smith to argue that markets may produce 'social' rationality, even when the individuals within that market are not necessarily rational. In the context of real markets, however, this claim is not meaningful if at least one side of the market involves agents who do not have quasi-linear preferences. Fundamentally, with markets efficiency means Pareto efficiency which is obviously defined with respect to (subjective) preferences. If, within this market, marginal rates of substitution are not equal for active participants (i.e. those who trade at least one unit of the good), then the outcome is not efficient by definition. Generally, in real-world consumer markets it is not clear how a separate concept of 'social rationality' is defined. One possibility is that the market is socially efficient if it is efficient for all individuals. But if this is defined as the equalisation of all marginal rates of substitution, then many experimental markets fail to exhibit social rationality because of the variation in marginal valuations across individuals. An alternative approach would be to define the degree of social rationality as the extent of the departure from the market clearing quantity. This too has its fundamental problems. Consider a market for a single egg with two potential buyers and one seller. A does not like eggs at all, B does. Suppose that in a market equilibrium one egg would be bought by B and none by A. Now consider the situation where it is A who gets the egg while B does not. The market clearing quantity is traded, but A ends up with an unwanted egg while B goes hungry, an outcome which is hardly Pareto efficient. Empirically such misallocation may represent a large portion of the costs of poorly functioning markets – as Glaeser and Luttmer (2003), argue in their recent analysis of the efficiency losses of rent ceilings in New York. So, in general the fact that experimental markets produce social rationality tells us little about the efficiency of real markets with subjective preferences.

There is a second and related issue with regard to market experiments of this kind. As already noted, individuals are typically induced to maximize a function (profit) of just one variable. Much of the evidence on anomalies discussed in Chapter 2 suggests that it is precisely when individuals are asked to conduct trade-offs between different objectives that framing effects begin to emerge. In other words, if we are

testing the rationality of individuals in markets, it is their allocation of resources between markets that is most vulnerable to framing.

So what experiments should count as tests of the market learning begets rationality hypothesis? One important element of course is repetition – repeated plays of the same choices in near-identical environments. A second element that could be considered important is real rather than hypothetical choices and a third element could be exchange – either of goods for money or the reverse, since this is a typical feature of most markets.

And what constitutes evidence of rationality in such environments? A necessary, though by no means sufficient requirement is convergence of behaviour by individuals – either to a unique choice or to a set of choices, with each of the options played with some stable frequency. Most work on repetition has concentrated on this issue. A stronger test is equalisation of marginal rates of substitution, provided each person can trade multiple units. A third possible test is the absence of explicit anomalies, such as immunity from money-pumping or Dutch Books, though individuals may pass such tests without having complete and coherent preferences. We would also expect no implicit anomalies – in other words, the long-run individual behaviour should be independent of framing effects. For instance, if two incentive mechanisms are both incentive compatible, then they should both produce the same long-run results. Although the property of invariance has been widely tested for one shot decisions, it has received far less attention for repeated decisions in market-like environments.

As was noted in Chapter 2, many experiments that explore the WTA-WTP gap allow a degree of repetition, yet the eradication of reference-point effects was limited. In some experiments, such as that reported by Shogren et al. (1994), repeated play led to the gap between WTP and WTA disappearing for one type of good (chocolate), but not for another (pathogen risk). Coursey et al. (1987), find a substantial reduction in the median and mean WTA – WTP gap over the course of several repetitions, though even at the end of their trials the gap remains. Meanwhile, Cox and Grether (1996), report an experiment which tests for convergence using preference reversals. In each round, subjects were required to value two lotteries; after five rounds, they were also asked to choose between the lotteries. Comparing the results of the first round valuations with choices, there was a preponderance of preference reversals in the standard direction. However, using the fifth round valuations as the basis of comparison the preference reversals were more symmetric. Braga et al. (2006), report a similar experiment but one where markets, in the shape of nth price auctions, play a central role in generating the prices. They consider two theories about how feedback influences behaviour when lotteries are repeatedly valued: one idea is that subjects take their cue from the equilibrium price and adjust their own reservation prices towards it. Under reasonable conditions the adjustment is likely to be larger with valuations of the $-bet. In the alternative theory, the prime driver of revision is the experience of playing a lottery and losing. By varying the value of n between treatments and giving subjects in some treatments round by round feedback on the outcome of their choices, Braga et al. are able to show that it is the second effect which seems to have the large impact on stated reservation prices. Indeed,

when it is sufficiently large it leads not just to the eradication of most standard preference reversals, but to a preponderance of reversals in which the P-bet is valued more highly yet the $-bet is chosen. Braga et al. thus conclude that the market may eradicate some anomalies, but it can also create some of its own.

Where convergence in behaviour occurs it may also affect decisions made in parallel contexts. In an experiment conducted with students at the University of Wyoming, Cherry et al. (2003), set their subjects two sets of tasks. In the first set, individuals chose between two monetary lotteries and placed a value on each; the second set mimicked the first, but the goods involved were environmental public goods. The other essential difference between the two sets of tasks was that the first was played out for real, while for the second set, statements about preference and valuation were purely hypothetical. As might be expected high levels of preference reversal were found in both sets of tasks. When the tasks were repeated the reversals persisted until, after the tenth round, the experiments used an arbitrage mechanism to 'money pump' those subjects who had reversed their preferences in the monetary lotteries. In subsequent rounds, the number of reversals fell in the monetary lotteries. But what was more significant was the drop in the number of reversals in the hypothetical tasks – despite the absence of feedback specific to the environmental public goods. Cherry et al. (2003), interpreted this phenomenon as an example of rationality spillovers, whereby feedback from market behaviour corrects or removes anomalous behaviour.

The concept of rationality spillovers is potentially very important. After all, many markets are similar in structure, while rarely identical. The 'take it or leave it' pricing strategy is common to many retail environments, so that even if an individual buys, say, luxury chocolates on an infrequent basis, it is highly likely that she or he can apply a central lesson to be learnt from their toothpaste-purchasing experience: that there are no incentives to misrepresent preferences in the typical 'take it or leave it' retail setting. In this context, Kagel (1995), exposes subjects first to an English auction and then to a first price auction (reversing the sequence in a second treatment). While, having played the first price auction has a positive impact on subsequent English auction strategies, the reverse is not true – subjects who face a first-price auction do have strategies that are closer to optimality if they have previously undergone an English auction. Kagel's result suggests that rationality may only spill-over in certain directions.

In the Cherry et al. (2003), experiment subjects were 'taught' by the use of the money pump. Slembeck and Tyran (2004), use team decision-making and a rank-order tournament to stimulate learning in an important experiment on the 'Monty Hall' or 'three doors' problem. In this well-known decision problem, based on a game show hosted in the USA by Monty Hall, subjects must choose one of three doors, behind one of which lurks a monetary prize. The subject is given an opportunity to amend their choice after the experimenter opens one of the other doors and shows its contents to be empty. If the subject's original choice is correct, then the experimenter opens one of the other two doors at random; if the subject's first suggestion is wrong, then the experiment opens the other incorrect door. Under these conditions it is always optimal for the subject to switch choices when given

the chance to do so, because the probability of the prize being behind the original door is 1/3 compared to a chance of 2/3 that the prize is behind the alternative door. Previous experiments suggest that individuals instead generally stick to their original choice. For example in Friedman's, 1998, repeated experiment, even after 15 rounds subjects chose to switch in less than 60% of cases, despite being give lots of feedback on their choices and outcomes. Slembeck and Tyran (2004), repeat the choices for 40 rounds, but still find that individual decision-makers choose to switch in only 40% of occasions when payoffs are based purely on their own performance. However, with the rank order tournament or when groups rather than individuals made the decisions, convergence to the rational choice was nearly achieved by the end of the experiment. Moreover, even at the start of the experiment participants were more likely to switch compared to the baseline design. They conclude that working in groups or acting in competition with other subjects gave the subjects incentives to overcome the biases (e.g. anticipation of regret) which otherwise tended to hinder learning.

Little work has been done on the issue of whether other forms of framing affects the convergence behaviour. One exception is Loomes et al. (2002), which question some of the evidence of convergence in market experiments and suggests that some implicit anomalies survive the effects of repetition. In their experiment, subjects within a group bidding to buy or sell a good are not all valuing the same good. Some subjects are valuing one type of lottery while the remainder value another lottery. The experimenters vary the proportions of each type. This should have an impact on the equilibrium price in the market (set by the median individual), but with an incentive compatible mechanism, it should not affect individual values. However, Loomes et al. find that individual values are correlated with equilibrium prices (and hence with the proportion of types), suggesting that convergence in some market experiments may be due to herding effects rather than the discovery of preferences.

In a similar vein, Ariely et al. (2003), conduct a novel experiment using repeated exposure to an annoying noise as the good. The framing effect is provided by anchors, created by asking subjects if they were willing to accept x in return for exposure to the noise prior to the main body of the experiment. For some sessions, x was set at one of two values (high or low) across the sample; in other sessions students were asked if they were willing to accept a level of compensation set by the last three digits of their (US) social security number. For the actual repeated rounds of the experiment, either nth price auctions or Becker de Groot Marschak mechanisms were used to elicit reservation selling prices. In all variants of the experiment, longer exposure to the noise required more compensation and across the rounds WTA converged, but the convergence was systematically linked to the initial anchor, leading the authors to describe the behaviour of their subjects as 'coherent arbitrariness'. Further experiments were conducted using marketed goods as the objects in the auctions. As Fig. 3.1 illustrates the case of cordless keyboards where valuations were correlated with social security numbers.

Field studies on learning in market environments also provide mixed evidence. Several of the results in the latter parts of Table 2.1 refer to market data (e.g. from supermarket choices), where subjects have plenty of opportunities to learn through

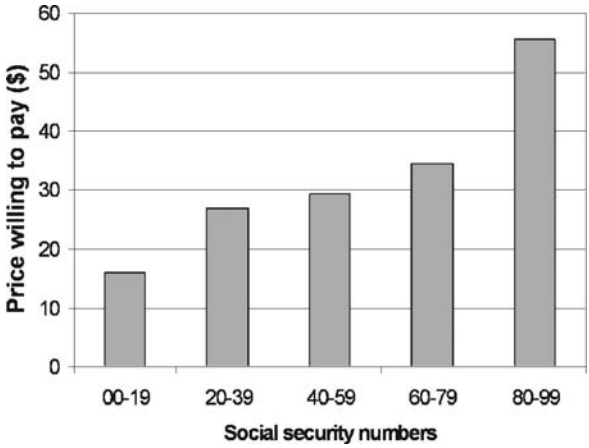


Fig. 3.1 Willingness to Pay and Social Security Numbers

repeated purchase. Despite this, reference point effects remain. In the field, studies reported by List (2003), for instance, professional traders showed little evidence of an endowment effect, but private individuals with fifty plus trades per year still showed a marked discrepancy between willingness to pay and willingness to accept values.

3.4.1 Summing Up Market Learning

In his survey on the subject of individual choice experiments, Camerer (1995), concluded:

> Overall, the data suggest that competition or learning in markets reduces the buying-selling gap somewhat, in some settings, but does not eliminate it. The gap is large for environmental and consumer goods (like wetlands and mugs), but small for lottery tickets (p. 667).

This still seems to be the current situation with regard to the available evidence, but we might add the possibility raised by some of the experiments discussed that simple monetary incentives to learn may not always be enough. Subjects might be more motivated in a social context, though as the Loomes et al. paper suggests, not always in a manner which is consistent with rational behaviour.

The mixed position on convergence and the elimination of framing effects within markets suggests that what is required is a more sophisticated model of how subjects adapt their behaviour in market environments. Formal and informal theories of learning would suggest a variety of factors affect the degree to which behaviour converges in markets. Among these factors I would include,

- Degree of frequency of purchase – how often the consumer gets to evaluate and change their decision.

- Degree of frequency of use – how frequently the consumer receives payoffs from the decision.
- Complexity of interaction with other goods – how easy it is to disentangle the marginal payoff from the consumption of one particular good given variations in the consumption of other goods.
- Timing of payoffs – how closely they are related to the time of consumption.
- Uncertainty – how much noise there is in the relationship between consumption and payoff.

It would be extremely useful to point to a robust body of work which had investigated the impact of all of these elements on learning in markets. However, as we saw above, to a large degree, most of the empirical work concentrates on only the first two factors, and then combining them by choosing consumption goods rather than goods which yield a flow of payoffs. So it remains an open question how the other forces actually influence learning.

3.5 Learning

So why does individual choice behaviour not always converge quickly to the rational? In this section, I cover briefly some of the issues involved.

It is useful to begin by classifying learning problems into two types. The first type of learning is environmental learning: the individual knows his or her preferences over outcomes but does not have a complete understanding of the technology which links choices and outcomes. The classic one arm bandit problem is a typical example of this type. The second type is where an individual has rational preferences, but does not have a complete understanding of their nature. Learning then is devoted to uncovering the true nature of preferences, hence the term discovered preferences (Plott, 1996). In practice, the boundary between the two types of problem is blurred. Consider an individual who knows that they like bitter chocolate, but does not know the taste of individual chocolate brands. They try different bars, eventually settling on one which satisfies their snack requirements. On the one hand, this may be presented as a typical example of the first type of learning, since the individual is clear about their preferences between the characteristics of chocolate bars, but is uncertain about the technology of the relationship between brands and taste. On the other hand, we could model this as uncertainty over preferences between brands, in which case the example would seem to fall into the second category of discovered preferences. The distinction between the types of learning problem therefore largely turns on whether preferences are defined over characteristics or over choices.

In both these types of learning problems, preferences are defined and rational at one level or another. Consequently the issue of rationality is confined to the question of whether the strategy for discovering preferences is in some sense rational. Such a strategy can be viewed as meeting a standard of rationality if, through its repeated application, an individual discovers his or her true preferences. But one might legitimately ask where these underlying, rational preferences come from. How does

one acquire them? What are the forces inducing individuals to be rational in their discovered preferences?

Preferences might satisfy principles of weak rationality for two kinds of reasons: first that such preferences are beneficial and secondly that they represent an equilibrium. In the first case, if weakly rational preferences are beneficial then there must be some underlying metric of betterness against which different kinds of preferences can be evaluated. For instance, it could be as the utilitarians supposed, that all drives can be reduced to a single metric, such as pleasure minus pain and that each experience can be evaluated against this common standard. We could call this standard, utility. Given it, it is possible to imagine learning rules which would lead agents to choose consistently from choice sets, with the options chosen yielding the highest value of the utility index. Consequently, in the limit choice would be rational. Within this picture, learning to be rational is reducible to the first or the second models of learning. Just as characteristics are associated with particular choices, so different combinations of characteristics would be associated with different levels of pleasure or pain, which would constitute the fundamental characteristics of the consumer's experience.

Alternatively, individuals with consistent preferences may have higher evolutionary fitness. However, there are many examples, such as the Prisoner's dilemma and dynamic choice, where some apparently non-rational agents can have higher expected payoffs than weakly rational individuals in the right circumstances. This kind of example is easy to construct, especially in a world of repeated social interactions. As a consequence, there can be no completely general model in which rationality is always optimal. Nevertheless, there are some arguments proposing that, in general, it is better to be rational. One such argument is that of the money pump. An individual with cyclical preferences, it is argued, can end up with a bundle which is dominated by his or her original bundle. But as Conlisk (1996), observed failure of weak rationality and immunity to money pumps (or its cousin, the Dutch book) is perfectly consistent. Someone who always refused to trade, for instance, no matter what their endowment, could not be money pumped, but would also have either incomplete or inconsistent preferences over goods.

In biological evolution it is relative fitness which matters for the success of an individual, not absolute fitness. A case can be made that the type of individual who always refuses to trade may also be subject to evolutionary pressures. Suppose for instance, society consists of individual foragers who come in three types: non-traders; traders with inconsistent preferences and individuals with consistent preferences. On any given day a hunter gatherer gathers a random catch. Suppose also that nutritionally, a balanced diet is better, in the sense that individuals with a diet balanced on a daily basis have more offspring. The trading individual who gathers only one type of food can always trade with others with complementary endowments, but the non-trader is reliant on catching or gathering a mixed bag. If the catch is actually random then the non-trader will have a lower relative fitness compared to the trader. Note that the same environment may also work against the trading individual with inconsistent preferences, since he or she will be money-pumpable by individuals with consistent preferences.

Though this is an improved argument why there might evolutionary pressures favouring rational choice, it is not without its own weaknesses. As noted above there are some strategic contexts where apparently rational choice is not advantageous to the individual. Moreover, evolutionary arguments suffer from the weakness common to all such theorising: selection pressures which were relevant in the context of the development of early humans may have little or no relevance in 21st century urban society.

Rational preferences might also be a by-product of rationality in other aspects of information processing and decision-making. It is useful and advantageous to know that many relationships are transitive. If I am comparing the sizes of three cake slices, A, B and C, it saves some time to know that if A is bigger than B and B is bigger than C, then I do not have to measure C against A.

So, rationality may be the result of bounded rationality. This sounds a little paradoxical, but the idea is simple. Consider the individual of limited memory who must faces potential choices amongst n objects. Storing information about potential choices is costly. If each object is associated with a utility number then only n such numbers need be memorised in order to make a choice on any subset of the n objects. On the other hand if transitivity is not a property of preferences, then up to $n(n-1)/2$ comparisons must be remembered for pairwise choice and if preferences are not choice set independent then $2n$ choices must be recorded. Transitivity can therefore be cost minimising. Of course though the argument for transitivity rests on it being a good guide to underlying preferences or for the costs of storage and retrieval being high. If, for instance, preference storage costs are low and the individual has context contingent preferences, then transitivity will be a poor decision rule.

To sum up, all the cases share one common theme: there is a standard against which any individual's choices can be measured. In the case of environmental learning or with discovered preferences, that standard is the individual's underlying preference structure. In the case of the innate and evolutionary arguments, the standard is provided by utility and fitness respectively. In the absence of a standard it is not clear why learning should lead to rationality. So one reason why learning might not converge to the rational is because it may serve no good for the individual concerned.

But suppose that rationality does produce benefits in terms of some pre-existing standard. What we now need to explore is whether, given such a standard, learning will tend to produce rational behaviour.

Let us begin with the reference point effects which were the focus of the first half of the previous chapter. The explanatory models offered are essentially static and struggle to accommodate the (limited) evidence on convergence of WTA and WTP in repeated choices. Two models have been put forward to explain learning. They assume the existence of underlying well-behaved preferences, but incorporate imperfect information on the part of the chooser or unfamiliarity with the incentive mechanism and some caution as a result. Kolstad and Guzman (1995), for instance, consider a first price auction, where some subjects may not know the true value of the good to them. As the proportion of uninformed subjects rises so does the incentive to under-report WTP and overstate WTA. They argue that this can explain

observed experimental data on valuation, but their model only applies to first price auctions. Most of the data referred to in Table 2.1 comes from environments with second or nth price auctions, where given their assumptions the incentive is always to report true WTA and WTP. Hence their model cannot explain most of the available evidence.

Zhao and Kling (2001) use option theory to explain the WTA-WTP gap. They consider a model in which consumers can either purchase or sell the good immediately or wait, possibly at some cost, until they are certain of the good's true value to them. Their model makes a crucial assumption about WTP (and WTA): WTP is defined as the price of the good which makes the individual indifferent between buying now and delaying purchase to a time when the good is available at this same price. Consider an individual for whom delaying consumption is costless. Suppose that they are uncertain of their true valuation now, but they know that its lowest possible value is zero. That individual will have a WTP of zero for the good today, preferring to delay purchase at any higher price. Conversely WTA equals the highest figure in the support of possible values. If experience (e.g. through repeated exposure to buying or selling) narrows the support of possible values then WTP and WTA converge. This result though depends on the assumption made that the WTP (or WTA) equals the price relevant for future decisions. To see this consider the example shown below, where I assume that agents have preferences of the form $v + y$, where v is the value of the good and y is income.

Value	0	4	8
Case I	0.3	0.4	0.3
Case II	0	1	0

There are two scenarios illustrated. In case I, which we can identify with inexperienced consumers there is a 0.3 chance the good is worth 0 etc. In case II, which represents the experienced consumer, there is no uncertainty. Suppose there are two time periods, perfect information in the second period, transaction costs do not allow selling in the second period and no discounting, in which case we have WTPI[2] $= 0$ and WTAI $= 8$ and WTPII $=$ WTAII $= 4$. In other words, experience is associated with the convergence of buying and selling prices.

The problem with this model is that it would be normal and more reasonable to suppose that consumer valuation takes the fallback price (i.e. the price offered to the consumer in the second period) as a given when computing WTP, WTA etc. This, after all, mirrors the situation in the laboratory when consumer goods are traded. If the good has a real-world counterpart (e.g. a mug bought from the campus shop or chocolate from a local specialist), then WTP and WTA would be determined in part by the consumer's beliefs about those prices.

[2] To see this note that at a price of zero in the second period the subject will always buy and so expected utility from the second period is 4. This must be equal to 4 – WTPI, so WTPI=0.

Consider for the sake of an example that the consumer believes second period prices will lie on the interval [6, 8], with a uniform distribution. In all other respects the example is unchanged. In case II, WTPII = 4 = WTAII because the second period price is irrelevant for the decisions of the individual with known preferences. Meanwhile, WTPI = 3.7=WTAI. The equality arises because the consumer can always sell their endowment in period 1 to the experimenter and purchase a replacement if the true valuation turns out to be 3.7. So, WTA/WTP does not diverge with lack of information and more significantly, WTAII > WTAI – in other words more diffuse priors about the true valuation leads to a lower value for willingness to accept. This is contrary to the experience of experimenters, where typically WTA declines with experience. Of course the result depends on the prior beliefs about future prices, but there is no reason to believe that in all relevant experiments subjects do have priors which guarantee that WTAII $\leq$ WTAI.

The discussion of the Kolstad and Guzman model and the Zhao and Kling alternative suggests that models based on the rational choice model struggle to accommodate the data on reference point effects. It is possible that some other model grounded in auction or options theory (for instance) can explain the existing body of evidence. By itself, though, uncertainty about preferences is an unlikely explanation. After all, it is not reasonable to suppose that students and other subjects have poor information about every day goods such as cola, chocolate and coffee mugs.

3.5.1 Melioration

Within the psychological literature on individual learning, a key model is that of melioration, advanced by the late Richard Herrnstein and his collaborators. Melioration is an apparently simple notion: if choice A tends to produce better outcomes than choice B, then, in the future the probability that A is chosen will rise, compared to the probability that B is chosen. Melioration is not a complete theory of learning, but it is hard to see how some such assumption can be omitted from a credible model of the learning process. We will see that by itself it need not yield optimising behaviour.

Consider an environment in which, at each instant, an individual chooses one of two options, A and B.[3] The choice yields an instantaneous outcome. It is known that preferences are strictly increasing in the value of the outcome. It is straightforward to see that if the payoffs for A and B are constant, then for reasonable probability updating rules, melioration will tend to lead to either A or B being played all the time.

Now consider a slightly different environment, where the payoff to the choices depends on the choices made in recent periods. Specifically, the payoffs are as in Table 3.3:

[3] This example is taken from a PhD thesis, still unfortunately unpublished, Gorter (2000).

Table 3.3 Payoffs in a Learning Environment

Payoff to	A played in period t–1	A not played in period t–1
A played in t	1	2
B played in t	1.2	1.2

If we want to give an economic context to the table we could imagine that it represents the planting decisions for a farmer each year. One crop, A, is potentially of higher value, but it depletes the soil, perhaps by removing nitrogen. The other crop, say a legume, fixes its own nitrogen, replenishing the soil for crop A, but has a lower, constant value. Alternatively the table could represent a trawler choosing between two fishing grounds, one of which is large and subject to rapid replenishment from the open sea (B) and one which is potentially of higher value, but which can be rapidly fished out, if the trawler returns too frequently.

In the absence of discounting, the optimal pattern of play is to alternate A and B. This yields an average payoff of 1.6. If melioration is employed as the learning rule, then the only approximately stable outcome is where A is chosen with probability 0.8 and B is chosen with 0.2 probability. This yields an average outcome of 1.2 – equivalent to that obtained by playing B permanently. This is not optimal. It is not even the highest expected outcome which can be obtained by playing A with a constant probability – that is 1.36, obtainable if A is played with probability 0.4.

Melioration fails to deliver optimality in this instance; essentially because the player is meliorating over the wrong choices. Suppose the player chooses between AA, BA, and BB, where xy means 'play x, followed by play y'. Now each time AA is played it yields either 3 or 2 over the two periods; option BA yields 3.2 and BB yields 2.4. Playing BA therefore dominates the other two choices and will always increase in probability until, in the limit, it is chosen with probability one.

Herrnstein conjectured that, for the correct framing of the choice set, melioration will always lead to optimality. Though reasonable I have not seen a formal proof of the conditions under which this conjecture is true, but what the example makes clear is the importance of the choice frame in this conjecture. Some act of imagination has to occur prior to the actual experience of A and B, so that information learnt about payoffs can be coded in the most useful fashion. Without it, melioration may need not lead to rational behaviour. In fact when Herrnstein ran experiments on pigeons, rationality was not achieved. Instead in equilibrium the participants' behaviour was closer to the predictions of the matching law – that average returns from choices are equated. Subsequent experimental work by Gorter (2000), using human subjects confirmed that many people are also susceptible to the same imperfect learning and match rather than maximize.

Melioration may not eliminate violations of independence when the payoff received is context dependent. Consider an individual who chooses repeatedly from a set consisting of three elements, A, B, and C. A yields higher direct utility compared to B and B yields greater utility compared to C. However, suppose also that this individual suffers from feelings of regret and rejoicing when the true state of the world is revealed. Table 3.4 sets out the details of a particular example in

Table 3.4 Learning with Regret

Prob	0.2	0.2	0.6	EU	0.2	0.2	0.6	ER	
A	100	0	10	26	5	–45	–5	–11	15
B	80	45	0	25	–20	45	–15	–4	21
C	95	–5	15	27	–5	–50	5	–8	19

which overall expected pleasure is the sum of two components: Expected utility and expected regret/rejoicing. When state i occurs the individual receives the direct utility attached to that state (given their previous choice) plus or minus the difference between the utility obtained from their actual choice versus the utility obtainable in that state from the best of the other choices available. So in the three state example below, the individual who chooses A experiences rejoicing of 5 in state of the world 1 (where A is the best choice and C is the second best) and regret of 45 if state of the world 2 occurs (since B is the better choice in this state). In the example, choice B has the highest expected pleasure when three choices are available. When all three options are available, melioration will therefore lead to B eventually being chosen with probability 1. When only options A and B are available though, then the pattern of regret/rejoicing is different and expected pleasure follows the ranking of expected utility. Consequently, when only these two options are available, then melioration will lead to A being chosen with probability 1.

The lesson is that learning based on feedback need not lead to expected utility behaviour if the underlying relationship between choices and reward is itself non-standard. Humphrey (2006), finds some evidence of this in an experiment where subjects must value lotteries and then, for each subject, a pair of lotteries is chosen at random and the one valued most highly is played out of for real. Half of his subjects are exposed to 'learning ' in the shape of information about the payoffs from ten hypothetical plays of a lottery. The other half of the subjects are not given an information other than the probabilities of particular outcomes and their associated monetary rewards. Humphrey finds evidence that the common consequence effect is stronger amongst those subjects exposed to the information, compared to those who have had no opportunity for learning.

An important experiment on the clash between Bayesian learning and melioration is provided by Charness and Levin (2005) who face subjects with the problem in Table 3.5:

Subjects are told that they must make two sequential choices. The state of the world is unknown but the same in both periods. Both states are equally likely. For

Table 3.5 Bayesian Learning versus Reinforcement

	Left urn	Right urn
State of the world	Probability of win	
Up	2/3	1
Down	1/2	0
Payoff for win	1	7/6

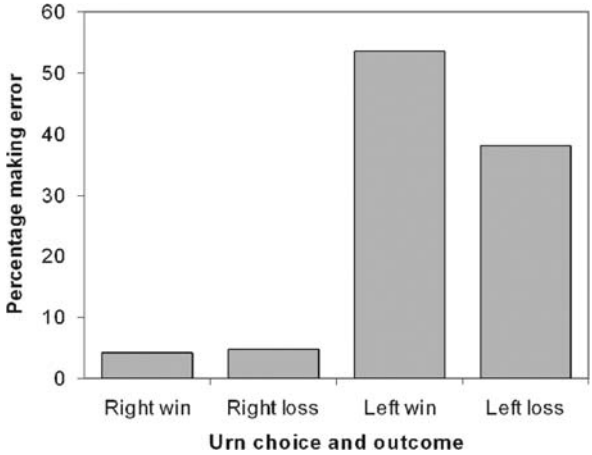


Fig. 3.2 Bayesian Error Rates in an Experiment

the subject who chooses 'Right' in the first period and who wins nothing, the state of the world is clearly 'Down'. Thus the optimal choice in the second period is the Left urn. If the subject wins, then the optimal strategy for a risk neutral or risk averse player is to choose Right again. Choosing Right initially is optimal for the risk neutral player because it is more informative as to the state of the world, but a subject who chooses Left also receives some information. In fact this information is 'perverse' in the sense that a player who chooses Left and then wins should switch to the Right urn, while a player who loses should stick to Left. For many players this feels counter-intuitive: one is used to sticking to the choice that comes good and switching when a choice fails. The issue here is that a win suggests that Up is the more likely state and then Right is the better choice. Conversely a loss suggests that Down is more likely in which case Right would not be a clever choice.

Despite playing the game 60 times and being forced on some occasions to begin with the Right or Left urn, it appears that most subjects shared this unreliable intuition. Figure 3.2 summarises the key data, showing how the majority of subjects who chose the Left urn responded sub-optimally to the information provided by the first draw.

Melioration raises the possibility of other psychological mechanisms by which feedback will fail. First, future experiences and responses to experiences may not be correctly anticipated. Loewenstein and Adler (1995), report an experiment in which subjects consistently underestimated reservation prices at which they would sell back to the experimenters, an object with which they were not currently endowed. Meanwhile, subjects in Simonson (1990), overestimate their taste for variety in snack choice. In part, failure to predict may be the result of failure to remember actually the pleasure or pain of an experience. Kahneman (2000a), lists a number of experiments in which subjects' moment by moment assessment of pleasure or pain ('experienced utility') is compared to ex post assessments of the same experience. ('remembered utility') and concludes that remembered utility is most clearly modelled as a function of the peak pleasure (or pain as it is in most of the experiments he

considers) and the intensity of the final experience before the stimulus is withdrawn. As a result, a shorter negative experience (e.g. a minor surgical procedure) may be remembered as being more unpleasant than a longer experience that is the same as the shorter procedure except for the addition of a period of less pain at the end. Assessed on a moment by moment basis, the shorter experience dominates, but in retrospect it is the longer procedure that is preferred. With both failure to predict tastes and failure to remember accurately, the conditions by which learning can lead to optimality are subverted.

3.6 Conclusions

Experiments with repeated decisions represent part of a programme for testing the external validity of one-shot experiments. Mostly, they suggest that repetition leads to behaviour which converges to something closer to that predicted by rational models, compared to one-shot experiments. However, convergence is often not complete and the long-run behaviour often appears to be context dependent. In terms of the theories that opened the chapter, each of them is deficient. Information processing is often not Bayesian. Repeated choice may lead to convergence but to a limit that is affected by the frame in which preferences are elicited. Finally, market-based evidence also points to some degree of convergence, but as we have seen, this can be incomplete and, to a degree the value to which behaviour converges appears to potential malleable by the choice of frame.

For public policy, an implication of the evolution of behaviour as a result of repetition is that one shot experiments may not provide a reliable guide to long-run behaviour in markets. This is most graphically seen with the sharp changes in WTA observed in repeated selling environments. Conversely, learning failures may mean that markets will not always represent a gold standard for the revelation of true preferences and the elimination of irrational behaviour.

Chapter 4
Markets and Reference Dependent Preferences

4.1 Introduction

The previous chapter discussed the impact of markets on anomalies. This chapter reverses the direction of causation. In particular I offer three examples of how reference-dependent preferences may be expected to affect market outcomes. The first example focuses on the single, competitive market; the second example explores monopolistic competition, while the third considers a general equilibrium world. In all cases preferences are assumed to be complete and transitive for a given reference point, but not necessarily reference point independent. The influence of the reference point on consumption follows the empirical work summarised in Chapter 2. In the first example preferences exhibit a status quo bias. In the second example, preferences are affected by decoy bundles. Consequently consumer choices are not consumption set independent. In the third example, which is built on work jointly undertaken with Robert Sugden, a full theory of reference dependent preferences is put forward and then the implications of that theory for the existence and stability of general equilibrium are examined.

The three parts of the chapter can be seen as extended examples designed to emphasise three basic points.

- First, markets with preference anomalies need not be well-behaved in the sense of gravitating towards a long-run equilibrium. Of course, markets without preference anomalies need not be stable, but the presence of reference points adds to the complexity of the dynamics.
- Second, profit maximizing firms may seek to exploit anomalies and this can have ramifications for nature and the optimality of market outcomes.
- Finally, for some reasonable models of non-Hicksian preferences it is possible to make some claims about the existence of a general equilibrium and about its welfare properties. So, it is not the case that theorising about welfare economics is impossible where there are widespread anomalies in preference.

A. Munro, *Bounded Rationality and Public Policy*, The Economics of Non-Market Goods and Resources 12, DOI 10.1007/978-1-4020-9473-6_4,
© Springer Science+Business Media B.V. 2009

4.2 Partial Competitive Equilibrium

The simple supply and demand model is one of the basic tools in the economist's kit. In this section we consider some possible implications of reference dependent preferences for a competitive market. Of particular interest are whether the possibility of reference dependent preferences make the market equilibrium non-unique and how the dynamics of the market are affected. In addition, in Chapters 5 and 11 we encounter the view that the design of non-market valuation methods should take their cue from markets. This only make sense if markets themselves are reliable vehicles for eliciting consistent estimates of marginal values. What we shall see is that this need not be the case, even in fairly straightforward environments.

The basic model of a competitive market has four ingredients:

- Demand by existing customers
- The flow of consumers into and out of the market
- Supply
- Price adjustment

It is standard to combine the first of these two ingredients together, especially in a static analysis, but it is worth separating them out because, with reference dependent behaviour, preferences evolve and experienced consumers may have quite different patterns of behaviour compared to their novice counterparts.

The basic models is as follows: time is discrete; in each period, a proportion r of the population leave the market to be replaced by r newcomers. Each individuals lasts 1/r periods in the market, where in what follows I take $1/r$ to be an integer. There are no income effects and demand for the good by new comers depends only on its price. Total demand by newcomers is given by rf(p), where $f(p)$ is termed the entry demand function. The entry inverse demand function, v(.) is defined such that, for a quantity q, v is the price at which $f(v) = q$. Experienced consumers have, at least potentially, reference point effects. Demand therefore depends not only on price, but also on consumption in the previous period. In period t, the experienced demand function, is $g(p, q_{t-1})$, where q_{t-1} is the quantity consumed in the previous time period.

To introduce an endowment effect, assume that for all q_{t-1} there is $p^1$and p^2 (with $p^1 < p^2$), such that $g(p, q_{t-1}) = q_{t-1}$ for $p^1 \leq p \leq p^2$ and with $p^1 < v(q_{t-1}) < p^2$. Suppose also that p^1 and p^2 are both continuous in q_{t-1}.

Figure 4.1 illustrates the story, showing the demand function in period t for a cohort who consume q_{t-1} in the previous period. For these agents there is a range of prices within which demand does not change from its value in the previous period.

I assume that the competitive supply function, $q^s(p_{t-1})$ is the same in each period and either upward-sloping or vertical. In addition, assume that there is a price such that supply is greater than demand, no matter what the history of consumption and that there is a price at which demand is greater than supply no matter what the history of consumption.

Now consider the dynamics of this model. One option is to suppose that the market clears in each period, but as indicated above, this may not be true if reference points adjust very quickly. Consider the following market adjustment rules:

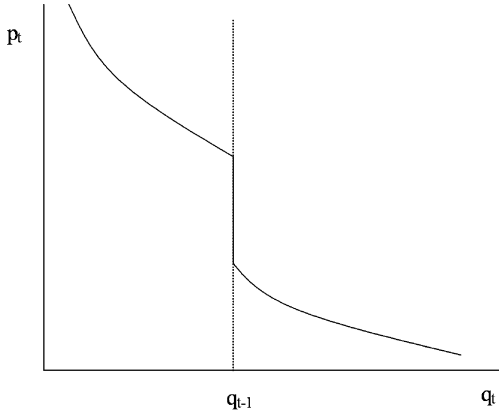


Fig. 4.1 Experienced Demand

$$\begin{aligned} q_t &= \min(q_t^s, q_t^d) \\ p_t &= p_{t-1} + \lambda(q_t^d - q_t^s) \end{aligned} \tag{4.1}$$

The first equation simply states that the allocation in period t is governed by the short side of the market: if supply is greater than demand, then consumers receive all they wish to receive at the prevailing price; if demand is greater than demand, suppliers produce all they wish to produce. The second equation is a standard rule for price adjustment.

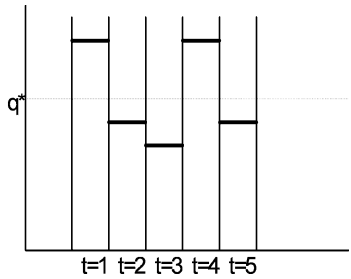
Define a long-run equilibrium, as a price p such that $p_t = p_{t-1} = p$ and for all experienced consumers, $q_t = q_{t-1}$ and the consumption of an inexperienced consumer in period t equals the consumption of an inexperienced consumer in period t–1. Let p^* be the price which would clear the market if all experienced consumers were inexperienced and let q^* be the corresponding level of demand. A long-run equilibrium exists at p^*; what is unclear is whether it is stable.

Consider first a case where market clearing is complete in every time period and supply is fixed at q^*. Define the set $P = \{p \mid g(p^*, f(p)) = f(p)\}$. The facts that $p^2 > p^1$ and that both these prices are continuous in q_{t-1} implies three things: (i) P is non-empty; (ii) it is a convex subset of the positive real numbers and (iii) it includes p^*, as well as prices above and below p^*. Thus it is possible to construct a set R with r elements, some of which are different and such that R is a subset of P and such that $r \sum_{i \in R} f(p^i) = q^*$. We then have the following:

Result 4.1 *Any r element set $R \subseteq P$ which satisfies $r \sum_{i \in R} f(p^i) = q^*$ is consistent with a cycle of length r.*

The idea behind the proposition is illustrated by Fig. 4.2, which shows five time periods and a cycle of length three. The dotted horizontal line is q^* – total and average demand. The heavier horizontal bars represent inexperienced demand in

Fig. 4.2 Equilibrium Cycles



each of the periods. For two cohorts demand is below the average in equilibrium, but this is counterbalanced by demand from one cohort which is above the average. Conversely, the market clearing price is below p* in one period out of three and above p* in two periods out of each cycle. Notice that the number of cycles of this sort will be uncountable, so that, within a range, almost any price can appear during a cycle. All that can be said is that the range of prices which may appear in a cycle includes the long run equilibrium price.

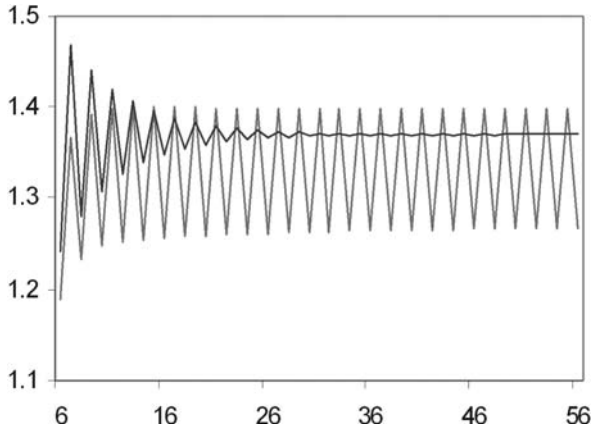
Cyclical behaviour is generic, in the following sense. First note that existence of a price at which demand is always greater than supply and the existence of a price at which supply is always greater than demand rule imply that the system is not unstable. Now suppose the long run equilibrium was strictly stable, in the sense that it always converges to p*. By implication therefore, there is some time t beyond which all prices lie in some neighbourhood of p* which itself is a subset of P. Consider the first r prices after time t. Since each of them is in P, it follows that the $t + r + 1$th price is equal to the $t + 1$th and we therefore have a cycle, unless all of the first r prices after time t are equal to p*. In turn, that can only happen if the $t - 1$th equals p* and so on until the price at $t = 1$.

What about the case when supply is not fixed and prices do not adjust immediately to equilibrium? Formal results are harder to find, but it is still the case that there may be equilibrium cycles rather than a situation in which the economy gravitates to the long run equilibrium. Figure 4.3 shows one such case. Beginning with period 10, time is shown along the horizontal axis; price is on the vertical. The equation governing supply is $q = 3p^2$; the value of λ is 0.2; inexperienced demand is $q = 7-p$; experienced demand is completely inelastic and $r = 0.2$, so that each individual has two periods in the market. Two time paths are plotted. The damped cycle is for comparison and shows the outcome when there are no reference dependent preferences. Consequently demand for all individuals is given by $q = 7-p$ and the market gradually converges to $p^* \approx 1.37$. The two period cycle is for reference dependent preferences. The price in the market flips between 1.27 and 1.40.

An alternative model is presented in Sugden, 2003. He makes the following assumptions.

A1 In each period the market clears.
A2 In each period, a proportion r of the population are replaced at random.
A3 Supply is fixed, at q*.

Fig. 4.3 Evolution of Prices



Note that A3 implies that average consumption is always equal to q*. Sugden proves the following proposition: a long-run equilibrium exists at $p = p^*$, with all consumers consuming q*. Moreover, it is stable, in the sense that, for any initial price p_0, and allocation of goods compatible with that price, the economy tends towards the long-run equilibrium.

Note that $g(p^*, q^*) = q^*$, so that existence is clear. Now consider the dynamics of price in this market. Suppose that $p_0 > p^*$. This means that for inexperienced consumers, $q_0 < q^*$, while for the average experienced consumer, $q_0 > q^*$. Suppose that $p_1 = p_0$. If this is the case, then inexperienced consumers will demand the same quantity as their counterparts in the previous period. Each experienced consumer will also demand the same quantity as in the previous period. However, average consumption by experienced consumers will now have fallen, because a proportion of experienced consumers will have retired, while the consumers who were inexperienced in period 0 consumed less than the average taken by the consumers who were experienced in period 1. Thus total demand must fall if $p_1 = p_0$, implying that $p_1 = p_0$ does not clear the market in period 1 and in fact $p_1 < p_0$. Note though that $p_1 \geq p^*$, because if $p_1 < p^*$, then total demand would exceed the quantity available. Thus the long-run equilibrium is stable from above and a similar argument shows that it is stable from below.

It is clear from this simple model that equilibrium can easily exist in models with non-standard preferences and that markets can be stable with reference dependent preferences. Sugden, though, draws additional and important conclusion. He argues that the framing of valuation should, as far as is possible, seek to mimic the market frame. In support of this line, he notes that if a project appraiser were to evaluate the surplus to be gained from increasing or decreasing the supply of the good by asking consumers to value such changes (and the answers were honest) from the vantage point of a long-run equilibrium, then the average WTP would be below the market price by a non-trivial amount, while the average WTA would be above the market price, again by a non-trivial amount. However, the valuation of inexperienced consumers (i.e. consumers lacking reference points) would produce

exactly the market price, p^* which would transpire. Thus, he argues, the most appropriate measure of value is one that avoids undue emphasis on the status quo in the framing of choices.

Yet, the previous model shows how sensitive this conclusion is to the dynamics of the market place. For a fairly small change to the process by which consumers leave and join the market, there is a major shift in the long-run behaviour of the system. Now the values elicited from novice consumers or consumers without a firm status quo bias may be no more reliable predictors of future market prices than the values obtained from other market participants. The previous model therefore illustrates the fundamental lesson of this section: reference point effects are complicating factors in competitive markets. In particular, the price signals produced in markets may not be reliable signals of underlying welfare values, if they exist.

4.3 Imperfect Competition

Frey and Eichberger (1994), argue that firms will devise policies to exploit preference anomalies and consumers will adapt, at least to some extent. In this section we examine an example embodying the first part of this insight.

In marketing and politics, decoys are attempts to alter the choices of consumers and citizens through the establishment of anchoring effects. Decoy effects were introduced in Chapter 1 and discussed in Chapter 3 in the context of the anchoring and adjustment heuristic. Recall that with anchoring consumer choices and valuations are correlated with information which would usually be viewed as irrelevant in a rational choice context. With decoys, the addition of a new option into the consumption set adds a new set of between-bundle comparisons.

In the context of marketing, McFadden (1999), summarises this as,

> . . . restaurateurs know that consumers use price as a guide to quality of wines, and that the second lowest priced wine is usually the best seller on the wine list. This effect is enhanced if a clearly lower quality wine is offered at a price only slightly below the targeted seller, making the targeted wine appear to be a good buy. Similarly, supermarkets will typically carry a low-quality generic brand priced slightly below the house brand, making the latter seem like a better deal. (p. 30).

In this section I examine how one of the standard models of imperfect competition is affected by the possibility of producing decoys. By producing a decoy that is inferior to its main brand a firm can raise the perceived advantages of its main brand, inducing consumers away from other competing producers. Of course if all firms produce decoys then the net effect is simply raised costs of production and hence, in a free entry equilibrium, a reduced number of firms.

We begin by briefly reviewing the standard Salop (1979), model of product differentiation. In this model, individuals are located around a circle (representing different tastes) with uniform density. Firms make three, sequential decisions:

1. whether to enter the market;
2. where to position themselves around the circle and then
3. what price to charge.

For a consumer located a distance x away from firm i the total cost of purchasing the good is $p_i + tx$, where p_i is the factory price of the good and tx represents the cost of travel. More generally, tx represents the cost to the consumer of not having the brand that exactly matches their tastes. Consumers buy the brand that has the lowest total cost for them and buy exactly one unit.

In this context we look for a symmetric, perfect Nash equilibrium.[1] An equilibrium is then a price vector p and a number of firms, N such that

1. all firms are equally-spaced;
2. p is a Nash equilibrium in prices given firms' locations and
3. $\pi(N) \geq 0 > \pi(N+1)$, where $\pi(N)$ is the profit of a typical firm in the symmetric Nash equilibrium for prices, given N firms in total. It is often easier to think of N as a continuous variable in which case N is the value for which $\pi(N) = 0$.

To find the equilibrium, consider first the firms' optimal pricing strategies given the N firms are equally spaced round the circle. Then firm i will serve all the customers between it and firm i+1 for which, $p_i + tx \leq p_i + 1 + t(1/N - x)$. Since all firms have two neighbours, firm i's total demand is therefore $(p_{i-1} + p_{i+1} - 2p_i + 2t/N)/2t$. Profits are

$$-F + (p_i - c)[p_{i-1} + p_{i+1} - 2p_i + 2t/N]/2t \tag{4.2}$$

where c is the marginal cost of production and F is the firm's fixed cost. Maximizing profits with respect to p_i and using the symmetry of firms yields an equilibrium price of $c + t/N$. Profits for a typical firm are then $t/N^2 - F$, so that in the zero-profit world of free entry, $N = (t/F)^{1/2}$.

Suppose we add a fourth stage between entry and the choice of price: each firm can decide whether to offer a decoy brand, the fixed cost of which is βF where $0 < \beta \leq 1$. There are a number of ways the impact of decoys could be modelled. One option would be as a form of vertical production differentiation, along of the lines of Shaked and Sutton (1982). An alternative, and possibly the simplest way to model the effects of the decoy is to suppose that it raises the attractiveness of the main brand. Formally, the cost of travelling to purchase the associated brand is lowered to αtx for some $0 < \alpha \leq 1$. In this simple framework, no consumers actually purchase the decoy (so its price is immaterial). If all firms produce a decoy, profits are,

$$\pi(N) = \frac{\alpha t}{N^2} - (1 + \beta)F \tag{4.3}$$

[1] There may well be asymmetric equilibria.

Table 4.1 Equilibrium Outcomes with a Decoy Brand

	Price	Number of firms
Equilibrium with decoy	$c + ((1+\beta)\alpha t F)^{1/2}$	$(t\alpha/((1+\beta)F))^{1/2}$
Equilibrium without decoy	$c + (tF)^{1/2}$	$(t/F)^{1/2}$

We can set $\pi(N) = 0$ to find the equilibrium number of firms conditional on the use of decoy brands. The prices and equilibrium number of firms for the two types of equilibria are shown in Table 4.1.

For any given N it may not be profit maximizing for a firm to produce the decoy, so we need to check when each of the possible outcomes represent equilibria. Consider first the equilibrium with all firms producing a decoy. Suppose firm i does not produce a decoy, given that its neighbouring firms produce the decoy. Its profits are then,

$$-F + (p_i - c)[p_{i-1} + p_{i+1} - 2p_i + 2\alpha t/N]/2t(1+\alpha) \tag{4.4}$$

In this situation, the equilibrium value of p is $\alpha t/N + c$, unchanged by the ith firm's deviation. Profits for the ith firm are then $-F + 2\alpha t/((1+\alpha)N^2)$. Using the zero-profit condition for N yields the result that deviating from the decoy equilibrium is not profit-maximizing for the ith firm when $2\alpha/(1+\alpha) < 1/(1+\beta)$. So when adding a decoy creates a large rise in demand, for a low fixed cost, the outcome with decoys for all N firms is an equilibrium.

Now consider the other, non-decoy outcome When is this an equilibrium? Suppose all firms but the ith do not produce a decoy and consider the ith firm's decision. This, it turns out, will be unprofitable provided $1 + \beta > 2/(1+\alpha)$.

Figure 4.4 illustrates the possible outcomes: In region C, when the decoy is expensive and its effect on demand is small (α, β are both large), then only the non-decoy outcome is an equilibrium. On the other hand, when the effect of the decoy on demand is larger and the cost of the decoy is small (region A), then only the decoy outcome is an equilibrium. However, in region B there are two equilibria – one with and one without the decoy. Note that, in region B, $2 > (1+\alpha)(1+\beta)$, so the price is lower with the decoy equilibrium. It follows

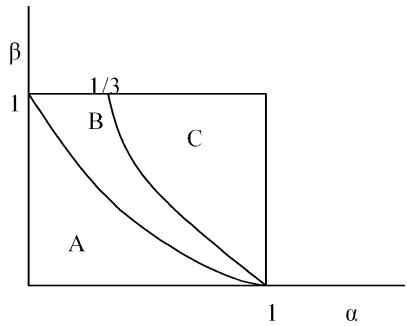


Fig. 4.4 Markets with Decoys

that there are also fewer firms in the decoy equilibrium. Now, we do not have a well-specified model of welfare for this situation. And as the next chapter makes clear, it is by no means simple to devise one. However, it is worth noting that in the Salop model without an outside option, the equilibrium number of firms is higher than is surplus maximizing. Here the decoy equilibrium has fewer firms than the non-decoy case and lower prices. So if welfare is only a function of the item chosen then allowing the decoy is welfare-enhancing, even though it leads to additional costs for each firm and a significant number of unused brands.

4.4 General Equilibrium

One problem not so far faced is the issue of whether a general competitive equilibrium can exist in the absence of full rationality and whether it is stable. The partial equilibrium model discussed above suggests that stability is not a simple issue. Moreover, if we consider the case of reference dependent preferences there are some reasons to suppose that there may be no general equilibrium, because what may appear to be an equilibrium when viewed from one vantage point (e.g. initial endowments) may not be when viewed from the 'equilibrium' itself.

There are at least two different understandings of reference points within a market setting. The first is that of customary consumption. Consider a consumer who must choose a pattern of within-period consumption repeatedly from the same consumption set over several time periods. His or her endowment might be money or, in a model with labour supply, it might include a set of goods which can be sold. In period t, customary consumption is some weighted average of consumption in periods 1 to t–1. In this context when developing a theory of how reference points might change over time, it seems most natural to think of customary consumption as the reference point from which options are judged in each period. An alternative candidate for the reference point is the endowment of the individual. Trades would be coded as gains and losses relative to this endowment point. Such a notion might be more appropriate in the case of an individual who buys or sells goods over a period of time, with a view to consuming them later on.

Formally, the customary consumption concept of reference dependent preferences is similar to the concept of habit formation investigated for instance in Pollak (1970), El-Safty (1976) and Hammond (1976). As with the customary consumption model, within the habit formation literature preferences are assumed to be shaped by previous consumption experiences, either through a process of addiction or through the develop of consumption capital. In either case, the value of good x is an increasing function of its past consumption, creating a bias towards the status quo which can be behaviourally equivalent to that found in models of reference dependent preferences.

Arguably, what distinguishes the two types of theory is the speed of adjustment of reference points which is extremely rapid, at least in experimental settings. Since there is little comparable evidence on the rate of habituation, one might question whether this is a sustainable distinction between habit formation and reference

dependent models. For our purposes what matters is that, in a typical market setting, prices and supply adjust dynamically. If the speed of adjustment of preferences is slow compared to the rate of adjustment of prices, then the outcome may be very different to that which obtains when the speed of adjustment of preferences is faster than the rate of adjustment of other market variables. Thus, in considering markets and general equilibrium there are four cases to consider which arise because reference points may be customary consumption or endowment points and preferences may adjust faster or slower than the market, in the following sense. Suppose that there are a sequence of trading and/or consumption periods. If the market clears in t periods and reference points are constant throughout the t periods, then we say that the rate of adjustment of reference points is slow. If reference points adjust prior to the market clearing then we say that the rate of adjustment is fast. Table 4.2 lays out the options, together with some relevant general equilibrium models.

If we interpret reference points as customary consumption then the relevant general equilibrium literature is that of general equilibrium with nontransitive preferences (GENP). This literature drops the assumption of weak rationality: instead a notion of local preference, dependent on some reference point is assumed (e.g. Georgescu-Roegen, 1936; Gale and Mas-Collel, 1975; and Shafer and Sonnenschein, 1975). The second and third of these papers, for instance assume only that the strict preference relationship, is irreflexive. Formally, they assume that for each bundle r the set $\{x|r \prec x\}$ is nonempty and convex. Then, optimisation is defined as the choice of x^* such that $\{x|x^* \prec x\} \cap S = \varnothing$, where S is the feasible set for that individual. A competitive general equilibrium (a GENP equilibrium) is defined as a price vector and a profile of individuals' consumption bundles such that, when all agents optimise on their feasible sets, all markets clear. Gale and Mas-Collel and Shafer and Sonnenschein provide an existence proof for such an equilibrium.

The concept of optimising within GENP corresponds with the maximisation of reference-dependent utility when the optimal bundle is also the reference point. Thus the GENP notion of equilibrium corresponds with a profile of consumption bundles and a price vector, such that each individual's bundle is optimal when viewed from itself. Although the GENP literature establishes existence, it is unsatisfactory in a number of ways. For a given profile of endowments, consider the allocation of goods at the general equilibrium within the standard Arrow-Debreu model. Given the relevant equilibrium price vector, excess demand will be non-positive in all markets. Moreover each individual will wish to consume exactly the bundle of goods which assigned to them in the equilibrium. The same is only necessarily true within the GENP equilibrium if the reference point for each individual is their

Table 4.2 Reference Dependent Preferences

Reference point		
Rate of adjustment of reference points, relative to market clearing	Endowment	Customary Consumption
Fast	Munro-Sugden (2003)	GENP (for existence)
Slow	Arrow-Debreu	GENP (for existence)

equilibrium bundle of goods. If their reference point is some other bundle, then at the 'equilibrium' price vector they may wish to consume some bundle of goods other than their 'equilibrium' bundle. Consequently, excess demand need not be non-positive in all markets. In short, the GENP equilibrium corresponds to a model in which individuals have reference-dependent preferences, where the reference points are provided by current consumption rather than the fixed endowments or historical consumption.

When reference points are interpreted as current endowments a different notion of equilibrium is needed. On this latter interpretation, endowments change as trade proceeds. Trade has to be modelled as a sequential process; at any given time, an individual takes his or her current holdings of goods as the point of reference when choosing what trades to make; but these trades then induce changes in the holdings of goods, and thus induce changes in his or her reference point. If all adjustment in the market place takes place prior to any changes in the reference point and if the reference points determining this adjustment are given by the agents' original endowments, then the relevant model of existence is the standard Arrow-Debreu one, at least in the short run. Once, though, we accept the possibility that reference points adjust out of market equilibrium, then the question of whether an equilibrium exists for this dynamic process becomes important.

In recent papers (Munro and Sugden, 1997, 2003) we tackle this issue. Our starting point is Hicksian consumer theory which we amend just as much as is necessary to accommodate the evidence on reference point effects discussed in Chapter 2, but avoiding some of the problems for consumer behaviour which are posed by Kahneman and Tversky 's theory of riskless choice .

4.5 Reference Dependent Preferences

We consider an individual's preferences over all possible bundles of n goods (with $n \geq 2$), represented by the set $\mathbb{R}^n_+$. For simplicity I shall use the same notation and definitions introduced into Chapter 2 to describe reference dependent preference structures and the reference dependent preference relationship, $\preceq_r$. In addition, to simplify notation somewhat, I use the conventions xRy, xPy, xIy to denote $x \preceq_y y, x \prec_y y, x \sim_y y$ respectively. Thus, for example, xRy signifies that if the individual were endowed with y and viewed this as her reference point, she would be willing to exchange y for x. Notice that the relations R, P and I do not necessarily have the typical properties of $\preceq$, and $\sim$. For example, R is not necessarily complete (an individual might be willing neither to move from x to y nor to move from y to x); x I y does not entail y I x (an individual who is just willing to move from y to x might be unwilling to move from x to y). The notation x RR y means that there exists some sequence of bundles $z^1, \ldots, z^m$ in $\mathbb{R}^n_+$ such that $z^1 Ry$, z^2 R $z^1, \ldots, z^m$ R z^{m-1}, and xRz^m. Thus, $xRRy$ signifies that if the individual is endowed with y, and if she always views her current endowment as her reference point, she is willing to engage in each of a series of exchanges leading from y to x.

With this notation, we take C1-C3 (completeness, transitivity and increasingness for a given reference point) from Chapter 2 and propose the following additional assumptions for each individual:

Axiom 4.1 (C4 Strict convexity) *For all* $r, x : \{y|y \succeq_r x\}$ *is strictly convex.*

Axiom 4.2 (C5 Continuity for a given reference point) *For all r, x:* $\{y|y \succeq_r x\}$ *and* $\{z|x \succeq_r z\}$ *are closed.*

Axiom 4.3 (C6 Continuity for changes in reference points) *For all r, x, y:* $\{r|x \succeq_r y\}$ *is closed.*

Axiom 4.4 (C7 Weak acyclicity) *For all x, y:* $\neg(yRRx \wedge xPy)$.

C1–C5 are standard assumptions for reference independent theories of the consumer. The next two assumptions add restrictions on preference that apply across reference points. Without the continuity condition, C6, small changes in the reference point could lead to large changes in behaviour. C7 meanwhile, is responsible for dictating how reference-dependent preferences can vary with reference points. There are two underlying motives for adopting C7. First, by assumption it rules out the problems of cycles which can occur in Kahneman and Tversky's theory. C2 (transitivity) rules out cycles for a fixed reference point, but it says nothing about the consequences of having preferences that alter with reference points. C7 extends C2 by ruling out cycles of the following form: x^2 is chosen when x^1 is feasible and is the reference point, x^3 is chosen when x^2 is feasible and is the reference point, . . ., and x^1 is chosen (and strictly preferred) when x^m is feasible and is the reference point.

As an assumption acyclicity has some psychological plausibility. For a start, the classic form of status quo bias rules out cycles with two elements. More generally, the same sort of psychological effects which would tend to create reference point effects will also tend to militate against cycles. Once an individual has moved from x to y, the shift of the reference point will imply a reluctance to move back to x. C7 implies this regularity, in the following weak form: changes in the reference point from x to y either leave preferences between x and y unchanged or shift them in favour of y.

In Munro and Sugden (2003), we also refer to two stronger variants of C7. Letting $d(y, z)$ denote the Euclidian distance between y and z, these variant conditions are:

Axiom 4.5 ((C7*) Strict acyclicity) *For all distinct x, y,* $\neg(yRRx \wedge xRy)$.

Axiom 4.6 ((C7) Limit acyclicity)** *For all x, and for all* $\delta > 0$, $\exists\, \epsilon > 0$ *such that,* $\forall\, x^1, x^2, x^3$ *which satisfy* $d(x^1, x) < \epsilon$, $d(x^3, x) < \epsilon$ *and* $d(x^2, x) > \delta$, $\neg(x^2RRx^1 \wedge x^3RRx^2)$.

C7* strengthens C7 by ruling out the possibility of a sequence of willing exchanges in which the individual starts with x, moves to y, and then returns to x. It has though the draw back of making the set of axioms incompatible with Hicksian preferences, which allows cycles of indifference. C7** is a still stronger condition as it rules out the possibility of a sequence of willing exchanges in which the individual

starts with a bundle which is arbitrarily close to x, moves to y, and then returns to some other bundle arbitrarily close to x. Thus, while C7 merely allows preferences to vary with reference points in the direction that corresponds with status quo bias, C7* and C7** postulate that such bias exists.

In addition we propose the following may hold for some individuals:

Axiom 4.7 ((C8) Reference Independence) *For all x, r, s* : $\{y|y \preceq_r x\} = \{y|x \preceq_s y\}$.

With C8 preferences become Hicksian – in other words there are no reference point effects. So this assumption is compatible with assumptions C1–C7 (though it makes the last two assumptions redundant), but as the discussion of the previous paragraph makes clear, it is not compatible with C7* or C7**.

4.5.1 Equilibrium and General Equilibrium

Having set out the model, we can show that there are preference structures that satisfy C1–C7. We also need to consider the issue of market equilibrium. In particular we analyse the trading behaviour of an individual who, over a series of periods, faces a fixed set of trading opportunities and has a fixed preference structure P that satisfies C1–C7. The first issue is whether there is an equilibrium choice for the individual.

Define a feasible set as a nonempty, closed, bounded and convex set S $\subseteq \mathbb{R}^n_+$, satisfying the restriction that, for all x, y $\in \mathbb{R}^n_+$: $(x \in S \wedge x > y) \rightarrow y \in S$. For a given S, the corresponding exchange set X is the set of nondominated elements of S, i.e. $X = \{y|y \in S \wedge \exists z(z \in S \wedge z > y)\}$. We interpret the exchange set as a set of bundles, every one of which is in every period exchangeable for any other. For example, suppose the individual enters a market with certain initial holdings of goods y, and faces a fixed vector of prices $p = (p_1, \ldots, p_n)$ at which he can buy and sell goods 1, …, n. Then $S = \{x|\sum_i p_i(x_i - y_i) \leq 0\}$ is a feasible set in our sense, and $X = \{x|\sum_i p_i(x_i - y_i) = 0\}$ is an exchange set.

Given these definitions, one can fix a particular feasible set S, and thereby an exchange set X and then define a trading sequence as a sequence $< x(0), x(1), \ldots >$ such that each x(t) is a member of X. We interpret each x(t) as the individual's endowment in period t; the difference between x(t) and x(t + 1) represents the exchanges carried out in period t.

A reflexive optimum is defined as a bundle x^* that is optimal on X, when viewed from itself. It can be viewed as an equilibrium state in which, were the individual to be endowed with that bundle of goods, he or she would not wish to engage in exchange. The assumptions of increasingness and strict convexity mean that x^* is uniquely optimal on X and on S. That a reflexive optimum exists follows from a simple application of Brouwer's fixed-point result and using the assumptions of continuity of preferences for a given reference point and continuity of preferences in the reference point.

Having defined the reflexive optimum we go on to consider whether a trading sequence will converge to a reflexive optimum, for any initial endowment. In doing

so we make the assumption that each individual is sufficiently rational to seek to exploit some of the gains from trade, given their current preferences. Specifically we impose the following:

Axiom 4.8 ((D1) Local improvement) *For all $t > 0$: (i) x(t) R x(t – 1) and (ii) if x(t – 1) is not a reflexive optimum, then $x(t)\ P\ x(t-1)$.*

D1 is not sufficient to guarantee convergence to optimality even in the case of reference-independent preferences. Consider someone with well-behaved preferences, with the initial bundle (4,0) and for whom the optimum is (1,3). If in period t they surrender $1/2^{t-1}$ units of x_1 in return for $1/2^{t-1}$ units of x_2, then D1 is always satisfied, but in the limit as t tends to infinity, their endowment tends to (2,2) rather than the optimum. Thus we need the individual to capture a significant part of the currently available gains if convergence to the optimum is to be guaranteed.

We therefore proceed as follows. Let L(Z) denote the Lebesgue measure of Z $\subseteq \mathbb{R}^n_+$. Taking S, X and P as given, define a set-valued function Φ (., .) as follows: For all x, z $\in$ S, Φ(z, x) = $\{y|yRx \wedge z > y\}$. To interpret this set, suppose that in some period the individual is endowed with x and is considering exchanging this endowment for z. Φ (x, z) contains all those bundles in S that when viewed from x are weakly preferred to x, but unambiguously inferior to z. Thus, if z R x is true, $L(\Phi[z, x])$ can be interpreted as an index, in terms of preferences viewed from x, of how far z lies above the indifference surface that passes through x. On the other hand, if z R x is false, $L(\Phi[z, x]) = 0$, so $L(.)$ can be viewed as a measure of the gains from trade that the individual realises in moving from x to z, evaluated from the perspective of x.

In addition to D1 we also place the following restriction on trading sequences:

Axiom 4.9 ((D2) Non-trivial improvement) *If $L(\Phi[x(t+1), x(t)]) \to 0$ as $t \to \infty$, then also $max_{z \in X}[L(\Phi[z, x(t)])] \to 0$ as $t \to \infty$.*

D2 requires that actual gains from trade in each period do not become vanishingly small unless this is also true of potential gains from trade.

The following result can then be proved (see Munro and Sugden, 2003):

Result 4.2 *For all preference structures P which satisfy C1–C7, and for all $x(0) \in X$: every trading sequence that satisfies D1 and D2 converges to the set of reflexive optima.*

The Result does not show that every trading sequence converges to a reflexive optimum. However, if there is a unique reflexive optimum, then every trading sequence satisfying D1 and D2 must converge to it. A reflexive equilibrium might be unique, but it is also possible to construct examples such that a trading sequence that satisfies D1 and D2 can converge towards a cyclical path that is always close to some reflexive optimum yet does not converge to any particular one. The problem is not eliminated by adopting C7* rather than C7. If reference-dependent preferences are smooth preferences in the immediate neighbourhood of a reference point are subject only to second order reference point effects. Thus, the reference point effects implied by C7* do not necessarily force a trading sequence to converge to a point.

However, because it imposes a status quo bias even in the immediate neighbourhood of reference points, the adoption of C7** does guarantee convergence to an optimum.

Result 4.3 *For all preference structures P which satisfy C1–C5 and which either satisfy C6 and C7** or C8, and for all x(0) ∈ S: every trading sequence that satisfies D1 and D2 converges to some reflexive optimum.*

Having set out conditions which guarantee convergence for an individual faced with an unchanging exchange set, we then go on to delve into the more general problem of whether an exchange economy composed of individuals with reference-dependent preferences will converge to some equilibrium.

To do this, we picture an exchange economy with n goods, N individuals (with $N \geq 2$) and no external effects. A typical bundle of goods possessed by individual j is denoted by $x^j = (x_1^j, \ldots, x_n^j)$. For each person j, preferences over such bundles are described by a preference structure P^j; the relations R, P, I and RR relevant to person j are written as R^j, P^j, I^j and RR^j. An allocation is then an Nn-dimensional vector $(x_1^1, \ldots, x_n^1; x_1^2, \ldots, x_n^2; \ldots; x_1^N, \ldots, x_n^N)$; **x** denotes a typical allocation. Feasibility for the economy is defined in terms of a resource constraint $q = (q_1, \ldots, q_n)$ such that, for each good i, q_i is strictly positive and finite. Taking some resource constraint q as given, an allocation x is feasible if $\sum_j x_i^j \leq q_i$ for all goods i; the set of feasible allocations is A. The set of exchangeable allocations Y is the set of non-dominated elements of A. Thus, an allocation x is exchangeable if and only if $\sum_j x_i^j \leq q_i$ for all i. The underlying idea is that each of the exchangeable allocations can be transformed into any other by means of exchanges of goods between individuals.

We set q, and hence also A and Y. We also set the preference structures $P^1, \ldots, P^N$ of the N individuals, and require that each P^j satisfies C1–C7. For all allocations x and y, y is a Pareto-improving move from x (denoted x P* y) if y^j R^j x^j is true for all persons j and if y^j P^j x^j is true for at least one j. If $y^j I^j$ x^j is true for all j, y is a Pareto-indifferent move from x (denoted x I* y). A feasible allocation x is a reflexive Pareto optimum if there exists no feasible y such that y P* x. Notice that if individuals are endowed with an allocation which is a reflexive Pareto optimum, no further voluntary trade can occur. In this sense, a reflexive Pareto optimum in an exchange economy is analogous with a reflexive optimum in the context of individual choice: it provides a one natural concept of efficiency in an exchange economy.

We define a collective trading as a sequence $< x(0), x(1), \ldots >$ such that each x(t) is an exchangeable allocation. Each x(t) is to be interpreted as a specification of individuals' holdings at the start of period t and thus $x(t+1) - x(t)$ represents exchanges carried out between individuals in period t. In the paper we do not propose any explicit model of the trading mechanism which determines each x(t + 1), given its preceding x(t). Rather our approach is more general, since it involves specifying two minimal restrictions on collective trading sequences – restrictions that are analogous with D1 and D2, and that are consistent with a wide range of trading mechanisms.

The first restrictions is:

Axiom 4.10 ((E1) Local improvement) *For all t ¿ 0: (i) either x(t) P* x(t–1) or x(t) I* x(t–1), and (ii) if x(t–1) is not a reflexive Pareto optimum, then x(t) P* x(t–1).*

The first part of E1 requires that the overall effect of the exchanges carried out in any period is weakly beneficial to all individuals, viewed from their current reference points. Given the fundamental assumption of myopia, this is an essential property of any trading mechanism in which individual participation is voluntary. The second part of E1 requires that, if potential gains from trade exist in period t, some part of those gains are realised in that period.

The second restriction is analogous to D2 in that it requires that in each period a non-trivial part of current potential gains from trade are realised. First, we extend $\Phi(.,.)$ to allocations as follows. For all z, x in Y, we define $\Phi(z,x) = \{y | yP^*x \wedge z > y\}$. Thus, Φ (z, x) contains all those allocations y that are Pareto-improving moves from x but that are unambiguously inferior to z – in the sense that z gives every individual at least as much of each good as y does, and gives at least one individual more of at least one good. (Notice that, because each allocation in $\Phi(x,z)$ is dominated by z, and because z is an element of A, each allocation is itself in A, i.e. $\Phi(x,z) \subseteq A$.) So, if zP^*x is true, $L(\Phi[z,x])$ can be interpreted as an index of the gains from trade that the N individuals together achieve in moving from x to z, evaluated with respect to their reference points in the relevant trading period. If z P* x is false, $L(\Phi[z,x]) = 0$. Hence our second restriction:

Axiom 4.11 ((E2) Non-trivial improvement) *If* $L(\Phi[x(t+1), x(t)]) \to 0$ *as* $t \to \infty$*, then also max* $_{z \in Y}$ $L(\Phi[z, x(t)]) \to 0$ *as* $t \to \infty$.

E2 requires that actual gains from trade in each period do not become vanishingly small unless this is also true of potential gains from trade.

Give all this, the following convergence results, analogous with Results 4.1 and 4.2 can then be proved (see Munro and Sugden, 2003 for the proofs):

Result 4.4 *For all profiles of preference structures* $P^1, \ldots, P^N$ *which satisfy C1–C7, and for all* $x(0) \in Y$*: every collective trading sequence which satisfies E1 and E2 converges to the set of reflexive Pareto optima.*

Result 4.5 *For all profiles of preference structures* $P^1, \ldots, P^N$*, such that each* P^i *satisfies* $C1 - C5$ *and either satisfies C6 and C7** or C8, and for all* $x(0) \in Y$*: every collective trading sequence which satisfies E1 and E2 converges to some reflexive Pareto optimum.*

4.5.2 Discussion

The results of the previous section, when combined with the GENP literature suggest that general equilibrium theorising is not impossible when preferences are reference dependent. The notion of equilibrium can be defined in a natural way and there are concepts analogous to standard definitions of Pareto efficiency.

The preference structure we adopt in Munro and Sugden (2003) represents a trade off between different goals. Acyclicity or something close to it is required to ensure that the economy gravitates to an equilibrium. On the other hand, though acyclicity is consistent with the experimental evidence on reference dependent preferences, it is not implied by it. Consequently it could be that evidence will be produced that leads to the rejection of acyclicity as a reasonable restriction on choices. It is also possible to obtain stronger results on existence and the properties of the equilibrium if one is prepared to ignore some of the less consistent evidence from the experimental literature. As we saw in Chapter 2, diminishing sensitivity plays an important role in Tversky and Kahneman's model of choice under uncertainty. It is less central to their riskless choice model and less well established in that context. If preferences between x and y are not-sensitive to changes in reference points outside the box defined by these two bundles, then an individual who chooses y from a linear consumption set when x is the reference point will also choose y on the same consumption set when y is the reference point. In other words the first step in the trading sequence is also a reflexive optimum for all agents. If one is prepared to accept this as a reasonable story then the competitive equilibrium dynamics are therefore much simpler. I exploit this simplicity in the model of tax policy developed in Chapter 9.

4.6 Conclusions

This chapter has illustrated how one class of anomalies – those associated with reference point effects – can have implications for the existence, stability and efficiency of markets. The chapter which follows pursues the implications of anomalies for measures of individual welfare. As a prelude, it is worthwhile reflecting briefly on the contribution of this chapter to that issue.

The claims one can make for a reflexive Pareto welfare optimum are rather limited. Within the Hicksian model the Pareto ranking of allocations need not be complete, but it is transitive. Assumption C7 means that the same is true of the Pareto relationship defined for reference dependent preferences. However, in the Hicksian model if an allocation x is preferred to allocation y by all individuals when x is the reference point then it will always be true that y is strictly preferred to x by all individuals when y is the reference point. In other words the Paretian relationship is asymmetric. With the reflexive Pareto welfare optimum, the claim need no longer be true. This makes welfare issues harder to resolve because they are not just about trading off the interests of different individuals, but also about trading off the interests of the same individual at different reference points.

On the other hand, this chapter has shown that some welfare comparisons are possible without complete and transitive preferences, if one is willing to accept that choices or preferences provide some information on the underlying welfare ordering. And once this position is accepted then something like the first

fundamental result of welfare economics is almost bound to follow: if individuals are free to choose in competitive markets and will choose a better option to their existing bundle when it is available then an equilibrium will be a position in which no one individual can choose a better bundle without some other individual losing out.

Chapter 5
Welfare

5.1 Introduction

The central propositions of welfare economics rest on the twin assumptions that each individual makes choices so as to maximize his or her own welfare and that social welfare is an increasing function of individual welfare alone. This approach, denoted standard welfare economics in Chapter 1, most starkly set out in microeconomic theory texts (e.g. Varian, 1978; Mas Colell et al., 1995) and in books on welfare economics, has attracted a long list of critics. Much of the work for which Amartya Sen won the Nobel prize for economics, for instance, is centred on his 'rational fools' critique of standard welfare theory and in the articulation of his own chosen alternative. Meanwhile, Rights theorists have stressed the problems of consequentialist theories, such as utilitarianism. The evidence surveyed in Chapters 2 and 4 adds a class of empirical challenges for standard welfare economics. This chapter is devoted to considering the implications of it for theories of welfare.

Before embarking on the main topics of the chapter it is worth taking up an important preliminary issue. The evidence of the previous chapters can be interpreted in two ways: measurement and coherence. The measurement interpretation is that, with the tools available to social scientists, it seems that we are unable to elicit reliable indicators of individual preferences for a wide variety of goods. The coherence interpretation is that individuals really do not have complete and coherent preferences. If the measurement interpretation is pursued, then this would limit the advice social scientists could give to decision-makers about the value to individuals of public policies. Beyond this though, there would be no presumption that individuals themselves did not know what they preferred and, when given the option, would choose accordingly. This is the interpretation of the evidence on anomalies adopted by some critics of contingent valuation, for instance Diamond and Hausman (1994), but it could also be applied to other methods of preference elicitation.

Under the alternative interpretation of incoherence or incompleteness, the problem remains of obtaining reliable indicators of individual preference. However, there is now the additional issue of deciding what goods individuals should actually consume. I shall call this the allocation problem.

A. Munro, *Bounded Rationality and Public Policy*, The Economics of Non-Market Goods and Resources 12, DOI 10.1007/978-1-4020-9473-6_5,
© Springer Science+Business Media B.V. 2009

It is also worth making a distinction between marketed and non-marketed goods. Under the faulty measurement interpretation of the data, one might take the pragmatic view that evidence of the widespread failure of the rational choice model for goods bought and sold in the marketplace has no necessary significance for public policy. Puzzling thought it may be, it is simply poor quality data. On the other hand, with non-marketed goods the question of what is optimal is inescapable and, in the absence of reliable data about individual welfare, potentially intractable. One solution to the problem of preference or welfare revelation might then be privatization, where feasible. This would solve the preference revelation problem, since individuals would then have incentives to make choices according to their underlying preferences. Simultaneously it would solve the allocation problem, since again individuals would be choosing levels of consumption where the marginal benefits to them equated with marginal costs. With the incoherence interpretation, privatization does not offer a panacea; on the other hand it may still be the optimal policy response if it can be proved that the market provides the best incentives for maximisation of individual welfare.

The evidence presented earlier in the book included many examples from market behaviour – and more generally, from situations where individuals had clear incentives to choose rationally if they were capable of doing so. Consequently, it is not reasonable to think that the evidence on anomalies is simply evidence of the limitations of the tools of preference elicitation available to the economist. As a result, much of the focus in this and subsequent chapters will be on the incoherence interpretation. Yet, as already indicated this does not remove the measurement issue; in fact it remains fundamental. As a result I shall return to it in Chapter 11.

Before going on, it is also worth recalling that bounded rationality and merit wants are not the same thing. An individual may have perfectly consistent and complete preferences which involve always preferring coal over food for consumption and food over coal for heating fuel. At least according to some notions of welfare, and according to many people's intuition, this individual's choices would not maximize his or her individual welfare. Most of the evidence of the previous chapter does not have much to say about the case of consistent, but (from a welfare perspective) sub-optimal preferences, although some data from financial markets does point to the inefficiency of many saving strategies (see Chapter 10). My main concern in this chapter is with the failure of the weak rationality postulate, rather than the strong one. Nevertheless I shall draw on some of the paternalism and merit wants literature because, as I shall argue, it has useful things to say about the implications for welfare economics in the context of the incoherence problem.

The line pursued through the chapter is rather circuitous in places, but the basic message is that preference elicitation provides information on welfare, but as we have seen, much of that information is confused and often contradictory. As is usually the case with contradictory information, the key question is then how this information is aggregated in order for a decision to be made.

5.2 Frames and Paretian Welfare Economics

Within a context of complete and transitive preferences, Paretian welfare economics as it is typically presented consists of four assumptions on the social welfare relationship, $\preceq_w$ and its relationship to individual welfare, and individual preferences, $\preceq_i$. The first assumption embodies the view that all social states can be consistently ranked. The second states an accordance between preference and individual welfare. The final two properties make welfare economics Paretian. The first of these is the well-known Pareto principle, which comes in various strengths but which I present below in its strong form. The second is perhaps less often explicitly stated (though see Sen, 1986; Boadway and Bruce, 1984 and, for a recent revival of the issue, Kaplow and Shavell, 2001 and a critique by Chang, 2000), but it is typically implicit in discussions of Paretian welfare: it says that only the preference aspect of social states count towards social welfare (Sen, 1986).[1] Together the principles represent *welfarism* (Sen, 1986).

More formally, suppose that each individual $i \in H$ has *individual welfare* $\preceq_{wi}$ which is a binary relationship[2] on X, the set of social states. An n-tuple of individual welfare rankings is a *profile,* w and the set of all possible profiles is $\mathcal{U}$. Given the profile w, the social welfare relationship is denoted $\preceq_w$. Standard welfare economics is then summarised by the following four axioms:

Axiom 5.1 *Social welfare (W)* $\preceq_w$ *is an ordering.*

Axiom 5.2 *Welfare Equivalence (WE).* $\forall x,\ y \in X,\ i \in H,\ (i)\ (x \preceq_i y) \longleftrightarrow (x \preceq_{wi} y)$.

Axiom 5.3 *Strong Pareto Principle. (Pareto)* $\forall x,\ y \in X,\ (i)\ (\forall i \in H,\ x \preceq_{wi} y) \rightarrow x \preceq_w y$; *(ii)* $(\forall i \in H, x \preceq_{wi} y\ \&\ \exists i : x \prec_{wi} y) \rightarrow x \prec_w y$.

Axiom 5.4 *Strong Neutrality (SN).* $\forall x,\ y,\ a,\ b \in X$, *if* $\forall i \in H\ \ (x \preceq_{wi} y) \longleftrightarrow (a \preceq_{wi} b)$ *then* $(x \preceq_w y) \longleftrightarrow (a \preceq_w b)$.

How can these core principles of Paretian welfare economics can be set out in a manner that is potentially compatible with a world where framing affects preferences? One way is as follows. For the formal statement of the principles, as in previous chapters, the subscript f denotes a frame dependent relationship. The absence

[1] Note that without the axiom of completeness on preferences however, non-preference information can count towards social welfare, for those comparisons where one or more people do not have defined preferences.

[2] In this chapter, in order to consider issues of transitivity and completeness, it is useful to work with the binary relationship for individual welfare $\preceq_{fwi}$ rather than an individual welfare function $w_{fi}(.)$. Impossibility theorems are not the focus here, but it is worth acknowledging one lurking problem with working with $\preceq_{fwi}$ in a social choice context: once strong neutrality and the Pareto principle are added to a universal domain assumption then a complete and transitive $\preceq_{fw}$ must represent a dictatorship. The problem disappears if there is sufficient information on the comparability of individual welfare.

of an 'f' subscript indicates that the relationship is frame neutral. With F as the set of frames, the domain of welfare profiles is $F \times \mathcal{U}$, with typical elements identified by fw or gw′. The individual welfare relationship of individual i within frame f is then $\preceq_{fwi}$. The set $F \times \mathcal{U}$ must be interpreted with some care. For any given frame, f, all welfare profiles are possible. However, when moving between frames f and g, a particular profile w_f in $\mathcal{U}$ will be linked with exactly one profile w'_g with the link determined by whatever difference in framing effects occurs between f and g.

Definition 5.1 A reframing $f \rightarrow g$ is a mapping from $\mathcal{U}$ to $\mathcal{U}$ such that a profile w_f maps to w'_g. A reframing is described as neutral if $\forall w_f \in \mathcal{U}$, $w'_g = w_f$.

An example illustrates the idea:

Example 5.1 There are two social states, x and y, two individuals, 1, 2, and two frames, f, g. Ignore the possibility of indifference, then there are two possible rankings of the states for each individual and therefore four combinations in each frame, each of which produces one welfare profile from 1–4, (see below):

Profile 1	Profile 2	Profile 3	Profile 4
1 2	1 2	1 2	1 2
x x	x y	y x	y y
y y	y x	x y	x x

The first entry is welfare profile 1, in which both person 1 and person 2 rank x above y. In the second entry, profile 2, person 2 now ranks y above x etc.

Suppose that f corresponds to being endowed with x and g represents being endowed with y and the only framing effect relevant for welfare is an endowment effect (i.e. moving the reference point from x to y favours y). Then only certain relationships between welfare orderings in the two frames are compatible with the underlying theory of the effect of framing. For instance in the table below, there are two examples of possible relationships, both of which are incompatible with the underlying theory of how reframing alters rankings. In each case the f frame represents the domain.[3] In case 1, if profile 1 maps to 2, then profile 3 must map to 4, because the apparent effect of moving from frame f to g is to transform agent 2's ranking of x above y into one where y is ranked above x. The second example is inconsistent because moving from f to g appears to favour x, whereas by assumption the reverse is true.

Case 1		Case 2	
f	g	f	g
1	2	1	1
2	2	2	1
3	3	3	3
4	4	4	3

[3] To complete the check for compatibility we also need to consider the effect of changing between frames when the domain is the set of welfare profiles with g as the frame.

Given the definition of reframing we are now in a position to state a set of axioms for a frame dependent Paretian welfare economics.[4]

Axiom 5.5 *Preference (PF).* $\forall i \in H,\ f \in F,\ \preceq_{fi}$ *is an ordering.*

Axiom 5.6 *Individual frame neutrality (IFN)* $\forall f,\ g \in F,\ x, y \in X,\ x \preceq_{fi} y \rightarrow x \preceq_{gi} y$.

Axiom 5.7 *Welfare Equivalence (WEF).* $\forall x, y \in X, f \in F, i \in H, \left(x \preceq_{fi} y\right) \longleftrightarrow \left(x \preceq_{fwi} y\right)$.

Axiom 5.8 *Individual welfare (IW)* $\forall i \in H,\ f \in F,\ \preceq_{fwi}$ *is an ordering.*

Axiom 5.9 *Social welfare (WF)* $\forall f, \preceq_{fw}$ *is an ordering.*

Axiom 5.10 *Strong Pareto Principle. (ParetoF)* $\forall x, y \in X, f \in F, (i)\ (\forall i \in H, x \preceq_{fwi} y) \rightarrow x \preceq_{fw} y$; *(ii)* $\left(\forall i \in H, x \preceq_{fwi} y\ \&\ \exists i : x \prec_{fwi} y\right) \rightarrow x \prec_{fw} y$.

Axiom 5.11 *Strong Neutrality (SNF).* $\forall x, y, a, b \in X,\ fw, gw' \in F \times \mathcal{U}$, *if* $\forall i \in H$ $\left(x \preceq_{fwi} y \longleftrightarrow a \preceq_{gw'i} b\right)$ *then* $\left(x \preceq_{fw} y \longleftrightarrow a \preceq_{gw'} b\right)$.

The first three axioms are the behavioural postulates encountered in Chapter 1, repeated here for clarity. The remaining axioms are concerned with what is good for the individual and society as a whole. IW and W are familiar claims that the welfare relationship for an individual is complete, transitive and reflexive, as is the social welfare relationship (see Broome, 1997 for example). The final two axioms are reformulations of the Pareto and Strong Neutrality principles; the latter is particularly strong since the property runs across rather than just within frames. Note that on its own the Strong Neutrality property allows non-preference information to affect social choice through the distinction between preference and individual welfare; WEF removes this possibility.

For this collection of axioms, the four properties of conventional Paretian welfare economics as listed above hold:

Proposition 5.1 $PF + IFN + WEF + WF + ParetoF + SNF \rightarrow$ *(i) WE, (ii), Pareto, (iii) SN and (iv) W.*

Proof (i) First, PF+IFN means $\preceq_{fi}$ is independent of f, so that $\preceq_i$ exists – we can talk about a preference relationship that is complete and transitive for each individual. WEF then means that $\preceq_{fwi}$ is independent of f and so $\preceq_{wi}$ is an ordering that accords with $\preceq$; (ii) Pareto follows from ParetoF (iii) similarly SN follows from SNF; (iv) that the social welfare relationship, $\preceq_{fw}$ is independent of f follows from IFN, WEF and SNF, if we simply allow $a = x$ and $b = y$.

Figure 5.1 summarises the relationships. Within the diagram we see that in a frame dependent world SP and SN do not require IFN etc. to hold. However without a degree of frame neutrality, SP and SN become vacuous statements.

[4] In stating these axioms I am not making any ethical claims for them.

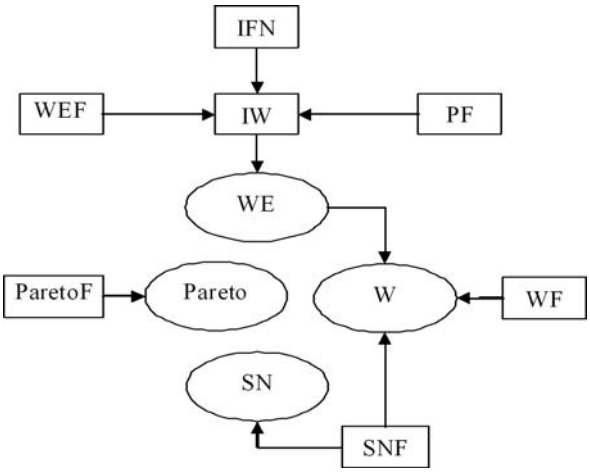


Fig. 5.1 Framing and Welfare Economics

If so desired, there are different ways of shoring up welfare economics so that $\preceq_w$ remains an ordering. Figure 5.1 indicates one method but we may also wish to consider the following axiom, which is superfluous if PF, IFN, WEF hold, but may have a role in any discussion of welfare economics without IFN. As above, when comparing frames f and g, we are fixing the profile in one frame and considering how that profile alters when the frame is g:

Axiom 5.12 *Individual welfare frame neutrality (IWFN)* $\forall f, g \in F,\ x, y \in X,$ $x \preceq_{fwi} y \rightarrow x \preceq_{gwi} y.$

Example 5.2 Continuing the example, the first illustration represents frame neutrality, while the second illustration provides a mapping which is compatible with an endowment effect, but in which the frame is not neutral for welfare.

Compatible Reframings

Frame Neutrality

f	g
1	1
2	2
3	3
4	4

Compatible but not Frame Neutral

f	g
1	2
2	2
3	4
4	4

.

The extended group of axioms (i.e. including IWFN) is coherent, but relaxing some of them causes serious problems for welfare economics. In particular, without IFN the extended set of axioms is inconsistent.

Proposition 5.2 *IWFN+WEF*$\rightarrow IFN.$

Proof Suppose $x \preceq_{fi} y$. By WEF then $(x \preceq_{fwi} y)$, but then by IWFN, $\forall g \in F, x \preceq_{gwi} y$. By WEF once more, $x \preceq_{gi} y$.

An alternative to IWFN is to impose consistency at the higher level of aggregation represented by the social welfare relationship. There are different ways to think of consistency at this level. One route is to make $\preceq_{fw}$ independent of f and it is this notion that is embodied in the original definition of $\preceq_w$. An alternative route is to suppose that the relationship $\preceq_{fw}$ is not affected by any reframing of the welfare profile. The axiom WFN does that:

Axiom 5.13 *Welfare frame neutrality (WFN). Let w'_g be a reframing of w_f then* $\forall x, y \in X, x \preceq_{fw} y \rightarrow x \preceq_{gw'} y;\ x \prec_{fw} y \rightarrow x \prec_{gw'} y$.

WFN is illustrated by the following continuation of the example.

Example 5.3 Suppose, reframing maps welfare profiles in the manner shown below

Reframing

f	g	g	f
1	2	1	1
2	2	2	1
3	4	3	3
4	4	4	3

.

If $\preceq_w$ exists, the welfare ranking for a given profile is independent of the frame. If instead WFN is imposed then the welfare ranking of x versus y given profile 1 in frame f must produce the same ranking in frame g given profile 2. If ParetoF applies then because x dominates y for both individuals in profile 1, then we know that $y \prec_{g2} x$ – i.e. x is ranked above y in g. If $\preceq_w$ exists, WFN and ParetoF are imposed then x is ranked above y whenever the profile is 1 or 2. Similarly, x is ranked below y whenever the profile is 3 or 4.

The difference between W and WFN is that, with the former, a given profile produces the same welfare ranking whether the frame is f or g. However with WFN, the welfare ranking is linked to the reframed profile, so that, in general the same profile will produce different welfare rankings in different frames. Consequently we have:

Proposition 5.3 *There exist reframings such that $\neg IWFN$, WFN, ParetoF$\rightarrow \neg W$.*

Proof Suppose W, so that $\preceq_{fw}$ is independent of the frame. From the example it can be seen that WFN+W implies that $\mathcal{U}$ can be partitioned into groups of profiles within each of which the welfare ranking of all x, y $\in X$ is the same. By $\neg IWFN$, at least one group contains more than one profile. Consider a society in which the reframing is such that the group contains one profile where x strictly Pareto dominates y and another profile in which y strictly Pareto dominates x. By ParetoF therefore $x \prec_w y$ and $y \prec_w x$ – a contradiction.

Thus the practical distinction between W and WFN is to do with possible reframings. If any possible reframing maps each profile onto itself – i.e. all reframings are neutral ≈ then W and WFN are equivalent. Conversely, as the impact of reframing becomes more and more diverse, then at some point W and WFN become

incompatible. A happiness maximizer might view W as the superior axiom – if a reframing of the choice makes an individual happier with x rather than with y, then the social choice rule should respect that pleasure. On the other hand, WFN might be justifiable if there are 'true' preferences, which may deviate from actual expressions of preference because of the way that options are framed. This though would suggest that IWFN is an appropriate axiom in the first place. But IWFN when combined with the strong neutrality axiom implies that reframing is always neutral. As a result with IWFN and SNF, W becomes equivalent to WFN.

5.3 Responses to the Evidence

The existence of anomalies represents a rejection of the IFN property. Without IWFN the remaining axioms are not necessarily inconsistent, but they may be empty of content. Assuming that emptiness is not desirable, then, broadly speaking there are six possible routes a theory of welfare might take in the face of the evidence on anomalies. I shall consider them briefly in turn, though it is worth noting that they are not all mutually exclusive. These approaches are:

- Switch to an alternative, competitor theory of welfare (this amounts to dropping WEF, as well as possibly SNF and ParetoF).
- Allow welfare to be frame dependent (Keep WEF but drop W).
- Declare that no ordering is possible (Drop W and WF).
- Choose an 'optimal frame' in which to elicit preferences (Drop WEF and add WFN).
- Aggregate the contradictory evidence to produce a consistent welfare ranking (Drop WEF and add WFN).
- Search for supplementary evidence on welfare and then aggregate the information (Replace WEF and SNF).

5.3.1 Switch Theories

First, it might be possible to switch to some other, competitor theory of welfare that is not associated with preference satisfaction. For some alternative theories this is not logical. For an anti-consequentialist, for instance, welfarism is not wrong because it is difficult to elicit reliable indicators of welfare, but because of its focus on ends rather than means. On the other hand, there are some alternative approaches, especially those used in applied welfare analysis which might be viewed as superior to flawed welfarism. QALYS (Quality Adjusted Life Years), as used in healthcare appraisal, provide an example (Kind et al., 1982), as do indicators of development, such as those used by the United Nations. In both cases we might justify the use of the alternative by viewing them as crude indicators of welfare. Direct approaches to measuring welfare are, as already indicated, shot through with difficulties, so that although in an ideal world we would use a direct approach to measuring welfare,

given the problems this poses for the applied economist, in fact QALYS etc. provide a more accurate albeit indirect guide. In other words, there may be an instrumental value in using alternative welfare indicators.

This is a strong claim, one that needs supporting evidence of that kind that shows QALYS or some other alternative can come closer to ranking individual welfare than stated preference methods. The coherence of the alternative welfare measures is not in itself a sufficient nor even a necessary part of the evidence. To see this, suppose that from a welfare point of view an individual's options can be ranked A, B, C, D, in decreasing order. Suppose that preference information ranks A above B in some context, but B above A in other contexts. C and D are always ranked below both A and B. Compare this to an alternative system of welfare measurement which always offers the ranking: C, D, A, B. The ranking is coherent, but it is not necessarily better as it places the two worst outcomes at the top of the list. As a result, depending on the available choice sets and how one resolves the conflicting preference information, it is not inherently superior to some technique of ordering the alternatives which uses the flawed preference information.

The possibility of using an alternative measure of welfare for instrumental reasons, involves some judgement about the relative inaccuracies of direct information from choice and indirect information from, for instance, stated preference. This is something I take up in detail in Chapter 11.

5.3.2 Allow Individual Welfare to be Frame Dependent

One possible route is to allow the ranking of social states to be affected by the frames of elicitation. This is often the underlying notion for instance, in advocates for the maximization of happiness (e.g. see Ng, 1997, or Layard, 2006).

Once individual welfare is conceded as frame dependent the argument for W makes less sense. Consider for instance two frames, which I label pejoratively, transparency (t) and lies (l). In the second frame information is misleading or obscure. In one profile everyone prefers option x to option y in both t and l. In another possible profile, everyone prefers y in t, but x is preferred by all in l. W asks that we ignore the nature of the frames and if the Pareto principle is respected then for the second profile, we rank x above y. In this context, WFN appears slightly more attractive: it implies that for the second profile, x and y are ranked consistently.

One particular difficulty with frame dependent welfare is the ephemeral nature of many framing effects: if loss aversion can be undone rapidly through repeated exposure to exchange, then policy may be unstable. A good example is perhaps provided by the Asian disease example encountered in Chapter 1, where a small, rephrasing of the problem leads typically to a change in rankings for a significant subset of individuals. Leaving aside the ridiculousness of supposing that welfare in two societies could differ significantly simply because of this linguistic subtlety, any decision-maker which sought to treat the two treatments as representing distinct states would rapidly create incoherence in policy making. W may therefore imply policy instability, whereas WFN does not.

A specific model which allows frame dependent welfare is that proposed by Daniel Kahneman (Kahneman, 2000a,b). The basis for his notion of 'objective happiness' is an integral over time of each moment of happiness experienced by an individual. This integration is subject to some normative restrictions, notably an assumption of separability between time periods. Yet, such restrictions make it hard to see how moments of objective happiness can be calculated from behavioural data since it is normally the case (as Kahneman readily acknowledges) that feelings of happiness in one moment are often closely tied to feelings in preceding moments. Objective happiness is potentially consistent with W, but inconsistent with WFN. Objective happiness is therefore also associated with policy instability.

Perhaps the greatest difficulty that flows from allowing $\precsim_{fw}$ to depend on f, is that it leaves unresolved the choice of f in which welfare (and possibly choices) is to be elicited. This is true whether or not WFN is imposed. For the transparency and lies example, WFN only requires that profile 2 be ranked consistently between frames, but it does not prioritise the information obtained from any particular frame. Once f becomes an aspect of the social state that is relevant for welfare, then a ranking of social states requires some discussion of ranking frames. Allowing welfare to be frame dependent does not therefore solve the conundrum posed by anomalies; it merely shifts the focus.

5.3.3 No Ordering

When questioning the value of stated preference methods, Peter Diamond and Daniel Hausman (1994) ask 'is some number better than none?' They argue that,

> In short, we think that the evidence supports the conclusion that to date, contingent valuation surveys do not measure the preferences they attempt to measure. ... Viewed alternatively as opinion polls on possible government actions, we think that these surveys do not have much information to contribute to informed policy-making. Thus we conclude that reliance on contingent valuation surveys in either damage assessments or in government decision making is basically misguided. (p. 46).

So, a third possible response to the evidence of the last few chapters would be to confess that perhaps sometimes no number is appropriate. The welfare economist might then agree with Diamond and Hausman and say that such is the doubt over the validity of a particular method of preference elicitation, it might be better to stay silent on the subject of welfare when no more reliable instrument is to hand.

One drawback with the no-ordering approach arises out of the borderline – or its absence – between rationality and bounded rationality discussed in Chapter 1. There it was pointed out that it was only meaningful to separate out goods for which behaviour was rational and goods for which behaviour was boundedly rational, if the two classes of goods were separable in a fairly strict sense. In the absence of separability, there may be no watertight boundary between the two types of goods. Moreover, as we saw in Chapter 2, many market goods appear prone to framing effects. It may therefore be impossible to say anything about welfare, even for

marketed goods, so that the silence of economists on the subject of preferences would be complete.

Now perhaps such humility is overdue amongst economists but obviously, with public goods in particular, and more generally with non-marketed goods this position leaves normative public policy in somewhat of a void. Diamond and Hausman argue that, instead of relying on expressions of individual preferences,

> . . . it may be more informative to have expert evaluation of the consequences of an environmental change than to consult the public directly about environmental change. (p. 56).

This side-steps the issue. Experts may well be the source of information about the consequences of a policy, but they may have no special insight into the value of such consequences. Secondly, experts seem to be prone to many of the same anomalies as lay persons (Kahneman et al., 1982). So, as a general solution to the problem, the 'no ordering' option appears to be a non-starter.

5.4 Optimal or Golden Frames

Some of the frames one might use to elicit preferences might themselves be viewed as unacceptable or inferior to other frames. For instance, it is common to suppose that researchers should be truthful in their description of the objects of choice. If that indeed is true, then any preference elicitation mechanism which was untruthful would be ranked below all mechanisms where the choices were accurately described. Suppose one could go beyond such simple comparisons of frames and construct a ranking of possible frames. At the top of this ranking would be the frame, f*. Whether A was truly preferred to B would then be judged according to whether A was preferred to B when f* was the frame. In other words a principle of dictatorship would be invoked in which f* was the deciding perspective.

One could therefore replace IFN by:

Axiom 5.14 *Gold Standard Frame (GF).* $\exists\ f^* \in F, s.t. \forall i \in H, x, y \in X,\ x \preceq_{f^*i} y \rightarrow x \preceq_{wi} y$ *and* $x \prec_{f^*i} y \rightarrow x \prec_{wi} y$.

Note that GF 'bypasses' the welfare equivalence and individual frame neutrality and implies WFN. As a result we have:

Proposition 5.4 *GF + WF + SNF* $\rightarrow$ *W*

Proof Consider any frames f, g $\in F$, $x, y \in X$. By SNF, $\forall$x, y, $\in X$, $fw, gw \in F \times \mathcal{U}$, if $\forall i \in H$ $\left(x \preceq_{fwi} y \longleftrightarrow x \preceq_{gwi} y\right)$ then $\left(x \preceq_{fw} y \longleftrightarrow x \preceq_{gw} y\right)$. By GF, $x \preceq_{fwi} y \longleftrightarrow x \preceq_{gwi} y$, so $x \preceq_w y$. WF states that $\preceq_{fw}$ is an ordering, so $\preceq_w$ is also complete and transitive.

Having a complete ordering over frames appears as unachievable as having an complete ordering over options. Nevertheless there may be 'better' and 'worse' frames and frames which are acceptable and unacceptable according to some criteria (an *unacceptable* frame is defined as one in which $\preceq_{fi}$ has no impact on $\preceq_w$). Using

such criteria could then result in a set of undominated frames, any of which would be acceptable for the elicitation of preferences or might be viewed as the domain upon which aggregation rules would be meaningful.[5]

Of course if the set of acceptable frames was not a singleton and if different frames within the acceptable set produced different preference orderings, then we might still, ultimately, have only a quasi-ordering of options or something else which falls short of an ordering.

How then do we rank frames? There seems to me three different kinds of approach. One can evaluate frames directly, according to whether they represent good processes. Alternatively, one can rank frames according to whether they yield good outcomes; finally one can assess frames according to whether they yield good decisions rules. For instance, the requirement that frames be truthful would be a feature of the process of elicitation itself; the requirement that the resultant rankings respect monotonicity would be an outcome-orientated measure of the goodness of a frame, while the requirement that a framing technique yield transitive rankings would be a decision-based yardstick for frame evaluation. Many proposed schemata for frames, which are often proposed as representing 'good practice' in valuation mix up elements from the three different approaches. For instance, the NOAA panel on contingent valuation, Arrow et al. (1993), suggests that the interviews should normally be conducted face-to-face and also that the results should show sufficient sensitivity to scope. The first of these is a process requirement; the second is an outcomes-based yardstick.

Ranking frames is not without its own difficulties. The first problem is that of practicality – is it possible to define frames in a sufficiently precise way that is amenable to ranking and then, are the definitions sufficiently broad to be widely applicable? For instance, let us return to Tversky and Kahneman's Asian disease example, where the different responses to the choice between A and B turned on whether the options were framed as choices between saving lives or preventing deaths. One can view the difference between frames either as quite a small change in a very specific part of the English language or as an exemplar of the difference between 'losses' and 'gains' frames. If we were to define the frame in the first way, we might end up with a very detailed taxonomy for frames. But such detail might represent a barrier to implementing the ranking system in new contexts. On the other hand, a broader brush taxonomy would be easier to apply to newer contexts, but we might have much less confidence in its applicability.

The second issue is that of existence – given some criteria for acceptability does there exist at least one frame which satisfies all criteria? In the NOAA list of criteria for contingent valuation, there are over twenty suggested guidelines many of which

[5] If the latter one might end up with an axiom such as $\exists$ F* $\subset F, s.t. \forall i \in H, x, y \in X$, if $\forall f \in F^*, x \preceq_{fi} y$ then $x \preceq_{wi} y$ and if $\forall f \in F^*\ x \prec_{fi} y$ then $x \prec_{wi} y$. Such an axiom still allows frames outside F* to influence welfare ranking when there is no unanimity amongst the acceptable frames. One could strengthen the axiom by imposing a form of neutrality, whereby the welfare ranking of x and y is unaffected by their ranking in frames that do not belong to F*.

have direct implications for the framing of options. Applying all of them simultaneously might yield an empty set of acceptable frames.

The third problem is that of non-uniqueness mentioned above – if a group of frames are all acceptable, but some of them lead to different rankings of the options, how should the options be ranked?

Finally there is a fundamental the issue of value consistency. Ultimately, the criteria for ranking frames rest on value judgements. If these value judgements contradict the theory of welfare, then the ranking of options produced by the optimal frame may not be optimal as viewed from the theory of welfare. As an example suppose that we employ the criterion 'show visual information wherever it is possible' perhaps on the grounds that this satisfies some deeper requirement for transparency, which for the purposes of the example we take to be the ultimate good. Two frames, X and Y are available; only X satisfies the criterion, by offering photographs of the options to subjects. Otherwise the frames are identical. In a particular case, survey respondents are asked to choose between investing in a new treatment for disease A or disease B. Disease A is transient, rarely life-threatening, involves little pain for the sufferer and is not particularly contagious. Disease B is contagious, painful and frequently fatal, but whereas the sufferer from A is covered by large and weeping sores, disease B does not alter the appearance of the sufferer at all. Now we might anticipate that, faced with the visual information in frame X, many more respondents choose A for the investment compared to frame Y. On the other hand, some underlying theories of normative choice (e.g. QALYs) would prioritise B rather than A. The principle 'maximize welfare' would then be in contradiction with the principle 'choose a transparent frame'. To the extent that welfare maximization represented the more basic value judgement, then 'choosing a transparent frame' would have to give way to some other principle for judging frames. Note, that I am not suggesting that welfare maximisation is necessarily the more basic value judgement; some might argue that maximising transparency is a more fundamental principle in the context of public policy decisions. All I am pointing out is that within a consequentialist framework the criteria used for ranking frames need to be consistent with the criteria used for ranking options.

5.4.1 Markets and Referenda as Optimal Frames

Within the non-market valuation literature, two types of frames have attracted significant attention. The first is the notion of a market frame and the second is the referendum.

In a market, individuals buy and sell goods, usually exchanging commodities for money. A market frame is a frame which uses the language and constructs of the market place in order to elicit preferences. Thus subjects are asked to choose rather than to rank, to value rather than match. In a referendum meanwhile, citizens face a choice between two options. Whichever option receives the most votes will be implemented. A referendum-based frame mimics the format of the political

referendum, offering each subject two choices and using the language of referenda to suggest that voting for option A rather than B makes higher the probability that society will option A. Both forms therefore share a language of choice and importantly have an underlying notion of consequence: that choosing A rather than B raising the probability of securing A.

It is probably a mistake to think of *the* market frame. Market environments differ: an online ascending price auction is a market frame as it the take-it-or-leave it pricing of the supermarket or the bargaining environment of the souk or used car showroom. So what is *not* a market frame? Exchange in markets is typically conducted by the individuals concerned or by individuals on behalf of their households or companies. In the last case, the maximand is usually defined in a decentralised manner (e.g. maximize profit; minimize the cost of buying x) making it unnecessary for the agent acting as buyer or seller to consider the individual preferences of all of the principals. It is rarely that an individual trades for the citizenry as a whole. Secondly, anthropologists stress that much non-market exchange is symbolic, designed to create and sustain reciprocating relationships (Malinowski, 1961) or to express status (Yan, 2005). In the context of this book, frames that are non-market occur most frequently in stated preference exercises, where hypothetical questions are asked and whether techniques used to elicit values can include approaches such as matching, ranking and attitude expressions that are uncommon in real markets.

To my mind the case for a market-based frame or its relative, the referendum-based frame rely on three potentially separable arguments (Sugden, 2004). The first kind of argument that might be proposed is one from familiarity. In outline, it states that some choices or expressions of value are mistakes, in the sense that they arise from a lack of understanding of the options and their consequences. The possibility of mistakes is increased when agents are unfamiliar with the framework within which the information is presented, as well as when they are unfamiliar with the objects of choice. Presenting information in a familiar manner eases the undoubted problems of information processing, thereby lowering the chances of mistakes and increasing the probability that agents will make expressions of value consistent with their preferences. Markets (and referenda) are familiar frames of reference (so the argument concludes), and so in the absence of clearer alternatives, they should be considered as the optimal frames for preference elicitation in general.

The second argument for the use of market based frames relates to the opportunities provided by markets for repeated choice and learning in an environment where sub-optimal choice is punished (by a failure to achieve maximum utility). It is a preference-revelation-based argument in that these are usually considered to be exactly the kinds of conditions most likely to uncover preferences. Consequently the decisions made in markets yield good outcomes in the sense that they are true to any underlying preferences. Thus if a frame uncovers preferences that contradict those inferred or revealed by markets, then that frame should be rejected – only frames which yield preference information compatible with market-revealed preferences are acceptable. Note though that this argument does not imply that it is market-like frames which are necessarily optimal. Rather it is the market which provides the test against which other frames are to be measured.

Formally, consider the finite set S of outcomes, let P(S) be its power set and let $M \subseteq P$ represent the portion of the power set for which there is preference information revealed in markets. Let $\preceq_{mi}$ be the preference quasi-ordering on M revealed by market behaviour and let $N = P/M$ be the non-market subset of P, for which there is no direct revealed preference information. Suppose we use the known elements of $\preceq_{mi}$ plus transitivity and, possibly, some other imposed properties of $\preceq_{mi}$, such as monotonicity or normality of particular goods, to make inferences about the ranking of elements of S. Meanwhile, we also use some method of preference elicitation to obtain preference information about some elements of N. Denote the quasi-ordering uncovered by this procedure by $\preceq_{fi}$, where f refers to the frame employed. Then we say that the frame has the property strong market acceptability (SMA) if the following is true:

Axiom 5.15 *SMA* $\forall x, y \in N, i \in H$ *if* $x \preceq_{fi} y$ *then* $x \preceq_{mi} y$.

The property is 'strong' in the sense that if there is a single contradiction between the ranking under f and m, then the frame is rejected as unacceptable. On the other hand, one might demand still higher conformity between f and m by, for instance, requiring the 'only if' property as well. For now I shall take it an exemplar of the sort of standard by which frames might be assessed. How reasonable is it?

First SMA is problematic from the viewpoint of the empirical evidence. We saw in previous chapters that preferences might not be consistent between frames. Hence, when preferences are elicited over a sufficiently large subset of N, there may be no f which yields a quasi-ordering that passes the SMA test.[6]

SMA might also be limited in its ability to rank frames. In many preference elicitation exercises the range of goods for which preferences are found is limited and as a result the inferences one might make are restricted. Thus many frames might satisfy SMA, which provides no means of choosing between them. Alternatively, as discussed above, none of the frames might pass the SMA test, yet the failure of one frame might be much more severe (in an as yet undefined sense) than a competitor.

The final argument for the market frame rests on the notion of consent: the individual concerned may offer some sign that they find one method of preference elicitation more acceptable than another. This touches on the arguments concerning paternalistic intervention which we explore in later sections of the chapter. Consent to the market frame is most likely to arise from one of the first two reasons.[7] One might argue that the second argument for market frames is prior. For the first argument (familiarity) the acceptability of market frames will depend in part on their

[6] We have set aside the issue of merit wants, but a further problem with SMA is its confusion of 'market-based preference' with welfare. Suppose in the market we obtain information that an individual prefers 10 grams of heroin to £ 20 and £ 20 to a night at the theatre. In a non-market elicitation exercise the 10 grams of heroin is preferred to the night at the theatre in one frame (f_1), while under another frame (f_2) the night at the theatre is preferred. In this case SMA would imply rejection of f_2, while f_1 passes the test, but the observer might be tempted to conclude that it is the information obtained from the market frame which is flawed and not necessarily f_2.

[7] One might prefer the market for instance because of its anonymity or the freedom it provides, but these is not necessarily relevant in the choice of how to design stated preference.

frequency. There may be multiple equilibria – in some of which market frames are rare, unfamiliar and therefore have limited acceptance.

What links all three of the arguments for market frames is their limited nature: they are expressed with reservation. To a degree this reflects limitations on the acceptability of markets. Some forms of markets (e.g. a descending price auction) are not familiar to a large section of the population. As we saw in Chapter 3, learning may not be optimal in all markets. Most obviously consent to the market is heavily circumscribed in many OECD countries. Votes are rarely alienable; children cannot be traded for cash and while the pattern of restrictions differs markedly from country to country, the delivery of healthcare is often outside of the marketplace. It is not though inconceivable that the citizens of a society, having opted not to have a particular good allocated through the market, should still find acceptable the use of a market frame to determine the allocation. For instance, proponents of specific egalitarianism (Tobin, 1970) might be against general redistribution in cash, but for equal distribution of healthcare or educational opportunity. Within healthcare, rationing decisions might still be made according to preferences elicited in market-like frames. It is not inconceivable, but it is presumptive.

5.5 Information Aggregation

The previous option rendered welfare rankings by placing all its weight on information garnered under one frame and dispensing with the evidence on preferences from all other frames. It is therefore a special case of our next alternative: some sort of aggregation of the contradictory, but frame-dependent information available to the decision-maker. Formally this route means dropping IFN and WEF, and adopting IWFN. However, unlike the first option of choosing another welfare standard, initially in this aggregation process we confine ourselves to the situation where the only available information is from the frame specific preference relationships, $\preceq_{f_i}$. Thus a property of neutrality is imposed on the process by which individual, frame-specific rankings are aggregated to produce individual welfare. Presumably also something akin to a Pareto principle is also a feature of desirable aggregation.

If the problem is how to construct a welfare relationship, on the basis of $\preceq_{f_i}$, are there rules which the process of aggregation should always obey?

From a purely technical point of view it is possible to see this problem as analogous to the problem of social choice, where the f_is represent individuals rather than frames. Now, social choice has generated a huge literature on the implication of rules of aggregation, so it would be useful if the analogy was a close one. Kenneth Arrow's famous impossibility theorem makes it clear that it is not in general possible to construct a welfare relationship from individual preferences if the relationship is to be an ordering and if it is to satisfy a number of fairly minimal requirements: non-dictatorship (D), universal domain (U), independent of irrelevant alternatives (I) and the Pareto property (P). If the same properties are sensible for aggregation of then the same kinds of impossibility theorem applies.

However, U, P, I and D are not necessarily desirable features for the individual welfare relationship. In the absence of merit wants, P is probably unarguable, but none of the other three properties which make sense in a social choice setting are compelling when we deal with a single individual. Consider the principle of dictatorship: someone who advocates market-based frames as the true means of assessing individual interest argues for dictatorship. So D need not be a necessary property. Similarly U might be rejected because the preferences in some frames might break a principle of monotonocity. Finally, I is not uncontentious. For instance, let x = a ticket to a football match, $y = £10$ and $z = £5$ and two frames, f_1 and f_2 one of which elicits the ranking $z \preceq_{f_1} y \preceq_{f_1} x$, while the other produces $x \preceq_{f_2} y \preceq_{f_2} z$. The principle of independence of irrelevant alternatives would dictate that the choice between x and y should depend only their pairwise relationship, which is inconclusive. However, one might argue that the relative positions of y and z are informative. In frame 2, the dominated option (z) is ranked above the dominating option (y). On these grounds, an observer might be suspicious of the reliability of frame 2 and therefore rank x above y on the basis of the information provided by the frame in which monotonicity is not violated.[8]

In this example two things stand out. First, there is the fundamental difficulty of squaring contradictory rankings if the ranking information is all that is available. Secondly, therefore, there is the potential value of information other than that embodied in the frame dependent orderings. It is this information (e.g. that frame 2 is unreliable or that monotonocity is a feature of preferences) that makes the properties borrowed from social choice theory less than compelling.

Aggregation though need not mean completeness in the aggregate relationship. It may also be possible to relax the requirement that the resulting welfare relationship is an order. To be more specific, consider the following properties (Sen, 1970):

Acyclicity (A): $\forall j, \forall x_1, \ldots, x_j \in X, \quad [x_j \preceq x_{j-1}$ and $x_{j-1} \preceq x_{j-2} \ldots$ and $x_2 \preceq x_1 \rightarrow x_j \preceq x_1$.

Anti-Symmetry (AS): $(x \preceq y \; and \; y \preceq x) \rightarrow x = y$.

A relationship which satisfies reflexivity, completeness and transitivity is an ordering. This is the standard picture of the welfare relationship; we can weaken it by dropping completeness or by replacing transitivity by acyclicity or something similar. A relationship which satisfies reflexivity and transitivity is then a quasi-ordering; one which satisfies reflexivity, transitivity and AS is a partial ordering.[9]

If we are prepared to go down this route, then possibility theorems can be constructed in a number of ways. Sen (1970), for instance shows that if all that is

[8] Given U, and a welfare ordering summarised by $\preceq_{wi}$, I + P is equivalent to strong neutrality (see Blackorby et al., 1984, for instance) so the margin for avoiding impossibility theorems is small if strong neutrality and P are seen as desirable.The ability to avoid dictatorship or impossibility then rests on the size of F and its effects on individual preference.

[9] I am using the language of Sen (1970). As he points out there is a wide variety of uses in the social choice literature.

required is that, for any non-empty subset of X there is a set of best elements (a 'choice set')[10] then the following aggregation rule satisfies U, P, I, D:

Definition 5.2 Consensus rule (CON). $\neg\left[(\forall f_i\ x \preceq_{f_i} y) \text{ and } (\exists i : x \prec_{f_i} y)\right] \leftrightarrow y \preceq_{wi} x$.

The consensus aggregation rule states that x is welfare-preferred to y if and only if it is not the case that y is frame-preferred to x in all frames and strictly frame-preferred in at least one frame. When combined with a completeness assumption on $\preceq_{wi}$ this implies welfare indifference if there are at least two frames where the rankings of x and y are contradictory. Other rules are possible, but they share this unappealing feature of failing to offer a definitive ranking of two elements whenever the evidence is contradictory. Only if all frames yield the same ranking of x and y will the consensus rule produce something other than indifference.

There is another feature of the consensus rule worth pointing out: it does not necessarily produce consistency. To illustrate this, suppose we have two frames. In one x is strictly preferred to y, which in turn is strictly preferred to z; in the other z is strictly preferred to x which is strictly preferred to y. The rule then yields $y \prec_{wi} x$, $z \sim_{wi} x$ but $z \sim_{wi} y$.[11] In other words the welfare relationship is not transitive. This may actually be a reasonable conclusion in some circumstances, as Fig. 5.2 illustrates. In the figure, we might be certain that x is welfare-ranked above y, but it might be that depending on the frame, x could be preferred to z or vice versa and similarly for the relationship between y and z. If we applied the consensus rule but then also imposed transitivity on welfare-indifference we would then end up with the conclusion that x was strictly welfare-preferred and welfare-indifferent to

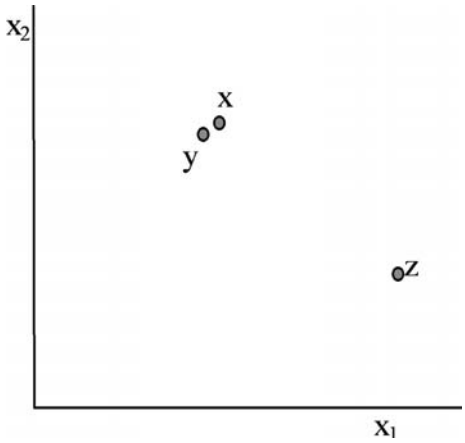


Fig. 5.2 Intransitivity of Welfare Indifference

[10] In the sense that for any S ⊆X, x belongs to the choice set for S implies that y $\preceq_w$ x, $\forall y \in S$.

[11] The focus here is on aggregating across f for given i, but obviously inconsistency here produces problems for the subsequent aggregation to the welfare relationship, $\preceq_w$ which I shall not pursue.

y, simultaneously, an obviously contradictory position which suggests that welfare indifference need not be assumed to possess the transitivity property.

To sum up, there are limited but useful analogies between aggregation over frame-dependent preferences for an individual and aggregation over individual welfare profiles for social choice. The social choice literature shows that the existence of a welfare-ordering can be problematic if restrictions are put on the relationship between that ordering and frame-dependent preferences. It also shows the kind of aggregation rules (such as the consensus rule) which might, in principle be applied. However, most of the standard restrictions which define the acceptability of a social choice rule have limited relevance for aggregation within an individual. Rather, the kind of restrictions which might naturally shape an aggregation rule come from extra-preference information.

5.5.1 Extra-Preference Information – Relaxing WEF

Figure 5.3 shows the wider data on individual welfare that might be to hand. As can be seen, there are several potential sources of information other than that obtainable from expressions of preference or choice behaviour. This information may either be used to exclude certain kinds of frames or to place relative weights on conflicting information gained from different frames.

Sympathy involves judgements of the kind 'if I was person A I would prefer x to y'. A higher order facility, usually known as extended sympathy (see Arrow, 1977 or Elster and Roemer, 1991 for instance), is the ability to make judgements of the kind 'it is better to be person A with x rather than person B with y'. The possibility that humans can make sensible judgements about the preferences of their fellow humans is unremarkable. Sympathy and its companion, empathy are features common to nearly all of us from at an early age. Both sympathy and extended

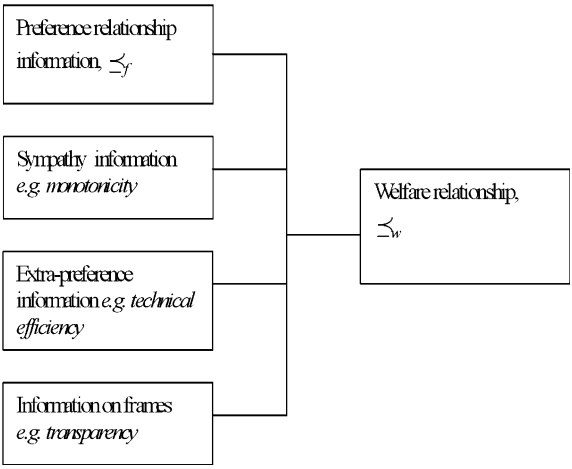


Fig. 5.3 Information and Aggregation

sympathy are facilities which provide information about welfare that need not come directly from the individual concerned. Neither need to be universal, in the sense that to be potentially useful information sources, it is not necessary that individual i be able to rank X accurately, for all other individuals j. Our ability to say much about a stranger's intertemporal discount rate, for instance, or about their preference between bananas and (non-edible) wildfowl may be limited, but it may still be possible to make statements about the preferences of other individuals which are generally true. I shall call this kind of information, sympathy information. One commonly employed candidate statement is that of monotonicity in consumption (with the usual caveats about excess attached).

There has to be some worry about the possibility of consistently accurate sympathy information. Suppose we are trying to judge whether A is better than B for person X. Expressions of preference by X are sensitive to frames and contradictory, so we fall back on sympathetic information from other agents. Ideally we could find some agents who are similar to X in important respects, but who do not share his or her frame dependency when it comes to A versus B. There is no guarantee that this is possible. Similar agents might have similar cognitive biases and use similar heuristics and if we found agents without such biases we might wonder if they were sufficiently similar to X.

The third item in Fig. 5.3 is the non-preference information that is relevant for welfare, such as the exact relationship between inputs and outputs that yields technical efficiency. I shall call this extra-preference information. In some areas of public sector appraisal, such as healthcare and transport, information on technical and cost efficiency may remove the need for preference information.

Finally, there may be frame information; that is, reasons for choosing the information provided in one frame over that provided by another. One frame may deceive individuals for instance, or provide stated preference information that clearly conflicts with extra-preference information.

5.5.2 Bayesian

As with conflicting preference information, the great problem with wider sources of information lies in their consistent aggregation. One route towards aggregate is provided by the Bayesian. For him or her there is some prior that option A is better than option B. Preference elicitation then gives some new information and from that the posterior is calculated. Write P(A) to mean the prior probability that A is better than B and write $P(A|i)$, $i = a, b$ to mean the posterior probability that A is better than B, conditional on the preference elicitation method sending the signal that i is the better of the two options, then Bayes' theorem states that,

$$P(A \mid i) = \frac{P(i \mid A)P(A)}{P(i \mid A)P(A) + P(i \mid B)P(B)} \tag{5.1}$$

The Bayesian approach can accommodate many of the previous approaches. From the 'no-ordering' viewpoint, the signals of preference are uninformative. In

other words, $P(i|A) = P(i|B) = \frac{1}{2}$ and so $P(A|i) = P(A)$. Nothing is therefore learnt from flawed preference elicitation methods. Within the optimal frames approach, there is one frame at least within which $P(b|A) = 0 = P(a|B)$, so that the signal is completely informative and prior information is unnecessary. And within the general information aggregation approach, the weights are provided by the reliability of the signals.

Although the Bayesian framework is conceptually neat, as with the other options it raises some important questions which might lead one to doubt its usefulness as a practical tool:

- What sort of information is embodied in the prior? At worse, prior information is uninformative – $P(A) = P(B) = \frac{1}{2}$. In this context, it is the reliability of the signal which matters. More typically, especially given the large numbers of preference elicitation exercises that have now been conducted, we might have prior information on whether A is best, either from other, similar studies or from the behaviour of the individual concerned.
- How are we to understand the conditional probability terms? $P(b|A)$ is the probability of receiving the signal that B is preferred to A given that A is actually welfare-preferred to B. Imagine that we have some sort of testing ground, where we can observe both welfare-preferences and signals for a given frame. Under such conditions $P(b|A)$ might still differ from zero for a number of good reasons. First, with merit wants, preferences may be systematically at variance with welfare, hence $P(b|A)$ may be strictly positive and could even be equal to one. Secondly, a consumer may be close to indifference, in which case small errors in understanding or in attention may lead to B being apparently preferred, even when on deeper reflection the individual would choose in accordance with their welfare-preference. Thirdly, the testing method itself might be imperfect, either because of the inadequacies of the researcher or through some irreducible, random element in the process of measurement.
- How can we rank frames? In practice, an entirely satisfactory testing ground is unlikely to be available. Moreover, the reliability of a particular signal may also depend on the frame. For typical frames f (or g) re-write the conditional probability of observing signal i in frame f given that J is welfare preferred as $P(i|J\&f)$. The information matrix for frame i is then,

$$I_f = \begin{bmatrix} P(a \mid A\&f) & P(b \mid A\&f) \\ P(a \mid B\&f) & P(b \mid B\&f) \end{bmatrix} \tag{5.2}$$

The informativeness of particular frames is a useful means of ranking them, but as there are some practical difficulties. First, some pairs of matrices need not be comparable according to the informativeness criterion, so that only a partial ordering is possible. Secondly, the informativeness of a frame is almost certainly conditional on the options being compared and hence the ranking of the frames cannot be made unconditionally. Last and by no means least, to compute I_f we need information on both welfare-preferences and on the signals of preference. But if we

have information on welfare-preferences that is reliable, then we do not need to go through an elicitation procedure.

Despite these drawback, particularly the last one, the information matrix might have some practical value as a means of ranking frames. Methodologically we could proceed as follows: first, find some sets of options where we are certain of the welfare orderings. Secondly, for each candidate frame and set of options compute I. For each set of options we then have an ordering of the informativeness of the frames. Since the ordering potentially differs across the different option sets we can construct an overall ranking system either by accepting its incompleteness (if f is more informative than g for one set of options, but less informative for another set) or by allowing indifference (but not necessarily transitivity of the indifference relationship) or by employing a weighting scheme.[12]

5.6 Paternalism

In this section I consider whether concepts from the paternalism literature can be useful in the context of the choice of optimal or acceptable frames. Since we know that framing affects choice then selecting which frames to allow when preferences are elicited for the purposes of public policy involves tackling many of the same dilemmas created when one considers direct interference in choice.[13] Paternalism, is usually justified on the basis of more fundamental arguments, such as consent in some form. We can borrow such arguments and apply them to the selection of frames and the weighting of information garnered from different frames.

5.6.1 The Justification of Paternalistic Intervention

Overriding individual preferences or choices has often criticised or justified on the grounds of paternalism, an encompassing term for theories of welfare which involve intervention by A in B's affairs to promote B's welfare. Weak paternalism is usually defined as applying to cases of 'lunacy, idiocy and childhood', to be contrasted with strong paternalism which involves intervention in the lives of rational adults. Paternalism has been defined in a variety of ways by modern writers, most of whom write against a backdrop provided by Mill and his views on individual liberty. Dworkin, for instance, defines paternalism as:

---

[12] We could summarise informativeness by using the distance between an information matrix and the identify matrix. Then the comparison between frames could be done on the basis of a weighted sum of the distances.

[13] Preference elicitation does not at first blush appear to be fall within any sensible definition of the scope of 'interference in actions or opportunity sets' which is at the heart of the definitions of paternalism advanced above. But that this initial view is wrong is most obviously seen when one considers the extended demand function x(p,m,f). If changing the budget set to influence behaviour may obviously arise from paternalistic motives, then so can changes in f.

> The interference with a person's liberty of action justified by reasons referring exclusively to the welfare, good, happiness, needs, interests or values of the persons being coerced. (1971, p. 108).

There are a number of problems with this definition which has been summarised by Kleinig (1984) as 'freedom subjugated in the name of benevolence'. For an economist at least, coercion is not a necessary feature of paternalistic behaviour. As we shall see later on this chapter, subsidies may be optimal from the viewpoint of the paternalistically minded government, and subsidies are not normally viewed as coercive. Moreover, many apparently paternalistic actions do not interfere with the liberty of the subject's actions. Information, advice and advertising are routinely used by governments to persuade individuals of the optimality of certain actions or the risk associated with undesirable choices such as smoking or driving dangerously, but these actions do not actually stop individuals from smoking or driving at high speeds. Thus, Kleinig suggests that it may be better to think of a paternalistic rationale rather than simply in terms of actions.

As Amartya Sen (1970), points out with his 'impossibility of a Paretian liberal paradox', for the confirmed welfarist, there are no ethical barriers to intervention in the affairs of others. What matters is only whether end state x yields higher or lower welfare compared to end-state y. From within the welfarist perspective, the principle source of reluctance to interfere in the choices made by other adults is therefore empirical (McClure, 1990): such a course of action may reduce rather than enhance welfare if the true welfare of the affected individual is unclear or if there are agency problems associated with the intervening actor. Consequently, for a welfarist, the main task is to define clearly those unfettered choices which are not welfare-maximizing and where it is apparent to some rational bystander what the optimal choice should be.

For a liberal the issue of intervention is thornier, because it involves a compromise of the principle of autonomy or autarchy, a central tenet of the liberal perspective. John Stuart Mill's classic book, On Liberty, remains the clearest view on the relationship between the individual and the state and has been the principle source of later arguments on the subject of both merit wants and paternalism from the liberal perspective. In that book he is unequivocal in opposition to the acceptability of intervention in someone else's choices when those choices do not directly affect others. In Chapter IV, 'Of the Limits to the Authority of Society over the Individual' he asserts that '... there is no room for entertaining any such questions when a person's conduct affects the interests of no person besides himself, or needs not affect them unless they like (all the persons concerned being of full age and the ordinary amongst of understanding). In all such cases there should be perfect freedom, legal and social, to do the action and stand the consequences.' p. 142. Shortly afterwards comes the key passage:

> But neither one person, nor any number of person, is warranted in saying to another person of ripe years that he shall not do with his life for his own benefit what he chooses to do with it. He is the person most interested in his own well-being: the personal attachment, can have in it is trifling compared with the that which he himself has; the interest which society has in him individually (except as to his conduct to others) is fractional and altogether

> indirect, while with respect to his own feelings and circumstances the most ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by anyone else. The interference of society to overrule his judgement and purposes in what only regards himself must be grounded on general presumptions which may be altogether wrong and, even if right, are as likely as not to be misapplied to individual cases, by persons no better acquainted with the circumstances of such cases than those who look at them merely from without. In this department, therefore, of human affairs, individuality has its proper field of action. (pp. 142–143).

Mill provides a more reflective summary of his position in the Principles of Political Economy and in particular in Book V Chapter XI 'On the Grounds and Limits of the Laissez-Faire or Non-Interference Principle'. This chapter is largely concerned with what we would now call merit wants. Externalities, income distribution and so forth are not covered except in passing. Mill stresses that he is not producing a general theory, instead offering four areas of 'some of the most indispensable and unanimously recognized of the duties of government' p. 941. He begins by making a sharp distinction between authoritative and advice-giving/subsidizing behaviour by governments, stressing that the former is only justifiable in the most extreme circumstances. Five objections to government intervention are offered: the damage done by incursion into individual freedom; the fact that power begets power, so that a government once enfranchised to do x cannot resist also doing y; diseconomies of scope; information asymmetries and agency problems.

> All the facilities which a government enjoys of access to information; all the means it possesses of remunerating, and therefore of commanding, the best available talent in the market – are not an equivalent for the one great disadvantage of an inferior interest in the result. (p. 947).

The final (and in his view most powerful objection) is that 'the business of life is an essential part of the practical education of a people' (p. 947).

Despite the strong line on individual sovereignty, Mill allows some important exceptions to his doctrine. The first, and most important of these, is that 'the uncultivated cannot be competent judges of cultivation' p. 953. He argues that individuals are probably best judges of material wants but not those things which create a better character. But beyond that there are basic elements of knowledge which should be inculcated in all citizens and for the achievement of which it is acceptable to impose restraints on parents. The quid-pro quo is that such education should be free or at most provided 'at a trifling expense' p. 954.

The second exception concerns 'lunacy, infancy and idiocy'. In these cases he recognises that there are parents, relatives and so on who could be expected to take a keen and caring interest in the welfare of their subjects, but there is still a role for the state: first to recognize the sovereignty of the care-takers and secondly to limit its abuse.

He is also reluctant to allow individuals to bind themselves irrevocably, arguing that the presumption that an individual knows best must be set aside when the enter into any perpetual contract. Selling oneself into slavery is therefore unacceptable.[14]

---

[14] The other cases are closer to modern notions of market failure.

The summary given above makes it clear that in common with many writers on the topic, Mill begins by accepting the incontrovertible nature of weak paternalism, then sets it aside and consider arguments which support strong paternalism. For Mill, freedom from intervention is primarily instrumental. Freedom provides the best prospects for the full and rounded development of a human being, a person who has 'autarchy' – the 'developed capacity for rational choice' as Kleinig puts it (p. 21). Conversely, only the individual who has attained this ideal of individuality merits full freedom from interference in their self-regarding decisions. As a consequence, it is not that interference in the life of another individual is wrong in itself, only that it is wrong if the individual has attained autarchy or, possibly, if intervention hinders the attainment of autarchy.

Intervention in an individual's choices is also acceptable when those choices are prompted by 'other-regarding' rather than 'self-regarding' behaviour. Thus there are two principle reasons for intervention, and later writers on the subject have concentrated on delineating them.

Kleinig, for instance, lists several supporting rationales for paternalism, some of which he views as more credible than others. The first argument, interconnectedness, is based on a denial of Mill's distinction between self-regarding and other-regarding acts. According to Mill, principles of liberty and non-intervention were only to involve the former – there is no reason to respect the individual's freedom to hit another person for instance, since that act is quite clearly other-regarding. The interconnected argument views the boundary between self and other as arbitrary. For instance A might read a book that B disapproves of because of its racy tone. Thus B is affected by A's action. Viewed in this way, there may be no limit to what constitutes an other-regarding action, hence there may be no particular reason to respect individual choices (at least not for the reasons proposed by Mill). One might therefore reject the interconnectedness argument in its most universal form, but Kleinig gives a number of specific examples of interconnectedness that might be applied even if the general argument is not accepted. For instance, others may be necessary for our survival, but this would suggest that we avoid self-harm simply to be available to help others. Alternatively other individuals might act as role models or partners, and so the well being of other human beings might be viewed as necessary for us to flourish as individuals. Overall, Kleinig is sceptical:

> In sum, the various Arguments for Interconnectedness have at best a limited validity, carrying some weight in contexts where an assignable duty to others can be established, but otherwise coming into contention only where self-regarding failure occurs on a large scale. It is not surprising that liberal writers have been able to acknowledge the importance of such interconnections while remaining generally suspicious of paternalism. (p. 43).

Further arguments in favour of the principle of paternalistic interference considered by Kleinig include the idea that actions taken by an individual today may have consequences for that person in the future. If the contemporary self and the future clash in their interests then actions taken by the current self cannot be purely self-regarding. Alternatively, the actions taken by the current self, such as use of an addictive dry may limit the ability of future selves to achieve autarchy. Either way, interference in the actions of the current self may be justified in the name of future selves.

As noted above in the discussion of Mill, another class of arguments in favour of some paternalism stems from the view that freedom is fundamentally an instrumental value, freedom is an instrument to help us achieve some good, such as well-being. If some other instrument is superior, such as paternalistic intervention, then so the argument goes, why not choose it. Other arguments, such as for the ban on personal slavery, are based on instances where restrictions on certain kinds of behaviour can be seen as freedom enhancing. Possibly, though, the class of reasons most widely offered in defence of paternalism are those arguments based on some notion of consent and it is these I shall discuss in slightly greater detail.

We can, at least in principle, identify five sorts of consent:

1. Prior or ex ante consent. In the case of the professional athlete who is tempted by drink, removing alcohol from the house may be an acceptable infringement of their liberty, because he or she might know that they suffer from a weakness of will and give specific permission – either to specific friends or more generally – for others to stop them drinking (see also Marschak, 1978).
2. Real will. Saving an individual from a collapsing bridge in circumstances where there is no time to obtain explicit consent is an example where the individual at danger is presumably making a choice which is at variance with any fundamental plans for their life they might have made. These fundamental plans represent their 'real will' in a way that the decision to walk onto the unsafe bridge does not. As Mill puts it, 'liberty consists in doing what one desires' and the individual does not usually desire to fall into the river. The notion of real will is therefore that even when someone chooses x rather than y, there may be a core self that wishes otherwise. Arguments for real will rest on following the preferences of this other self. Obviously the difficulty of the real will argument lies in elucidating the preferences of the core self, when only the day to day decisions of the individual are observed and the individual makes no explicit statement that there is a self which objects to the choices he or she makes and wishes them otherwise.
3. Anticipated consent. In the case of the collapsing bridge, intervention might also be justified by the notion that were it possible to slow time down, ask the endangered individual their preferences they would be quite clear that they wished to live. This is anticipated consent. It differs from the cases of actual consent (prior or subsequent) in that there is no assumption that consent is actually obtained. As such it shares with the real will argument (to which it appears very similar) the weakness that the intervener is making choices on behalf of the subject when there is no explicit statement by the subject that he or she, at least in part, shares the preferences implicit in the choices made on their behalf. As such it provides ample grounds for moral hazard by the intervener.
4. Subsequent or Ex-post Consent differs from the previous two cases; it rests on some explicit consent, albeit after the event. An example is provided by the child forced to take violin lessons who in later life thanks the parents who kept them to the task. Though this makes it more acceptable than the previous two cases, it sounds like an argument for the acceptance of brain washing. Consequently Carter (1977), delineates three situations in which she argues that subsequent

consent is not an acceptable argument for intervention: brainwashing, distortion of the subject's values and manipulation of the information set.

5. The final form of consent is hypothetical rational consent. Here justification is by recourse to the notion of a (strongly) rational alter ego, a more enlightened and reasonable version of the subject of intervention, someone who is able to weigh up the pros and cons of a particular choice.

As an example of the kind of choices which might be the subject of hypothetical rational consent, Dworkin (1971), gives the following example:

> Goods such as health which an y person would want to have in order to pursue his own good - no matter how that good is conceived . . . [T]he attainment of such goods should be promoted even when not recognised as such, at the moment, by the individuals concerned. (1971, pp. 120–121).

It is not immediately obvious what is meant by rational in this case, except that it goes beyond weak rationality as defined in Chapter 1. Some external standard is apparently called for and it is not clear who sets that standard.

A final line of reasoning is the argument from personal integrity, which Kleinig sees as the most acceptable justification for paternalism. It begins by making a distinction between those aims in life which are central and those which are peripheral. Where actions associated with the periphery threaten long term central aims then intervention is justified.

> Where our conduct or choices place our more permanent, stable, and central projects in jeopardy, and where what comes to expression in this conduct or these choices manifests aspects of our personality that do not rank highly in our constellation of desires, dispositions, etc. benevolent interference will constitute no violation of integrity. Indeed it helps to preserve it. (Kleinig, 1984, p. 68).

For instance, for the alchohol-tempted athlete, intervention to remove drink from the house would be best done by a close friend or loving relative, on the grounds that the distinction between central and peripheral goals is likely to be a highly personal one, and therefore only known to intimate acquaintances.

Kleinig lists four major objections to paternalism, which to a degree overlap with Mill's discussion of the subject. First and foremost is the view that paternalistic actions are freedom-reducing, which is undesirable because freedom is the best means available for an individual to achieve their own goals. Secondly, freedom-reduction is wrong because autarchy is an implication of the categorical imperative which implies the equal standing of rational choosers. It would contradict this principle if one person, A, was allowed to interfere with the choices of person B, without the reciprocal right being granted to B. If therefore, we accept the categorical imperative and also do not want others to interfere in our choices, then we must accept non-interference in their choices. Next, we have the Arguments from Paternalistic Distance – because humans are diverse, this means that we cannot know one another's tastes with any accuracy and so we are not competent to interfere in their affairs. The final objection relates to the developmental value of choice – the notion that it is only through making our own decisions that we understand our preferences and develop as individuals. Each of these objections clearly has some weight: it is

notable that even proponents of paternalism are reluctant to allow wholesale and unregulated intervention by one person in the choices made by another, but none of them is decisive. The first objection rests on the assumption that unrestricted freedom is the best means for an individual to achieve their own goals, but if some external restrictions support goal achievement then paternalistic intervention is potentially justifiable. Similarly the second objection takes as its starting point a world of rational choosers, in the absence of which it loses much of its force. The third objection can be deflected when there are individuals with insight into our true interests and when there are incentives to put those interests first. Finally, when choice is not developmental, then there is one less argument in its favour.

Putting together arguments from consent and integrity we are in a position to create a tentative hierarchy of preference information with regards to individual welfare (see Fig. 5.4). Respecting preference information means avoiding interference in choice. Hence arguments for respecting preference are stronger when arguments for paternalistic control are weaker. As the figure illustrates, the information we have on preferences will vary in its quality. An individual who states she prefers x to y, but has no experience of either and will not have to live with the consequences of her decision, has a lower ability to provide accurate information on preference as well as a weaker incentive, compared to someone who knows both x and y and is making a choice that she knows will lead to either x or y. As a result, greater

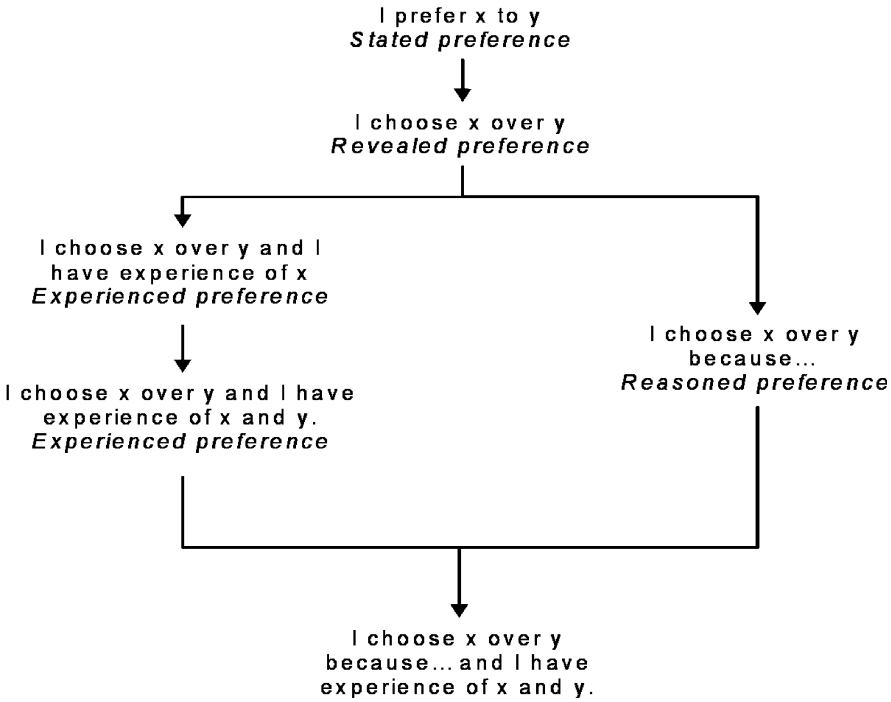


Fig. 5.4 A Hierarchy of Preference Information

experience and incentive leads to consent to having x that is more clear cut. Real choices also provide more developmental opportunities. Meanwhile, according to arguments from integrity, reasoned preference and choice is to be accorded greater weight than unjustified choices because it provides evidence that decisions are at least in part in tune with notions of long term interest.[15]

5.6.2 Paternalism: The Contractarian View

Unlike liberalism, where paternalism is at the heart of any analysis of the role of the state, discussion of the acceptability of paternalistic action has been less common within the contractarian tradition of welfare. The most significant contribution is that of John Rawls who, when arguing about individuals behind his famous veil of ignorance states:

> they will want to insure themselves against the possibility that their powers are undeveloped and they cannot rationally advance their interest, as in the case of children or that through some misfortune or accident they are unable to make decisions for their good, as in the case of those seriously injured or mentally disturbed. It is also rational for them to protect themselves against their own irrational inclinations by consenting to a scheme of penalties that may give them a sufficient motive to avoid foolish actions and by accepting certain impositions designed to undo the unfortunate consequences of the imprudent behaviour. (1971, p. 249).

The first two instances mentioned, childhood and mental disturbance, correspond closely to the concept of weak paternalism. The second half of the quotation is concerned with strong paternalism. In Rawl's formulation, paternalistic interference is justified by prior consent, but it is a special prior consent by a rational agent embodied with clear foresight about their own possible frailties. In that sense, Rawls' argument is much closer to that of hypothetical rational consent, which is how Kleinig categorises it, but it has the advantage over Dworkin's argument in that we can see who sets the standard of what is and is not rational.

For Rawls, there are four factors which guide paternalistic interference: first and foremost, 'intervention must be justified by the evident failure of absence of reason and will' p. 250, but then given this fundamental the form of intervention should be,

> Guided by the principles of justice and what is known about the subject's more permanent aims and preferences, or by the account of primary goods. (p. 250).

These three elements form a hierarchy – justice first, then preferences then primary goods. In the case of preferences, 'as we know less and less about a person, we act for him as we would act for ourselves from the standpoint of the original position' (Rawls, 1971, p. 249). Clearly this may be a difficult task: if we suffer from some deficiencies of reason then we may lack the ability to place ourselves in the standpoint of the original position. On the other hand, in the original position,

[15] Of course we need to be somewhat sceptical of reasoned choice: if the reasons are ex-post they could simply be serving the current self.

there must be limits on the ability of the individual to see deeply enough into the hearts of all possible selves, in such a manner as to be able to write down a full plan of contingencies for all cases where reason fails. In other words, the veil of ignorance must obscure more than just identity, otherwise it would not be necessary to delegate decisions to actual selves.

5.6.3 Summing Up

Central to most arguments for strong paternalism is integrity: the notion of a core self which persists and which has some preferences about the future development of the life in question. This core self is potentially in conflict with the day to day selves which may be foolish, impulsive, myopic or easily confused. Because two or more selves are engaged in choices, not all actions taken by a contemporary day self can be viewed as purely self-regarding – they may affect the well-being of future selves. Consequently, Mill's first principle is breached. The argument then gives some priority to the views of the planning self and therefore grants some rights to other individuals and possibly the state, to intervene in the actions of the day-to-day self.

Most arguments summarised are not universal; their authors recognise that the case for intervention may be incremental and that there are costs to restricting individual freedom. I have outlined some of the costs above, but it is worth considering other important factors which raise the case for intervention. One possible such factor is the risks involved in the decision. For instance, it may be more acceptable to interfere in the decision of a person facing choices involving possible death, compared to a situation where only a few pence either way are involved. Certainly this is Kleinig's claim. Yet, McLean et al. (2000), in particular is highly critical of risk as the defining or even an important factor. He argues that if the individual has less freedom to choose with more risky decisions, then this amounts to an outcome-related restriction, meaning that they can make any decision as long as it is the correct one, as viewed by the principal, for instance the doctor.

Instead MacLean favours a metric based on complexity because higher complexity means a higher possibility of cognitive error (and this he sees as providing legitimate grounds for intervention). For instance suppose I am risk neutral and need to compare two lotteries, A and B. A is £10 for sure; B is a 71% chance of £14 and a 29% chance of nothing. Part of my assessment might involve comparing the expected value of the two lotteries. Now consider two other lotteries, A′ and B′. A′ is £1,000 for sure; B′ involves a 71% chance of winning £1,400, and a 29% chance of nothing. The risk involved with lottery B′ is larger than the risk with B. According to the risk view, therefore, my competence to deal with the decision between A′ and B′ is lower than my competence for the A versus B choice. According to the complexity view the comparisons have roughly equal complexity and therefore there is no case for intervention in the latter which does not also apply to the former.

From an economic viewpoint the case against a risk-based criterion has some strength: as the size of the wagers increases, an individual who economises on

cognitive resources would be expected to put greater effort into making the right choice. Thus, all other things being equal, the rate of error should be smaller with larger payoffs. But the rate of error is only one factor in the equation. If there are more or less fixed costs of intervention in another agent's decision, then only cases involving large payoff are likely to merit intervention. Moreover, the case for complexity can produce some troubling examples: consider a relative who through some eye defect consistently believes that objects are 10 inches to the right of where they are. We may hesitate to intervene in their putting on the golf course; on the other hand we might be very reluctant to let them drive a car along crowded or narrow streets. In the two instances the complexity of the decision (where to drive, where to put) might well be the same; the risk though is very different.

The crucial point here is the existence of an unbiased or more perfect viewpoint. If the bystander or expert has no information advantage over the subject, then increasing the risk of a choice does not provide an argument for greater intervention in that choice. If though there is a pre-existing bias, in the sense that the observer has a superior and different perspective to the subject, then, to the extent that this bias is not removed by raising the stakes, the case for intervention is higher with higher risk.

5.6.4 Ranking Frames Using Consent and Other Arguments from Paternalism

The strongest cases for the acceptability of intervention in the choices of others concerned consent and other-regarding behaviour. In the first of these there were several different types of consent arguments, of greater or lesser validity, including prior consent, ex post consent and hypothetical rational consent. In the other-regarding case, the argument for intervention was made on the grounds that the decision of the self today may conflict with the views of the self tomorrow and there may be grounds for limiting the power of today's self to constrain the choices of the future self, in the same way that other kinds of externalities can yield possibilities for Pareto-improving modification of the choice set.

To illustrate what bite, if any, the arguments have, let us consider a class of framing effects commonly encountered: reference dependent preferences. This class was discussed extensively in Chapter 2, where we encountered the notion of a referendence dependent preference system (RDPS).

For the case of reference dependent preferences, ex ante or prior consent is perhaps the easiest of the arguments to apply. It says that in comparing two or more options on a choice set X the status quo perspective carries all the weight. There is a problem though in establishing what is the correct reference point. It might be current consumption, the current endowment, some weighted average of past and current consumption or some other simple point of reference. Out of all candidates, given loss aversion, perhaps the strongest case can be made for the principle that for y to be ranked no lower than x, then y must be preferred to x when x is the

reference point. There are two reasons why this has some claim. First, as a general rule, individuals know more about their current consumption than about alternative options, so if despite this superior information, an individual is still willing to surrender the current bundle in favour of another, this suggests some strong reasons for doing so. Secondly, the most successful models of reference dependent preferences (i.e. Tversky and Kahneman, 1991; Munro and Sugden, 1997, 2003) imply that as the reference moves towards x, it is increasingly favoured over other alternatives. In other words, if x is not preferred to y when x is the reference point, then y will be preferred to x from all other reference points. So, taking x as the crucial reference point will provide a reliable guide for rankings involving x, in the sense that if x is not preferred in a pairwise comparison from x, it is not likely to be preferred from any other point that we could take to be the status quo.

These arguments are not irresistible. The fact that x is or is not preferred from x may have little resonance when an individual is currently at r and he or she is being asked to compare x and y. The main aim of the following paragraphs is therefore to explore the implications for the aggregation of information, conditional on one particular view of the concept of consent, in order to illustrate how links can be constructed between philosophical perspective and economic decision-making.

Let us use the label $\preceq_{ri}$ to mean preferences for person i when the frame is defined by reference point r (or s, x or y). For prior or ex ante consent, the basic principle is stated as:

Definition 5.3 Principle of Ex Ante Consent (EAC): If $[x \prec_{xi} y]$ then $\neg[y \preceq_{wi} x]$.

It is possible to strengthen or indeed weaken this principle. If we had $[x \preceq_{xi} y] \leftrightarrow [x \preceq_{wi} y]$, for instance, then this would make welfare synonymous with the ranking as viewed from x. One implication of this alternative formulation is that $\preceq_w$ would be complete, since each reference point yields an ordering. However, the strict relationship might have some awkward properties. To be more specific, if $[y \prec_{xi} x]$ and $\left[x \prec_{yi} y\right]$, then both $[x \prec_{wi} y]$ and $[y \prec_{wi} x]$! The stated formulation for the principle does not run afoul of this problem; it also allows incompleteness which may be considered a virtue when information about the relative ranking of x and y conflicts. It also allows other information, from other reference points to affect the welfare ranking of x and y.

With ex-post consent the acceptability of any outcome must be judged from that outcome. Suppose that, as viewed from y, x is superior to y and both x and y are available, then a principle that chose y from the available set would break with the notion of ex-post consent, since the individual when choosing from the perspective of y would not choose y over x. As with prior consent we can have different degrees of strictness, but for the purposes of this exercise, I shall stick to:

Definition 5.4 Principle of Ex Post Consent (EPC). If $\left[x \prec_{yi} y\right]$ then $\neg[y \preceq_{wi} x]$.

We have outlined some principles by which the information obtained under different frames might be aggregated. What has not yet been demonstrated is that it is possible to use such principles to form a coherent welfare ranking of some sort. First, it is useful to quote the following result concerning the consensus rule discussed earlier:

Result 5.1 *CON satisfies completeness, reflexivity, quasi-transitivity,*[16] *P and D but not necessarily transitivity.*

Proof See Sen (1970).

An example shows the problem with transitivity. In all the examples which follow the reference point is shown in the first row. Alternatives are then shown in descending order within the columns. Indifference is indicated when two alternatives are placed in the same cell; otherwise all preferences are strict, except where otherwise indicated. For instance in the first example, as viewed from x, y is strictly preferred to x.

Example 5.4

x	y	z
z	y	x
x	z	x
y	x	y

Here, in some cases x is ranked above y and sometimes the converse so we have x as welfare indifferent to y. Similarly we have y is welfare indifferent to z, but z is always ranked above x, so we do not have x welfare indifferent to z.

It is worth noting that this conclusion is not modified if we impose some restrictions on the form of reference dependent preferences. For instance, the example is compatible with the stylised facts of endowment effects discussed in the previous chapters. So, restricting the domain of RDPS in an empirically reasonable way provides little extra in terms of obtaining transitive welfare rankings.

However, invoking some of the stylised facts of reference dependent preferences does have some value: with standard models of social choice, relaxing the assumption of universal domain by, for instance, supposing single peaked preferences, allows possibility theorems to be true which would fail if there was no structure to preferences. The same is true here.

Based on the evidence presented in Chapter 2, consider the following two regularities as candidates for stylised facts. The first states that a move towards y always favours y; it is a familiar feature of laboratory and field tests of preference theories.

Definition 5.5 Kahneman-Tversky (KT): $\forall x, y, r, (x \preceq_{ri} y \rightarrow x \preceq_{yi} y)\&(x \prec_{ri} y \rightarrow x \prec_{yi} y)$

The second regularity is much stronger; I shall say more about it a little further on.

Definition 5.6 No Reverses (NR):$\forall x, y, r, \leftarrow \neg(x \prec_{ri} y \ \& \ y \prec_{ri} x)$

The next Result 5.2 demonstrates the fact that a range of aggregation rules may satisfy a desirable principle such as EAC.

[16] Quasi-transitivity is the property that the strict relationship, $\prec$, is transitive, but not necessarily the weak relationship.

Result 5.2 *If the RDPS satisfies KT then CON satisfies EAC.*

Proof Suppose CON but not EAC then for some x, y, $x \prec_{xi} y$ but also $y \preceq_{wi} x$. By CON this implies $\neg(\forall s, x \preceq_{si} y \ \& \ \exists\, s : x \prec_{si} y)$. Since $x \prec_{xi} y$, then it must be the case that $\exists s : y \prec_{si} x$, but by KT, if x$\prec_{ri} y$, then $\forall s, x \prec_{si} y$. So we have a contradiction and therefore EAC is satisfied.

Two further examples illustrate the issues. They show why we need the behavioural principle embodied in KT in order to obtain the result and why the same result is not available for EPC.

Example 5.5

$$\frac{\text{x y}}{\begin{array}{c}\text{x y}\\ \text{y x}\end{array}}$$

In this example CON implies that $x \sim_{wi} y$, but EPC would imply that x and y are non-comparable within w. Hence EPC is not satisfied by CON even though, as in the example, KT may be satisfied. Simply, KT has no bite with EPC, because it makes no predictions about preferences for other reference points when all that we have is conditional information that some bundle is preferred from itself to another bundle.

In the second example, EAC is not satisfied:

Example 5.6

$$\frac{\text{x y}}{\begin{array}{c}\text{y x}\\ \text{x y}\end{array}}$$

According to CON, $x \sim_{wi} y$, but EAC implies that x is not weakly welfare-preferred to y and y is not weakly welfare-preferred to x, so again x and y are non-comparable.

However, it is clear that in this last example, preferences do not exhibit any kind of status quo bias. Imposing the behavioural assumption, KT rules out preferences of the kind depicted in the second example and allows CON to be compatible with EAC.

Obviously both EAC and EPC are candidate principles. Ideally we might wish both to be satisfied. At the same time we might want our welfare ranking to be complete and reflexive and transitive. We might therefore wonder what it would take to have a complete and reflexive individual welfare ranking that obeys both EAC and EPC principles. The following result provides a necessary condition.

Result 5.3 *Suppose (i) $\preceq_{wi}$ satisfies C and EPC and the RDPS satisfies KT, then the RDPS also satisfies NR.(ii) $\preceq_{wi}$ is a quasi-ordering and satisfies EAC and EPC and the RDPS satisfies KT, then also $\preceq_{wi}$ is transitive.*

Proof (i) First, completeness. Suppose not NR, then $\exists x, y, r, s$, s.t.$x \prec_{ri} y \ \& \ y \prec_{si} x$. By KT we therefore also have $x \prec_{yi} y \ \& \ y \prec_{xi} x$. But then by EPC we must

have $\neg x \preceq_{wi} y$ and $\neg y \preceq_{wi} x$ which is not possible given the assumption that $\preceq_{wi}$ is complete. (ii) Note first that NR is not required for to satisfy R and that part (i) shows how NR is required for completeness. Consider any x, y and z s.t. $x \preceq_{wi} y$ and $y \preceq_{wi} z$. By C and EAC we must have $x \preceq_{yi} y$ and $y \preceq_{zi} z$, while EPC and C together imply that, $x \preceq_{xi} y$ and $y \preceq_{yi} z$. Using the transitivity of the frame dependent preferences, together with KT yields the following table of relationships (where now the rankings represent weak orderings):

x	y	z
z	z	z
y	y	y
x	x	x

Hence $x \preceq_{wi} z$.

To sum up: if the RDPS satisfies KT and any welfare ranking must satisfy EAC and EPC, then in order for such a ranking to be complete, the RDPS must in turn be NR. If this is satisfied, then any resulting welfare ranking is also transitive.

It is worth emphasising that NR is a very strong assumption. Hicksian preferences satisfy NR, but the only non-Hicksian reference dependent preferences that are compatible with NR are those where (for instance) y is strictly preferred to x from x, but viewed from y, x and y are indifferent. What this means is that if we require a welfare ordering then there is very little leeway in the sense that almost any regularly observed deviation from the Hicksian precepts implies the non-existence of a welfare ordering that satisfies EPC and EAC. Of these two principles, it seems to be EPC which causes the most trouble for the construction of an ordering; this is largely because KT places extensive restrictions on the set of possible frame dependent preferences. These restrictions mean that if a option y is preferred to x from x, then there will not be contradictory evidence on the ranking of x and y from other reference points. No such general implication can be drawn when y is viewed as superior to x when y is the reference point. Hence KT restricts the chances of contradictory evidence on welfare when EAC is the welfare principle.

5.6.5 Hypothetical Rational Consent

Let us now turn briefly to hypothetical rational consent (HRC) which is a concept fundamentally different to both EAC and EPC in that, quite obviously, it is both hypothetical and rational. Thus it will lead to an ordering of options, but that ordering might not receive the acquiescence either of the 'before' self or of the 'after' self. As we saw above, to be acceptable as a principle for judging options, HRC has to have some grounding in the evidence available about an individual's welfare. We can distinguish two kinds of situation: a general case, for instance when the quality of a public good must be decided and an individual-specific case, for instance about whether a senile pensioner should go into a nursing home or stay with the rest of their family. The situations are distinguished by the amount of information

we have about the individual's preferences. In the general case, that information might be fairly limited. We might argue for instance that any ordering derived from HRC should satisfy monotonicity. Moreover, if an individual prefers x to y from the perspective of both y and x, we might be more likely to view x as preferred to y by the individual's hypothetical, rational counterpart. So the general case looks rather similar to some of the instances encountered above. With the specific case, more information might be to hand, either because explicit consent has been previously given by the individual or through the intimate knowledge of someone close to them.

An alternative to grounding any HRC-derived ranking in direct evidence about individual preferences might come from viewing frames as the objects of consent rather than choices.

Consider the following problem. In the future a decision will be made, meaning that one element will be chosen from a, possibly unknown, choice set. A rational individual has to agree the frame which will be used to elicit preferences which will then be invoked to make the decision. Given a set of possible frames, F, how should the individual select a sub-set of frames acceptable for preference elicitation? If the individual knows their true preferences and knows how framing effects would effect the process of their elicitation, then the problem is straightforward. He or she chooses the frame which provides the outcome closest to their clear-sighted perspective. It is harder to say what their choice should be if they do not know their future preferences exactly. The individual might reject frames which encourage the choice of dominated options. Should the frame be truthful in the information it provides? Not necessarily if the individual is non-Bayesian in their ability to process data. Should the choice of frame be such as to favour the status quo? Only if we have some reason to believe that the status quo would provide the best platform from which to survey all other options. Again therefore, it seems unlikely that a definitive choice of frame can be made from within the HRC perspective. Some frames might be ruled unacceptable, but an ordering does not appear practicable.

5.6.6 Contractarian Consent

In its basic form the notion of consent proposed by Rawls shares many of the problems identified with the related concept of hypothetical rational consent. Suppose for instance, an individual lies behind the veil of ignorance knowing that his individual welfare will vary according to the state of world. He does not know which state will obtain, but he knows that in a certain state of the world s, y will be better for him than x. Meanwhile, in frame 1, x will appear worse than y while in frame 2, x will appear more attractive. Thus the individual will favour a social contract in which, in state s, policy should override preference when the frame of reference is 2.

The difficulty with this neat solution is that it presumes too much. The veil of ignorance is full of holes, sufficiently transparent for individuals to see through it to the true preference of their future selves. An alternative degree of ignorance would suppose that the prior self does not know their welfare ranking in state s. They may have some information but not everything.

In this situation, the contractarians may wish to agree on acceptable frames rather than acceptable consumption sets.[17] More acceptable frames are likely to those where as argued above, subjects have greater opportunities for accurate learning or greater incentives to make an informed choice. Let us illustrate the point through two examples which follow on from the learning models of Chapter 3.

Example 5.7 Suppose the individual is unsure of her preference between x and y in state s. She knows that the utility difference between the two is ΔU. With equal probability ΔU is either positive or negative, but with cost c she will be able to find out its true sign. Two possible frames will be available for eliciting preferences. In frame 1 ('stated preference'), an expression of preference for one option over the other brings about that option with probability p (where $1/2 < p < 1$). In frame 2 ('choice') meanwhile, an expression of preference brings about the that option with probability 1. The net gain from becoming informed in frame 1 is $-c + p\,|\Delta U|$, whereas in frame 2 the corresponding expression is $-c+|\Delta U|$. Obviously therefore the choice frame dominates the stated preference frame, because for no values of c or ΔU does it produce a worse outcome. Behind the veil of ignorance, therefore, the potential citizen will favour frame 2.

This example therefore provides an argument for markets over state provision of goods. In the second example we see that it is not just about market and non-market provision: the initial endowment might alter opportunities for learning. The contractarian may therefore prefer one kind of endowment over another.

Example 5.8 As above there are two frames, which we can think of as representing two possible endowments: with x or with y. There is no uncertainty over the value of y, which in a slight abuse of notation is always y, but the value of x can be x^1 (probability p) or x^0. We suppose that $x^1 > y > x^0$. and write the expected value of x viewed from behind the veil of ignorance as E(x). Consumption takes place twice and in the first period the consumer must consume their endowment. Subsequently however, they may switch. Learning comes only from experience. An individual who is endowed with one option learns the true value of that option, but nothing about the other option. Thus, an individual endowed with y, will only switch to x if E(x) > y, while a consumer endowed with x will switch to y after one period if the true value of x is x^0. Because of this, if E(x) > y, then endowing the consumer with x is always the optimal decision. On the other hand, if E(x) < y, it is still optimal to endow with x provided that $E(x) - y > \frac{1}{2}(1-p)(x^0 - y)$ (see Fig. 5.5). In other words, for a given value of E(x)-y, the higher the uncertainty over the value of x, then the greater the advantage of endowing the consumer with x.

Judging frames according to their capacity to incentivize choice or induce learning does not mean that one frame or type of frame is best for all binary rankings.

[17] There may still be arguments for restricting consumption sets on the basis of issues of risk and justice. Moreover, following Rawl's line, there may also be a case for delegating decision-making to a third party who has some good understanding of the subject's long term interests.

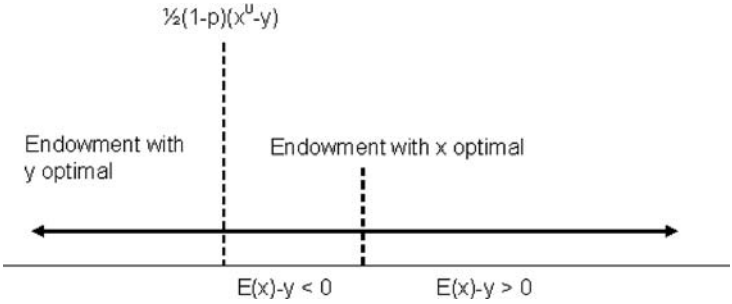


Fig. 5.5 Optimal Endowment with One-Sided Learning

In other words it does not imply that there is an optimal frame. In the last example endowment with x was more likely to be optimal when the value of x was uncertain and its true value could only be learnt through experience. But when the value of a good can be fully understood prior to purchase, having a choice with a no status quo option might be superior if it avoids the creation of an endowment effect that would limit market research.

5.7 Conclusions

This chapter has provided a tour around some of the arguments which could be invoked in the face of evidence on incomplete or inconsistent preferences. Several themes have emerged in the course of it. The first and central theme of the chapter has been that of information: once we view preference elicitation methods as providing potentially useful, but not wholly reliable information, then the key issue is how to place weights on different kinds of information from different kinds of sources.

Refusing to rank alternatives was seen as an unattractive option and one that is likely to be incompatible with the requirement to make social choices. Accepting that welfare is frame dependent also had some unsatisfactory consequences, notably policy instability and the underlying need to define a ranking of frames for welfare purposes.

The optimal frame approach was also shown to have its limitations. Choosing a single frame as the context for all welfare comparisons means losing potentially valuable information from other sources and frames. It is also by no means clear that one frame or one type of frame even, can be singled out as universally superior to all others.

The problem with using information from multiple frames or from mixing preference information with non-preference information lies in aggregation. One way to rank sources of preference information is to identify their capacity to incentivise choice and to guide optimal learning. If patterns of framing effects are simple (as in endowment effects), the set of possible relationships between frames and

preferences can be sufficiently straightforward such that transitive individual welfare rankings can be established for reasonable aggregation rules.

Although this has been a long chapter, it has really only scratched the surface of the problem of welfare economics with bounded rationality . There still remains a large amount of research to do, much of it beyond the scope of this book, particularly on the issue of how to aggregate conflicting preference information. However, some issues will be taken up in subsequent chapters. In the next chapter for instance, I consider the general issue of public policy in the context of bounded rationality, while the last chapter of the book focuses on non-market valuation.

Chapter 6
Public Policy and Bounded Rationality

6.1 Introduction

This chapter examines the broad implications of anomalies for the economics of public policy. A central aim is to set out a general framework for the kinds of models which might be usefully employed to explore public policy issues. Parts of the chapter provide specific illustrations of these different types of models; later chapters provide other instances. A second goal is to relate the existing theory of market failure to bounded rationality. The main argument here is that market failure concepts are typically defined with respect to the technology of production rather than by consumer preferences and hence they survive transition to a world of bounded rationality, provided that the economy is sufficiently smooth in the sense that behaviour and accompanying notions of welfare are continuous in some measure of bounded rationality. A third aim is to reflect on the relationship between bounded rationality and the optimal boundaries of the state – does bounded rationality make the case for state organisation of economic relationships stronger or weaker? My argument here is that the issue is not clear-cut. Markets provide learning opportunities for citizens. Moreover, information on what is right may come from citizens (see the contingent valuation literature surveyed in Chapter 11), so as individuals become less rational, the information available to the government on welfare may become less and less reliable.

6.1.1 Bounded Rationality: Identification and Valued Added

What is added to the public policy toolbook by the addition of the bounded rationality concept?

As we saw in Chapter 1, models of individual behaviour are essentially theories of neutrality and non-neutrality. In the traditional Hicksian model of the consumer, market behaviour is summarised by the demand function. In this function $x(p, m, f)$, the demand vector x is affected only by changes in the vector of prices, p, and income m. Moreover, it is homogeneous of degree zero in prices and income taken together. The Hicksian model also has the property that x is unaffected by changes in the relevant frames, f. With bounded rationality, for any given frame,

A. Munro, *Bounded Rationality and Public Policy*, The Economics of Non-Market Goods and Resources 12, DOI 10.1007/978-1-4020-9473-6_6,

© Springer Science+Business Media B.V. 2009

demand functions are still likely to be affected by p and m in the standard Hicksian manner. On the other hand, it is usually the case that a change in f will either affect x or affect the response of x to changes in prices and income. For instance, if f is the reference point, then as we have seen in Chapter 2 a change in the reference point can affect demand or it can affect the response of demand in the face of changes in prices.

So, having some understanding of the effects of frames on behaviour adds greater determinacy to the model of individual behaviour and adds new potential tools to the public policy toolkit.

On the other hand, the value of models of bounded rationality as an aid to understanding of social decisions is less clear, precisely because they are social decisions, made as the result of the aggregation of preferences of many individuals. Arrow's theorem is of fundamental relevance here. Since it states that individual rationality does not generally imply that society as a whole make its decisions in a rationalisable fashion, then it follows that apparently irrational social decisions may not be due to boundedly rational individual decision-making.

For instance, consider the issue of sunk costs. Casual empiricism suggests that, in public policy, sunk costs are frequently not sunk. Major public works are rarely abandoned half finished, even if it is abundantly clear that the net benefit of completion is negative. In the UK, the Millenium Dome, initiated by a Conservative government, completed then abandoned by a Labour government, provides a recent and expensive example. The mental accounting literature discussed in Chapter 2 could be employed to provide an explanation of the Dome and other political fiascos, but there may be other reasons which are entirely rational, particularly when there is an element of asymmetric information between the government and citizenry. Cancelling a project may send a signal that the government is incompetent because it overestimated the returns from the project. Or other groups may have invested sunk costs in anticipation of the successful completion of the project. Householders for instance, may have installed double glazing in anticipation of an airport expansion.

Or consider the issue of flypaper effects – where the elasticity of local government expenditure differs according to the source of the increase in income. So, for instance, a one percent rise in per capita lump-sum grants from central government typically leads to a higher rise in local government expenditure than is the case with a one percent rise in locally generated income per capita. In other words, some sources of increased local government income 'stick' more readily than others. Often flypaper effects are identified as an anomaly in government behaviour [Hines and Thaler 1995], explicable in similar terms to endowment effects and status quo bias. However, there are alternative and potentially convincing explanations which do not invoke bounded rationality. For instance, asymmetries of information between citizens and political actors are one possible source of the difference. Or flypaper effects may sensibly arise out of the fact that local governments are engaged in a repeated game with central government. In such a context, grants from the centre may not be formally tagged as 'additional' or 'matching', but significant reductions in local contributions to expenditure following a rise in grants may lead to subsequent punishment.

Table 6.1 Political Economy Models and Bounded Rationality

			Rulers		
			Rational	Boundedly rational	
				Aware	Unaware
	Rational		Standard political economy/normative	–	Strategic delegation
Citizens	Boundedly rational	Aware	Delegation	Self-control	Strategic delegation
		Unaware	Merit wants	One-eyed ruler	Myopia

The point is that it is hard, if not impossible to find government policies that cannot be explained without invoking bounded rationality. Often unverifiable asymmetric information does the trick, but the social aggregation of preferences can always provide a handy alternative. As a consequence, the role of models of bounded rationality in explaining the actions of government is likely to be most important in those cases where market or non-market individual behaviour is at variance with the predictions of standard theory.

Table 6.1 sets out a number of possible public policy models involving bounded and unbounded rationality. It divides individuals into two groups: ruled and rulers. Rulers are politicians and their agents – the individuals used called bureaucrats in the public choice literature. The 'ruled' are the citizens, those responsible for electing political rulers, though of course it is worthwhile noting that in a significant strand of the political economy literature (namely the 'citizen candidate' models of Osborne and Silvinski, 1996 and Besley and Coate, 1997) ruled and ruling are drawn from the same population and hence this division can be seen as a split between different roles rather than necessarily between different individuals.

In the table I also make the distinction between boundedly rational individuals who are aware and those who are not aware. The aware individual knows that he or she is prone to cognitive biases or intertemporal inconsistency, but is not necessarily able to control for these biases on a day to day basis. Conversely the unaware individual blithely assumes that he or she maximizes individual welfare and makes no errors in judgements of any kind. The aware individual may be happy to set up institutions or to delegate power to other individuals who provide corrections to behaviour or provide some constraints on choice. The aware individual may even be willing to delegate such powers to others even when he or she is aware of potential agency problems [Frey and Stutzer 2004], provided the reductions in welfare introduced by agency are less than the gains to be had from delegating choices.

Standard political economy/normative models. Standard models of political economy typically make the assumption that both citizens and their rulers are rational agents. The same is true in most normative models.

Merit Wants. One prominent case is where there is a sharp division between the cognitive abilities of the electorate and that of its leaders. Although it does not have a formal theory of the political process, the traditional merit wants literature (see Chapter 7) would be categorised under this heading, since in

those models, the benevolent dictator is able to see and potentially correct the perceptual defects of the citizens.

Delegation. When individuals are boundedly rational and aware of their cognitive limitations, theoretically they may be willing to delegate responsibility to a cognitive elite.

Self-Control. Possibly the most interesting cases are those where voters are aware of their own cognitive limitations, but not sufficiently aware or capable to adjust for those limitations in particular decisions (Simon 1985). As a result individuals might wish to use the political system to police their own behaviour. For instance, a consumer might vote to restrict the framing of advertising information or place legal restrictions on their ability to avoid saving for a pension. However, given these same cognitive limitations, the consumer might also wish to evade such restrictions once they are in place or see them rescinded. Just as with private decisions, therefore internal conflict between the 'planning' and 'doing' self (Thaler, 1981) can limit the set of sustainable plans, so I have termed these 'models of self-control'. The prime difference between this and the previous case is that the self-controlling electorate is cognitively constrained in its actions in a way which would not apply with delegation.

Strategic Delegation. It is theoretically possible that citizens will wish to elect rulers who are boundedly rational. Models of reputation, commitment and deterrence suggest that in a country surrounded by hostile neighbours or facing international negotiations voters may strategically elect a non-rational leader.

One-eyed Ruler. The most extreme cases in the table, though not necessarily the most unrealistic, are towards the bottom right hand corner where all individuals have bounded rationality. In this particular case, the leaders at least have some awareness of their cognitive deficiencies.

Myopic. In the final case everyone is boundedly rational and no-one is aware of this possibility. Individuals may prefer leaders with the same cognitive defects as themselves or simply prefer the choices made by such leaders. Either way, they are not aware that their choices are boundedly rational. Of all the categories, this is the least explored and offers the greatest potential for a deeper understanding of the role of bounded rationality in political economy.

For many of the categories within the table, rulers and ruled have different cognitive abilities, but will leaders tend to be rational or to share the cognitive biases of voters? To the extent that voters are unaware of their own biases, then the elected may well be boundedly rational even if rational leaders are potentially available.[1]

[1] In part the answer to this question depends on how cunning the rational leader can be, on their desire for power and on their ability to imagine the world through the eyes of a typical citizen. Consider a world where citizens ignore base rate information and where a political leader faces lobbying to move policing resources from a small city to a larger one on the grounds that there

The countervailing forces are based on efficiency: we might expect a more rational (in some sense) leader to deliver consistently better results and therefore be more likely to be re-elected. Nevertheless, at first sight, a distinction between the governed and elite makes an unsatisfactory assumption but there are some reasons why it may be reasonable to accept the hypothesis, though it is not obvious that the arguments are in any way compelling. I shall present each of them briefly.

The first kind of reason is that there are greater opportunities for learning for rulers can explain why there may be a difference between the decision-making capabilities of the government and the electorate. It has already been argued that if individuals repeatedly face similar problems, then feedback may lead to an improvement in the quality of decisions made. Consequently a cadre of civil servants or a class of experience politicians may be more adept at taking public decisions, compared to the average citizen. The second reason is associated with the nature of government with its access to pools of experts, unavailable to individuals. More fundamentally, if cognitive limitations are the result of an economic approach to decision-making, then a professional decision-maker choosing the allocation of a public good will find it optimal to devote a much higher level of investment in making the right decision, compared to the case where he or she is choosing on behalf of one individual. Thus Downs' 'rational ignorance' view of the typical voter has its counterpart in the 'rationally informed' decision maker.

A third line of reasoning is based on drawing parallels with the process of selection amongst competing firms, which may induce profit maximization in equilibrium. Suppose that there is some selection mechanisms (unspecified) which selects behaviours over time within the governing elite, such that over time the quality of decision-making improves, even if the leaders themselves do not become 'smarter'. Then even though the individual agents are not themselves rational they may choose alternatives for society which would be those of a rational agent.

A fourth argument is that proposed by Caplan (2001), and discussed above: the 'rationality of irrationality'. This is based on the view that rationality requires the exertion of costly cognitive effort and that consequently, rational individuals will allocate their brain power so as to maximize its value. In market-based decisions, there are clear incentives to make the correct choices, but when individual vote or deal with government, they typically have little influence over the final outcome. As a result, it is not just Downsian ignorance that is optimal for the citizen, but also irrationality, in the sense that individuals will process information in a non-Bayesian manner. However, for decision-makers and their aides, a rational understanding of information is likely to be more crucial and hence there may be a 'cognitive gap'

---

are more murders in the big city. The rational politician knows that the murder rate is lower in the big city and knows that as a consequence moving resources to the big city would not lower total murders (or any other accepted measure of policy success). The boundedly rational politician shares the citizen's bias and favours the switch. If the rational politician is sincere then defeat may loom. Alternatively, the rational politician may have the capacity to understand voter irrationality and may seek only power, in which case she or he would happily mimic limited understanding of statistical information and support the switch of resources.

between the able and informed rulers and the irrational and ignorant populace at large.

The final argument, and one that that may be least convincing – both from a theoretical viewpoint, but also from empirical observation, is that the voters may elect the decision-making cream of society. In addition, recruitment for the civil service may be designed to select individuals with the greatest capacity for making 'correct' decisions.

These arguments differ in what they require of voters, in terms of awareness of their limitations, but each of them throws up a number of problems in the context of political economy. First how do 'stupid' individuals identify their need for 'smart' leaders? How do individuals who fall prey to perceptual and preference anomalies avoid those self-same traps when electing leaders. Second, the analogy with rational profit-maximizing firms may be false. Within markets, there is a clear metric – namely profit – which provides the benchmark for competition amongst firms. The nearest analogy for government is popularity, but popularity as an index of fitness may simply lead to successful politicians serving the cognitive biases of voters, and not correcting for them. This may be an accurate model of the political process, but it does not conform to a world of the definition of 'smart leaders' given above.

6.2 Market Failure

A central question is this: if individuals do not maximize their own welfare in what sense can the market be said to be efficient or inefficient? In other words, what is the status of the Fundamental Theorems of welfare economics in the context of bounded rationality? The traditional market failures of increasing returns, public goods and externalities are generally defined in technological terms, rather than in terms of preferences. Increasing returns occur when inputs of λx ($\lambda > 1$) yield more than λ times the output produced by inputs of x; public goods are commodities where the cost of exclusion is prohibitive and where the marginal cost is zero for supplying the good to an extra individual; externalities are goods where the consumption or production possibility sets depend directly on the choices of other agents.

As a consequence, the concepts of public goods, increasing returns and externalities can be defined without mentioning rationality and hence there is no reason to suppose that they are not relevant in the world of boundedly rational individuals. However, that does not mean that Pareto inefficiency results when they are present. The example below provides a counter-example, albeit possibly an unlikely one. Part A depicts a standard Prisoners' dilemma. In it the Nash equilibrium is Pareto dominated by the outcome when both players choose to contribute (to the public good). Suppose though that the individuals perceive the payoffs as in part B of the figure, but then still act so as to maximize their own perceived payoffs. The perceived dominant strategy for each player is to contribute and the unique Nash equilibrium is also the one that represents the Pareto efficient outcome. In this instance, if the players perceive that the game is as in part B, there is no inefficiency.

	Part A			Part B	
	Contribute	Free-ride		Contribute	Free-ride
Contribute	2,2	−1,3	Contribute	2,2	1,1
Free-ride	3,–1	0.0	Free-ride	1,1	0,0

6.2.1 Asymmetric Information

The market failures mentioned above were presented as technological relationships. Arguably, asymmetric information belongs to a different family, one in which it is asymmetries in the stock of a good that are the source of the problem.

Asymmetric information is also different because some of the most important anomalies act directly on information processing. As a result, bounded rationality may be a source of apparent asymmetric information. For instance, the same information might be available to two individuals, but given confirmation bias, say, each might ignore or discount that part of the information set which contradicts his or her view of the world. Or the availability of the information set might differ between individuals depending on which part is salient.

The consequences of this for market efficiency may be difficult to predict in general. In the familiar market for lemons problem, agents use deductive reasoning to make inferences about the behaviour of agents on the other side of the market and hence are able to make the associations between price and quality that potentially destroy the existence of market equilibrium. Suppose individuals are not able to deduce the other agents' optimal strategies given common knowledge of the game. Potentially, buyers may then fail to associate low price with low quality, restoring the downward sloping demand curve and producing market equilibrium – and therefore some degree of efficiency. Alternatively agents might use rules of thumb in which price and quality are positively associated, which hands market power to firms where they might not have any in the absence of bounded rationality.

So, the presence of both standard market failures and bounded rationality creates a problem similar to the familiar one of the second best, in which intuitive policy interventions which might be optimal when only one failure is present, cease to be so when there are multiple sources of inefficiency. One cannot therefore automatically claim that traditional market failures are completely unaffected by the introduction of bounded rationality. Can something weaker be claimed?

6.2.2 Near-Optimality

Suppose that the equilibrium of an economy is given by the set $E(\varepsilon, y) \subseteq X \subset \mathbb{R}^n_+$, where X is closed and bounded. Elements of E will typically be vectors of consumption and production values for the finite number of agents in the economy, while the parameter,[2] ε, is a measure of the degree of bounded rationality in the

[2] In fact we could treat ε as a vector.

economy. For instance, ε could be the maximum degree to which any agent deviates from maximizing their own individual welfare. E will also typically depend on other factors, such as endowments, technology, preferences etc. These other factors are given the label y. Anderlini and Canning (2001), show that if behaviour responds smoothly to changes in ε and if the economy is stable – in the sense that the outcome E is continuous in the parameters y for all values of ε, then it will also be the case that E is continuous in ε. It follows that if E is discontinuous in ε, then it must be that behaviour is discontinuous in ε as well. Let us apply their approach to the issue of market failure.

Let $P(\varepsilon, y)$ be the (set of) Pareto efficient outcomes given the bounded rationality parameter and the other factors in the economy, and let $P(y)$ be the set of Pareto efficient outcomes when all agents are rational. Market failure is about the relationship between P and E. When there is no bounded rationality, market failure is said to occur when members of the set $E(0, y)$ are not members of the set $P(y)$.[3] With bounded rationality, one possible way to extend the definition is, 'not all members of $E(\varepsilon, y)$ are members of $P(y)$'. Alternatively we could define it as, '$E(\varepsilon, y) \nsubseteq P(\varepsilon, y)$'. With the first definition we identify the set of Pareto efficient allocations with those that occur when rationality is unbounded. The advantage of this definition is such allocations are desirable in the sense that they represent Pareto-efficient allocations involving rational agents. However, even when $E(0, y)$ is a subset of $P(y)$, $E(\varepsilon, y)$ need not be a subset of $P(y)$ if for instance bounded rationality reduces the set of feasible outcomes. Figure 6.1 illustrates the point for a Robinson Crusoe economy

In the diagram, the locus PP represents the production possibility frontier when Robinson Crusoe is fully rational. E^1 is then the Pareto efficient outcome and also the competitive equilibrium – that is $P(y)=\{E^1\}$. With bounded rationality, suppose Crusoe is unable to work so efficiently and the production possibility frontier becomes $P'P'$. The new equilibrium is at E^2 and clearly this is not in $P(y)$. If we

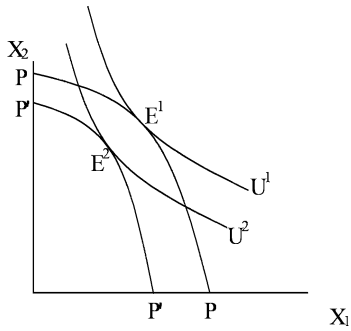


Fig. 6.1 Bounded Rationality and Pareto Optimality

[3] We might opt for a weaker definition, such as E(0, y) does not overlap P(y). One argument in favour of the formulation given arises from the first fundamental theorem of welfare economics which states that E(0, y) is a subset of P(y), when E is the set of competitive equilibria.

further suppose that Crusoe's preferences are unaffected by his bounded rationality – only his productive abilities, then interpreting E^2 as an instance of market failure is potentially unappealing. Given Crusoe's reduced productivity, the market achieves the best outcome for him. Not achieving E^1 is therefore a failure of rationality rather than the market.

The alternative extension of market failure requires a definition of $P(\varepsilon, y)$. As we saw in the previous chapter, a compelling definition of welfare is not easily available given bounded rationality and that is the weakness of the alternative approach. However, although we may not be able to specify P exactly here, we may be able to claim some sensible properties for it. For instance given that X is closed and bounded, the existence of $P(\varepsilon, y)$ appears inescapable, while we would also expect the definition of Pareto efficiency to accord with the standard notion in the absence of bounded rationality. A third possible assumption is that small changes in bounded rationality create only small changes in the set of Pareto efficient points. These potential assumptions are set out below:

1. $P(\varepsilon, y) \neq \varnothing$.
2. $P(\varepsilon, y) = P(y)$ when $\varepsilon = 0$.
3. $P(\varepsilon, y)$ is upper hemi-continuous in ε.

Suppose we know that $E(0, y)$ is not a subset of $P(y)$. What can we say about the relationship between $E(\varepsilon, y)$ and $P(\varepsilon, y)$? Given properties 2 and 3, if $E(0, y) \nsubseteq P(y)$ and E exists and is continuous in ε, then there must exist values of ε for which $E(\varepsilon, y) \nsubseteq P(\varepsilon, y)$. In other words, if the equilibrium of the economy is not Pareto efficient in the absent of bounded rationality, then it is not Pareto efficient with bounded rationality. We state this as,

Result 6.1 *Given properties* 2 *and* 3, $E(0, y) \nsubseteq P(0, y)$, $E(0, y) \neq \varnothing$ *and continuity of E, then* $\exists\, \varepsilon_0$ *such that for any* $0 < \varepsilon < \varepsilon_0$, $E(\varepsilon, y) \nsubseteq P(\varepsilon, y)$.

Proof see chapter appendix.

Note that we cannot weaken the continuity of E to upper hemi-continuity, a point illustrated in Fig. 6.2 where the solid line on the vertical axis and the curve connected to it show the set $E(\varepsilon, y)$ for different values of ε. Meanwhile, the shaded area represents $P(\varepsilon, y)$. In this Fig. E is upper hemi-continuous but not continuous at $\varepsilon = 0$. At $\varepsilon = 0$, $E(\varepsilon, y) \nsubseteq P(\varepsilon, y)$, but for any positive value of ε, E is a subset of P. In other words, with bounded rationality there is no market failure.

To sum up, concepts of market failure that apply in the absence of bounded rationality will also apply with some bounded rationality provided that the economy and measures of welfare are sufficiently continuous. However it is worth sounding two notes of caution: first and most obviously the result applies for 'small' deviations from rationality, but not necessarily to larger departures. Second, even if at the individual level a departure from rationality is small, in a market or interactive context the net effect can be large. In other words, though continuity of E in ε sounds reasonable it is not a wholly innocuous assumption. A classic instance of small individual deviations from rationality having large consequences in aggregate is found in the macroeconomic literature on menu costs Mankiw, 1985.

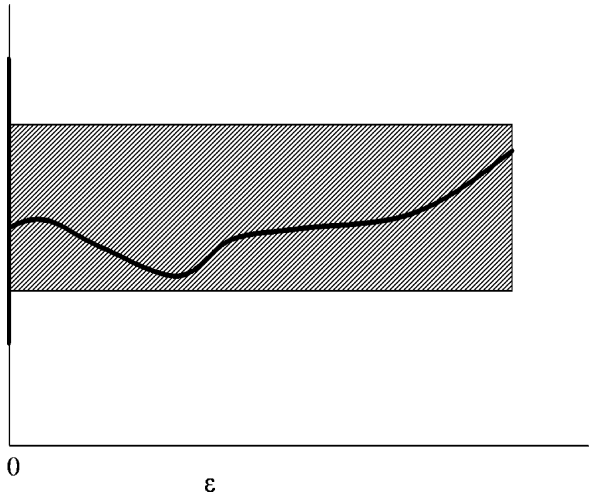


Fig. 6.2 Upper Hemi-Continuity of E and No Market Failure

6.2.3 The Boundaries of the State

Does bounded rationality also have implications for the boundaries of the state? According to Caplan (2001), irrationality becomes more irrational when the state takes over the role of chooser. In other words, the individual has less incentive to invest in the accumulation of information and decision-making skills when they have little or no influence over the decision. As a result, he argues that provision by the state creates anomalies. Against this and in keeping with the traditional merit wants literature, Cullis et al. (2000), view the anomalies literature as providing justification for greater regulation of personal choice. In part, these different views represent fundamentally opposed views on the nature of bounded rationality. For Caplan, rationality is endogenous and economic in character in the sense that the individual will develop cognitive strengths in those areas of his or her life where the returns are greatest. Conversely, where the returns to cognitive effort are minimal, he or she will slip into irrational behaviour. In this view, government regulation of choice alters the incentives to be rational in a standard moral hazard fashion. On the other hand, Cullis and Jones view the size and strength of anomalies as largely unaffected by incentives. As a result, the state may be able to limit their damaging impact. Beyond these contrasting positions, there is a further issue of information in that any government which seeks to correct for anomalies has to have information on individual welfare. The greater the degree of bounded rationality the poorer the data may be.

Let us explore elements of the trade off by considering two simple examples. In the first model, suppose that there are two kinds of agents, type 1 and type 2, and two choices, A and B. We have $u_1^A > u_2^A$, but $u_2^B > u_1^B$. In other words option 1 is best for type 1 agent and option 2 is best for type B. Let the proportions of type 1 in the population be q. Suppose that government knows this proportion, but not each individual's type. Left to make their own choices each type 1 individual

would choose option A with probability p_1 and each type 2 individual would choose option 1 with probability p_2. In this sense individuals are not strongly rational – each type chooses their best option with a probability less than 1. Suppose welfare is utilitarian, then the free choice outcome yields expected welfare of,

$$q(p_1 u_1^A + (1 - p_1)u_1^B) + (1 - q)(p_2 u_2^A + (1 - p_2)u_2^B) \tag{6.1}$$

Now consider a process of government provision, by a welfare maximizing government. Because the government does not know individual preferences it supplies only one option for all agents. If the government dictates choice then it will choose option A provided,

$$q u_1^A + (1 - q)u_2^A > q u_1^B + (1 - q)u_2^B$$

Otherwise it will opt for B. In what follows I shall suppose that the inequality in this equation is satisfied. Define Δ as the net gain to government action, then,

$$\Delta = q(1 - p_1)\Delta_1 + (1 - q)\Delta_2(1 - p_2) \tag{6.2}$$

where $\Delta_i = u_i^A - u_i^B$, $i = 1, 2$. Thus government action, defined as imposing A on all its citizens, is more likely to be welfare enhancing when

1. the proportion of type 1 individuals is large;
2. when the gain to these individuals from having A over B is large;
3. when the proportion of type 1 who would otherwise have chosen A is small and
4. when the cost of imposing A on type 2 agents is small, which in turn occurs when

   a. the proportion of type 2 agents is small,
   b. when they were likely to have chosen incorrectly anyway and
   c. when the costs to them of making the wrong choice are small.

Now, as Caplan suggests, the probabilities p_1 and p_2 are likely to depend on experience and the incentives individuals have to make welfare-maximizing choices. Let us consider the first of these. Suppose that individuals are infinitely-lived and that future welfare is discounted at the rate ρ. Write p_{it} as the probability that given a free choice, an individual of type i would opt for choice A in time period t. Supposing once more that the government would make all consumers take option A, then the net gain to government action is now,

$$q\Delta_1 \sum_t \frac{(1 - p_{1t})}{(1 + \rho)^t} + (1 - q)\Delta_2 \sum_t \frac{(1 - p_{2t})}{(1 + \rho)^t} \tag{6.3}$$

It is clear that (6.3) may be negative, while (6.2) is positive, so that while government intervention may make sense in the short run, in the long-run it would be better for consumers if they were left to learn their own preferences. Such a claim

rests on the time path for the p_{it}s as a more specific example may make clearer. Suppose that $p_{1t} = 1 - x^{t+1}$ for some x, such that $0 \leq x < 1$ and $p_{2t} = y^{t+1}$ for some y, such that $0 \leq y < 1$. Under these conditions we obtain,

$$\Delta = (1+\rho)\left[q\Delta_1 \frac{x}{1+\rho-x} + (1-q)\Delta_2 \frac{1-y}{1+\rho-y}\right] \tag{6.4}$$

It is apparent therefore that the gains to government intervention are more likely to be positive when the discount rate is high, when initially, type 1 consumers choose A with a low probability, while consumers of type 2 choose A with a high probability.

In practice, there are not just two but many types of individuals who learn in many forms. Individuals learn from their own past experience, through reflection and imagination and from the experiences of others. So let us now develop a second example in order to explore the theory further.

In the next model individuals choose from an uncountable subset of the real line, rather than between two options. Such a choice might represent the proportion of a budget to devote to a particular good, assets to a particular form of investment or time to an activity. Suppose in particular that preferences can be represented by the quadratic form, $u = 2bx - x^2$, $b \in [b^l, b^u]$ where $b^u > b^l$. Individual welfare has the form: $u = 2b^*x - x^2$, $b^* \in [b^l, b^u]$. Normalise the population to 1. Let the joint density function of b and b^* be $f(b, b^*)$ and as before suppose a utilitarian government which maximizes the expected value of W. Let us suppose that irrespective of any imperfection in individual choices, there are potential gains of θ from government provision of the good. Control by government may give it the opportunity to correct an externality or reap economies of scale; on the other hand it may that government is inefficient in its activities. In the first case θ is positive; in the second case θ is negative and when there are neither gains nor losses, then $\theta = 0$.

The key question is whether choice of x should be left to individuals or whether the government should choose a single value of x for all. If x is chosen by the government then its optimal value is $\overline{b}^*$, which is the mean value of b^*, that is $\int\int b^* f dbdb^*$. Welfare is then $W^g = \overline{b}^{*2} + \theta$ If, on the other hand, individuals choose freely, then someone with a preference parameter b will choose $x = b$. Consequently under market provision overall welfare is given by:

$$W^m = \int\int (2bb^* - b^2) f(b, b^*) dbdb^* = 2\rho\sigma_b\sigma_{b^*} - \sigma_b^2 - \overline{b}^2 + 2\overline{b}\overline{b}^* \tag{6.5}$$

where σ_i denotes the standard deviation of i, ρ is the correlation between b and b^*, and $\overline{b}$ is the mean value of b. It follows that, in a one period world, the net advantage of market supply over government is:

$$W^m - W^g = 2\rho\sigma_b\sigma_{b^*} - \sigma_b^2 - \overline{b}^2 + 2\overline{b}\overline{b}^* - \overline{b}^{*2} - \theta \tag{6.6}$$
$$= 2\rho\sigma_b\sigma_{b^*} - \sigma_b^2 - (\overline{b} - \overline{b}^*)^2 - \theta$$

Equation 6.6 shows that:

- When there is no variation in b* across the population and θ is positive, then non-market provision is never inferior to the market. In other words, greater uniformity in need raises the advantages of government provision.
- When there is no difference between preference and welfare ($b = b^*$ for all agents) and θ is non-positive, then the market solution is never inferior.
- A rise in the variance of b* raises the advantages of the market. In other words, greater diversity in tastes raises the advantages of the market.
- A increase in the gap between the mean values of b and b^* raises the advantage of government supply. In other words, the greater the systematic difference between choice and welfare-optimal consumption, the greater the disadvantage of the market.
- A reduction in the correlation of preferences and welfare raises the advantage of government supply. In other words, the greater the randomness of choices (when viewed from a welfare perspective), the weaker the advantage of the market.
- Higher values of θ raise the gains to non-market provision.

Now consider a two period model in which there is the possibility of learning for individuals in the market. In the calculation of overall welfare I weight the second period by a. This parameter allows the possibility that the welfare of the second period might be discounted relative to the first, or that the 'second' period is actually much longer (or shorter) than the first, but I shall always suppose that $a > 0$. Let b denote an individual's preference parameter in period and 1, with b′ denoting the preference parameter in period 2. Given a deterministic relationship between these variables we can write $b' = (1 - \lambda(b, b^*))b + \lambda(b, b^*)b^*$. Now, learning can take many forms and as we saw in Chapter 3 it is not necessarily the case that experience or choice gives an individual keener insight into their welfare. Let us suppose though that learning has some benefits in the following manner:

Definition 6.1 Universally beneficial learning (UBL). UBL occurs if, $\forall\ b, b^*, 0 < \lambda(b, b^*) < 1$.

In other words, with UBL, agents' preferences move towards their true welfare.

Result 6.2 *Suppose* $f(b, b^*) > 0$ *for some* $b \neq b^*$. *Then the net advantage of the market over non-market provision is higher when there is UBL compared to the situation when there is no learning.*

Proof see chapter appendix.

Not surprisingly, UBL provides support for market provision rather than government provision and this support is higher when:

- The higher is the weight placed on the second period.

- The degree of learning is higher (i.e. when $\lambda(b, b^*)$ is higher).

Figure 6.3 summarises the argument. The solid, 45 degree line marked $\lambda = 0$, represents the locus of values for which the value of market and government provision are equal. We shall call this the boundary of the state. Points above and to the right of the locus represent parts of the space where government provision is superior; to the left of the locus is the region where market provision is optimal. The impact of learning is to shift this boundary to the right, for instance to the broken line marked by $\lambda > 0$. Rises in ρ (the correlation between welfare and preference) and σ_b also push the boundary to the right, with increases in σ_{b^*} having the opposite effect. As a result, for a point such as A, government provision is optimal in the absence of learning, but market provision is superior for sufficiently high rates of learning.

At this point we have established that the possibility of learning reduces the gains to non-market provision. It is still true though that a rise in irrationality (defined here as a reduction in the correlation between b and b^* or a rise in the gap between their respective means) raises the case for government intervention. Essentially this occurs because a rise in irrationality has no implications for the government's knowledge of the distribution of b^*. That seems unreasonable for instances where it is citizens that supply the government with information on welfare.

Suppose the government has only approximate insight into the true distribution of b^*. To be more specific, suppose that its belief is given by $b^{*\prime} = \mu b + (1 - \mu)b^*$. In other words, for any given individual, the government's belief about their true type is some weighted average of what the individual believes and the true value. Such a belief may be justifiable on the basis that a part of the evidence on welfare will come from individuals. For simplicity, assume now that λ is constant across the population. Under these circumstances, government provision will be at $x = \overline{b}^* + \mu(\overline{b} - \overline{b}^*)$. W^g is now $\overline{b}^{*2} + \mu(\overline{b} - \overline{b}^*)^2$. The equation for the boundary of the state is now given by,

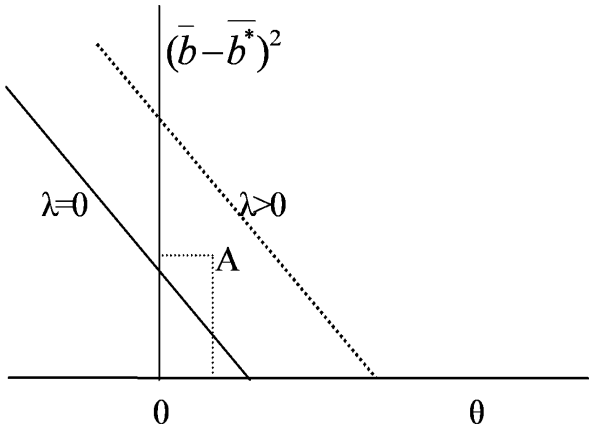


Fig. 6.3 Learning and the State

$$(\bar{b} - \bar{b}^*)^2 = \frac{(1+a)(2\rho\sigma_b\sigma_{b^*} - \sigma_b^2 - \theta) + (a(1-(1-\lambda)^2)(2\rho\sigma_b\sigma_{b^*} + \sigma_b^2 + \sigma_{b^*}^2)}{1 + a((1-\lambda)^2 - \mu^2)} \tag{6.7}$$

The crucial part of this equation is the denominator on the right hand side. When a is smaller than 1, so that little weight is attached to the future, then the denominator must be positive and so the boundary slopes down as in Fig. 6.3. But if a is large then it is the relative values of λ and μ that matter. When μ is higher than $1-\lambda$ then even with long run learning, the government has a more accurate view of individual preferences than the individuals themselves. However, if the converse is the case, then through long run learning individuals will have greater insight into their own preferences than the government. Under these circumstances, the boundary of the state may have the shape depicted by the solid 45 degree line in Fig. 6.4.

In this diagram, points to the right of the boundary represent situations in which government provision is optimal. Rises in θ still raise the comparative advantage of government provision but now rises in $(\bar{b} - \bar{b}^*)^2$ support market provision, as a rise in the gap between the average belief about b and the average true b undermines the government's knowledge about the true distribution of preferences. Moreover, a weakening of the relationship between preference and welfare in the shape of a reduction in ρ pushes the boundary downwards, expanding the domain in which the market is superior to the state.

In Caplan's story irrationality is, at least in part, endogenous. We can incorporate some endogeneity into our model. Suppose that we interpret the model as follows. Individuals begin with the belief that their correct choice of x is b. They receive some hedonistic feedback from their choice, but learning how to interpret that information is costly as it requires reflection and potential realignment of choice. Thus to choose optimally in the second period given the updated beliefs about b, requires an investment of cognitive effort which costs $c > 0$. Learning is still potentially imperfect in that just as above an individual adjusts their beliefs to $b' = (1 - \lambda(b, b*))b + \lambda(b, b^*)b^*$. The gain from investment is therefore $(b' - b)^2 = \lambda^2(b^* - b)^2$. Thus for c sufficiently small, any individual for whom $b^* \neq b$ will invest in learning. Meanwhile individuals with government provision

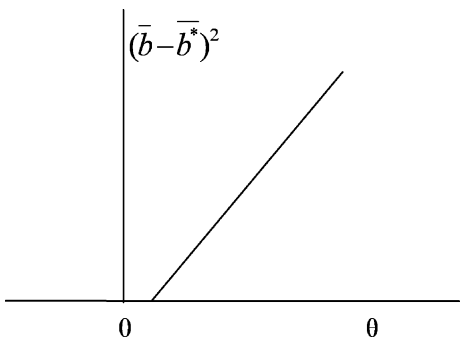


Fig. 6.4 The Impact of Learning on the Boundaries of the State

will not invest because there is no point.[4] The relative comparison between government and marketplace (for sufficiently small c and $b^* \neq b$ for all individuals) is now,

$$W^m - W^g = 2\rho\sigma_b\sigma_{b^*} - \sigma_b^2 - (\overline{b} - \overline{b}^*)^2 - \theta - c$$

In other words c acts like a rise in θ. The boundary condition is therefore only marginally changed and the essence of the original argument remains. In particular the relationship between the boundary and the state and changes in bounded rationality hinges on the relative values of μ and $(1 - \lambda)$.

What is more reasonable: a positive or negative sign for $(1 - \lambda - \mu)$? There are some reasons why governments might have access to superior knowledge, at least for particular kinds of information. In such a case $(1 - \lambda - \mu)$ will be positive and government provision optimal. Safe doses of medicine, the probability of road accidents when drunk, the levels of saving required to maintain a target level of consumption in retirement are examples of areas where the superior resources of the state might yield to it advantages over the individual. In these kinds of cases, the optimal choice does not depend critically on the idiosyncracies of the individual or at least, a straightforward and reliable statistical model can relate optimal behaviour to easily observable characteristics. For instance, the safe dose of most medicines can be related to weight, gender etc. However, for other cases, where optimal choice is highly individual specific (μ high relative to $1 - \lambda$), the state might have no advantages in provision. In other words, the 'Hayekian critique' of government control would extend to the case where individuals have imperfect insight into their own preferences. As a result, therefore, bounded rationality does not necessarily extend the boundaries of the state.

6.3 Voting for the Nanny State: Regulation of Markets by Self Aware Consumers

Having established that there is no necessary relationship between bounded rationality and the argument for state control of the economy, it is nevertheless worth acknowledging that government intervention may not only be desirable, but also desired when individuals are aware of their own limited rationality. In fact, regulation of consumer markets is an extensive feature of many societies. As an introduction to it, let us consider four examples:

[4] There is a third option that I do not pursue here: the government could choose in the first period and then let individuals choose in the second period. Presumably individuals would learn something about their true welfare from consumption of the government good and in the second period they would have an incentive to incorporate this learning into their choices. The third way is more likely to be superior to the alternatives when, on the one hand, learning is potentially effective and on the other when initially individuals hold highly inaccurate views on their own optimal choices.

- Limitations on advertising. In many societies, there are strict limitations on what can be advertised, while advertisers are limited as to how they describe their product. In the UK, for instance, prescription drugs cannot be advertised; unlike say the USA, which has seen a boom in advertising of this kind in the wake of regulatory relaxations. In many countries of the European Union, tobacco advertising is illegal or severely circumscribed as is advertising directed at children.
- Small print protection. In the European Union, citizens (but not usually businesses) are at least to some extent protected from their own failure to read the small print on contracts if those clauses are especially onerous.
- Safety standards. Cars must meet crash standards; medicines are regulated for their side effects; children's toys may be compelled to be withdrawn from the market if they are deemed unsafe. In these and in many other cases, government goes beyond the provision of information about risk and actually alters the set of available products. Alternatively, governments might regulate the form of information given to consumers. In Japan for instance, the flat-pack furniture retailer, IKEA, was asked to provide assembly instructions more detailed than its standard language free cartoons on the grounds that Japanese consumers were used to detailed and even exhaustive manuals.
- Special kinds of restrictions are often encountered with financial products, where information about interest rates and payment periods must be stated in a standard format. Statutory 'cooling off' periods are common for major decisions involving extended periods of credit. In addition, minimum levels of pension saving are often mandated as are maximum levels of risk.

Regulation of consumer markets might be founded on a number of forms of bounded rationality [Earl 2005]: in Chapter 10 we discuss issues of self-control (see also below) which provide one motive. Alternatively, as we have seen in the previous section, when government has a good understanding of welfare, while citizens are slow to learn what is best, then the state can have a comparative advantage in the allocation of goods.[5] However in keeping with the theme of the chapter, we also need to consider how the regulation of consumer markets might be explained by means other than through a bounded rationality framework. Four standard reasons are: rent seeking behaviour; regulation of monopoly; reduction of asymmetric information and overcoming the public good problems associated with information gathering.

Rent seeking models are quite general, in that they are potentially capable of explaining all economic and political behaviour. The key actors are politicians who supply the regulation. They face consumers, businesses and government producers -more usually called bureaucrats – who make possibly competing demands for rents.

[5] In the previous section the government chooses a fixed level for x. Alternative policies that more closely resemble actual practices in regulation include setting upper and lower limits on x (see Chapter 8 for example). One argument for this kind of regulation arises when certain choices of x change the probability of death or major injury (the parameter a in the model) and hence lower the educational benefits of free choice.

In theory, regulations in consumer markets might be advantageous to members of each of the groups on the demand side. Safety standards might raise entry costs, thereby increasing the profits of incumbent firms; regulatory red-tape will usually increase the demand for civil servants and raise the incomes of government employees. Meanwhile product standards might be demanded either all consumers (given that product knowledge has some of the properties of a public good) or especially by those with high search costs. They might also be demanded as a means of avoiding 'markets for lemons' type problems of market failure. In the case of consumers, though, it is hard to see why the product itself rather than information should be the subject of regulation.

There are also other political economy explanations for restrictions on trade. In one particular class of 'good Samaritan' models, they are explained on the basis of a time consistency problem, usually allied to altruism on the part of one or more agents (see Coate, 1995 for example). One class of agents (the 'Samaritans') wish to improve the life-time well-being of another group (the 'targets'). The latter group, given a transfer in period t will fail to save any portion of it, either because of a high discount rate or simply because they correctly anticipate that, faced with their poverty in period $t + 1$, the Samaritans will wish to transfer more income. On the other hand, if the targets are given guaranteed payments in period $t + 1$, then, with well-functioning capital markets and a high rate of time preference, they make seek to borrow against future income and once again end up poor in period $t + 1$ (Andolfatto 2002). Government enters this context in two or three ways. First, in a standard manner, voting for a compulsory tax/transfer system can reduce the free-riding associated with individual giving. Secondly, Andolfatto argues that government can make certain assets inalienable. Consequently, spendthrift individuals cannot borrow against them and therefore any transfers committed to for $t + 1$ will be consumed in that period. Finally, transfers can be made in the form of goods which are naturally hard to resell or borrow against (e.g. education), which while creating a distortion in consumption, can still lead to Pareto improvements when no other instruments are available to the altruist.

The models used to explore this issue are typically stark and therefore it is easy to criticise them on the grounds of realism. However even taking such a fact into consideration they have some important limitations as the basis for more general models. In Andolfatto's model, for instance, the spendthrift minority place no weight on future consumption. As a result policies such as consumption subsidies are ruled out. Equally important is the fact that his central example of an inalienable asset, namely social security, may actually be sold or borrowed against using other, substitute assets. Because of this, it is quite difficult to find examples of assets which match the requirements of his model in terms of simultaneously being inalienable while providing a suitable vehicle for poverty alleviation. Finally, restrictions on market behaviour are usually universal in tone and not income or wealth contingent. If we allow these criticisms then what is left of this class of models are two general insights: first there may be reasons to redistribute income intertemporally when individuals are myopic and second, implementing redistributive and/or social insurance policies may be difficult when pre-commitment is impossible.

To sum up, in many cases, rent seeking and other standard kinds of explanations offer a satisfactory model for restrictions on trade (see e.g. Barzel, 1997, for a fascinating examination of the subject). However, they do have their deficiencies. Quantitative restrictions on advertising such as those which used to be in place for spectacles are explicable in terms of the gains to imperfectly competitive firms from avoiding competitive behaviour. Qualitative restrictions on marketing which do not limit advertising expenditure are much harder to understand without having a model of how different kinds of frames might affect consumer behaviour.[6] Secondly, it is often hard to see how the models which are used to explain behaviour in market X do not apply equally well in market Y without invoking notions of bounded rationality. Why are health good markets subject to so many restrictions on advertising, for instance, when food markets are not?[7]

6.4 Political Economy – An Illustration

Political economy has a variety of meanings within economics, so it is probably wise to begin this section with a clarification of the term. In some cases it is identified with traditions at variance with the dominant neoclassical paradigm, for instance Marxism or Institutional economics; in other situations it simply signifies the importance of the political sphere as well as the economic. Most recently the label has been attached to the kind of general equilibrium model in which rational agents maximize both in the economic and political sphere (Persson and Tabellini 2000). The goal of this literature, as I see it, is to derive political outcomes as a function of tastes and technology in much the same way that competitive general equilibrium theory derives economic outcomes as a function of the same variables. As such political economy has some links with public choice in that both attempt to explain political outcomes in terms of the rational agent model. It also has antecedents in the formal literature of incentive mechanisms and optimal taxation.

One central issue in Political Economy is whether the political system delivers efficiency (Wittman 1995) or whether there is *political failure*. As with market failure, we might suppose that bounded rationality would fuel inefficiency in the political sphere, but that need not be the case as the following extended example shows.

---

[6] Though advertising restrictions may raise the cost function for effective communication with potential customers, thereby creating entry barriers.

[7] In making this criticism of the rational choice explanation, I am not assuming that the policy maker is a benevolent maximizer of social welfare, in the case of the boundedly rational consumer. Willingly and through the electoral system, boundedly rational consumers may erect constraints on firm behaviour that raise their ex ante welfare. They may also erect constraints in the belief that these restraints are welfare enhancing, though, in fact they limit welfare.

6.4.1 Efficiency and Bias

In this section, I use a simple model of confirmation bias (see Chapter 2) to explore efficiency. Recall that for this particular anomaly, individuals overly weight initial impressions and then tend to ignore disconfirming evidence (Rabin and Schrag 1999).

In the basic overlapping generations model that begins the example, there are $n+1$ voters, where for simplicity I shall take n to be odd. Each voter has a life-span lasting for $n+1$ periods, and there are elections in every period. In the first period of their life each voter simply observes the performance of one of two infinitely-lived candidates who is elected on the basis of votes cast by the other n voters. At the end of the period all voters receive a signal about the performance of the elected candidate. If the signal is positive, the new voter votes for that candidate for the rest of their life. In other words, no subsequent evidence can disconfirm their first impression of the candidate. If the signal is negative, the new voter thereafter votes for the other candidate.

The candidates themselves differ only in their efficiency and therefore in the probability of sending a positive or negative signal. I shall suppose that candidate A performs well with probability p and that candidate B performs well with probability q. Without loss of generality, let $p > q$. On average therefore A performs better than B and so the political system is said to be efficient if A is continually voted into power.

The rationality of these voters is heavily circumscribed. They ignore evidence in a number of ways: only the outcome in one period influences their decision; they fail to make inferences from the fact that A or B is in power (which says something about past performance) and they fail to take into account the fact that they have information only on one candidate. Nevertheless, as we shall see, the outcome tends towards the efficient as n becomes large.

With n generations of actual voters, there are 2^n possible combinations of votes. Treat each such combination as a state of the world, then the electoral process can be modelled as a Markov process. Each state of the world can be denoted by a n-vector, with the first entry denoting the vote of the youngest voter and the nth entry representing the vote of the nth and oldest voter. For simplicity, let 1 represent a vote for A and let 0 represent a vote for B. It is apparent that each state can only transform into two other states and that each state is only accessible from two other states. For instance, consider a four period model where (1,1,1) means that in the previous period all eligible voters voted for candidate A. Given this, (1,1,1) can transform into (1,1,1) with probability p or it can be transformed into (0,1,1) with probability $1-p$. Note also that by construction, the process is ergodic.

Let $v \in \{0, \ldots, n\}$ be the number of votes cast for A in any given period and let v_i denote the number of votes cast for A in state i. Order the possible states such $v_i < v_j$ implies $i > j$. So, (1,...,1) is state 1 and (0,...,0) is state $2^n = N$. The pattern of possible transitions between states falls into a limited number of cases. If, for instance, the last entry in a state vector is 1, then the number of votes cast for A cannot rise in any immediate successor state.

Let x_i denote the expected proportion of periods spent in state i and let x be the vector of proportions. If M is the Markov transition matrix then x is the solution to the equation $Mx = x$, but rather than inverting this (potentially huge) matrix we can exploit the sparseness of its entries to obtain the solution directly. So, for instance we know that, $x_1 = px_1 + px_j$ for some j $\neq 1$, where $v_j = n - 1$. In other words, $x_j = (1 - p)x_1/p$.

More generally, describe a state as borderline if either of its successor states involves a change of winning candidate or either of its preceding states involves a different winning candidate. Call all other states interior. For all states, whether interior or borderline we have the following conjecture: if $v_i = v_j$ then $x_i = x_j$. In other words, if two states have the same total votes then the expected proportions of time spent in those states are equal. The conjecture is proved implicitly: it is used to find a solution to $Mx = x$ and we can then invoke the uniqueness of this solution[8] to show that the conjecture is true.

Consider state i where i belongs to the class of interior states which involve a win for candidate A. That is $v_i > (n + 1)/2$. If the vector for i begins with a 1 then, $x_i = px_j + px_k$ for two states j and k where $v_j = v_i$ and $v_k = v_i - 1$. If it begins with a zero then, $x_i = (1 - p)x_j + (1 - p)x_k$ where $v_j = v_i$ and $v_k = v_i + 1$. In either case, invoking the conjecture we have, $v_i = v_j + 1$ implies $x_j = (1 - p)x_i/p$.

Next consider interior states involving a win for candidate B. If the vector for i begins with a 1 (i.e. a vote for A) then, $x_i = (1-q)x_j + (1-q)x_k$ for two states j and k where $v_j = v_i$ and $v_k = v_i - 1$. If it begins with a zero then, $x_i = qx_j + qx_k$ where $v_j = v_i$ and $v_k = v_i$. In either case, invoking the conjecture we have, $v_i = v_j + 1$ implies $x_j = (1 - q)x_i/q$.

For borderline states involving a win for A, if the vector for i begins with a 1, then, $x_i = px_j + (1 - q)x_k$ for two states j and k where $v_j = v_i$ and $v_k = v_i - 1$ (so that in state k, candidate B wins). If it begins with a zero then, $x_i = (1 - p)x_j + (1 - p)x_k$ where $v_j = v_i$ and $v_k = v_i + 1$. Invoking the conjecture, we have $v_i = v_j + 1$ implies $x_i = (1 - q)x_j/(1 - p)$ for all states where $v_i = (n + 1)/2$.

Putting it all together:

$$x_i = \begin{cases} x_1 \left(\dfrac{1-p}{p}\right)^{n-v_i} & \forall\, i\ s.t.\ v_i = \dfrac{n+1}{2}, \ldots, n. \\ x_1 \left(\dfrac{1-q}{1-p}\right)\left(\dfrac{q}{1-q}\right)^{n-1^2-i}\left(\dfrac{1-p}{p}\right)^{n-1^2} & \forall\, i\ s.t.\ v_i = 0, .., \dfrac{n-1}{2} \end{cases} \tag{6.8}$$

From this we can see that the conjecture is true. Note also that there are $n!/(n-v)!v!$ possible states of the world where the number of votes for A is v, hence,

[8] The solution to $Mx = x$ is only unique up to a scalar, but of course x must also be a vector of proportions.

$$1 = x_1 \left(\begin{array}{c} \sum_{i=0}^{i=\frac{n-1}{2}} \left(\frac{1-q}{1-p}\right)\left(\frac{q}{1-q}\right)^{n-1^2-i}\left(\frac{1-p}{p}\right)^{n-1^2}\left(\frac{n!}{(n-i)!i!}\right) \\ + \sum_{i=n+1^2}^{i=n}\left(\frac{1-p}{p}\right)^{n-i}\left(\frac{n!}{(n-i)!i!}\right) \end{array} \right) \tag{6.9}$$

The equations for the individual x_is are not particularly diverting, but if we gather terms we can formulate an expression for the expected proportion of periods in which A is the winning candidate, which is all states of the world in which v is at least $(n+1)/2$:

$$\frac{\sum_{i=\frac{n+1}{2}}^{i=n}\left(\frac{1-p}{p}\right)^{n-v_i}\left(\frac{n!}{(n-i)!i!}\right)}{\left(\sum_{i=0}^{i=\frac{n-1}{2}}\left(\frac{1-q}{1-p}\right)\left(\frac{q}{1-q}\right)^{\frac{n-1}{2}-i}\left(\frac{1-p}{p}\right)^{\frac{n-1}{2}}\left(\frac{n!}{(n-i)!i!}\right)+\sum_{i=\frac{n+1}{2}}^{i=n}\left(\frac{1-p}{p}\right)^{n-i}\left(\frac{n!}{(n-i)!i!}\right)\right)} \tag{6.10}$$

or, more simply,

$$\frac{P(n,p)}{\left(\left(\frac{(1-p)p}{(1-q)q}\right)^{n+1^2}(1-P(n,q))+P(n,p)\right)} \tag{6.11}$$

where $P(n, p)$ is the probability that in at least half of n trials (with replacement) the event with probability p will occur. The following proposition can then be proved:

Result 6.3 *If $p > q$ then for a sufficiently large n, the expected proportion of periods in which the efficient candidate is elected is arbitrarily close to* 1.

Proof It suffices to show that, in the denominator of the expression given above $\left(\frac{(1-p)p}{(1-q)q}\right)^{\frac{n+1}{2}}(1-P(n,q)) < P(n,p)$. By exploiting the symmetry of the binomial rewrite this condition as,

$\sum_{i=\frac{n+1}{2}}^{i=n}\left(p^i(1-p)^{n-i} - \left(\frac{(1-p)p}{(1-q)q}\right)^{\frac{n+1}{2}} q^i(1-q)^{n-i}\right)\frac{n!}{(n-i)!i!}$ or,

$$\sum_{i=\frac{n+1}{2}}^{i=n}\left(p^{i-\frac{n+1}{2}}(1-p)^{\frac{n-1}{2}-i} - q^{i-\frac{n+1}{2}}(1-q)^{\frac{n-1}{2}-i}\right)(p(1-p))^{\frac{n+1}{2}}\frac{n!}{(n-i)!i!}.$$

Now note that $p^{i-\frac{n+1}{2}}(1-p)^{\frac{n-1}{2}-i}$ is increasing in p for the relevant values of i and we have the result required.

It is interesting to note that the result holds even when p is less than $\frac{1}{2}$. When this is true, the incumbent is judged inefficient more often than not and so there is a tendency for narrow majorities and relatively frequent changes of regime. However, because q < p it will still be the case that the majority will usually be for the efficient candidate and that, as n increases, the proportion of time in which the relatively efficient candidate wins rises as well. When $p > \frac{1}{2}$ then the dynamics of the system depend in part on the relationship between q and $\frac{1}{2}$. If $q < \frac{1}{2}$ then as n increases

large and stable majorities for the efficient candidate will be the norm. However when $q > \frac{1}{2}$ though stability is still a feature of the system we would now expect to see extended periods of majority for each candidate with relatively infrequent changes of regime.

In this simple environment, citizens obtain useful, albeit imperfectly understood, signals of the true worth of candidates. In what sort of systems would this not be the case? If, for instance, citizens place too much weight on the judgement of their fellow voters, then party X could end up being repeatedly re-elected simply because it was placed into power in the past. For such a society, increasing the value of n would not lead to greater efficiency.

6.4.2 Term Limits

As noted above, the 'candidates' can be interpreted as long-lived political parties with consistent patterns of efficiency. In other contexts, this makes less sense: individual politicians have only a limited lifespan and in some countries there are limits on the terms that political leaders can serve. Let us consider the case where a politician can serve a maximum of m periods in office. When faced with unknown candidates, citizens cast their votes for either with probability $\frac{1}{2}$. Once a leader has served one period in office, each existing voter makes a judgement about the incumbent's efficiency and sticks to that view for the rest of their voting life. In such a framework, there are two important cases to consider: $m < (n + 1)/2$ and $m \geq (n + 1)/2$. In the first case, since all living voters make the same judgement in the period following the candidate's initial election, then either they immediately vote the incumbent out of office, or he or she serves a full term of m periods. Even if, during that period, the incumbent consistently sends signals of inefficiency there will not be enough new voters during the remaining $m - 1$ periods for the office-holder's majority to be overturned. This is where the condition $m < (n+1)/2$ comes from.

Beginning at a date where an election involving fresh candidates is held, four possible sequences of events are possible for the case where $m < (n + 1)/2$:

- With probability $p/2$, the efficient candidate is elected; they produce a signal of efficiency in their first term of office and serve out a full term of m periods.
- With probability $q/2$, the inefficient candidate is elected; they produce a signal of efficiency in their first term and serve out the m periods;
- With probability $(1-p)/2$ the efficient candidate is elected; they produce a signal of inefficiency in their first period and are replaced by the alternative, inefficient candidate who serves for m periods.
- With probability $(1 - q)/2$ the inefficient candidate is elected; they produce a signal of inefficiency in their first period and are replaced in office by the efficient candidate who serves for m periods.

Putting this together, the expected proportion of time in which the office is occupied by the efficient candidate is therefore, $\frac{(1-p)+mp+(1-q)m}{2(1+m)-q-p}$. The variable n does not

appear in this equation, hence there is no tendency for the changes in the lifespan of voters to affect the probability of efficiency. Meanwhile, as m increases, the proportion of time where the incumbent is the efficient candidate tends to $(1 + p - q)/2$, meaning that there is no strong tendency for the efficient to supplant the inefficient.

When m is sufficiently large then we move to the second case, where $m \geq (n + 1)/2$ and dismissal for the poorly performing leader becomes feasible. Though it aids the chances of efficiency this case does not converge to the previous model (of infinitely lived parties) because no candidate or potential candidate can live longer than n+1 periods.

6.4.3 Candidacy

Although I have presented the model in the context of a chapter on political economy, the technology by which candidates are chosen is not so far clear. In the typical citizen candidate model (Osborne and Slivinski 1996) and (Besley and Coate 1997), individuals face a fixed cost of presenting themselves for election. Once elected they choose policies according to their own preferences. These policies determine the payoffs to citizens. In the election itself, each individual votes so as to maximize his or her own net payoff, and takes as given the voting strategies of other citizens. The citizen elected is the one with the most votes. If two or more candidates tie for first place, then the elected leader is chosen at random from the set with equal votes. A political equilibrium then consists of two parts:

1. a Bayes-Nash equilibrium in voting, given a set of candidates;
2. a set of probabilities of standing for election, one for each citizen.

Applying the same notions here, we must firstly relax the assumption of Bayesian information processing. We also need to flesh out some aspects of the beliefs of citizens. Consider for example a candidate with previous experience of power. On some occasions the individual has sent a positive signal about their ability; on other occasions he or she has sent a negative signal. If a Bayesian citizen observes that this individual is re-elected, then he or she might infer something about the beliefs of other individuals and hence might come to revise their own beliefs. However, it is more in keeping with the notion of confirmation bias to suppose that this information is not employed. In other words the citizen sticks to his or her beliefs come what may possibly because he or she believes that any citizen who has observed the same event thinks in the same way. Note that this means that, even in equilibrium different citizens might hold different and contradictory beliefs. That leaves unresolved the issue of the new citizen who observes their first election without fixed views and we also have to suggest beliefs for those citizens too young to observe previous terms in office for some of the candidates.

One possibility is that the young have no view, perhaps because only personal observation counts. In this case the young will attach the same prior to this experienced individual as they would to someone who had never been in power. However,

another possibility is that they make some inference on the basis of candidate's current status. Someone who once was in office, but is currently not, must at some time have been perceived as inefficient by a majority of the then living. Similarly, someone who is an incumbent must either have just served their first term or they must have been successfully re-elected - meaning that in the previous period a majority of the population viewed them as efficient.

To be more specific, suppose a new voter observes (a) the election of a candidate with a history and (b) a negative signal from their period in office. Confirmatory bias might suggest that they subsequently view the candidate as inefficient, but it is also possible that they might attach some weight to the views of other, experienced, voters. In what follows I shall assume that the individual forms his or her initial views only on the basis of observed performance. After all, confirmatory bias arises in part out of a tendency to give undue weight to personal experience, at the expense of the experience of others, (Greenwald 1980).

So, in this context, a political equilibrium is: (1) a Nash equilibrium in voting, given a set of candidates and given the beliefs of the citizens and (2) a set of probabilities of standing for election, one for each citizen.

Consider the case where no individual has private information about their type, then all citizens will be in a symmetric position with regard to their candidacy. Suppose the cost of standing for election is $e > 0$. Let U^0 be the per period payoff when no candidate stands and let U^I, U^E and U^U be the expected per period payoffs when, respectively an inefficient, an efficient and a citizen of unknown efficiency are in office. I shall suppose that, $U^0 < U^I - e < U^U - 2e < U^E - 3e$, meaning that an inefficient individual will always wish to run for office when the alternative is no candidate, a candidate of unknown efficiency will prefer to be in power rather than having an inefficient leader and so on. Given this assumption, when there is no individual standing for re-election there is a mixed strategy Nash equilibrium in which all citizens stand with some probability p, and one is voted into office at random.

Now consider the situation where there are living individuals with a history of serving in office. Call such an individual a $veteran$. In theory there may be up to n such individuals at any given time. If there is one citizen who has been previously elected, then there is some public information about their efficiency. Given that individuals do not differ in their tastes, only in their efficiency, then this citizen will share the views of his or her fellows who observed their performance. In other words, if the outcome of their previous term of office was positive, then along with all their fellow citizens who are as old or older, they will believe themselves to be efficient. Similarly if the outcome in their previous term was negative, then they will believe themselves inefficient.

Consider an individual who was elected for the first time in the previous round and then performed poorly. Call this individual a $failure$ and use the term *distribution of outcomes* to mean the distribution of probabilities of election over the set of citizens.

We have the following:

Result 6.4 *(i) In the view of the failure the addition of the failure to the set of candidates does not alter the distribution of outcomes provided that there is at least one candidate who is not inefficient; (ii) if, in the view of the failure, all members of the set of candidates are inefficient then adding the failure to the set does not change the expected payoff as viewed by that individual; (iii) if at least one other candidate stands for sure, then the failure will not stand; (iv) there is no political equilibrium in which the failure stands with positive probability.*

Proof For (i) consider the failure's viewpoint: Novice citizens do not vote. Experienced citizens will share the failure's rating of candidates. If there is at least one novice candidate or efficient candidate then no voter will vote for an inefficient candidate. Since the failure believes that other voters share the failure's perspective, then he or she believes that no individual will vote for the failure were he or she to stand or alter the distribution of their voting probabilities over the other candidates. For (ii), since all the candidates are inefficient, each has the same expected value which is equal to the expected value of the failure. Part (iii) follows from (i) and (ii) and the fact that the cost of standing is strictly positive. For part (iv), suppose otherwise, in which case $pU^0 - e + U' > U''$ where p is the strictly positive probability that no other candidate stands apart from the failure, U' is the expected payoff given that some other candidate stands and U'' is the expected payoff if the failure does not stand. Given 1 and 2, $U'' = U'$. Hence, the failure stands if pU^0-e is non-negative. Consider the viewpoint of a potential novice candidate. From his or her perspective, she or he should stand if $pU^U - e + (1-p)U^U > (1-p)U^U$ – that is, she or he will stand if $pU^U - e$ is non-negative. But $U^U > U^0$, hence if the failure stands then so must at least one other, non-inefficient candidate, which contradicts (iii). Hence, the failure does not stand.

Now consider a candidate who, after being elected for the first time, gives a positive signal about their performance. Call this individual a *success*.

Result 6.5 *After the first election of a success, unless he or she dies first, there is a sequence of at least n/2 periods in which there exists an equilibrium where (i) the success stands for sure; (ii) no other experienced candidates stand; (iii) the success wins. This minimum sequence of wins is extended by 1 period every time the success sends a positive signal.*

Proof When the candidate wins for the first time and provides a positive signal, all n+1 living cohorts label him or her as efficient. This includes any experienced candidates who will therefore never wish to stand when the success stands. The only candidates who might wish to stand against the success must therefore be born after his or her initial victory. Since such candidates are novices, when running against the success they will only attract the votes of individuals who, in their first period of life, observe a negative signal from the winning candidate. It follows that for at least n/2 periods the success will win against any other candidates.

Alternatives might be proposed within this model: each individual might know their own ability, in which case, for this basic model, only the efficient would stand

for election. With private information of that kind therefore efficiency is maximised by citizen candidates. More generally, it is clear that the 'first impressions count' feature of confirmation bias tends to carry over to the nature of the political equilibrium. The paths of the failure and the success diverge rapidly. For the former, there is no point in standing for office again; for the latter at least half a lifetime in office beckons. As such therefore, just as in markets (e.g. Smith 2003) bounded rationality at the individual level is compatible with some degree of efficiency in the outcomes produced by a democratic political system.

6.5 Conclusions: Possible Worlds

In this chapter we have covered a number of issues relevant to the economics of public policy given bounded rationality. In its attempts to explain human behaviour and institutions, economic theory is not usually hampered by a lack of degrees of freedom. By altering the game form or placing restrictions on information sets or strategies, arguably any social phenomena can be explained within the rational choice framework. With collective decision making, where rationality in the set of agents does not imply rationality in joint decisions, this 'freedom to explain' is even greater. As a result, in this chapter I argued that bounded rationality models have their work cut out, if they are to be useful within political economy. One role is to explore the robustness of rational choice predictions and we have done this in a number of ways. Elsewhere, for instance Kollman et al. (1992), use an adaptive model of political party behaviour to show the robustness of the Downsian hypothesis that party positions will converge towards the median voter in a two party system. Alternatively, they may be used to explain behaviour that seems difficult to reconcile with rational choice models. A prime example of this is the phenomenon of voting in a large democracy, Palfrey and Rosenthal (1985). Bendor et al. (2003), use a reinforcement learning model with an aspirational model to explain patterns of voting.

We have also seen that within normative theories of the state, the concept of market failure can survive translation to a world of bounded rationality. However, the strength of the connection between the rational and boundedly rational worlds, depends on the closeness of the relationship between notions of Pareto efficiency with and without rationality. Moreover, the theory of the second best implies that in the presence of bounded rationality, standard solutions and correctives for market failures may not always be welfare enhancing. There is therefore need for more enquiry into notions of market failure, when behaviour is in at least some senses, far from rational.

Finally, we have examined the relationship between rationality and the optimal boundaries of the state. The crucial ingredients here are learning and information: if there are good opportunities for citizens to learn from their experience and bounded rationality lowers the state's ability to acquire superior information about what is good for individuals, then bounded rationality may imply a need for less government

rather than more. As with market failure, though, this is an issue that could benefit from more research.

6.6 Appendix

Proof (Result 6.1) Define $d(x, y)$ as the Euclidean distance between x and y. Then in the standard manner, for the non-empty set, A, $d(x, A) = \min_{y \in A} d(x, y)$ and for the non-empty sets A and B, $d(B, A) = \max_{x \in B} d(x, B)$. Consider the strictly declining infinite sequence $\{\varepsilon_n\}$ such that the limit is 0. Let $\{E_n\}$ be the corresponding sequence of equilibria and let $\{P_n\}$ be the sequence of Pareto sets. By continuity of E and upper hemi-continuity of P, we know that the limit of these two sequences are E and a closed subset of P, $\widehat{P}$ respectively. Note that $E \subsetneqq P$ implies $E \subsetneqq \widehat{P}$. Let $d(\widehat{P}, E) = \delta$, which is strictly positive by virtue of the fact that $E \subsetneqq \widehat{P}$. Consider $d(E, E_n) = \max_{e' \in E_n} d(e', E)$. For any $\delta/3 > 0$, there is n_1 such that for any $n' > n_1$, $d(E, E_{n'}) < \delta/3$. Suppose not, then for some $\delta/3 > 0$, for any n there is some $n' > n$, such that $d(E, E_{n'}) \geq \delta/3$. Take the resulting infinite sub-sequence of e' that have this property, namely $d(e', E) \geq \delta/3$ and relabel it $\{e'_n\}$. We know that this series is bounded. Hence from the Bolzano-Wierstrass theorem it contains a sub-sequence $\{e'_m\}$, that converges and given upper hemi-continuity, it converges to $e \in E$. Hence, there exists m, such that for all $m' > m$, we have $d(e'_{m'}, E) < \delta/3$, which contradicts the construction of the series $\{e'_n\}$. Hence, for $n' > n_1$, $d(E, E_{n'}) < \delta/3$. Repeat this argument for $\widehat{P}$, so that for $n' > n_2$ we have $d(\widehat{P}, P_{n'}) < \delta/3$ and then let $n_3 = \max(n_1, n_2)$. By the triangle inequality, $d(\widehat{P}, E) \leq d(E, E_n) + d(\widehat{P}, P_n) + d(P_n, E_n)$, so for all $n' > n_3$, $\delta \leq \delta/3 + \delta/3 + d(P_{n'}, E_{n'})$ or $\delta/3 \leq d(P_{n'}, E_{n'})$. Hence $E_{n'} \subsetneqq P_{n'}$.

Proof (Result 6.2) Let W^{mi} denote within period welfare from market based provision for periods $i = 1, 2$. We need to show that UBL implies $W^{m1} < W^{m2}$. That is,

$$\int\int (2b'b^* - b'^2) f(b', b^*) db' db^* > \int\int (2bb^* - b^2) f(b, b^*) dbdb^* > 0$$

Abbreviate $\lambda(b, b^*)$ to λ and define $g(b|b^*)$ as the density function for b, conditional on b^*. Note that $g(b|b^*) = f(b, b^*)/\int_b f(b, b^*)db$, so $f(b', b^*) = g(b'|b^*)\int_{b'} f(b', b^*)db' = g(b|b^*)\int_b f(b, b^*)db = f(b, b^*)$. This gives,

$$\begin{aligned} W^{m2} &= \int\int (2((1-\lambda)b + \lambda b^*)b^* - ((1-\lambda)b + \lambda b^*)^2) f(b, b^*) dbdb^* \\ &= \int\int (2((1-\lambda)b + \lambda b^*)b^* - ((1-\lambda)b + \lambda b^*)^2) f(b, b^*) dbdb^* \end{aligned}$$

Define $\underline{\lambda} = \min_{b,b^*}[\lambda(b, b^*)]$. Note that by construction $1 > \underline{\lambda} > 0$. Since the integrand of W_2^m is increasing in $\lambda(b, b^*)$, then, $W^{m2} \geq \int\int(2((1-\underline{\lambda})b + \underline{\lambda}b^*)b^* - ((1-\underline{\lambda})b + \underline{\lambda}b^*)^2)f(b, b^*)dbdb^*$. So, $W^{m2} - W^{m1} \geq \int\int(2((1-\underline{\lambda})b + \underline{\lambda}b^*)b^* - ((1-\underline{\lambda})b + \underline{\lambda}b)^2 - (2bb^* - b^2))f(b, b^*)dbdb^*$ or, $W^{m2} - W^{m1} \geq \int\int([1-(1-\underline{\lambda})^2](b - b^{*2}))f(b, b^*)dbdb^* > 0$ where the last step follows from the supposition that $f(b, b^*) > 0$ for some $b \neq b^*$.

Chapter 7
Standard Fiscal Policy and Merit Wants

7.1 Introduction

In the opening chapter, I defined merit-worthy in a manner which, while quite orthodox, masked the degree of disagreement which has characterised attempts to define the concept in a form useful to economists. As John Head put it in his retrospective of 1990:

> As compared with social wants, however, the merit wants concept raises methodologically much more difficult and controversial issues symbolising as it does, for the public finance literature, many of the doubts and reservations which have been expressed over generations by economists of varying political persuasions regarding the ultimate normative authority of the consumer sovereignty principle. (p. 211)

One of the tasks of this chapter is provide an overview of alternative definitions of merit wants. Apart from various attempts to delineate the concept, the existing literature has been marked by two other features: its small size and the acknowleged lack of progress in formal modelling of the implications of the existence of merit goods for public policy. In part, the deficiencies of the policy literature arise from the absence of a good understanding of the psychology of economic behaviour. As a result, most of the writers on the subject have been content to assume that consumers systematically undervalue one (or more) goods, relative to the weight which would be placed on the merit good by a benevolent social planner. This approach has its uses. Policies based on tax incentives or direct allocation of the goods involved can be assessed without requiring a detailed examination of the source of the gap between preference and welfare. As a result, the tax-based models (e.g. Sandmo, 1983; Besley, 1988; Munro, 1992 or Racionero, 2000) probably give a good guide to the optimal use of taxes and provision in-kind. However, they cannot say much about the effect of alternative policy instruments, such as information provision nor how the policies ought to be framed. In addition, as long as the source of the difference between preference and welfare is unspecified then, unlike standard optimal tax models where estimates of elasticities and preferences can be employed to simulate optimal policy, models of merit want tax policy cannot be used to calibrate the optimal tax system.

A. Munro, *Bounded Rationality and Public Policy*, The Economics of Non-Market Goods and Resources 12, DOI 10.1007/978-1-4020-9473-6_7,
© Springer Science+Business Media B.V. 2009

A tour of most of the merit want cases is provided within Section 2 of this chapter. At various times in the lifespan of the concept, some examples of 'merit wants' have clearly trespassed on the other forms of market failure. For completeness, the more common of these examples are included here. The classification system used follows John Head's taxonomy (see also Hillman, 1980 or ver Eecke, 2001) – at the individual level, separating out the rankings created by preference, choice and welfare, then, at the social level distinguishing between individualistic and non-individualistic social welfare functions.

7.2 Individualistic Welfare Functions, Private Failure

An individualistic social welfare function is one that depends only the individual welfare of the citizens who make up society. Thus it conforms to the neutrality property of Paretian welfare economics discussed in the Chapter 1.

In his classification of possible merit want arguments, Head begins with the distinction between preference and welfare. An operational definition of this distinction may not be possible, since both preference and welfare are potentially unobservable, but conceptually the idea is clear. As pointed out in Chapter 1, failures of preference and welfare to accord can be attributable to two kinds of problems: incoherent or incomplete preferences on the one hand or, on the other hand, mis-aligned preferences – which occur when individuals have complete and transitive preference rankings and choose according to them, but the preference rankings do not accord with their welfare counterparts.

Pigou (1920) was one of the first economists to point out one reason why preference may not match welfare: intertemporal choices may be distorted by our limited ability to imagine future consumption with the same intensity as the consideration of immediate gratification.

> But this preference for present pleasure does not – the idea is self-contradictory – imply that a recent pleasure of given magnitude is any greater than a future pleasure of the same magnitude. It implies only that our telescopic faculty is defective, and that we, therefore, see future pleasures, as it were, on a diminished scale. Pigou (1920) 96–97.

As a result there would be a systematic tendency for the over allocation of resources to immediate gratification, compared to the allocation which would maximise welfare.

The possibility of intertemporal preferences which do not match welfare is of obvious importance in any discussion of pensions and savings, so my discussion of this concept is deferred to Chapter 10, but it is also the case that what may appear to be a gap between preference and welfare is in fact a gap between preference (which may accord with welfare) and choice, because of a failure of will (Baumeister 2003).

One of the main battlegrounds in debate over what constitutes a merit want has centred on the position of individuals who make choices based upon the wrong information. One view is that only choices made on the basis of full information can be viewed as welfare-maximising. Musgrave (1987) for instance, takes it as obvious

that rational choice requires correct information, yet it is not obvious that this is in line with standard models of the consumer. An individual who makes a choice in a condition of uncertainty may, prior to the actual choice, invest in the acquisition of information. The rational consumer equates the anticipated marginal benefits of the investment with its marginal cost, but unless the marginal cost of information is zero, and the time costs of information processing are also zero, then it will not normally be optimal to be fully informed. There is therefore nothing per se about information that makes it a merit want – it is a good along with any other good. Of course the reasons for its purchase are often instrumental, as a means to making an informed choice amongst other goods. But then this just places it in the category of production rather than consumption goods.

So for imperfect information to be a source of market imperfection, either the social benefit (or cost) of information must differ from its private value or, as McClure puts it, 'one must accept the existence of widespread inabilities to gather and process the information necessary for making self-interested decisions.' (p. 181). Both are possible, though the first would arise because of quite standard market failures, such as external benefits from information production (see Munro and Hanley, 1999, for instance).

As the quote from McClure indicates, the second feature of information processing failures – 'widespread' – is also an important part of the argument. John Stuart Mill, hardly an advocate of large government, was in favour of compulsory consumption in certain cases, mostly on the grounds of imperfect information. Beginning by observing that 'the uncultivated cannot be competent judges of cultivation' (Mill, 1891, p. 953), he argues that individuals are probably best judges of material wants but not those things which create a better character. But beyond that there are basic elements of knowledge, which should be inculcated in all citizens and for the achievement of which it is acceptable to impose restraints on parents. The quid-pro quo is that such education should be free or at most provided 'at a trifling expense' (p. 954). Ignorance therefore provides some reason for regulation because of the bootstrap nature of knowledge: the ignorant cannot know what they do not know and thus cannot place an opportunity cost on the deficiency.

Many modern writers have been more sceptical about a general argument that ignorance implies the desirability of compulsion. McClure (1968) is one of those who favours the provision of information or subsidising its provision, rather than transfers in-kind or any other direct manipulation of consumption. In a later argument, McClure (1990) he also points out that the government or its appointed experts is not always better informed than the public it is supposed to serve, as the historical cases of thalidomide or asbestos illustrate. Head (1966) on the other hand, is readier to accept other dimensions to policy. His list of possible examples of state intervention includes information provision, but he adds taxation of advertising, regulation of food and drug safety, controls on the professions and subsidies for meritorious goods.

The absence of mental capacity can arise in two profoundly different ways: either through learning problems (mental deficiency) or through madness. The observation that preference and welfare differ for children, the insane and the mentally deficient

is the starting point for many discussions of merit wants, but Musgrave (1987) classifies these cases as pathological, arguing that their exceptional nature means that they should not be seen as 'part of the essential merit good problem' (p. 452).

This neat separation of the mentally capable from the incapable is, however, much harder to sustain in real life. Chapter 3 took us on a tour through some of the most common and well-documented ways in which information processing falls short of Bayesian perfection. In many instances these deficiencies appear to be widespread, even in so-called experts. It therefore seems more sensible to think of a continuum of individuals and also a continuum of problems of differing complexity. Choosing whether to buy a shampoo with or without built-in conditioner may be straightforward; choosing sensibly whether to opt for a bypass operation or angioplasty may involve much higher levels of cognitive effort.

In the context of merit wants, children raise many of the same issues as mental deficiency. It is often taken as a given that consumer sovereignty does not apply to non-adults. In practice, even with young children there is in most families some respect for their views, whether it is only over the choice of sweets and the spending of pocket money. The law generally takes a pragmatic view of the boundary between child and adult, as it does between senility and mental competence. In the UK for instance, the person of 17 years and 11 months is not mature enough to buy alcohol. Another month and they may buy all they wish and consume it alongside the cigarettes they were allowed to purchase once they reached 16. This sharp, legal distinction between mentally fit and unfit is clearly a tidy solution to the question of when someone becomes an adult, competent to make their own choices, but it hides the underlying continuum. At an early age, children are able to express sensible preferences about favourite colours, books or playtime activities. In many families children may have some input into their schooling, whether to opt for the violin or the trumpet for instance, but other choices are reserved for the adults in the household.

Fundamentally therefore the picture is more like that depicted in Figure 7.1. The shaded area represents cases which are merit worthy. Individuals with strong cognitive powers, faced with goods which are simple to evaluate are less likely to be meritworthy compared to goods which are complex to evaluate[1] faced by individuals such as children whose abilities to choose sensibly are not yet developed. The difficulty, as McClure points out, is in drawing the boundary.

A further category of merit goods arises when individuals know what is best for them and, at least in some sense, prefer what is best, but nevertheless choices are at variance from preference and welfare. The best documented case of an apparent difference between preference and choice arises in the case of addiction, defined by the Oxford English Dictionary to be, 'doing or using something, esp. a drug as a habit or compulsively.' It comes within the more general category of failure to obey one's own preferences when making choices. Head (1990 p. 226) adds to this

[1] The huge variety of goods that have been suggested as merit goods includes education Arcelus and Levine (1986), organic foods Mann (2003), and family planning, Godwin (1991).

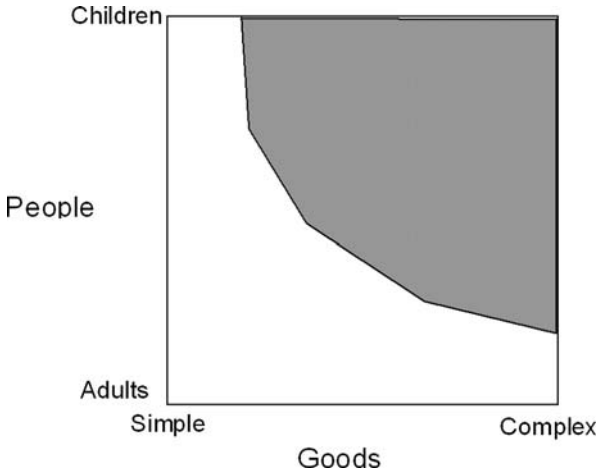


Fig. 7.1 Merit Wants and Merit People

list the case where 'tastes' differ from 'moral values'. E.g. prostitution or racial discrimination. He argues that, although policy-making in any society marked by heterogeneity of moral values may be nigh on impossible, state interference in the case of the examples just cited is acceptable because of the close analogy between them and the problems of weakness of will.

7.2.1 Individualistic Welfare Functions, Public Failure

This section deals with definitions of merit goods which involve some element of social interaction. The most obvious such case is that of public goods where individuals have mis-aligned preferences or the wrong information about the value of the good, what Head (1966) calls *social goods*. For mis-aligned preferences, the public goods context adds little to the definition although it may change the optimal policy dramatically. Consider for instance, two identical individuals with preferences summarised by the function $z - z^2 + \theta x$, where z is expenditure on level of the public good and x is expenditure on the private good and $\theta > 0$. Suppose that each individual acts non-cooperatively allocating their income between the private and public good, then the Nash equilibrium level of expenditure on the public good is $(1-\theta)/2$. Suppose that individual welfare is given by $z - z^2 + x$ and the government is utilitarian, then the optimal level of $z = \frac{1}{4}$. It follows that if the individuals undervalue the private good, in the sense that $\theta < 1$, then the non-cooperative outcome will be closer to the welfare optimum, compared to the situation where individuals value the private good accurately. In this example, though not always in general, subsidizing consumption of an undervalued good may therefore reduce welfare, because that undervaluation serves to correct another source of market failure in a manner familiar from the literature on second-best policy. Conversely, if it is the public good which is undervalued, then this may exacerbate another market failure,

while if the public good is overvalued, again one market defect may counterbalance the other.

For some situations, the idea of wrong information or ignorance about the value of public goods may be more credible than in the case of private goods. There are two reasons for this. The first is that information is very often a public good itself. Hence investment in its acquisition may be too low and ignorance could be rational. Of course this argument also applies to private goods, but it is likely to be stronger with public goods because all individuals consume the same good. With private goods, homogeneous consumption is much rarer and so the research done by one individual about the taste of toothpaste, for instance, is likely to be of far less general interest than research done about the effect of a new flood defence scheme on the likelihood of inundation.

Strictly speaking the second reason does not apply to public goods so much as to publicly provided or non-market goods. Many, but by no means all public goods fall into this category, while in many countries private goods such as education or healthcare are provided by the state or the state sets such tight regulations over their use that individuals have very little control over consumption levels. Where this is the case, there is little or no incentive for the individual to become informed about the value of the good. Ignorance truly is rational because there is no decision to be made. Consequently, if ignorance is seen as a source of merit-worthiness then public provision of a commodity may well be the cause of that ignorance.

7.2.2 Two Spurious Categories of Merit Wants

In addition to the general problems of wrong preferences, there are two specific arguments which crop up when merit wants and social interaction is discussed: psychic externalities and the inefficiency of Nash equilibrium.

Many individuals do not care about the general well-being of their fellow humans, but about their consumption of specific goods, such as healthcare, food and basic housing. Tobin (1970) called this specific egalitarianism. Hochman and Rogers (1969) produced a general rationale for redistribution based upon interdependent utility functions and later writers extended the analysis to the case of specific egalitarianism. In all cases the problem can be seen as one of externalities (and is correctly described and analysed as such by Hochman and Rogers). If the concern for person B's welfare or consumption pattern is unique to person A, then the interdependency comes under the category of depletable externalities – goods with private, external effects Baumol and Oates(1975). If the concern for B's welfare or consumption is shared amongst more than one person, then an element of non-excludability enters the equation and the external effect becomes non-depletable. But in both cases, the situation is covered by the existing notion of externality and the remedies for the Pareto inefficiencies, which might arise, are the familiar ones from the market failure literature.[2]

[2] Mulligan and Philipson (2000) argue that these paternalistic motives on behalf of some voters can create large deadweight losses for the recipients of transfers and subsidies in-kind.

In his 1987 summary of merit good arguments, Charles McClure Jr. offers the example of competitive pressures, for instance when stores selling alcohol stay open late, running the increased risk of late-night robberies. Mill (1891) has several similar examples in his fifth exception to the laissez-faire principle, including the incentives faced by individual workers who wish to cut their hours of work, and the case of colonialists impeding development of an empty land by claiming excessive stakes, thereby limited the population growth necessary for prosperity. Each of these is an example of a Prisoner's Dilemma, but in each case the yardstick of inefficiency is individual preference, so though these cases may involve market failure, it is not clear why they should come under the rubric of merit good.

Similar examples of Prisoner's dilemmas are created by the pursuit of status signals and choices based on peer pressure. Children, and adults for that matter, may feel constrained to indulge in activities because of a desire to be part of a group or to be ahead of a group. As a result, 'prohibitions that may appear to violate consumer sovereignty may actually be quite welcome if they help us to strengthen our backbones (or that of our children).' McClure (1990 p. 183). Seatbelt laws may be a good example of this kind of constraint.

Ireland (1994) sets out a lucid analysis of optimal taxes on status signals, but what he also makes clear about his model is the externality nature of the status signal or peer pressure. So, as with psychic externalities, status signals fall squarely into a standard market failure classification and he treats them as such.

7.2.3 Non-individualistic Welfare Functions

The situation so far treated in this chapter is that where preference and choice are potentially different from individual welfare, but social welfare is itself a function only of individual welfare. In other words the strong neutrality property of welfarism is respected. However, some proponents of merit goods – most notably Musgrave himself, argue that it is in the clash between individual and social values that merit wants most clearly arise. In his 1987 contribution to the Palgrave Dictionary of Economics, after considering the cases of public goods, children's consumption and the 'rule of fashion', Musgrave identifies the concept of community preferences thus:

> . . . consider a setting where individuals, as members of the community, accept certain community values or preferences even though their personal preferences might differ. Concern for maintenance of historical sites, respect for national holidays, regard for environment or for learning and the arts are cases in point. (p. 452)

Twice he states that community preferences are at the core of the merit wants concept, writing that 'as this author sees it, this [community preferences] is the setting in which the concept of merit or demerit goods is most clearly appropriate, and where consumer sovereignty is replaced by an alternative norm' (p. 452).

An example of such a community preference might occur in the case of Stonehenge. This Neolithic monument is largely inaccessible to the public, protected as it is by a wire fence. Many people might wish to touch the stones or to sit on some of

them for a picnic or to have their photograph taken. The stones also offer an ideal clambering opportunity for young children, but all this is forbidden. As individuals many Britons object to the restrictive policy, and yet at the same time support it, because it is only by limiting access that the Stonehenge may have a chance to survive another 3–4,000 years, to be enjoyed as a national monument by future generations.

Although many similar examples come readily to mind, it is not clear that there is really is a clash between individual and social welfare here. For, as Musgrave himself emphasises, it is the individuals of the community themselves who accept and promote the values. In other words, the social preferences are derived only from individual preferences. These preferences might be directly for the national monuments, or they might be for the welfare of future generations, they might even be higher level preferences (Harsanyi 1955). However, unless there is a failure of will problem, of the kind described above, or the social values are not derived from individual preferences the community values case does not appear to constitute a merit goods category.

7.3 Objective Functions

Since the 1960s there has been a unsteady trickle of papers which derive optimal corrective policies in the presence of merit wants. The welfare function is the starting point for any formal discussion of policy. Three broad classes of function can be found in the literature, each of them corresponding to one of the classes of merit wants identified above. We consider the implications of each of them in turn. Following that, I cover models of optimal fiscal policy in the presence of merit wants. For readers familiar with optimal tax models this section is fairly straightforward and the results largely intuitive. One issue that is frequently raised within the merit wants literature – particularly by critics hostile to the concept – is whether tax-based policies are inferior to those based on the provision of information. I show that, within the context of standard models of merit wants, information-based measures to correct the divergence between preference and welfare may be inferior to tax-based instruments and, in some cases positively counterproductive, in the sense that they reduce welfare. I end the chapter with a summary of the lessons to be learnt from the merit wants literature, together with a list of some its problems.

Let x denote the vector of non-merit goods, with y as the vector of merit goods. Suppose that preferences can be represented by the utility function u(x, y) and individual welfare by $w(x, y)$. The social welfare function is $W = W(w^1, w^2 \ldots, w^H)$, increasing in all its arguments. Thus the function is individualistic, but does not necessarily respect individual preferences.

As noted in the first chapter, in this general formulation, it is not possible to state which are the merit goods and which are the non-merit goods. The function $w = u(x, v(y))$ is a more specific form which in which the merit good are separable from non-merit goods. Besley (1988) for instance, provides one of the simplest

model of merit wants when preferences do not match welfare. In his story, welfare is given by $u(x, \theta y)$, where y is the single merit and θ is an index representing the degree to which an individual under or over-weights the merit good in their preferences.[3] A special case of this function, frequently found elsewhere, is the additively separable form, $w = u(x) + \theta v(y)$. In Racionero (2000) for instance, some agents are assumed to maximise $u(x)$, while others have a strictly positive, but still incorrect value of θ.

Non-individualistic social welfare functions actually predate the individualistic form (e.g. Pazner, 1972; Wenzel and Wiegard, 1981). The typical functional form is $W = W(u^1 \ldots u^H; y^1 \ldots y^H)$, so that social welfare is taken to be a function of individual welfare (here assumed to be identical to preference) plus extra arguments for the merit good. For a merit good, the function is increasing in the y^hs; for a merit bad the converse is true. Often some form of separability is assumed, as in the form, $W = W(u^1 \ldots u^H) + g(y^1 \ldots y^H)$. In Tobin's model of specific egalitarianism, for instance, the standard social welfare function is amended by the addition of the strictly concave function $g(y^1, \ldots, y^H)$. However, if it is the total consumption of the meritorious good which matters, and not its distribution, then $g(\sum y^h)$ would be appropriate.

Formally, the differences between the case of a non-individualistic welfare function and the case of an individualistic welfare function with a merit want are not always identifiable. For instance, if preferences are additively separable between the first $n - 1$ goods and the nth, the social welfare function, $\sum U^h(x^h) + u^h(y^h)$ – utilitarian, but with 'corrected preferences' is equal to $\sum U^h(x^h, \theta^h y^h) - \sum w^h(y^h)$ where $w^h(y^h) = [u^h(\theta^h y^h) - u^h(y^h)]$ – a function which respects preferences at the individual level, but then adds in (or subtracts) an extra term for the consumption of nth good at the social level. A more general proposition on this point is supplied by Baigent (1981). In his model, behaviour, $B(x, y)$ depends on preferences or tastes, T(x,y) and values, V(x,y). He then examines the conditions under which values can be written as a function of behaviour and the merit want, i.e. $V = f(B, x)$, showing that necessary and sufficient conditions are,

1. That there exists a sequence of functions $V_0, V_1, \ldots$ such that,

$$\frac{\partial V_\alpha}{\partial x} = V_{\alpha+1} \frac{\partial B}{\partial x} \forall \alpha = 1, 2 \ldots$$

2. The sequence in 1 converges uniformly on the range of B.

A special case of this is the additive form discussed above, which means that it is not always possible to determine from the social welfare function alone, whether the social welfare function is non-individualistic or whether it is individualistic but does not respect individual preferences.

[3] Feehan (1990) Capéau and Ooghe (2003) and Schroyen (2005) update and correct the Besley model.

7.3.1 Risk and Uncertainty

In principle, the distinction between preference and welfare functions used in the literature on merit wants policy under conditions of certainty could also be pursued once an element of risk enters choice. However, in practice, in the few papers that have tackled the subject of merit wants and uncertainty, the focus has been firmly on the 'wrong' probabilities concept. Sandmo (1983) begins his paper by clarifying the distinction between ex ante and ex post notions of social welfare. With the ex ante approach, the social welfare function is, $W = W(E^1[u^1], \ldots, E^H[u^H])$, where E is the expectations operator and the superscripts indicate that the expectation is taken relative to each individual's beliefs about the true state of the world. In the ex post framework, social welfare is defined as $W = E[W(u^1 \ldots u^H)]$, where the absence of a superscript on E indicates that the expectations is taken with regard to the social planner's beliefs. Hammond (1981) shows that these two formulations yield the same function only under two specific conditions, namely,

1. all individual beliefs are identical and equal to the social planner's.
2. the function W is linear.

Thus if either individual beliefs differ from those of the social planner or the social planner does not respect individual attitudes to risk, then individual choice will not necessarily maximise individual welfare, as viewed by the planner. Individual levels of risk aversion may not reflect individual welfare – for instance the reckless gambler. When discussing social welfare functions under uncertainty, it is more usual to refer to Harsanyi's (1955) paper which derives the additive form of the social welfare function from a contractarian perspective, in which a population of individuals must choose what society should maximise without knowing which identities they will take on within the economy which is to follow. The device makes future circumstances a lottery and leads to Harsanyi's position that attitudes to inequality should reflect attitudes to risk.

The convenient form of welfare function used by Sandmo (1983) and adapted by Munro (1992) takes preferences to be $u^h(x^h) + \sum_s \pi_s^h v_s^h(y^h)$, where π_s^h indicates the probability with which individual h believes state of the world, s ($s = 1, \ldots, S$) will occur. Thus only the value of consumption of the merit good or goods is subject to uncertainty. The planner, meanwhile, has beliefs π_s that may or may not accord with those of the individual. Social welfare is then, $\sum_h \left(u^h(x^h) + \sum_s \pi_s v_s^h(y^h)\right)$, where the absence of the h superscripts on the probabilities marks the difference between the ex ante and ex post approaches.

7.4 Optimal Policies

Almost all of the small formal literature on the subject of merit wants has taken place within the public economics tradition. Hence it is not surprising that the instruments considered are income and commodity taxation plus, on occasion, direct

instruments such as rations or transfers in-kind. The canonical model has the following features:

1. Households, with well-behaved preferences, $U^h(x^h, \theta^h y^h)\, h = 1, \ldots, H; \theta^h > 0$.
2. $n-1$ non merit goods, represented by the vector, x and an nth good, y – the merit good.
3. Fixed producer prices, p_i $i = 1, \ldots n$; price taking behaviour and no profits.
4. Welfare = $W(U^1(x^1, \theta y^1), \ldots, U^H(x^H, \theta y^H))$, strictly increasing and concave.

The assumption of fixed producer prices or perfectly elastic supply is for convenience. With a 100% profits tax, none of the results are changed as long as the production technology is convex. However, with less than 100% profits tax or non-convexities the conclusions of the model may alter, depending on who owns the property rights to profit income and the exact model of imperfect competition chosen.

A crucial issue is the degree to which the income and commodity taxes can be individualised. With more than one individual there are three sources of heterogeneity: in θ, in income, m and in preferences. Optimal taxation to correct the merit want problem must therefore also address issues of vertical equity, plus any other distortions present in the economy. If all taxes can be tailored to the individual – implying the absence of any kind of asymmetric information, then the redistributive and preference correction roles of taxes can be separated. Optimal taxes on merit good are therefore highly intuitive: first-best taxation involves setting the consumer price of the nth good at $\theta^h p_n/\theta$. Thus goods that are undervalued by consumers are subsidised; goods that are overvalued are taxed. Meanwhile other goods are untaxed. To see this consider the simplified model in which the government faces only one consumer with additively separable preferences of the form $u(x)+\theta^h v(y)$. The government's problem is to maximise:

$$u(x) + \theta v(y) + \lambda\,(tx + t_n y + T - R)$$

where λ is the shadow value of government revenue, R, T is a lump-sum tax, t is a vector of per-unit taxes on the non-merit goods and t_n is the per-unit tax on the merit good. To maximise this expression it is helpful to write it in a different way:

$$V(p + t,\, p_n + t_n, m - T) + (\theta - \theta^h)v(y) + \lambda\,(tx + t_n y + T - R)$$

where V(.) is the indirect utility function for the individual and m is income, while p is the vector of non-merit good producer prices and p_n is the producer price for the merit good. Maximising yields the following first order conditions:

$$-\alpha x_i + (\theta - \theta^h)v'(y)y_i + \lambda\left(\sum_j t_j x_{ji} + x_i + t_n y_i\right) = 0.\ i = 2, \ldots, n-1$$

$$-\alpha y + (\theta - \theta^h)v'(y)y_n + \lambda\left(\sum_j t_j x_{jn} + y + t_n y_n\right) = 0$$

$$-\alpha + (\theta - \theta^h)v'(y)y_T + \lambda\left(\sum_j t_j x_{jT} + 1 + t_n y_T\right) = 0$$

In these equations, α is the marginal utility of income, x_{ij} represents the derivative of the demand for the ith good with respect to the jth price, y_i is the derivative of the demand for y with respect to the ith price, x_{iT} is the derivative of the demand for the ith good with respect to T, y_T is the derivative of demand for the merit good with respect to T and $v\prime$ is the first derivative of v.

To understand these equations we can multiply the last of them by each x_i in turn and deduct the result from the first. For the first $n-1$ goods, this produces:

$$(\theta - \theta^h)v'(y)\left[y_i - x_i y_T\right] + \lambda\left(\sum_j t_j\left[x_{ji} - x_i x_{jT}\right] + t_n\left[y_i - x_i y_T\right]\right) = 0 \quad (7.1)$$

Repeating the procedure with y produces the additional equation:

$$(\theta - \theta^h)v'(y)\left[y_i - y y_T\right] + \lambda\left(\sum_j t_j\left[x_{ji} - y y_T\right] + t_n\left[y_i - y y_T\right]\right) = 0 \quad (7.2)$$

By inspection we can see that $t_i = 0, i = 2, \ldots, n-1$ and $t_n = -(\theta - \theta^h)v'(y)/\lambda$ solves (7.1) and (7.2). We can simplify further by noting that consumer optimisation implies that $p_n + t_n = \theta^h v'(y)/\alpha = \theta^h v\prime(y)/\lambda$. Rearranging produces $t_n = -(\theta - \theta^h)p_n/\theta$ or a tax which is proportionate to the difference between the value placed on the good by the individual and its true value. The result accords with standard ideas about first best policies in the face of market failure. The optimal tax or subsidy system internalises the divergence between preference and welfare for the merit goods and leaves the relative prices of other goods undistorted.

In general though individualised taxes are not feasible. If, for instance, income taxes can be individualised, but taxes on merit goods must be uniform, then the optimal tax will reflect mean levels of the distortion between preference and welfare, with individuals weighted by the social marginal utility of their income. It is still the case, however, that if all individuals under value a good (i.e. $\theta^h < \theta$), the optimal policy will be subsidised consumption of the nth good.

Once none of the taxes can be specific to individuals, then the price of the nth good must reflect the other aims of the government. If for instance, it is consumed

Table 7.1 Types in the Racionero Model

	$\theta^h = 0$	$\theta^h > 0$
Low income, m_1	n^{01}	$n^{\theta 1}$
High income, m_2	n^{02}	$n^{2\theta}$

most heavily with those with the lowest marginal weight in the social welfare function, then even if $\theta^h < \theta$ for all households, the optimal policy may even involve a tax on the nth good. These are the basic conclusions of linear taxation models.

Maria del Mar Racionero's article on non-linear taxation and merit wants provides a convenient starting point for a more technical examination of the issues. In her model, $H = 4$. Agents differ along two dimensions: income and preference, with two types along each dimension – giving four potential types in all. Table 7.1 lays out the frequency of preferences within the population in her story.

For simplicity, the numbers are normalised, so that $\sum_i n^i = 1$. There are only two goods, both normal and produced at a constant unit cost of one, and preferences are assumed to have the additively separable form, $U = u(x)+\theta^h v(y)$. Some consumers therefore have no taste for the second good, the merit want. She assumes infinite disposal costs however, so that consumers may be forced to consume some of the merit want as the price of getting the value of x they desire. Policy takes the form of income tax and merit want bundles with consumers unable to top up any merit want given to them, because of the transaction costs. The other key assumptions are that, for all consumers, the welfare function is $u(x)+\theta^* v(y)$, where $\theta^h < \theta^*$[4] and finally that the government cannot observe income or the taste parameter. Consequently, it is constrained in its actions by the requirement of incentive compatibility.

Define a bundle, (t^i, y^i), $i \in \{01, 02, \theta 1, \theta 2\}$ as an income tax and merit good transfer. Types with $\theta^h = 0$ are unconcerned with the amount of the merit want on offer and so will always opt for the same bundle – the one with the lowest income tax. It follows that at most three tax and merit want combinations will be offered by the government. If there are three combinations, then they can be ranked, in the sense that $t^{01} < t^{\theta 1} < t^{\theta 2}$ and $y^{01} < y^{\theta 1} < y^{\theta 2}$. Potentially, this still leaves two incentive compatibility constraints per type, or six in total, but Racionero (2000) shows, if there are three, distinct bundles offered by the government then very few constraints can bind simultaneously. To be more specific, the constraints for the $\theta^h = 0$ types will not bind, while for the $\theta^h > 0$ types, if the type with the lower income is indifferent between bundles, then the higher income type strictly prefers the bundle with the higher level of y. Conversely, if the higher income type is indifferent between two bundles then the lower income type will strictly prefer the bundle with the lower level of the merit want.

In fact, at the optimum there are only two distinct bundles, as can be shown by assuming that there are three distinct bundles and finding that a welfare improvement could then be found for any possible pattern of binding incentive compatibility

[4] The assumption that $\theta^h < \theta^*$ means that all consumers undervalue the good.

constraints. Suppose, for instance, that no constraints bind, then a reallocation of a small element of the resources from the $\theta 2$ type to the $\theta 1$ type could raise welfare – since the latter type has a lower income and hence a higher marginal welfare weighting. Alternatively the resources could be given to the 01 and 02 types, because they too have a higher marginal welfare weighting than the $\theta 2$ type. For the same reason, at least one type must be indifferent between the bundle aimed at the $\theta 2$ type and another bundle. Furthermore, the $\theta 2$ and $\theta 1$ types cannot both strictly prefer their own bundles to the 01 and 02 bundle, because again a transfer of resources to the $\theta = 0$ agents would raise total welfare.

Within the standard non-linear pricing problem (i.e. those without merit wants), the binding incentive constraints are 'downward pointing' in the sense that at the optimum, agents with higher marginal valuations for the good are indifferent between their own bundle and the bundle of the adjacent agent with a lower marginal valuation of the good. On the other hand, the agent strictly prefers his or her own bundle to that chosen by the adjacent agent with a higher marginal valuation. The same turns out to be true in this case. Suppose that, in equilibrium the $\theta 1$ types were indifferent between their bundle and that of the $\theta 2$ types, who in turn strictly prefer their own bundle. Figure 7.2 illustrates the situation.

In this figure, the income tax is shown along the horizontal axis, while the level of the merit good is shown on the vertical. Consequently, the indifference curves for the 01 and 02 agents are vertical lines, like the one shown. Indifference curves for the other two types slope up and to the right, with curves for the $\theta 2$ agents shown as shallower since they have higher incomes (and therefore greater willingness to pay for increments in y), compared to the $\theta 1$ types, and the goods are normal. The broken line at 45° is an iso-revenue line for the government. Three tax and merit good combinations are also shown in the figure, where, given the arguments of the previous paragraph, type $\theta 1$ agents are also presented as indifferent to the bundle

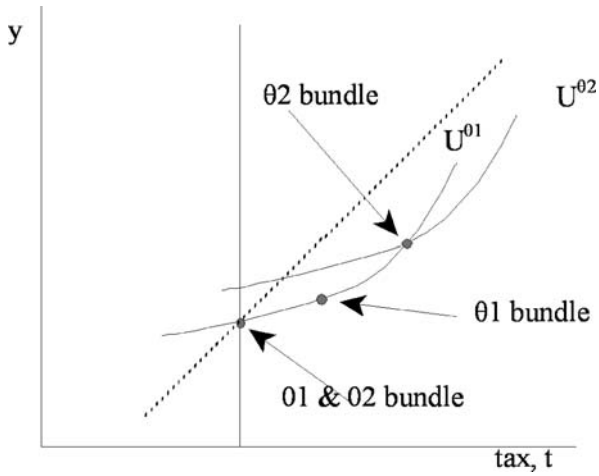


Fig. 7.2 Merit Wants and Incentive Compatibility

chosen by the type $\theta 2$ agents. In addition, they are depicted as indifferent to the bundle chosen by the other two types. This last property is imposed because it has already been seen that there must be some binding constraint that links the $\theta 1$ and $\theta 2$ bundles to the 01 and 02 bundle. If it was the $\theta 2$ agents who were indifferent, then the type $\theta 1$ agents would strictly prefer the bundle aimed at the 01 and 02 types, and so would not choose their own bundle.

Now the situation represented in Fig. 7.2 cannot be optimal. Consider an iso-revenue line, which passes through the $\theta 1$ bundle. If the indifference curve at this point has a slope less than one, then for the $\theta 1$ types, there are points above and to the right of the $\theta 1$ bundle which are feasible, preferred and do not induce any other types to switch bundles. Conversely if the indifference curve has a slope greater than one, then there are points down and to the left of the $\theta 1$ bundle which are feasible, preferred by the $\theta 1$ types and do not induce switching in other types. Hence the slope of the indifference curve at this point must be equal to the slope of the iso-revenue curve. At the $\theta 2$ bundle, the slope of the indifference curve must be greater than one, because if it were not then a Pareto improvement in the shape of a new bundle for the θ^2 types could be implemented, above and to the right of the $\theta 2$ bundle. However, more importantly, a small transfer of resources from the $\theta 2$ agents to the $\theta 1$ agents would raise overall welfare and not break any incentive compatibility constraints.

This leaves only one three-bundle case, in which the $\theta 2$ types are indifferent between their own bundle and that of the $\theta 1$ types, and the $\theta 1$ types are indifferent between their own bundle and that of the types with $\theta^h = 0$. This too is not an optimum. Consider Fig. 7.3. If the bundle aimed at the 01 and 02 types was withdrawn then these types would purchase the $\theta 1$ bundle. This is feasible since the $\theta 1$ bundle is on an iso-revenue line, which generates greater government revenue. It is also welfare-enhancing. Recall that the $\theta 1$ types are indifferent between their own bundle and the 01 and 02 bundle and since consumers undervalue the merit want, then from a welfare perspective, the $\theta 1$ bundle is strictly better for the $\theta 1$ types compared to the 01 and 02 bundle. Since the 01 individuals have the same welfare ranking, the $\theta 1$ bundle is better for them too and since the welfare-indifference curves of the 02 types are flatter than those for the 01 type, then the $\theta 1$ bundle is also better for them, compared to the 01 and 02 bundle. Hence a feasible, welfare improvement is enactable through the elimination of the 01 and 02 bundle.

In short therefore, there are only two bundles in the optimum and one binding incentive compatibility constraint, which applies to the $\theta 2$ agents. We can set out the optimisation problem as the following:

$$Max \sum_i n^i [u(m^i - t^i) + \theta^* v(y^i)] \tag{7.3}$$
$$+ \lambda_1 [u(m^{2\theta} - t^{2\theta}) + \theta v(y^{2\theta}) - u(m^{2\theta} - t^{1\theta}) + \theta v(y^{1\theta})] + \lambda_2 \sum_i n^i [t^i - y^i]$$

For the high θ, high income type, the first order conditions for the choice of tax and expenditure bundles yield,

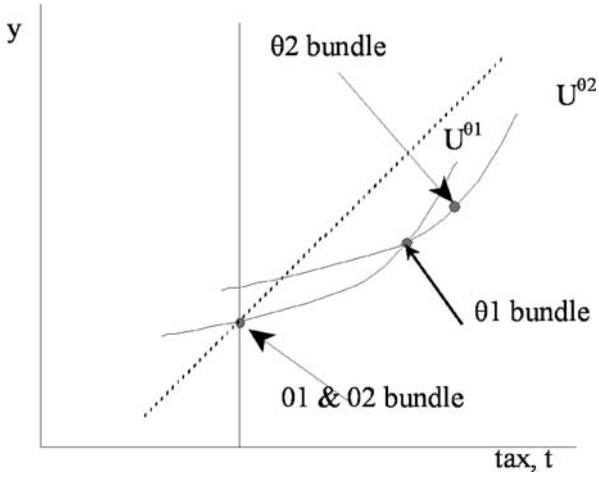


Fig. 7.3 Incentive Compatibility and Merit Wants II

$$\frac{\theta^* v'(y^{2\theta})}{u'(m^{2\theta} - t^{2\theta})} = 1 - \frac{\lambda_1(\theta - \theta^*)v'(y^{2\theta})}{(\lambda_1 + n^{2\theta})u'(m^{2\theta} - t^{2\theta})} \tag{7.4}$$

In the standard, Mirrleesian model of optimal income tax, the right hand side of this equation is equal to one. This is the well-known 'no distortion' rule, which states that the agent of highest income should face a zero marginal tax rate. In the absence of merit wants, the case where $\theta^* = \theta$, then the merit want formula reduces to the standard case. And, in the case of full information, where the incentive compatibility constraint is not relevant, $\lambda_1 = 0$, so again at the optimum the agent faces the marginal rate of substitution of the corrected utility function is equal to the marginal rate of technical substitution. But when the individual undervalues the merit want, then at the optimum, the $\theta 2$ agent should face a marginal tax, when viewed from the perspective of the principal.

However, the formula can be rearranged to yield,

$$\frac{\theta v'(y^{2\theta})}{u'(m^{2\theta} - t^{2\theta})} = 1 - \frac{n^{2\theta}(\theta - \theta^*)v'(y^{2\theta})}{(\lambda_1 + n^{2\theta})u'(m^{2\theta} - t^{2\theta})} \tag{7.5}$$

In other words, from the viewpoint of the agent, the merit good is subsidised at the optimum – and the agent overconsumes the good.[5]

These equations set out the optimal tax and merit good combination, but the results can be implemented by other means. Since all types but one pool, it is

[5] The relevant equations for the other types are less enlightening. All that can be said for sure is that the agents who do not value the good over-consume relative to their private optimum. For more details see Racionero (2000).

also possible to implement the outcome through a system in which the default is public provision of the merit good financed through the tax system. Consumers can choose to opt out of public provision, in which case they pay a higher amount of income tax, but purchase the merit good privately, albeit at a subsidised price. This interpretation of the optimal non-linear scheme is quite similar to the systems of provisions described in the in-kind transfer literature (Blackorby and Donaldson, 1988 or Blomquist and Micheletto, 1996), to which it obviously owes a deal of similarity. It is also similar to what happens in many OECD countries in the provision of healthcare, education or pensions, for instance. However, when there are more than two levels of income and two levels of preference for the merit good, the optimal non-linear tax scheme is likely to be a great deal more complicated and less easily reconciled with the stylised facts of public expenditure.

7.4.1 Wrong Information Models

In Agnar Sandmo's 1981 paper on merit wants discussed above, consumers choose consumption before the true state of the world is known. The value of consumption of the first $n - 1$ goods is unaffected by the state, by the value of the nth good is state contingent. Individuals have beliefs π_s ($s = 1, \ldots, S$) about the probability of the state s occurring, with the actual probability of the state is r_s. As long as $r_s \neq \pi_s$ for at least two values of s, then the consumer's expected utility-based choices will differ from those which maximise expected welfare.

As with the certainty model of the previous section, in a one consumer economy the first best outcome can be achieved through corrective commodity taxation of the nth good alone, coupled with a lump-sum tax. In a multi-agent economy, as long as taxes cannot be individualised, then taxes will also perform redistributive functions as well as correcting for the wrong belief structures. Because of this, it cannot be assumed that taxes should be positive on goods which are over-valued by consumers and negative on goods which are undervalued. For instance, healthcare is possibly undervalued by many individuals, but it is also a luxury good and therefore a prime target for redistributive commodity taxation. If this is the case then the optimal tax on healthcare may well be positive.

7.4.2 Limits of the Literature

If it is known what the source is of the difference between welfare and preference, then the models summarised above give some guide to the kind of fiscal policies which are likely to be optimal. What they do not provide is a means of modelling the impact of non-fiscal policies, such as information campaigns. Consequently, they cannot be used to assess the relative merits of the fiscal approach versus other behaviour-altering tools which are not traditionally considered within the economic

policy framework. Nor is it clear why, within the mis-aligned preferences argument, consumers would not wish to maximise individual welfare if they were informed of the problem and the nature of their 'correct' preferences.

A step in the direction of a more general framework is provided by supposing that preferences or information are, at least to some extent, malleable through government policy. I have already given the example of an information campaign, but the nature of policy in this area may be more subtle, involving regulations on how private sector adverts may be framed, for instance, or limiting the groups of individuals who can be the targets of television advertising. More fundamentally, in the introduction to his 1988 article, Tim Besley offers a reminder of McClure's point, that if the source of the merit wants problem is a lack of information, then the optimal solution is to provide that information, rather than to remedy the defect through taxation and subsidies. This particular criticism is a recurrent theme of much of the discussion of merit wants.

7.5 The Value of Information

The view on the provision of information mentioned above is one common refrain in discussions of merit wants. In this section we see that this view is erroneous in general, provided only that the provision of information is costly. Consider the first best case discussed above in which $W = u(x) + \theta v(y)$, but the consumer has preferences of the form $u(x) + v(y)$. Given a government budget constraint of the form $R = ty + T$, and a production price of p_2 for y, the optimal tax is $t = (1 - \theta)p_2/\theta$. At this tax rate, the discrepancy between welfare and preference is fully internalised by the consumer and hence, as viewed from the perspective of welfare, consumption is first-best optimal. It follows that the provision of information has no value in such circumstances.

What about second-best cases? It might be thought that, because providing information can add to the set of instruments, then its use must be welfare-enhancing in a second best environment. But this is not always the case as the next example shows. Consider a stripped down version of the Racionero model, with only two types of consumers, both of which have preferences of the form $u(x)+ \theta^i v(y), i = 1, 2$, but where the true value of θ is θ^* for both types. Consumers of type i, $i = 1, 2$ have income m^i, with $m^1 < m^2$. As before, the government can offer tax-good bundles of the form (t,y) and consumers choose whichever bundle they prefer. The government is utilitarian and numbers of each type are equal, so it maximizes $u(x^1) + \theta^*(y^1) + u(x^2) + \theta^*(y^2)$.

Focus first on the non-merit want case, where $\theta^i = \theta^*, i = 1, 2$. Call this the perfect information case. It is the standard non-linear taxation problem and its solution is illustrated in Fig. 7.4. Consumers of type 1 opt for bundle A and consumers of type 2 opt for bundle B and are indifferent between it and bundle A. Also shown in this figure is an iso-revenue line through A and A′. Now compare this to a situation where $\theta^1 = \theta^*$, but where $\theta^2 > \theta^*$. In other words type 2 agents overweight their

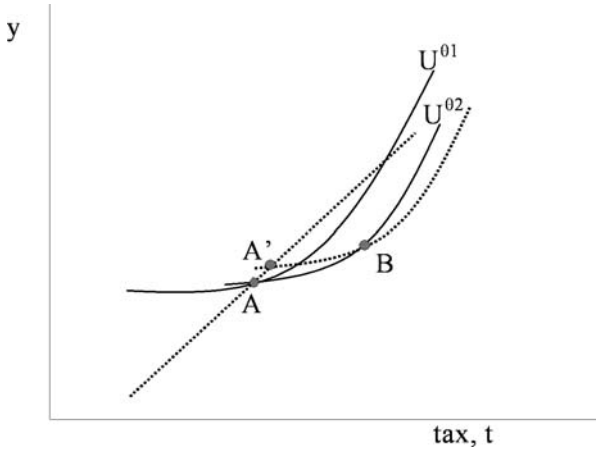


Fig. 7.4 The Value of (Mis)Information

consumption of the merit want, relative to its true value. For the purposes of this example, we can interpret this as a reduction in the information available to the agents, compared to the perfect information case.

The flatter, broken curve through B in Fig. 7.4 represents the preferences of the type 2 agents with $\theta^2 > \theta^*$. The welfare-indifference curves remain unchanged. It is apparent from the figure that the effect of the rise in θ^2 is to relax the incentive compatibility constraint. Consequently it is now possible to offer the tax-good bundles A′ and B rather than A and B. With this pair of bundles, individuals of type 2 continue to choose bundle B. Their welfare is therefore unchanged. Individuals of type 1 meanwhile choose A′, which represents equal revenue (and therefore feasibility), but higher welfare, compared to A for the type 1 individuals. Overall therefore, a Pareto improvement is possible, compared to the perfect information case. Put differently, starting from some values of θ^2 above θ^*, the provision of information such that θ^2 becomes equal to θ^* lowers welfare.

In this case the marginal social value of information is negative. While it may be viewed as special, it is also worth noting that in many instances within this kind of model, the marginal value of information will be zero. Consider, for example any values of the θ^is where they are both below θ^* and where, in the resulting optimal tax problem, the incentive compatibility constraint is binding only for type 2 agents. In this situation, any provision of information which raises only the value of θ^1 has no social value, because the preferences of the type 1 individuals do not represent a binding constraint on the activities of the government. Thus if provision of information is costly, its net social value is negative.

This is a counter-view to the standard argument that provision of information is the optimal solution to at least some class of merit want problems, so it is worthwhile considering the applicability of the argument. There are three kinds of criticism of the counter-view as I see it.

1. If information provision is costless, then in some models its provision may be desirable. However, as the second example demonstrated, even costless information may have a negative social value. On the other hand, just as information provision will typically be costly, so will the implementation of tax and expenditure analysis. A practical analysis of the merit wants problem should therefore consider this issue in a balanced way. It is not clear though that this would eliminate all examples of the kind put forward. Getting people to change their preferences through the provision of information is expensive and possibly infeasible in some circumstances, as many advertisers have found to their cost.
2. A more serious criticism is that the examples show the limitations of welfarism, with its emphasis on ends of a particular kind. Preferences in the welfarist model do not count except as constraints on policy. Perhaps preferences should count for welfare, as well as the views of the 'social planner' on what is good for individuals. Note though that as long as the views of the planner have some weight, and these views diverge from those of the individuals concerned, then examples of the kind produced can be readily produced. Arguments between taxes and the provision of information may not just be about ends; they may also be about the means by which ends are achieved. Arguments in favour of information provision may therefore invoke the notion that individual freedom is preserved in a way that it is not when individuals are manipulated via fiscal instruments. This is a difficult notion to make precise without resorting to an axiomatic assumption that information is a good to be provided above all others. If this assumption is not made, then there will be cases in which the cost of providing information are sufficiently high to make everyone better off with ignorance. In the second example, for instance, whatever the information set, the individuals concerned shared the same opportunity set, in the sense that two bundles of tax and good were always on offer. Similarly, in the first best case, the individual faces a budget set, whatever tax rate is set. For each example if therefore the costs of providing information are sufficiently high, then the opportunity set with under ignorance will be a superset of the opportunity set with no taxes but perfect information.
3. A third criticism of the counter-view relates to the model of information and preferences. In the model, there is a correct value of θ which is known to the planner. Consider instead a model where the planner knows the distribution of true values of θ, which may vary across the population, but it cannot identify the true value of θ for any individual, perhaps because this depends on information which is privately held. In the absence of information individuals do not know what is best for them, but information can be supplied which makes this clear. For instance, information about skin cancer might take the form 'if you are fair skinned then you are at risk and should avoid exposure to sunlight.' Information in this case may be socially valuable even with optimal taxes, but note that this is still not a universal conclusion, provided that individuals differ in other ways of concern to the social planner and first best solutions are not available.

7.5.1 The Value of Misinformation

It is worth being slightly more precise about the kind of model outlined in the third criticism given above, since it offers a reasonable alternative to the simple 'wrong preferences' type of story presented above.

The basic set-up in such a model can be similar to that previously described models from this section: Preferences and welfare have the form $u(x) + \theta^i v(y)$. A new feature is that each individual, h, has private information, a^h drawn from the finite set A with cardinality N. I shall assume that this private information is not directly welfare relevant. In other words the a^h do not enter the government's maximand. In the absence of further information, individuals are aware only of the possible values of θ^i drawn from the finite set Θ. Each individual has prior beliefs, π. Note the absence of a superscript; meaning that for the purposes of making a point it is assumed that all individuals have the same priors.

The government knows the distribution of the θs and the relationship between an individual's private information and their true preferences, but it does not know the identity of individuals. It has two kinds of instruments: messages, m, and tax/expenditure bundles. A message m is a mapping from A to Θ. We can order A, so that a message is simply a vector, a typical element of which is read as: 'if your private information is a', then your true value of θ is θ''. There is no presumption that all a^h will receive a distinct individual message. Let m^* be the honest message, giving the true mapping from A to Θ. Meanwhile, a tax and expenditure policy, T, takes the form of a list of (t, y) pairs for each type, some of which may not be distinct. So a *policy*, P, is a matrix $[m, T]$.

Let the post-message belief of individual h be given by γ^h, then the government's problem is to choose P to maximise $\sum \left[u(t_h) + \theta^h v(y_h)\right]$ subject to the revenue constraint $\sum [t_h - y_h] \geq 0$, and the incentive compatibility constraints $u(t_h) + \gamma^h v(y_h) - u(t_j) + \gamma^h v(y_j) \geq 0$, all $h = 1 \ldots, H$, $j = 1, \ldots, N$. Note that, to simplify interpretation, in the description of this problem I have taken the step of identifying the tax-expenditure bundles by the households who choose them, except in the latter part of the incentive compatibility constraint. Define $W(\theta^*, \gamma, T)$ as the value of welfare when the distribution of true values is given by θ^*, γ is the matrix of post-message beliefs, the message is m and T is the set of tax/expenditure bundles offered. Note that welfare does not depend directly on the message, but typically γ will be a function of both m and T. Let W(θ^*,γ) be the maximum value of welfare conditional on post-message beliefs γ.

One thing as yet undefined is the nature of post-message beliefs. One might expect them to depend in part on the messages sent by the government and possibly on the set of tax-expenditure options available. In a book devoted to bounded rationality, it seems contradictory to presume completely rational inference on the part of individuals and so let us explore different assumptions. In particular, call someone *naïve* if they always believe the message sent by the government; that is $\theta^h = m^h$. The following example shows that, in general, honesty is not the best policy for a benevolent government faced with naïve citizens.

Consider the case where there are two naïve individuals with true values of θ^{1*} and θ^{2*}, where $\theta^{1*} < \theta^{2*}$. Suppose T is chosen to maximise the following:

$$L = \sum_h \left[u(t^h) + \theta^{h*} v(y^h)\right] \tag{7.6}$$
$$+\lambda_1 \left[u(t^2) + m^2 v(y^2) - u(t^1) - m^2 v(y^1)\right] + \lambda_2 \left[t_1 - y_1 + t_2 - y_2\right]$$

Write the maximum value of welfare which results as $W(\theta^{1*}, \theta^{2*}, m^1, m^2)$. Then, using the Envelope theorem:

$$\frac{dW(\theta^{1*}, \theta^{2*}, m^1, m^2)}{dm^2} = \frac{\partial W(\theta^{1*}, \theta^{2*}, m^1, m^2)}{\partial m^2} = \lambda_1 \left[v(y^2) - v(y^1)\right] > 0 \tag{7.7}$$

In other words honesty is not the best policy, because overstating the value of θ^2 compared to its true value leads to a relaxation of the incentive compatibility constraint. This basic result has already been seen when we considered the case of reducing the quality of information in the previous section.

Extreme naivety can be unrealistic. A more sophisticated form of inference would be to believe a message only if the actions associated with the message are consistent with its veracity. Let $T^*(m)$ be the tax-expenditure policy which maximizes $\sum \left[u(t_h) + m^h v(y_h)\right]$ subject to the revenue constraint $\sum [t_h - y_h] \geq 0$, and the incentive compatibility constraints $u(t_h) + \gamma^h v(y_h) - u(t_j) + \gamma^h v(y_j) \geq 0$: $\forall$ $h = 1, \ldots, H,\ j = 1, \ldots, N$. In other words, $T^*(m)$ maximises welfare conditional on the message being believed and it being true. Call a policy, P, *potentially credible* if $T = T^*(m)$. Meanwhile say that a policy is *convincing*, if (a) it is potentially credible and (b) $W(\theta^*, \theta^*, T^*(m)) \geq W(\theta^*, m, T^*(m))\ \forall m \in \Theta$. In other words, a policy is convincing when the welfare achievable when an honest message is sent and when it is believed is greater than that possible when a false message is sent and believed, and where the accompanying tax/expenditure policy is one which would be optimal conditional on the false message being true and being believed.

Are all potentially credible policies convincing? The answer is no. The two-person counter example used above can be employed here. It showed that, for a set composed of two naïve agents, welfare can be increased by changing the message marginally away from the honest message, in the shape of a small increase in m^2. Now, consider a society where θ can take three ascending values: θ^1, θ^2 and θ^3. Suppose the government knows that one individual is type θ^1 and one individual is type θ^2 – in other words it knows there are no θ^3 types. Consider the resulting welfare if the government announces that the types are θ^1 and θ^3 and offers an accompanying policy which is potentially credible. If θ^3 is sufficiently close to θ^2, then using the earlier argument, welfare can be higher than that achievable if honesty is pursued. It follows that even those individuals with sophisticated beliefs, in the sense that they only believe messages when accompanied by their potentially credible policies, may receive messages which stray from the honest.

Of course, unboundedly rational agents would realise this possibility, thus there is no scope for rational agents to be fooled. Nevertheless it may still be optimal for

the benign government to be less than honest in the sense it may wish to hide its true information set. Consider a final, sketched example where only two configurations of types are possible: either agents are θ^1 and θ^2 or the economy is composed of θ^1 and θ^3 types. Call the first case the 'Medium' economy and the second case 'High'. With Bayes rational agents we have both a signalling and a screening game. The government signals whether the economy is Medium or High and then sets its instruments of tax policy to screen between types. The thrust of the previous sections is that it is sometimes advantageous for the government to mislead the highest θ type so as to relax the incentive compatibility constraint. Consequently, in the case of the High economy, it is optimal for High to be signalled honestly by the government and for the instruments to be set optimally given that the message is believed. However, consider the Medium economy. Here, either a pooling or a separating equilibrium may be optimal depending on circumstance. In a pooling equilibrium the government always signals that the economy is High by sending the message corresponding to the High case, accompanied by the configuration of instruments which is optimal in the High case. θ^1 agents will always know that they are θ^1, and hence choose the θ^1 policy. The other agents will not know if they are θ^2 or θ^3 – posterior probabilities will be the same as priors. If given these priors the non-θ^1 policy is the optimal choice for these non-θ^1 agents then the outcome, in terms of policies offered and chosen, will be the same whether the economy is Medium or High. It follows that if welfare is higher in the Medium case with pooling, than it is in the Medium case with an honest message sent (and optimal policies set, conditional on this message being believed), then overall welfare is higher with a pooling equilibrium. As we saw in the sophisticated case above, this may well be the case, provided that θ^2 and θ^3 are sufficiently close. Hence, even with weakly rational, Bayesian agents, it may be optimal for the government not to reveal all of the truth.

7.6 Conclusion

In this chapter I considered 'traditional' fiscal instruments and their role in ameliorating the welfare problems caused by merit wants. With merit wants, the optimal use of fiscal instruments can be analysed, in much the same way as they can with public goods or external effects. The results are largely intuitive, in that if sufficient instruments are available then optimal taxation involves re-aligning choice so that it accords with welfare. When there are second best constraints, such as the need to pursue other goals (e.g. redistribution) or where the policy-maker has only a limited set of instruments, then optimal taxes may not be positive when individuals over-value a good or negative when individuals under-value a good. However, contrary to many of the comments in the merit wants literature giving truthful information is not always optimal for the government because of the interaction between incentive compatibility constraints and the beliefs held by agents.

Because of the largely intuitive nature of the results, it is tempting to wonder whether bounded rationality makes much difference to the job of the public

economist. Some sort of division of labour could be proposed in which the behavioural economists or psychologists would specify the manner in which preference departed from welfare and then the public economist would devise the optimal instruments of tax and expenditure policy. However, consider the following example:

Negative Income Tax system A: All individuals receive a tax free sum of £2,000. All other income is taxed at a rate of 25%.

Negative Income Tax System B: Individuals with earnings of £8,000 neither pay nor receive funds from the government under the tax system. Individuals pay income tax at a rate of 25% on every £1 earned above £8,000. For every £1 earned less than £8,000 individuals receive a sum of £0.25 from the government.

Assume that no-one has negative income. In both the systems net income is equal to £8000 + 0.75(gross income – £8,000). So from a standard point of view the tax systems are equivalent. However, as we shall see in Chapter 10, the descriptions of the systems are likely to evoke different reference points and hence different behaviour. Indeed, Traub (1999) provides some evidence from the German tax system that this can be the case. Thus, the public economist who is interested in policy relevant analysis, needs an understanding of how policies which appear equivalent from a standard viewpoint, may in fact have quite different implications for behaviour.

One way to view the traditional merit wants literature is that the government is delegated by a boundedly rational but aware individual to regulate consumption of goods for which the individual has the wrong preferences. A key assumption is then the benevolence of the government. Of course in practice, agents who supply goods can have motives that depart from those of the principal and that is the subject of the next chapter.

Chapter 8
Agency and Dependency

8.1 Introduction

At the heart of the merit goods problem there is a fundamental issue of agency. If some individuals are to surrender at least part of their consumer sovereignty to other individuals then for individual welfare to improve it is not enough for the agents to have superior information or decision-making powers, they must also have the incentive to serve properly the true interests of their principal. Until now I have largely ignored the issue except in passing, but it will be the central focus of this chapter.

When individuals delegate some aspects of their choices, they pass responsibility to service providers such as their family, professionals or the state. There are at least five important dimensions along which service providers can potentially differ:

1. The degree to which the provider has concern for the welfare of the principal and/or the quality of the service provided;
2. The technical expertise of the provider, defined as a function of the diagnostic capabilities and of their technical efficiency in production;
3. The preference expertise of the provider, defined as their ability to understand the true interests of the principal, possibly in the absence of direct instruction;
4. The returns to scale in provision and finally,
5. The degree to which the providers are regulated by the state.

Potentially therefore there are a large number of institutional forms through which merit wants may be provided. However, by and large, in the context of merit wants, societies respond to the agency issue by three forms of delegated responsibility: first through the assignment of responsibilities to self-regulating 'professionals'; secondly to state-regulated organisations and thirdly to relatives and other close acquaintances of the individuals concerned. In addition, a large body of services are provided through the market place with little or no apparatus of state regulation, relying instead on the incentives offered by the market, through repeated purchase and reputation.

In this context friends or members of the family of an individual can be considered as those agents with expertise or knowledge of the principal's preferences based

A. Munro, *Bounded Rationality and Public Policy*, The Economics of Non-Market Goods and Resources 12, DOI 10.1007/978-1-4020-9473-6_8,

© Springer Science+Business Media B.V. 2009

on direct experience of his or her life as well as some concern for the Principal's preferences. We can think of a professional as someone who has technical expertise in a particular trade or service; has some concern for the quality of their provision and/or their clients' welfare and is subject to regulatory oversight by other members of the same profession. A state-regulated provider is someone who also has technical expertise, who may or may not have concern for the quality of their work or their clients' welfare and who is subject to regulatory oversight primarily by the state.

Thus, professionals and state-regulated providers are distinguished by the form of regulation to which they are subjected and to a degree, by their professional pride or altruistic concern. Nevertheless, the line between state-regulated provider and professional has always been an unclear one, especially since 'altruistic concern' is largely private information. Moreover, in many countries, professional self-regulation is limited by statute or by the intrusion of regulatory oversight by the state (Bartle and Vass, 2005; Scott, 2003).

Figure 8.1 summarises some of the important ways in which the typical examples of each of these types of institutions differ. Its purpose is primarily one of illustration, for in practice there is a great deal of overlap between different types of provision and with each type, individuals and institutions can vary greatly in their degree of altruism and in the regulation to which they are subjected. The main point is that, at least to some extent, regulation and altruism are alternative mechanisms for limiting the problems created by agency in the context of merit wants. Caveats aside, services provided by the family, such as personal care and attention of the fragile elderly or child rearing, are usually viewed as being produced under greater conditions of altruism towards the beneficiaries, compared to those produced by self-regulating professionals such as doctors or through the state. Family-provided services are less heavily regulated as well – an untrained parent can take their child rock-climbing, white-water rafting or surfing, whereas a teacher would have to have

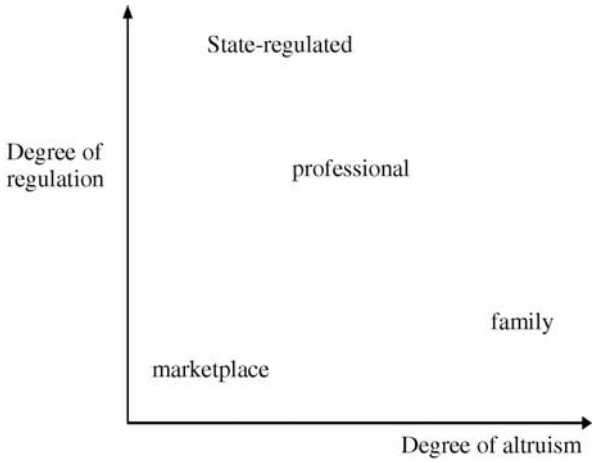


Fig. 8.1 Regulation and Altruism

prior training and possibly accreditation before being let loose with the child up a mountainside.

The three specialised institutions differ in more than alternative combinations of altruism and regulation. They also differ in the degree to which the services produced are tailored to the particular perceived needs of the principal and hence in their ability to reap economies of scale.

State regulation, for instance, has two principal drawbacks: first, there is the familiar fact that it too is subject to an agency problem – who regulates the regulators? Secondly, the superior knowledge of the state about true welfare is likely to be of a particular kind, concerned with population averages or typical behaviour. Compounding the latter are the undoubted diseconomies associated with product differentiation, which make it more cost efficient for the state to provide or mandate uniformity rather than diversity. Fundamentally therefore, the state is a rather clumsy supplier of goods, producing neither sledgehammers nor nutcrackers, but instead offering every customer an average kind of hammer which is too heavy for nuts and not massive enough for pounding fence-posts into the ground.

The plan for this chapter is as follows: in the next section I illustrate some of the issues surrounding dependency through the example of children. The focus is on the development of economic competency. Next we consider regulatory policy, given dependency and the various kinds of asymmetries in information that exist. Finally I use the example of adulthood to discuss briefly the question of when should the state restrict the independence of its citizens.

8.2 Dependency Relationships: Children

To illustrate some of the issues arising from dependency consider one important class of agents: children. In the economic sphere children are the largest single group of agents who are prohibited from entering into voluntary, enforceable contracts.[1] Their rights to economic activity are severely circumscribed: they are not allowed to smoke, to vote, to engage in sex or get married, to work more than a limited number of hours in paid employment as well as a host of other limitations. Most importantly, childrens' decision-making is usually delegated to their parents and, unlike their parents, children have few rights to end their membership of a family. Thus, to an extent they are denied exit and voice in their own lives.

As we saw in Chapter 5, two main justifications are often given for the principle of consumer sovereignty. The first is that individuals know what is best for themselves. The second, which is due to Mill (1859) and is a much more robust defence, is that while individuals might not know what is best for themselves, no-one else knows any better. Restrictions on childrens' ability to enter into contracts must therefore rest on a. that they do not, in general, know what is best for themselves and b. some one else knows better (note that b implies a). At the same time, with

[1] Others include prisoners and the mentally impaired.

most children there is some point where either b. is no longer true. To understand this further we need to consider the development of economic ideas in children.

8.2.1 The Child's Development of Economic Ideas

There is not a great deal of work on the development of economic ideas in children as opposed to thinking on development in general. Berti and Bombi (1988) provide a survey of some aspects, largely based upon their own research. It is helpful to think of development proceeding along two dimensions: reasoning and knowledge. In the first dimension Berti and Bombi (1988) analyse development along broadly Piagetian lines, distinguishing between concrete and formal reason. In the former, which develops first, children obtain local information about the world they live in but are unable to generalise information through formal reasoning. Their ideas remain unlinked: the fact that their parents work is not linked the family's ability to spend money. The spending of money is not associated with the income of those who receive the payment. Later on (in early teenage years for the majority of children), formal reasoning develops with children able to link and resolve separate parts of their knowledge and to generate formal chains of reasoning.

The economic consequences of this can be modelled as follows. To make 'correct' decisions, agents must have preferences satisfying reflexivity, completeness and transitivity. With small children none of these are satisfied (especially transitivity!). As they become older they acquire some local knowledge about patches of their preference map, but these patches can only become linked when formal reason (which includes the idea of transitivity) develops. For instance, according to Berti and Bombi (1988),

> Pre-schooler's knowledge is restricted to the identification of 'regularities' within situations of which they have first-hand knowledge (such as payments for goods or services by customers or consumers) and the acquisition of common information (such as 'grown ups work for money'). Thus, not all goods come to be recognised as such, but only those which are suitable for incorporation into the 'shop script'; a cow, land or a house are not things which can be bought. . ..a bus may belong to the passengers because they get on it, or a field might belong to some gypsies because they have camped on it. The reason why people pay for what they buy is not yet understood, it is done out of habit, or because of 'the law', or because otherwise people get taken to prison (p. 175).

While, at the other end of the developmental process,

> Finally, the isolated relationships become linked to each other so as to form a system of relations. We then have a conceptualisation of a totality wherein each part derives it significance from its position in the whole. At this point, a purely rational explanation becomes possible (Danziger, 1958, p. 240).

Something similar happens with a budget constraint: it is only gradually that children develop the ability to put together individual prices and overall resources. Harbaugh et al. (2001), demonstrate that 7 year old American children show substantially more violations of revealed preference than their 11 year old counterparts, whose performance in a choice experiment is similar to that of adult students.

The second dimension of development concerns the concepts of actions and consequences and the difference between 'what I like' and 'what is good for me'. While children often develop a clear understanding of the former at an early age, it is parents who are assumed to understand the relationship between inputs to child welfare and outputs. From an early age children are warned against doing things. Gradually rules, such as 'don't play with matches' or 'don't drink bleach' become rules with reasons (e.g. 'don't play with matches because you'll burn the house down').[2] As formal reasoning develops, children are able to place these rules within a systematic framework which enables them to reason about the consequences of their actions for their own well-being. For instance Strauss (1952) distinguishes 9 stages of development in the child's understanding of the function of money. According to Berti and Bombi (1988), in stage 1 (4 years and 8 months to 5 years, 11 months in the 'typical' child),

> Children. . . understand that money has to do with buying. This does not, however, indicate that they understand the function of money in buying and selling. They are unable to give any other justification for the rule imposing payment except that one must do it (p. 8).

By stage 5 (6 years, 9 months to 8 years and 9 months) children are able to compare coins and correctly identify their relative values, while by the final stages (8 years 4 months to 11 years), they are able to understand some economic relations, viz:

> Children understand for the first time that shopkeepers realise a profit by selling gods at a higher price than they bought them. This increase in price is also considered to be justified by the service which shopkeepers perform for their clients. (Berti and Bombi, 1988, p. 9).

Similarly, as their sense of time develops, children begin to value their future selves as well as their immediate gratification and understand the role of savings and strategies for avoiding temptation (Webley, 2005).

There are other important elements in child development for economic theory. Frank (1987) argues that is 'optimal' for humans to have in-built rules that hinder us from cheating at every available opportunity. Having a conscience enables the initial sense of trust to be built up in a relationship and helps avoid the destructiveness of 'tit for tat' behaviour. Guilt and embarrassment, shame and blushing are therefore of value because they encourage co-operation in adult life. A final aspect of economic development is the set of options that we perceive as open to us. One function of the developmental process is therefore to give children an idea of their own capabilities that extends beyond their immediate environment.

8.3 The Role of the State

The role of the state within dependency can be split into two questions. First, given a pre-existing relationship how or when should the state regulate that relationship,

[2] These reasons are also developed through play and experience. For example Alcohol makes you sick and playing with razors makes you bleed.

when the dependent has a limited ability to express their true preferences? Secondly, when should a relationship of dependent and carer be allowed?

A typical case is the child who is cared for by their parent or guardian or the elderly parent who is dependent on the care provided by one or more of their children. In such cases, usually the state has only a limited information advantage over the individual. And in many situations, two important factors militate against state intervention. First, there is the undoubted preference for one or both of the parties for personal care. Second, state-run operations are typically attuned to the needs of the average individual; the degree of personalisation of services can be limited. This feature stems from the diseconomies of scale and scope from individualising services, but perhaps more fundamentally, it arises through a lack of knowledge of the wants and preferences of the individuals placed under state care.

Against personal care are two kinds of arguments: first, there is the rogue or incompetent carer: the child abuser, the neglecter of the elderly relative. Secondly, there is the relief from duty. In a world where most individuals take responsibility for the care of their frail or senile relatives, there is a norm against passing on the duty to someone else, no matter how heavy the burden. Going against that norm can induce feelings of guilt or shame, but replacing or supplanting individual care with state provision can change the norm, with potential benefits to all carers. Let us deal with this argument first.

8.3.1 Relief from Duty

The argument is illustrated by the simple game in Table 8.1. In this game carers can select from one or two choices for their dependents. If they choose to care for the dependent themselves, then the net benefit is $1-c$, where c is the personal cost of the burden caring. On the other if they choose to allow someone else to care, they incur the price of this, p, but do not suffer the carer's burden. The term g is the guilt felt by individuals who do not do the caring themselves, when other carers are looking after their dependents. Thus g reflects a norm of caring that is social in nature.

Suppose first that $g = 0$. In the absence of the guilt term the carer will allow someone else to look after the dependent whenever $c > p$. The resulting Nash equilibrium is also Pareto efficient (assuming indifference on the part of the dependent about who does the caring).

When $g > 0$, there are potentially two pure-strategy Nash equilibria and a mixed strategy equilibrium as well. If $p < c$, there is still one pure strategy Nash

Table 8.1 The Game of Guilt

		Carer 2	
		Self-care	Allow others to care
Carer 1	Self-care	1–c, 1–c	1–c, 1–g–p
	Allow others to care	1–g–p, 1–c	1–p, 1–p

equilibrium in which others do the caring, but now there may also be a Pareto inferior pure strategy Nash equilibrium in which self-care is the norm, provided $g + p > c$. Hence one possible role of the state is to eliminate the inferior Nash equilibrium by compelling agents to allow others to care, or through some less coercive policy such as subsidising the provision of care by others. A 'guilt' subsidy of $s = g$, for example, eliminates the Pareto inefficient outcome from the set of Nash equilibria.[3]

8.3.2 Information Asymmetry

Let us set aside the relief from duty argument and concentrate on asymmetric information using the following extended example. Suppose that b is a possibly unknown preference parameter for an individual, that affects the optimal choice of x a consumption variable that is provided for the consumer by an agent.[4] There are many ways in which this problem can be formulated.

There are, for instance, three obvious classes of potential suppliers: the altruist, the profit maximizer and the service enjoyer (someone who obtains utility from x). There are also other possible preferences and there are different ways of formulating payoff functions for service enjoyers and altruists. In addition, there are mixture models. To be specific, suppose that the altruist maximizes $\lambda W + (1 - \lambda)\Pi$, $0 < \lambda < 1$, where Π are the profits from the operation and W is the welfare of the individual. Let us suppose that the service enjoyer maximizes, $\Pi + \gamma x$, $0 < \gamma$, where γx is the pleasure they get from supplying x. Thus the preferences of the supplier can be completely described by the two parameters, λ and γ in the general formulation, $V = \lambda W + (1 - \lambda)[\Pi + \gamma x]$.

There are also different information regimes concerning b:
(I0) the parameter b is observed only by the consumer;
(I1) the parameter b is observed by both players;
(I2) the parameter b is observed by the supplier, but not by the consumer who also does not observe W;
(I3) the parameter b is observed by the supplier, W is observed by the consumer (so x is an experience good);
(I4) the supplier has more accurate information on b than the consumer;
(I5) the supplier can, at some cost, obtain more accurate information than the consumer about b;
(I6) b is observed by the supplier and by an auditor with some positive probability, but not by the consumer;

[3] The mixed strategy equilibrium is also Pareto inefficient and is also eliminated by the subsidy.

[4] Wirl (1999) pursues a formal approach to a similar problem in which the principal has some paternalistic concern for the agent. Here though it is the agent who may have some concern for the principal.

(I7) the interaction is (possibly repeated), the parameter b is observed by the supplier, the consumer receives some information about b after each interaction.

For example, I0 might be considered as the standard consumer good case: the consumer can contract with the agent over the value of x, knowing exactly what is the best choice. I2 represents the opposite extreme – perhaps a medical procedure, where it is the supplier who knows what is good for the consumer. I3 is a special case of I2 – if the consumer's payment to the agent can be conditioned on their welfare (which might be contractually impossible), then they can agree a contract in which the agent will choose x optimally. I4–I6 are weaker versions of I2, in the sense that the agent has the upper hand in knowledge about b, but the information is not perfect, or it is costly to acquire, perhaps because medical tests must be conducted. Then, under I6, there is the possibility of this information being known to a third party, who given the correct incentives might be able to enforce optimality. Finally there is the issue of repetition. If the agent learns nothing from each choice, then repetition may have no extra value in providing incentives for optimality, but otherwise, through repeated choice the consumer may learn something about their own preferences or about the preferences of their chosen supplier.

Just as there may be asymmetric information about b, so there may be different information regimes for knowledge of the supplier's preferences. Typically it may not be possible for the agent to know whether their potential supplier is an altruist or a service enjoyer or someone who seeks only profit.

Many of these cases represent quite standard principal agent problems. As such they have quite standard solutions, which do not require government intervention. But for some of the cases, such as I2, the consumer is at the mercy of the supplier – there is no observable variable upon which to build incentives for the beneficial delivery of the service. Conceptually, there are two fundamentally different cases to consider that correspond to delegation and merit wants in Table 6.1:

- The consumer does not know what is optimal, but knows that she or he does not know and is otherwise able to make inferences about motives etc.
- The consumer does not know what is optimal, but does not know she or he does not know.

In the first case, the state may have a role to play, in providing auditing or quality assurance services or in producing information that can help the consumer know what is optimal. However, it is not obvious that the state has the advantage over privately-provided services in general. Perhaps, the state has a more useful role to play in the second case and in the difficult issue of deciding which case applies. Even here, there may be other agents who (e.g. family members) whose presence can obviate the need for state intervention. Given such agents, the fundamental role for the state is in regulating their actions, rather than in making choices for the principal. In particular, the state may have an efficiency-enhancing role if it can provide mechanisms for separating the altruist from the opportunist agent and in regulating the actions of the latter.

8.4 Regulating Care with Asymmetric Information

Let us be more specific using an example based on the case I2. The model used to illustrate the problem follows on from the learning model deployed in Chapter 6. Specifically, in any relationship there are two individuals: the personal carer (or agent) and the dependent (or principal). The welfare of the dependent is given by $U = 2bx - x^2$ where b is a preference parameter, drawn from the closed interval $[b^l, b^u]$, with $b^u > b^l > 0$. The variable x is the choice variable and so its optimal value, from the perspective of the dependent's welfare is $x^* = b$. The variable can be thought of in various ways, but as in the previous chapters it is probably simplest to think of it as reflecting some personal preference variable which varies across the population. Perhaps it is taste in food, or ideal room temperature or the degree of routine. Whatever it is, the dependent is not able to set it with any great insight, either because they are too young or too sick or unable to communicate and enforce their needs. As a result someone else must choose x. It is useful in what follows to think of low x values as being generally worse than higher x values, but to suppose that because of preference diversity there are some individual who would genuinely benefit from low values of x.

Personal carers have preferences described by the function $(1 - \lambda)U + \lambda V$, with $1 > \lambda > 0$ and where $V = 2x - x^2$. We can imagine V as the personal component of the carer's welfare. There are two ways to think about λ. One is to suppose that all carers are completely altruistic. Under this supposition, λ then reflects the degree of insight of the carer into the wishes of the dependent. An individual with $\lambda = 0$ can fulfil perfectly the needs of their dependent, whereas for $\lambda = 1$, the carer has no special knowledge of their wants, and treats the dependent according to how they themselves would wish to be treated.

The second way of viewing the model is that all carers know the wishes of their dependents but that the parameter λ is the weight that the carer puts on their own welfare. In other words carers vary in their degree of altruism rather than their facility for empathy. For ease of exposition, I shall generally use language appropriate to this case rather than the previous interpretation. I shall call $\lambda = 0$, the case where carers place no weight on their own welfare, complete altruism, while the opposite case is labelled as selfish. From the viewpoint of the altruistic carers, the optimal value is $x^a = (1 - \lambda)b + \lambda$. Selfish carers have preferences of the form V and thus the optimal value of x, from the viewpoint of the selfish carer, is $x = 1$. I make the assumption that, $b^l < 1$ which implies that the selfish individual cannot be identified from a completely altruistic agent purely on the basis of their unregulated choice of x. This makes the state's problem of distinguishing the altruistic from the selfish non trivial. In practice, let us suppose that there are just two types of carer: the selfish who form a fraction p of the population and the altruistic, with $\lambda > 0$ who form the remainder. Let us also suppose that the distributions of b and λ are independent.

One option in what follows will be state care, which involves two elements: a uniform value for x^g (defined as the level of government provision) for all dependents under its wing and a fixed utility cost of state care for dependents of $s > 0$. The state chooses a uniform value because it has no insight into the true value of b

for each agent, though it may have general information about the distribution. The term s is potentially positive because dependents may prefer their families or loved ones to be in charge of their day to day care. In order to keep the model simple there are no variations in resource cost associated with different values of x. Moreover, I shall suppose that if the state takes over the role of carer then it can force the personal carer to assume financial responsibility.

One issue is that of the personal carer's utility when the state takes over the role of carer. There are divergent possibilities here mainly because we might imagine the selfish and the altruistic agent feeling somewhat different if they were relieved of direct responsibility for the dependent. One option is that the personal carer evaluates the government provision according to $(1 - \lambda)U(x^g) + \lambda V(x^g)$. Under this interpretation, no personal carer, selfish or altruistic, would willingly surrender their charge to the government without compensation.

A second option arises from the interpretation of x as some good which must be more or less uniformly provided in a shared home, such as room temperature or destination for leisure trips; or where there are sharply increasing costs for differentiation (e.g. type of food consumed). The carer's utility would then be written as $(1 - \lambda)U(x) + \lambda V(y)$ with the constraint of $x = y$ for domestic provision of care. Provision by the state would then release the carer to set y differently from x^g while still gaining pleasure from the dependent's consumption. For the model which follows I take this to be the prevailing situation.[5,6] It follows that when the dependent is placed in the charge of the state, the carer will choose $x = 1$, whatever their degree of altruism. For the selfish carer, therefore, fallback utility is 1, while for the altruist, when the government assumes the mantle of provider, the utility obtained is, $\lambda + (1 - \lambda)[(2b - x^g)x^g - s]$.

I assume that the government maximizes the expected welfare of the dependents: $E(2bx - x^2)$. Thus the carers have no weight in the social welfare function. There are two reasons for this: simplicity and observation. The qualitative results which follow would not be dramatically altered by the addition of some weight for the carers in the welfare function. On the other hand, omitting them makes derivation simpler and makes the conclusions more transparent. Secondly, it often seems to be the case that, when making decisions about their care, the state attaches much more weight to the welfare of children and other dependents compared to the weight given to the carers. In the UK's Children's Act (1989), for instance, authorities are obligated to consider only the interests of the child when making their decisions.

---

[5] A further possibility is that, by passing off their charge onto the state, individuals may relieve themselves of the psychic burden of caring for others. For the selfish agent, x may be set so as to meet social norms of propriety; when the agent no longer has a dependent these norms may no longer be relevant.

[6] Recall the alternative interpretation of λ as an insight parameter. We would expect all types to maintain their feelings of altruism when their charges are placed in the care of the state. But if the high λ types are aware of their lack of insight, then they might feel relief at passing the burden onto others. So, the utility received by the personal carer when the state is the provider makes the natural model different for this situation.

With these preliminaries accomplished we now consider a number of simple policies for regulating the relationship between dependent and carer.

8.4.1 Compulsory State Provision

Perhaps the simplest policy is one in which the state takes over the caring relationship, provides the good and chooses one level for all individuals with no possibility of opt out. If the state chooses and knows the true distribution of b, then x^g will equal $\overline{b}$ – the expected value of b in the population. Expected welfare is then $\overline{b}^2 - s$. Clearly this is never optimal in a world populated only by complete altruists. However if all carers are selfish and if carers choose x then welfare will be $2\overline{b} - 1$. Hence a necessary condition for state care to be superior to the personal care by the selfish is $\overline{b}^2 - 2\overline{b} - 1 > 0$, or $\overline{b} > 1$ Note that if $b^l < 1$ is not satisfied then this condition is automatically satisfied, but that it is perfectly possible for the condition and $b^l < 1$ to co-exist.

8.4.2 Minimum Standards and Public Provision

Consider next the case where the state authorities know the true distribution of b and can monitor the value of x set for each dependent (but cannot directly identify either the true value of b or whether a carer is selfish or altruistic). The state sets a minimum level of the care variable, $\underline{x}$. Thus the state's objective function is:

$$\begin{aligned} &pE[b-1] + (1-p)E\left[2\underline{x}b - \underline{x}^2 \mid b(1-\lambda)+\lambda < \underline{x}\right] \\ &\quad + (1-p)E\left[b^2 - \lambda^2(b-1)^2 \mid b(1-\lambda)+\lambda \geq \underline{x}\right], \underline{x} \leq 1 \\ &pE[2\underline{x}b - \underline{x}^2] + (1-p)E\left[2\underline{x}b - \underline{x}^2 \mid b(1-\lambda)+\lambda < \underline{x}\right] \\ &\quad + (1-p)E\left[b^2 - \lambda^2(b-1)^2 \mid b(1-\lambda)+\lambda \geq \underline{x}\right], \underline{x} > 1 \end{aligned} \tag{8.1}$$

where E[.| a] means 'expectation conditional on a'. Differentiating the objective function with respect to x yields:

$$\begin{aligned} 2(1-p)E\left[b - \underline{x} \mid b(1-\lambda)+\lambda < \underline{x}\right], \quad \underline{x} \leq 1 \\ 2pE[\overline{b} - \underline{x}] + 2(1-p)E\left[b - \underline{x} \mid b(1-\lambda)+\lambda < \underline{x}\right], \quad \underline{x} > 1 \end{aligned} \tag{8.2}$$

Note the discontinuity at $\underline{x} = 1$. These derivatives are illustrated in Fig. 8.2. The horizontal axis measures the level of minimum provision, while the derivative is plotted on the vertical axis.

For the purposes of the diagram, the discontinuity is set below the point at which the derivative function first crosses the horizontal axis, thus in this case the optimal value of $\underline{x}$ exceeds 1. In practice though, it could be above or below. Either way the optimum is at the highest value of $\underline{x}$ where the derivative is equal to zero.

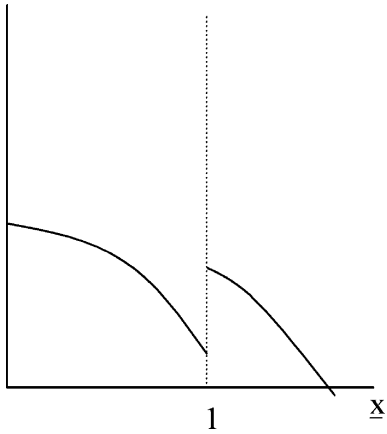


Fig. 8.2 Minimum State Provision

From the first order condition we can draw the following conclusions:

- If p is strictly positive, some constraint on x is optimal – the partial derivative with respect to $\underline{x}$ is always positive at b^l.
- If p is strictly positive, some altruistic carers may be constrained in their actions, forced to set a level of x above the point which is optimal for their dependent.
- The optimal minimum standard is always below the mean value of b.
- The optimal minimum standard is increasing in the proportion of selfish individuals, p, and in the degree to which altruists place weight on their own utility (λ).

8.4.3 Voluntary State Provision of Care

Another policy often pursued is state provision of care that is voluntary – no-one is forced to place their charge into the care of the state. Let the level of x chosen by the state for its operations be x^g. Selfish carers will not prefer to pass on their responsibilities to the state (formally they will be indifferent); neither will complete altruists who will strictly prefer domestic care unless $b = x^g$. Intermediately altruistic agents on the other hand, may wish to do so, provided the value of x^g is sufficiently close to b. More formally, they will prefer state care for their charge whenever,

$$(\lambda + b(1-\lambda))^2 < \lambda + (1-\lambda)[x^g(2b - x^{g^2}) - s]$$

For each carer, this inequality defines a convex set of values of x^g for which state care is preferred. If s is sufficiently high, then this set is empty, but if the set is non-empty then it always includes the value $x^g = b$. In particular, if s is zero, then the set is definitely non-empty for strictly positive values of λ. Let b^{lc} be the lowest value in the set and let b^{uc} be the upper value, then the maximand is,

$$pE[2b-1]+(1-p)E\left[-\lambda^2(b-1)^2+b^2 \mid b<b^{lc} \text{ or } b^{uc}\leq b\right]$$
$$+(1-p)E\left[2bx^g-x^g-s \mid b^{lc}\leq b<b^{uc}\right] \tag{8.3}$$

The first order condition for the optimal choice of x^g is,

$$2E\left[b-x^g \mid b^{lc}\leq b<b^{uc}\right]+\lambda f(b^{uc})\frac{db^{uc}}{dx^g}\left[2b^{uc}x^g-x^{g^2}-s-1\right] \tag{8.4}$$
$$-f(b^{lc})\frac{db^{lc}}{dx^g}\left[2b^{lc}x^g-x^{g^2}-s-1\right]=0 \tag{8.5}$$

where f(.) is the density function for the distribution of b. Examining this equation it is clear that the optimal value of x^g can be above, below or equal to the mean value of b, depending on the distribution of b. In other words, it is not necessarily the case that the government provided option should be of lower quality (if x is taken as quality) than the mean.

8.4.4 Mixed Policies

Now consider the use of both minimum standards and voluntary state provision simultaneously. If both policies are employed together, then for any $\underline{x}>1$, all selfish carers opt to place their dependents into state care.[7] For altruists, the choice depends on the relative values of $\underline{x}$ and x^g.

To make some of the issues simpler let us suppose that $\lambda=0$, so that we are only dealing with the completely selfish and the completely altruistic. There are potentially three groups of agents: the selfish; altruists who set $x=b$ and altruists who either opt for government provision or set $x=\underline{x}$. We get the following results:

- It is never optimal to set x above the level at which selfish carers opt to place their charges under state care. To do so must constrain the actions of altruistic carers.
- For large enough s, state care is never optimal.
- If only selfish carers place their charges in the care of the state, the optimal value of x^g is with $x=1$; otherwise no provision is optimal.

To understand these results better, let us consider a small number of important scenarios. First, suppose that there are no altruistic agents. In this situation the state delivers all care, set at the average figure for b. Now consider the other extreme: no selfish agents. With complete altruists, the best policy is no policy, in the sense of no limits on x and no government provision. Next consider a world where there are both complete altruists and selfish carers, but where s is small. Now there may be

[7] Note that there is nothing particularly special about the minimum standard as the cause of this effect, any infinitesimal gain to the selfish agents will lead to the same result. For instance, a tax on domestic care of one pence would have the same effect.

three groups at the optimum: the dependents of selfish agents who are all consuming government care, set at $\overline{b}$; a group of altruistic carers who are constrained to supply at $\underline{x}$ and finally another group of altruistic carers, who are unconstrained in their actions. Finally consider the last case again, but now with a high value of s. In this situation, the optimal policy may be no government provision, but potentially with $\underline{x} > 1$. It follows that for some increases in the psychic cost of government care,[8] there may be no changes in regulation and optimal levels of provision, but for higher values of s, there may be an abrupt change in the optimal policy, with a withdrawal of state provision and a discontinuity in the level of regulation upwards.

The model presented shows the dilemma faced by the welfarist planner: if the selfish cannot be successfully identified from the altruist, then some policies designed to protect the vulnerable will harm some of the agents they are supposed to protect. For most of the model, the selfish agents required only minimal incentives to pass on their charges to the state. In practice, selfish agents might be more reluctant to give up control, if, for instance, doing so means the loss of access to assets.[9] I have also concentrated on the case where carers differ in their degree of altruism rather than in the depth of their insight into their dependent's preferences. If it is the latter dimension across which carers really differ, then again it may be more difficult to design policies which encourage individuals to give up control voluntarily even when it is in the interests of the vulnerable.

This model can be extended in a number of different directions. One obvious limitation is that it assumes costless enforcement of state policies, but as many governments have found, maltreatment of dependents by carers can be expensive to detect. So, in practice, authorities must typically spend resources enforcing decisions. We now consider briefly an extension to the model in which this issue is faced.

8.4.5 Detection and Enforcement

Suppose that carers choose x, which is observed with probability $\pi > 0$, at a cost to the state of $c\pi$, for $c > 0$. If is found to be below x, then carers face a punishment $d > 0$. It is well known that if d and π can both be chosen then the optimal policy is one where π is infinitesimally small and d is infinitely large. This maximises enforcement for minimum cost. On the other hand, punishment is rarely infinite for a variety of reasons. One is the need for marginal deterrence if there are 'less' and 'worse' crimes. Another is the possibility of error in the justice system. Consequently, let d be positive and exogenous in what follows, concentrating on the choice of the probability of observation and its relationship to the optimal value of x.

For selfish carers, the expected payoff from deviating from the regulation is, $\left(2x - x^2 - \pi d\right)$ whereas if they satisfy the regulations then utility is $\left(2\underline{x} - \underline{x}^2\right)$.

---

[8] Changes in s might arise through changing perceptions about the stigma or otherwise attached to state provision.

[9] For instance, the benefits from living rent free in an elderly parent's home.

Conditional on not setting $x = \underline{x}$, the optimal choice for the selfish carer is either $x = 1$, or surrender of their dependent to the care of the state, if that is available. The optimal choices are:

1. (If available: Surrender dependent to state when $\underline{x} > 1$.)
2. Set $x = 1$ if $\underline{x} < 1$ or $1 - \pi d \geq 0$ and $1 - \pi d \geq (2\underline{x} - \underline{x}^2)$
3. Set $x = \underline{x}$ if $(2\underline{x} - \underline{x}^2) > 1 - \pi d$ and $1 \geq \underline{x}$.

The optimal choices are set out in Fig. 8.3 for different values of π and $\underline{x}$, on the assumption that $d > 1$ and that state provision is not available.

For low values of $\underline{x}$, the regulation is not binding and so the selfish carer sets $x = 1$; for intermediate values of $\underline{x}$, the carer obeys the regulations if the expected penalty is high enough, but if the probability of detection is low or the penalty is small, then the carer disobeys the regulation and sets x equal to the value which would be optimal for them in the absence of any government intervention.[10]

For the complete altruist, behaviour is unaffected by π and d. However, if part of the policy of punishment for failure to obey regulations involves removing the charge from the carer, then complete altruists may also set $x = \underline{x}$, even when it

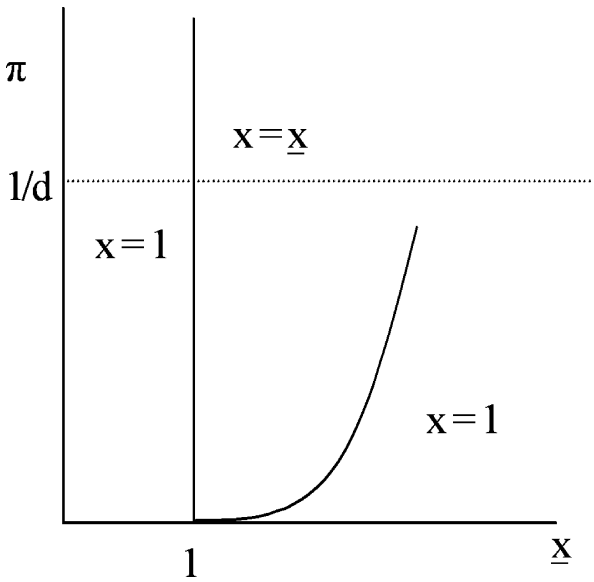


Fig. 8.3 Regulation of Carers

[10] In the example π is independent of the deviation from $\underline{x}$. If the probability of detection is increasing in the degree to which the carer deviates from the prescribed value for x, then as is standard in this kind of model, provided the carer's utility function is quasi-concave in the choice variables, then for intermediate values of $\underline{x}$, the optimal choice of x will be a continuous and increasing function of π.

would be better for their dependents if x was set at a lower value. Even in the absence of such a sanction, if carers are not completely altruistic, for large enough values of d and π, it may be optimal to surrender their dependents to the state rather than attempting to set a level of care below the regulatory $\underline{x}$.

So, to sum up, when policies are combined there is a basic dilemma for the state: whether to have no enforcement of the regulation or to have some probability of detection which means the dependents of altruists suffer.

8.5 The Role of the State in Regulating Adulthood

In the discussion so far, the state takes dependency as given, but of course defining who is free to choose and who is not is itself a fundamental issue. In the case of children we might begin by asking why the government should be interested in this question at all. After all, it is parents who have the greatest knowledge of their own offspring – it is unlikely that the state will know better.

Though it may not always feel like it to parents, fundamentally the relationship between parent and child is very unequal. Children, after all, do not ask to be born, while they have only limited power to influence their own upbringing. It is perhaps therefore surprising that one of the major ways in which governments regulate childcare is to actually strengthen the hand of parents. The various milestones that mark a child's passing into adulthood limit the ability of children to engage in economic activity. Up to the age of 16 it is very difficult in Britain for a child to be able to pass out from under the control of his or her parents without coming under the supervision of other adults or the state. The value of this legal restriction to society and to the child itself arises from the outline of child development given above. Full rationality involves self-knowledge. It means knowing what I know, but also knowing what I do not know. With immature children this is not the case and as we have seen, they may know what they like, but they do not necessarily know what is best for them. This includes whether they should subject themselves to the restrictions placed upon them by their parents. By reducing the gains to be made from leaving the family prematurely, the state strengthens the hand of parents and (it is to be hoped) thereby increases the well-being of their offspring.

It is interesting to consider what would happen in the absence of government restrictions – if exit from families was a purely private affair. Greven (1970), in his study of Andover (Massachusetts, USA) through four generations, argues that in the first generation of settlement parents had sufficient land to keep control over all their offspring with the promise of bequests. Children married late, suggesting that parents maintained economic power over their offspring and were able to control the date of adulthood (i.e. marriage).

Anderson's, 1971, study of Lancashire in the nineteenth century emphasises that as urbanization increased and the opportunities for children to earn a living

independently of their parents in the newly-emerging textile industries rose, so the balance of power shifted towards children.[11] He writes,

> Children's high individual wages allowed them to enter into relational bargains with their parents on terms of more or less precise equality. . ..The contrast between the choice element in these relationships between urban children and their parents, and the situation in rural areas is very marked. In the rural areas even in the short run, child and father entered a bargaining situation with the child at a very considerable disadvantage, because the father had complete control over the only really viable source of income. (pp. 131–132).

So adulthood (in the sense of decision-making independence) is an economic variable which may not be set optimally in the free market. Yet, if the problem were purely one of children leaving the family too early, then the optimal level of adulthood would be that set by the parents. In other words, the state would devolve the decision to parents rather than child. However, resource allocation in the family is a matter of bargaining: the parent weighs his or her own consumption against that of the child. To the extent that the child has no right to leave the family and pass into adulthood, then even if the parents care about the child, the distribution of welfare may be suboptimal. The gain to the parent of the child's going independent may be less than the loss of consumption that it implies for the parent (given the 'selfish kid' theorem the child will be maximising the income of the family) plus the loss caused by separation itself. For that reason the parent may resist adulthood even when it is desirable for society as a whole.

8.5.1 Welfare Functions with Children

Underlying this argument is a welfare function for society. There is little debate and even less consensus about the shape of well-being when children are involved. The problem chiefly arises because of the standard problem of constructing sensible measures of welfare when some individuals care about the welfare of others.[12]

Within the welfarist framework there are essentially two views. In the first welfare, W is a function of parental utility, U^p (suppose for simplicity that the economy consists of one child and one adult). In turn, parental utility is based on a vector of consumption goods, x, (some of which may be jointly consumed with the child) and the utility of the child, so it has the form, $W = W(U^p(U^c, x))$. This formulation, associated with the unitary model of the household (Becker, 1981), means that parental decisions are automatically correct from the viewpoint of social welfare, but suffers from two main failings. First it gives the child no identity – everything depends on the weight given by the parent to the child, so that in the extreme case where the

[11] It is interesting that the first major restrictions on child labour arose in this period, so shifting power back towards parents.

[12] Most of the work involving welfare functions with children has been concerned with the construction of equivalence scales i.e. the question of how many children equal one adult in cost of living indices. Fisher (1987), argues that these provide no basis for welfare comparisons, being merely technical comparisons of cost functions without a welfare basis.

parent does not value the child at all, then society does not value the child either. The second problem is that the function does not satisfy what could be called the 'principle of the irrelevance of adulthood for welfare'. It is normal to write welfare functions for adults as $W = W(U^p, U^c)$.[13] This formulation implies that everyone matters as an individual and not just as an element in the consumption of others. But if it is accepted for adults, but the previous formulation of the welfare function is the one relevant parents with children then welfare will be discontinuous at the point of adulthood. Yet immediately before and immediately after this point, presumably the consumption patterns and tastes of the individuals concerned are only infinitesimally changed, so on this basis the formulation is unacceptable. But the alternative implies that parents always give too little weight to their children from the point of view of society and so intervention in intra-family allocation is generally justified.[14] Thus double-counting of children appears equally unappealing.

8.5.2 Optimal Policy

I argued above that, from an early age, children tend to know something about their own preferences whereas parents often have a better idea of the relationship between actions and their eventual consequences. As children mature they become more aware of the nature of this relationship and hence the errors they would make through independent decision-making diminish. Suppose, for instance that child utility is given by $U(Bx, \gamma)$ where x is the consumption vector as before, B is a production relationship (here, for simplicity, a matrix) that converts consumption acts into utility consequences and γ is a parameter for preferences. Let G(B,γ) be the cumulative density function for these two factors, G_B and G_γ as the marginal density functions. Expected child utility is therefore, $\int u(Bx, \gamma)dG(B, \gamma)$.

If the child knows its own preferences, γ^* and the parent knows that $B = B^*$, then the child will maximize, $\int u(Bx, \gamma^*)dG(B, \gamma^*)$. The selfless parent, on the other hand would maximize $U' = \int u(B^*x, \gamma)dG(B^*, \gamma)$. The optimal age for adulthood is therefore a matter of choosing the age at which the outcome of maximizing the first of these provides a higher value for true welfare.

Consider the dilemma of the completely selfless parent deciding when to allow their child to make their own decisions. Once that decision is made it is not typically reversible. The essential problem for the parent who has full control is whether to pass control in period t, or whether to delay for one more period and assume responsibility for choices in the current period. In the short run the choices of the parent are likely to be superior. But if the child learns more effectively through making their own choices then passing responsibility to them can produce higher

[13] Note that U^p could still contain U^c as an argument (or vice versa).

[14] Ross (1991), gives an interesting analysis of the implications of this view. Not surprisingly, optimal taxes involve subsidizing those goods that the child consumes relative to those that the parent consumes. Transfers in-kind play a part here, too. If parents are unable to resell education for instance, public provision of it can raise welfare.

benefits in the long-run. The basic implication of the principle of optimality is that, if the parent and child share the same discount rate, assessment of utilities and time horizon, then the parent will not delay transferring power to the child beyond the optimum, even if the parent would make the better choice in the current period.

However, in general parents and child may not share the same maximand. Let V^t be the net present value of the child's welfare in period t as assessed by the parent. That is,

$$V_t = \sum_{s=t}^{s=T} \delta^{s-t} U_s = U_t + \delta V_{t+1} \tag{8.6}$$

where T is the time horizon for the parent and δ is the parent's discount rate for the child. From the child's point of view, the maximand is,

$$V'_t = \sum_{s=t}^{s=T'} \left(\delta'\right)^{s-t} U'_s = U'_t + \delta' V'_{t+1} \tag{8.7}$$

There are four important sources of difference between these two expressions:

- Parent and child may have different discount rates.
- Parent and child may make different period by period assessments of utility for the child.
- Parent and child may have different time horizons. For instance, the parent might only value the child's well-being during their own lifetime.
- The evolution of the child's preferences (including the discount rate) may depend on whether they are responsible for making their own choices in a manner which cannot be fully anticipated by the parent.

Because the expressions differ potentially in so many ways, it is not clearcut that the child will wish to have an earlier date of independence compared to the parent. However, for simplicity let us suppose that the only change from period to period is in the child's capacity to make sensible choices – so that the parent would always make the same choice for the child. Suppose also that, if the child makes choices the sequence of period by period utilities is increasing (i.e. they improve through practice), then if the parent discounts the future more heavily than the child or has a shorter time horizon, then the parent will wish to delay longer the time of independence compared to the child.[15]

So, in theory, the state has a role to play in regulating the date of adulthood, if parent's preferences are not aligned with the true interests of their child. Some additional considerations to be entered into the calculation include:

[15] As we noted in the discussion of the evolution of children's preferences, it is often supposed that children are too impatient for immediate satisfaction, compared to what is optimal for their well-being. The impatient child may wish to leave choices to their parents when those choices produce immediate consumption benefits, but wish for greater control over intertemporal choices.

- in the absence of a law tying the child to the family, the child may leave once its own views of its welfare contradicted those of its parents. An age for adulthood therefore may make the parent's views enforceable.
- Against this, an upper age limit on control may be required, because the parents, in choosing the allocation of resources within the family have to balance their own welfare against the welfare of the children. So, as argued in above, the children, lacking bargaining power (because the law prevents their departure) might therefore suffer, not because the internal allocation of resources did not correspond to their own preferences, but because too little weight was given to them as a whole in the family welfare function.

Does all this imply that the rights of parents over children should be unrestricted prior to the date of adulthood? The widespread presence of child abuse and neglect in most societies suggests that that the answer should be no. The theoretical arguments of this chapter also suggest otherwise, particularly since the activities of families are partially observable to outsiders. So, it cannot be presumed that parental preferences are such that non-interference in the internal operations in the family is unwarranted, especially when the legal system restricts the rights of children to exit voluntarily from families.

8.6 Conclusions

This chapter has covered some of the agency issues relating to individuals with bounded rationality of an extreme kind. Its major lesson is that, even when it is carers who have the most insight into their dependent's needs, the state may have some role to play in regulating care. In the presence of asymmetric information, standard policies such as minimum levels of provision or the provision of state care as an alternative may penalise altruists as well as providing incentives for more selfish carers to make choices more in line with the true interests of their dependents.

Chapter 9
Tax Policy

9.1 Introduction

We have touched on issues surrounding tax in several of the previous chapters. In Chapter 7, for instance, several of the standard merit wants models were analysed. The basic theme of this chapter is that even with an incomplete welfare ranking it is possible to make some claims about the nature of desirable tax policies. I support this argument in two ways:

- by considering the implications of monotonicity for the design of tax policy
- by examining the issue of tax reform when preferences are reference dependent

9.2 Monotonicity

Subject to the usual caveats about excess, 'more is better' appears to be the most obvious and most innocuous requirements to impose on the welfare relationship. Although it is possible to deduce some implications of monotonicity for tax policy not much mileage can be had out of it. As we shall see, in the absence of a complete set of policy instruments set optimally, monotonicity does not for instance imply that cost efficiency is a good thing or even that technical efficiency is desirable. For instance, if an improvement in the production of food releases agricultural resources for other uses, the smoker may simply smoke more. So, technical efficiency may not be welfare improving because of the use that individuals may make of the resources released when moving from an inefficient situation to an efficient one. Since cost efficiency requires technical efficiency, it follows that monotonicity does not imply cost efficiency either. In addition, price changes set in chain by a move from a cost inefficient allocation to one of efficiency may produce substitution effects which again could be welfare-reducing for the individual.

To gain more insight let us consider two kinds of model, in both of which a single agent makes choices of x and y (both vectors). For the simpler model I assume that the individual has preferences of the form $U = U(x, f(y))$, where x is a vector of n consumption goods and y is a vector of m inputs into a production function, f. A typical element of x is x_i, a typical element of y is y_k. For instance, f could

A. Munro, *Bounded Rationality and Public Policy*, The Economics of Non-Market Goods and Resources 12, DOI 10.1007/978-1-4020-9473-6_9,

© Springer Science+Business Media B.V. 2009

be health, which is the product of a number of inputs. U is assumed increasing in all its arguments, twice differentiable and strictly quasi-concave. Meanwhile f is also increasing in its arguments, twice differentiable and strictly concave. Welfare is given by $U(x\Theta, f(y))$ where Θ is a diagonal matrix. Welfare and preferences are therefore in accord when $\Theta = I$ where I is the identity matrix. Production technology for the economy is represented by the strictly convex, closed and bounded subset of $\mathbb{R}_+^{n+m}$, $G(x, y)$. The outer envelope of G is summarised by $g(x, y) = 0$, where I assume that g is twice differentiable. It follows that, except at boundary points, the assumptions of the implicit function are satisfied for g. I shall let p be a m+n vector of consumer prices, with p^x and p^y representing the n and m vectors of prices for x and y respectively; similarly t represents a m+n vector of unit taxes, with t^x and t^y having the obvious interpretation. T is a lump-sum tax (or lump-sum grant if negative) and t^f is an ad valorem tax on f. Finally $w = p - t$ will be a m+n vector of producer prices, with w^x and w^y defined analogously to p^x and p^y.

For the more general model, the welfare function and the production interpretation of y are maintained, but I shall not assume that the individual has complete and transitive preferences over x and f. What will be assumed is the existence of a choice mapping (MClennen, 1990; Sen, 1970) – i.e. that for any non-empty choice set, X, there is a non-empty set of choices. I shall assume some additional features for C; chiefly that it is single-valued and differentiable in some typical parameters of the budget set. Although it is useful to know that it is possible to state conditions under which productive efficiency is desirable, the added level of complexity brought on by weakening the assumption of utility maximization is not compensated to any great extent by the extra generality of the results which are on offer. Consequently, I shall confine description of the more general model to Appendix 2.

9.3 Technical Efficiency – Marketed Goods

We can model technical efficiency in the production of y, by writing g as $g(x, y\Lambda)$ where Λ is an mxm diagonal matrix with strictly positive diagonal elements, a typical member of which is $\lambda_k \geq 1$. An increase in λ_k represents a decrease in technical efficiency; $\Lambda = I$ represents technically efficient production. Say that efficiency is *welfare desirable* if W is strictly decreasing in λ_k for all k and similarly say that efficiency is *preference desirable* when U is strictly decreasing in λ_k for all k and when U is increasing in all its arguments we have *monotonicity in preferences*.

The question we wish to ask is whether efficiency is welfare desirable. The answer depends on three factors: the relationship between welfare and preferences, the instruments available to the government and the technological possibilities given by g. More formally,

$$\frac{dW}{d\lambda_k} = \sum W_i \theta_i x_{ik} + \sum W_f f_l y_{lk} \tag{9.1}$$

where subscripts indicate partial derivatives. In other words, W_i is the partial derivative of W with respect to the ith argument of $x\Theta$, x_{ik} is the derivative of x_i with

respect to λ_k, W_f is the partial derivative of W with respect to f, f_l is the partial derivative of f with respect to its lth argument and y_{lk} is the derivative of y_l with respect to λ_k. We also have,

$$0 = \sum g_i x_{ik} + \sum g_f \lambda_l y_{lk} + g_k y_k \tag{9.2}$$

where g_i is the partial derivative of g with respect to its ith argument etc.

Result 9.1 *Suppose the individual chooses x and y, maximizing U subject to the constraint given by g. Then (a) monotonicity in preferences implies that technical efficiency is preference desirable and (b) monotonicity in preferences and* $\Theta = I$ *implies welfare desirability.*

To understand the first part of (9.1), define α as the marginal utility of income then,

$$\frac{dU}{d\lambda_k} = \sum U_i x_{ik} + \sum U_f f_l y_{lk} = \alpha \sum g_i x_{ik} + \alpha \sum g_f \lambda_l y_{lk} = -\alpha g_k y_k \tag{9.3}$$

where I have used the first order conditions for utility maximization. Thus provided monotonicity in preferences is satisfied (so that $\alpha > 0$) then the consumer prefers technical efficiency over inefficiency.

What is the welfare perspective? The change in welfare following a decrease in technical efficiency is,

$$\frac{dW}{d\lambda_k} = \sum W_i \theta_i x_{ik} + \sum W_f f_l y_{lk} = \sum \left(W_i \theta_i - \frac{\alpha W_f}{U_f} \right) x_{ik} - \frac{\alpha W_f}{U_f} g_k y_k \tag{9.4}$$

If $\theta_i = 1$ all i, then $W_i = U_i$ and $W_f = U_f$, in which case the right hand side of this equation reduces to $-\alpha g_k y_k$k and we have Result 9.1b.

In general the welfare and utility equations for the effect of a marginal decrease in efficiency do not coincide, because the marginal welfare value of a change in consumption of the ith good is not equal to its price. Define π_i as the shadow price of good i with π^x as the shadow price vector which supports consumption x. Let α^* be the marginal welfare value of extra income, then the equation for $dW/d\lambda_k$ can be re-written as,

$$\frac{dW}{d\lambda_k} = \alpha^* \left[\sum \left(\pi_i - \frac{\alpha W_f}{\alpha^* U_f} \right) x_{ik} - \frac{\alpha W_f}{\alpha^* U_f} g_k y_k \right] \tag{9.5}$$

In this equation $W_f \alpha / U_f \alpha^*$ represents the degree to which the individual under or over-consumes f when viewed from the welfare perspective. If $W_f \alpha / U_f \alpha^* > 1$ then there is over-consumption; if $W_f \alpha / U_f \alpha^* < 1$ then f is under-consumed.

Result 9.1 applies when there is no difference between welfare and preference orderings. When welfare and preference diverge the situation is not so simple as the introduction to this section specified. Next I offer some special cases when technical

efficiency is definitely welfare desirable. The proofs are in the appendix. There are two underlying themes to these cases. First, following Diamond and Mirrlees (1971) we know that production efficiency is desirable in general if there is a government instrument for which welfare is unambiguously increasing. Secondly, production efficiency may also be desirable for particular patterns of preferences.

Result 9.2 *Suppose the government can choose a vector of taxes,* $\tau = (t, t^f, T)$ *where t is vector of unit taxes on x,* t^f *is an ad valorem tax on f and T is a lump-sum tax, then technical efficiency is welfare desirable.*

Result 9.3 *Suppose* $x_{il} = 0$*, all i, then technical efficiency is welfare desirable.*

Result 9.4 *Suppose (a) utility has the form* $U = U(h(x), f(y))$*, (b) welfare has the form* $U(\theta h(x), f(y))$*, for some* $\theta > 0$*, and (c) the government can choose an ad valorem tax* t^f *then technical efficiency is welfare desirable.*

Result 9.5 *Suppose (a) utility has the form* $U = U(h(x), f(y))$*, increasing in both h and f and (b) welfare has the form* $U(\theta h(x), f(y))$*, for some* $\theta > 0$*. If following a decrease in technical efficiency, (a) the value of the change in x is positive and (b) the value of the change in y is positive when measured at market prices, then technical efficiency is welfare desirable.*

Clearly a large number of similar results could be offered. Some remarks about them are probably in order. First, these conditions are sufficient not necessary; technical efficiency may be desirable in many other situations. Secondly, it is not clear which of them provides most comfort to the presumption that technical efficiency is generally a good thing. Result 9.2 requires the weakest assumptions on the nature of preferences, but then has the strongest assumptions about the instruments available to the government. Note too that the taxes required to ensure the welfare desirability of technical efficiency are, in general, individual specific and specific to the values of the θ_is. It follows that in a multi-agent economy it becomes less realistic to think of a government having the set of instruments sufficient to guarantee the welfare-desirability of technical efficiency. Arguably the most useful of the results is the last one, because it provides a simple test of the value of an improvement in efficiency, but here too, in general the ad valorem tax on f must be individual specific.

Results 9.2–9.5 concentrate on the combinations of preferences and instruments which make technical efficiency desirable. I have not said much about whether the technology of production might also throw useful special cases. Changes in efficiency create income effects and substitution effects in the consumption of other goods. Suppose that the price of the consumption goods is largely unaffected by a change in technical efficiency and suppose all of them are normal, then a rise in efficiency will increase consumption of all goods. If all goods are good, then again technical efficiency will be welfare desirable. This last example raises another point about the implementation of technically efficient changes. When welfare desirability and preference desirability diverge, then changes in efficiency might be viewed by consumers as bad, when a welfare perspective would see them as desirable.

The preferences set out above have a very specific form: the goods where preference diverges from welfare are not the goods which are the subject of technical

efficiency changes. What happens if we allow other patterns of preference? Suppose preferences are of the form $U(x, f(y\Theta))$ where Θ is now an mxm diagonal matrix. We can interpret this matrix as representing the agent's misunderstanding of the true technical relationship between y and f. In this situation $dW/d\lambda_k$ is given by,

$$\frac{dW}{d\lambda_k} = \alpha^* \left[\sum \left(\pi_i - \frac{\alpha W_f}{\alpha^* U_f} \right) x_{ik} - \frac{\alpha W_f}{\alpha^* U_f} g_k y_k \right] \tag{9.6}$$

As we can see from the form of this equation, the same basic issues arise, but some new results are possible if some of the levels of provision of some goods are controlled by government, as the next sub-section demonstrates.

9.3.1 Technical Efficiency – Non-Marketed Goods

With non-marketed goods, there is the possibility that some levels of consumption are set directly by the government. Suppose for instance that y is set by the government then we have the following results.

Result 9.6 *Suppose $W = U(x, f(y\Theta))$ and the government chooses y then technical efficiency is welfare desirable.*

Result 9.6 is probably the most interesting of those presented so far. It says that if the divergence between preference and welfare is confined to the goods under government control and those goods are weakly separable from the remaining goods consumed by the agent, then technical efficiency is desirable. We can obtain similar results for the different form of preferences presented in the previous sub-section:

Result 9.7 *Suppose $W = U(\theta h(x), f(y))$ and the government chooses y then technical efficiency is welfare desirable.*

To see the significance of Result 9.7, consider an example where consumers' preferences between health and other goods do not diverge from welfare, but their preferences over the choice of medical technique (or other aspect of consumption which affects only health) do diverge. The government cannot control total health directly, but it can control individual's choices of medical techniques. Provided health is weakly separable from other goods, then technical efficiency is desirable.

9.3.2 Cost Minimization

The previous section considered the issue of technical efficiency. Suppose now that we can assume technical efficiency and we wish to know whether a producer in a decentralised economy should be cost efficient. Formally, the analysis is very similar to that for technical efficiency and, as a consequence, the answer is also very similar: cost efficiency is not automatically desirable when preferences do not accord with welfare.

As before a representative consumer chooses x and f (i.e. the result of inputs y). He or she faces a given price vector (p^x, p^f), but the mix y is chosen on behalf of the consumer. The question is, should y be chosen so as to minimize the cost of producing f(y). Then the question, reformulated, is whether:

$y = \text{argmin}\{w^y y \mid f(y) \geq f\}$ is a generally desirable property?

Now clearly, for any choice of y on the convex production possibility frontier there must exist some vector of supporting shadow prices. So the real question reduces to whether this vector is equal to the shadow values of y in the production function. That is, given differentiability, whether,

$$\frac{f_k}{f_l} = \frac{g_k}{g_l} \; k = 1 \ldots m; l = 1 \ldots m.$$

If the consumer chooses x and f given the price vector (p^x, p^f) and an income constraint, m, then this leads to a constrained indirect utility function $V(p^x, p^f, m)$ and in turn this implies an constrained indirect welfare function of $W(\pi^x, \pi^f, z)$, where π^f is the shadow price vector which supports f and z is the corresponding virtual income, equal to $\pi^x x + \pi^f f$. The welfare maximizing planner's problem is then to choose y to maximize W subject to individual maximization by the agent and the resource constraint, that is:

$$\max W((\pi^x, \pi^f, z) - \mu g(x, y).$$

Here, μ is the shadow price of the resource constraint. Maximizing welfare under these conditions produces a typical first order condition,

$$\sum W_i \pi^x_{ik} + W_f \pi^f_k + W_z \sum(\pi^x_{ik} x_i + \pi^x_i x_{ik}) + W_z(\pi_f f_k + \pi^f_k f) - \mu(g_k + \sum g_i x_{ik}) = 0 \tag{9.7}$$

In this equation I have redefined the meaning of the subscript terms. W_i is the partial derivative of W with respect to the ith virtual price, π_{ik} is the change in the ith virtual price given a marginal increase in y_k, x_{ik} is the change in x_i given a marginal increase in y_k. Simplifying (9.7) using Roy's identity yields:

$$W_z \sum(\pi^x_i x_{ik} + \pi_f f_k) = \mu(g_k + \sum g_i x_{ik})$$

Now since changes in y_k only affect x_i via f, we have $x_{ik} = x_{if} f_k$. Simplifying we obtain,

$$f_k \left(\sum [W_z \pi^x_i - g_i \mu] x_{if} + \pi^f \right) = g_k$$

Since the term in parenthesis is constant across all goods we get the required result:

Result 9.8 *If consumers choose x and f then cost minimizing production of f is welfare desirable.*

9.3.3 Summary

The results of this section are standard in that they are closely related to the tradition in public economics created by Diamond and Mirrlees (1971). The conclusions are not entirely negative. On the other hand there are not entirely reassuring either. To see this, let us try to interpret them within a public policy context. Suppose we have two sectors of the economy: a preference sector (the 'x' goods) and a production sector (the 'y' goods). These goods are defined by the way they enter individual welfare functions. For the x goods there is variety in the mix which is good for individuals; for the y goods, there is an optimal mix which is the same for all individuals. For instance, the foods we like to eat might vary considerably from person to person, but there may be only one best way of dealing with a particular type of cancer. The results of this section say that if it is in the y goods that tastes depart from welfare and if the (benevolent) state controls the consumption of the y goods, then technical efficiency is desirable. Technical efficiency in the y good is also a good thing when tastes depart from welfare for the x goods, but only if there are sufficient policy tools available or there is sufficient separability between the x and y sets of goods. Of these two conditions, perhaps it is the second that is most likely to be satisfied in a multi-agent multi-good world.

9.4 Tax Reform Issues and Loss Aversion

We now turn attention to another case where the welfare ranking is potentially incomplete. Recall that consumer preferences are said to be reference dependent if an agent's ranking of two bundles is not independent of his or her vantage point. In empirical work, the most commonly encountered example of reference dependent preferences is loss aversion – said to occur when an individual places a greater weighting on changes in consumption when these changes represent losses rather than gains. This in turn produces 'endowment effects' or 'status quo bias' when an agent prefers bundle x over bundle y when holding x, but y over x when y is held.

9.4.1 The Problems Created by Reference Dependent Preferences

As in previous chapters, let x, y, r, s be vectors representing bundles of n goods and let $\prec_r$ be the strict preference relationship as viewed from reference point r. In the standard or Hicksian model of consumer preferences, if x $\prec_r$ y for some r, then $x \prec y$, because preferences are not dependent on a particular reference point. Thus if a reform leads to a consumer's bundle changing from y to x and the consumer

prefers x to y from the vantage point of y (the ex ante viewpoint), then it is also true that x is preferred when the reference point is x (the ex post view). Therefore, in the absence of uncertainty, ex ante and ex post views of welfare coincide, yielding three fundamental propositions: (1) there are no cycles in strict preferences; (2) if a reform alters budget sets, any reform which is feasible ex ante is also feasible ex post and (3) in a multi-agent economy, if an allocation is Pareto efficient ex ante, then it is also efficient when view ex post. None of these propositions need be true with reference dependent preferences. First, the endowment effect rules out cycles involving only two bundles, but it does not necessarily rule out more complex cycles, in which for instance, x' is preferred to x when x is the reference point, a third bundle is preferred to x' when it is the reference point and so on until the cycle is complete. Such cycles make the assessment of individual welfare difficult, if not impossible. Second, as long as preferences vary with the reference point, an allocation which is Pareto efficient when viewed from one set of reference points may not be when viewed from the allocation itself.

Third, reforms(such as alterations to prices) which appear feasible when viewed from the original reference point may not be ex post. Figure 9.1 shows an economy with the feasible set defined by RR. Price reform changes the budget constraint from AA to BB; as before, x represents the original bundle and reference point and x' is the interim bundle (i.e. the agent's optimal choice given the new price vector and preferences as viewed from x). The solid indifference curves represent preferences as viewed from x. The broken indifference curve through x' represents preferences as viewed from x'' – the point which is optimal viewed from itself, given the new price vector. Note that x'' may not be optimal as viewed from x' – a sequence of steps may be necessary to bring the consumer to x''. x'' is therefore preferred to x as viewed from x'' (and also as viewed from x', but this is not necessary for the point of the example), but it is not feasible. In short, the change in consumption which results from reform appears feasible ex-ante, but is not ex-post.

Loss aversion in particular and reference dependent preferences more generally therefore raise fundamental questions about appropriate measures of individual welfare. A constant theme of this book is that it is not immediately apparent how an index of individual welfare is constructed when preferences are apparently contradictory. In Chapter 5, we encountered the argument that it is perhaps better to view choices as providing evidence on welfare, without assuming that such evidence is definitive. Thus, when an individual views bundle y as superior to bundle x from x and also from y the evidence consistently points to the superiority of y over x. This is one reason for wishing to know whether the ex-ante and ex-post views are in accordance. A second reason for focussing on these two indicators of welfare is more practical. To be enacted, a reform has to be agreed, which, in a one person economy, means that the individual concerned must vote for the reform. Thus ex-ante preferences are important. But once the reform is enacted, if it is not to be reversed then the individual must also view it as ex-post desirable. Thus a reform which passes both tests is more likely to be enacted and persist than one which does not. The key question is then whether it is possible to say anything about ex post

welfare on the basis of ex ante judgements, since it is ex ante information which is available to the decision-maker.

Formally, let r be the initial reference point, from which all changes are assessed ex ante. A change in consumption from x to y is *ex-ante preferred* if $x \preceq_r y$.[1] Let B be a budget set and let S be the set of feasible consumption bundles. A reform yielding the budget set B is *ex-ante feasible* if $y \in S \cap B$, for any y such that $w \preceq_r$y $\forall w \in B$. A change in consumption from x to y is *ex-post preferred* if $x \preceq_y y$ and a reform yielding a budget set B is *ex-post feasible* if $y \in S \cap B$ and $w \preceq_y \forall w \in B$. A reform is *ex-ante desirable* if it is both ex-ante preferred and feasible; a reform is *ex-post desirable* if it is both ex-post preferred and ex-post feasible. The question is then whether ex ante desirability says anything about ex post desirability.

In Chapter 4 I put forward a set of axioms for reference dependent preferences, along the lines employed by Munro and Sugden (2003). Compared to the standard Hicksian model of the consumer, the key additional assumption was that of acyclicity (C7). In addition, continuity in reference points was also assumed. The assumption of acyclicity implies that a change of allocation or tax rate which is ex ante preferred will also be ex-post preferred. As Fig. 9.1 suggested, the issue of feasibility is not so easily dealt with. As viewed from x, y may be optimal on B and also feasible, but then from y another allocation may be preferred and so on, along a chain of allocations, until a point is reached which is still within the budget set, but no longer feasible. We are then faced with two modelling options. The first is to amend or strengthen the model of reference dependent preferences so as to rule out chains of this sort. Under this approach, acyclicity is strong enough to guarantee the existence of an general competitive equilibrium, but too weak to be entirely useful as an assumption for the purposes of understanding tax reform. The second approach

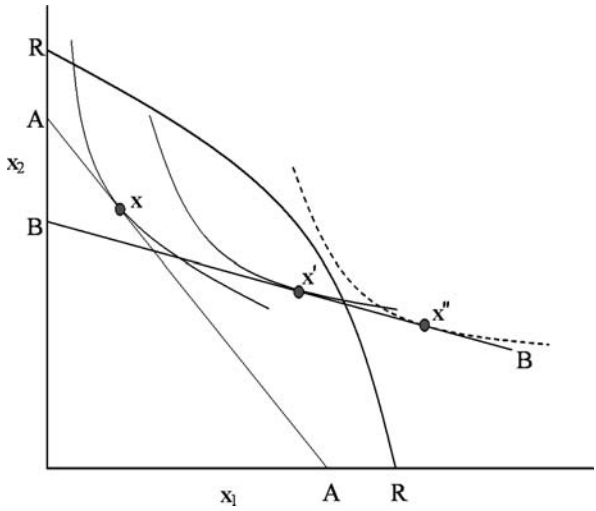


Fig. 9.1 Ex Post Feasibility of Tax Reforms

[1] It would be usual, but not necessary to think of the reference point as x in this case.

is to model production more deeply, so as to allow for the endogeneity of prices and guarantee feasibility. I shall consider both approaches. Initially, in the following sub-sections, I take a tack on reference dependent preferences slightly different to that pursued in Chapter 4. This has some considerable merit in simplifying the issue of equilibrium, but as we shall see it only takes us so far in understanding tax reform with reference dependent preferences. Subsequently, I extend the model of general equilibrium to allow for production.

9.4.2 Extensions to Consumer Theory

In Tversky and Kahneman's 1992 theory of riskless choice, they make the assumptions of loss aversion and diminishing sensitivity. As we remarked in Chapter 2, there is powerful evidence for loss aversion, but in the context of riskless choice there is very little evidence in favour of the second assumption. So, in this chapter I take C1–C6 (from Chapter 4) and then make the following assumptions about reference dependent preferences:

Axiom 9.1 *LA (Loss Aversion).* $\forall$ *r, s, such that* $r_i < s_i$ *and* $r_j = s_j \forall$, $j \neq i$, *if either*

1. $x_i \leq r_i \leq y_i$ or
2. $x_i \leq s_i \leq y_i$,

then the move from r to s favours y.

Axiom 9.2 *NS (Non-sensitivity).* $\forall$ *r, s, such that* $r_i < s_i$ *and* $r_j = s_j$, $j \neq i$, *if either*

1. $y_i \leq r_i$ or
2. $s_i \leq x_i$,

then $x \preceq_r y \leftrightarrow x \preceq_s y$.

These axioms are illustrated in Fig. 9.2.[2] Viewed from r, the move from x to y represents a loss in the x_2 dimension while x represents a loss and y represents a gain in the x_1 dimension. From s, the move from x to y represents the same reduction in gains on the x_2 dimension as from r. However, along the x_1 dimension the movement from x to y turns a larger loss into a small gain. Since losses loom larger than gains, then, compared to r, y is favoured over x from s. By the same token, a move from r' to s or s' or from r to s' also favours y over x, according to LA. However, NS implies that moving from r′ to s″, or from r′ to s′ has no effect on preferences (because from

[2] In Chapter 4 I stressed the point that a theory of reference dependent preferences not only requires a set of assumptions describing the nature of preferences contingent on a reference point, it also requires a theory of how reference points come to be. Two possibilities were considered for the 'standard' case: the consumer's endowment and their customary consumption. Here, I am mostly concerned with instances where a consumer buys a certain bundle of goods from out of a given budget set. Consequently it seems most appropriate to use the customary consumption notion of the reference point.

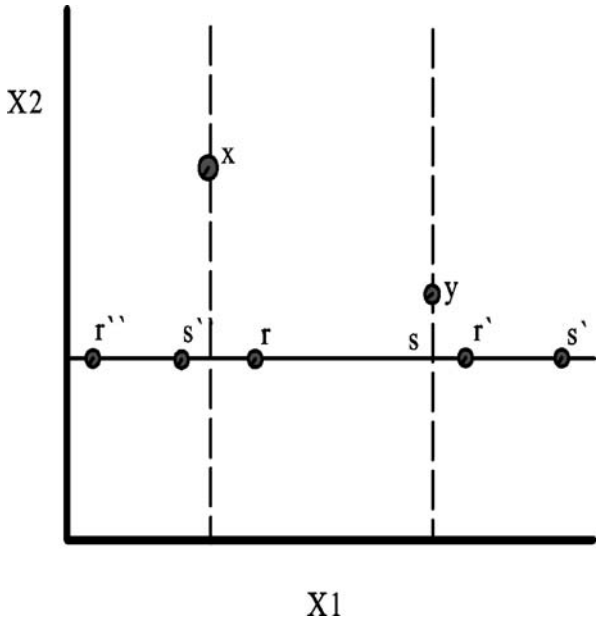


Fig. 9.2 Axioms on Loss Aversion

both r'' and s'', the moves to x and y involve gains on the x_1 dimension, while from r' and s', x and y both involve losses on the x_1 dimension). For simplicity, in what follows, I shall describe changes in reference point where either $y_i \leq r_i$ or $s_i \leq x_i$, with $x_i \leq y_i$ and $r_i \leq s_i$, as being *outside the box*.

NS is a strong assumption which makes this formulation of reference dependent preferences a special case of that developed in Munro and Sugden (2003). It shows the kind of assumptions required to produce the standard arguments of welfare economics when preferences are reference dependent. How though does it square with the evidence presented in Chapter 2? Most of the evidence on reference dependent preferences is concerned with changes where x and y are the reference points. This evidence is completely compatible with LA and NS, as is the evidence on equivalent loss and equivalent gain presented in Bateman et al. (1997b) (and in subsequent experiments by the same team). Table 9.1 compares the predictions of LA and NS with the remaining evidence for reference point changes which are outside the box. The reference points referred to are those in Fig. 2.3 on p. 32. Three of the four pieces of evidence are consistent with LA and NS. The remaining item refers to what Kaisa Herne calls a *relative closeness* effect for which she finds some weak evidence in her experimental work.

Thus there is some evidence against NS, although it is weak.

9.4.3 Implications of the Theory for Welfare Economics

The major consequence of axioms NS and LA is that if bundle y is preferred to bundle x when x is the reference point, then y is preferred to x for all possible reference points, as can be seen as follows.

Lemma 9.1 *Suppose LA and NS, then (i)* $x \preceq_x y \rightarrow x \preceq_r y \; \forall r \in \mathbb{R}^n$ *and (ii)* $x \prec_x y \rightarrow x \prec_r y \; \forall r \in \mathbb{R}^n$.

Proof Let x_i i=1,. . ., n be the ith component of x etc. Consider a move from x to r broken down into at most n steps, the ith of which keeps all components of the reference point constant except the ith, which is changed from x_i to r_i. Each step moves the reference point closer to y or further away or can be broken down into two sub-steps, the first of which moves the reference point closer to y and the second of which moves the reference point further away. Suppose the step (or sub-step) moves the reference point closer to y, then by LA the step favours y. Suppose the step (or sub-step) moves the reference point away from y, then by NS, since we also move away from x, there is no effect on the ranking of x and y. Thus the move favours y and this applies whether we are considering the strict or the weak preference relationship.

Remark Lemma 9.1 has a claim to be an alternative to LA and NS as an axiom for reference dependent preferences, given its intuitive attractions: if one bundle is not preferred over another when the first is the reference point then it seems plausible that it will not be preferred when the bundles are compared from elsewhere.

Lemma 9.2 *Given LA and NS, then (i)* $[\exists r$ *such that* $y \preceq_r x] \rightarrow y \preceq_x x$ *and (ii)* $[\exists r$ *such that* $y \prec_r x] \rightarrow y \prec_x x$.

Proof For (i), suppose not, then by completeness, x $\preceq_x$y, but then by Lemma 9.1 $x \preceq_r y$ – a contradiction. Similarly for (ii).

Recall that in Chapter 2 the key assumption additional to the standard set of Hicksian assumptions was that of acyclicity, C7. The next result establishes that LA and NS represents a special case of C7.

Result 9.9 *Given LA and NS, choice is acyclic.*

Proof Suppose false, then there must be some sequence of bundles $x_1, \ldots, x_m$ such that (i) $x_i \preceq_i x_{i+1} \forall i = 1, \ldots, m-1$; (ii) $x_1 \preceq_m x_m$ and (iii) the preference is strict

Table 9.1 LA and NS

Reference point change	Evidence	Source	LA + NS predict
r^2 to r^3	Favours y	Herne (1998)	No change
r^1 to r^{10}	Favours y	Tversky and Kahneman (1991)	Favours y
r^3 to r^8	Favours y	Herne (1998)	Favours y
$r^4$to r^6	Favours y	Tversky and Kahneman (1991), Herne (1998)	Favours y

for at least one $i = 1,\ldots, m$. Following Lemma 9.1, $x_i \preceq_i x_{i+1}$ implies $x_i \preceq_j x_j$ j = 1,. . ., m. But then we have $x_1 \preceq_1 x_2 \preceq_1 x_3 \preceq_1 \ldots x_m \preceq_1 x_1$ with at least one of the steps strict, which is a contradiction.

Acyclicity obviously means that ex ante preference implies ex post preference, thus reforms satisfying the Ex Ante Consent (EAC) principle also satisfy its ex-post sibling. We now move on to consider the link between ex ante and ex post feasibility. Let B be a non-empty budget set, then we have the following:

Lemma 9.3 *Assume LA & NS. (i) If, for some r,* $\exists x \in B$ *such that* $y \preceq_r x\ \forall y \in B$, *then,* $\forall y \in B$, $y \preceq_x x$ *and (ii) If, for some r,* $\exists x \in B$ *such that* $y \prec_r x\ \forall y \in B$, *then,* $\forall y \in B$, $y \prec_x x$.

Proof This follows immediately from Lemma 9.2.

Lemma 9.3 means that if x is optimal on some budget set, when viewed from reference point r, then x is also optimal when viewed from x. As a result, if a reform leads to x being chosen from the vantage point of another reference point, then the individual will still prefer x over all other possibilities open to them. Suppose that the feasible set is S and prior to a policy reform the reference point for an individual is r, then we have the following:

Result 9.10 *Suppose a policy reform leads to the budget set B and x is optimal on that set for some reference point r, then, given LA and NS: (i) if that reform is ex-ante desirable then it is ex-post desirable; (ii) if x is ex ante optimal on S, it is ex post optimal.*

Proof (i) The proof is straightforward, but we need to rule out the case where x is indifferent to another element in B, when viewed from x, but that element is not feasible. Let $O(r, B) = \{x | y \preceq_r x\ \forall y \in B\}$, in other words O is the set of optimal bundles in B as viewed from r. Clearly, if $x, x' \in O(r, B)$, then $x \preceq_r x'$ and by Lemma 9.3, if $y \preceq_r x\ \forall y \in B$ then, $\forall y \in B$, $y \preceq_x x$. Suppose for some $x \in O(r, B)$, $\exists x'' \in B$ such that $x \preceq_x x''$ but $x'' \notin O(r, B)$. If this is true then $x \preceq_r x''$, by Lemma 9.1, in which case $x'' \in O(r, B)$ – a contradiction. Since the reform is ex ante preferred, then, by Result 9.9 it is ex post preferred too and since it is ex ante feasible all elements in O(r,B) are feasible and therefore it is ex post feasible as well. (ii) If the reform is optimal then $y \preceq_r x\ \forall y \in S$, but then by Lemma 9.2 $y \preceq_x x\ \forall y \in S$. Hence x is optimal ex post.

This result shows that LA and NS imply that if a bundle is optimal on a set when viewed from a reference point r, then that bundle is also optimal when viewed from itself. Moreover, if the optimum bundle when viewed from r is not unique (perhaps because the budget set is not convex), but all the optimum bundles are feasible, then all the bundles which are optimal ex post are feasible too. Hence, for the individual, the ex ante approach to evaluating policy changes is in accordance with the ex post assessment.

We can also say something about welfare in a multi agent economy. Identify individual bundles and reference points by h, $h = 1,\ldots, H$, let X $(= (x_1,\ldots, x_H))$,

be a matrix representing an allocation (with Y, R etc similarly defined for other allocations and reference points).

Definition 9.1 An allocation X is *ex ante Pareto efficient* viewed from R when there exists no feasible allocation Y such that (i) $x^h \preceq_{r^h} y^h \ \ \forall h = 1,\ldots, H$ and (ii) $\exists h$ $x^h \prec_{r^h} y^h$. An allocation X is *ex post Pareto efficient* when there exists no feasible allocation Y such that (i) $x^h \preceq_{x^h} y^h \ \ \forall h = 1,\ldots, H$ and (ii) $\exists h$ $x^h \prec_{x^h} y^h$.

Result 9.11 *Given LA and NS, if X is ex ante Pareto efficient, it is ex post Pareto efficient.*

Proof Suppose false and let Y be an allocation ex post Pareto preferred to X and take an individual h such that $x^h \prec_{x^h} y^h$. By Lemma 9.1, this implies that $x^h \prec_{r^h} y^h$ $\forall r^h$, while $x^h \preceq_{x^h} y^h \ \ \forall h = 1,\ldots, H$ implies $x^h \preceq_{r^h} y^h \ \ \forall h = 1,\ldots, H$. Thus X is not ex ante Pareto efficient – a contradiction.

Note that as well as establishing that ex ante Pareto efficient implies ex post Pareto efficient, this proposition also provides an alternative basis for the proof of the existence of general equilibrium with reference dependent preferences, using the Fundamental Theorems of Welfare economics.

9.4.4 Implications of the Theory for Tax Reform Models

The propositions put forward so far are general, applying to a variety of situations in which welfare comparisons are made. Tax reform is a useful canvas for this illustration because it provides a context in which the status quo point is clearly a crucial part of the model, unlike many models of optimisation.

Let V(...) be the indirect utility function for an agent (which may be a function of the reference point, just like direct utility), let R be tax revenue and let t be the vector of commodity taxes. Consider a tax reform, dt (where dt is small). In standard theory, with linear production technology, and producer prices given by the vector p, such a reform is desirable in a single consumer economy if it is both welfare increasing and revenue enhancing, (see Ahmed and Stern, 1991, or Guesnerie, 1995, for summaries), that is:

$$dV_t dt > 0; \qquad (9.8)$$
$$dR_t dt > 0$$

The main justification for investigating marginal changes rather than the underlying optimisation problem is the idea that limited information precludes the kinds of calculations on which maximization is based. With Hicksian preferences, tax reform requires knowledge only of current demand ($x(p + t)$) and price elasticities local to $x(p + t)$ whereas optimization requires global knowledge of the functional form of V and x. Let $x(r, p + t)$ be a function relating the optimal bundle to the consumer price vector and the reference point and call this bundle x^*.

Corollary 9.1 *In a single consumer economy, if a tax reform is ex ante desirable, then it is ex post desirable.*

Proof Follows immediately from Result 9.10.

In summary, with the formulation of reference dependent preferences given by LA and NS, the local information needed to predict the ex-ante desirability of tax reforms is sufficient to predict the ex-post desirability of the reforms. This surprising aspect of the result is feasibility which arises because x^* is not just optimal when viewed from r, but also when viewed from x^*. In turn this arises because, with LA and NS, if bundle x^* is preferred to x from r, then it is preferred from all reference points. This means that the indifference curve through x^*, as viewed from x^*, is enveloped by the indifference curve through x^* from any other vantage point.

Perhaps the most famous result in tax theory concerns the superiority of a lump-sum tax over commodity taxation which raises the same revenue. This result arises because the bundle chosen under commodity taxation is also affordable, but not necessarily chosen, under lump-sum taxation. This result, too, is preserved under LA and NS. Clearly it is true for any individual reference point since the revealed preference argument can be made, but it is also true that the consumer ex post prefers the ex post outcome, and also prefers the ex post budget set. To see this, extend the tax model slightly by introducing T, a lump tax and write $x(r, p + t, T)$ as the consumer's demand function, with $B(p+t, T)$ as the corresponding budget set. Let $x^* = x(r, p + t, 0)$ and $x^{**} = x(r, p, T)$ where $T = tx^*$, then we have,

Result 9.12 *(i)* $x^{**} = x(x^{**}, p, T)$*; (ii)* $\forall y \in B(p + t, 0)$, $x^{**} \preceq_{x^{**}} y \rightarrow y = x^{**}$.

Proof (i) follows immediately from Lemma 9.3. (ii) By the standard arguments which establish the superiority of lump-sum taxation, $y \preceq_r x^{**}\ \forall y \in B(p + t, 0)$. Suppose that there did exist $z \in B(p + t, 0)$ such that $x^{**} \preceq_{x^{**}} z$, but then, from Lemma 9.2 we have $x^{**} \preceq_r z$ – a contradiction unless the claim holds.

Thus, ex post, the consumer facing the lump-sum tax of T would not prefer to swap it for the commodity taxes t, even though it is possible, when viewed from x^{**} that such commodity taxes would raise an amount of revenue less than T. Note also that acyclicity is not sufficiently strong to establish this result, but a slightly weaker proposition can be obtained. Define $x^{***} = x(x^*, p, T)$, with T defined as above, then acyclicity implies that the consumer does not prefer x^* to x^{***} when x^{***} is the reference point.

9.4.5 Optimal Taxation

It is worthwhile considering results which do not automatically extend to the case of reference dependent preferences. All the results derived so far rely on the argument that a bundle which is optimal on budget set B, when viewed from r, is also optimal on B when viewed from itself. However, when optimisation is over budget sets, generally results are not extendable. To see this examine Fig. 9.3. PP is the production

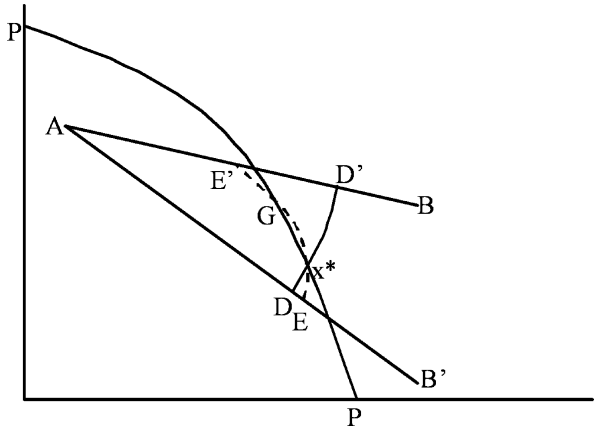


Fig. 9.3 Optimal Taxation and Reference Dependent Preferences

possibility frontier and the two lines AB and AB′ define a cone of possible prices. The price expansion path DD′ shows the locus of optimal choices, as viewed from the ex ante reference point r (not shown). Clearly Dx^* is the feasible section and therefore, given the reference point r, the optimal price reform leads to consumption bundle x^*. The broken line EE' shows the locus of optimal choices as viewed from x^*. In this case the set of feasible price reforms is not connected, but consists of the sections Ex^* and GE′. As viewed from x^*, the optimal price reform would lead to E′ as the outcome and not x^*.

Two points about this example are worth noting. First, observe that in Fig. 9.3, it is the Giffen-good nature of good 1 which prompts the problem. Thus, for suitable restrictions on reference-dependent preferences the accordance between ex ante and ex post viewpoints can be re-established. Alternatively, the problem in this particular example arises because some of the price reforms are not feasible when viewed from r, but are when viewed from x^*.

This last remark gives us two possibilities. First, there may be circumstances where the set of feasible price reforms as viewed from r and x^* coincide. Secondly, we might relax the assumption of a fixed producer price vector, p, and pursue the concept of tax equilibria as in Fuchs and Guesnerie (1983).

Let us take the first option initially. If x^* is interior to S then at least locally there can be no improvement available ex post. If there were, then any new preferred outcome would also be preferred to x^* from the viewpoint of r, which would negate the definition of x^*'s optimality.

Alternatively, the set of all points preferred to x^*, and representing optima (for a given budget set) as viewed from r might lie outside the box defined by r and x^*. If this is the case, then the set coincides with the equivalent set as viewed from x^*. For instance, in Fig. 9.3, if r is the origin (say), then x^*D' would also be a part of the price expansion path as viewed from x^*.

The alternative way of taking the model forward would be to close the model in a different manner by considering the endogeneity of prices. The model so far treats taxes as the instrument with producer prices as fixed. Consumer prices are then

equal to producer prices plus the tax rate and supply is implicitly determined as the minimum of what is feasible and what is demanded. Consider now an economy with a strictly convex technology, price taking behaviour by all agents, and 100% profits tax. Now when the government sets tax rates consumer and producer prices adjust so as equate demand and supply. For the case of reference independent preferences, Fuchs and Guesnerie (1983) (see also Guesnerie, 1995) show that with additional, technical conditions, a tax equilibrium exists in such an economy. If the government has a fixed revenue requirement, then we can define a set of equal-yield tax equilibria, all members of which produce the necessary net revenue. For suitable smoothness conditions, this set is a smooth manifold and if it is non-empty then we can characterise optimal taxes and their associated tax equilibrium. Suppose we consider reference-dependent preferences of the type defined above. Under the same conditions as those set out in Guesnerie (1995), for any given reference point, r, there exists a set of reference-dependent equal-yield tax equilibria (RDEYTE), a subset of which represents the reference-dependent tax optimum (RDTO). Let x' be the consumer's consumption vector at a member of the optimum set. Is it also the case that x' is necessarily in the optimum set as viewed from x'? Unfortunately we cannot say that. We can say that x' is a reference dependent tax equilibrium as viewed from x' for an unchanged tax vector. Consequently revenue is also unchanged, but as viewed from x', the set of RDEYTE is not necessarily the same as that when r is the reference point. As a result, the reference dependent tax optimum may also be different.

There are three positive comments worth making in this context. First, the arguments of Result 9.10 mean that there cannot be cycles in the reference dependent tax optimum, in the sense that if x_1 is consumption at a RDTO, then there is no sequence of budgets B_i, each member of which represents a RDTO when the reference point is x_{i-1} and where for some i, $x_i = x_1$. It could be the case though that there are cycles in the tax rates because, depending on the pattern of reference dependent preferences, there may be many reference points for which a particular pattern of taxes is optimal. Second, provided that the RDTO is unique and continuous in the reference point, we can use a simple fixed point argument to establish that there must exist some allocation which is an RDTO when viewed from itself. Finally, we could adapt the arguments of Chapter 4 to show that the sequence of allocations associated with an RDTO must converge to set of allocations which have this reflexive property.

9.4.6 Final Thoughts

We have seen that, within one theory of reference dependent preferences which accords with current evidence, many of the problems posed for welfare theory by the endowment effect can be dealt with. In particular, an ex-ante assessment of the desirability of a reform will not be contradicted ex-post. However, it need not be the case that all theories compatible with the evidence produce the same results. More generally, we might expect that theories of reference dependent preferences should predict acyclicity, on the grounds that cycles of choice are incompatible with basic

notions of economic rationality. If this is the case then ex-ante preference will lead to ex-post preference. In this sense, the 'preference' aspect of desirability is likely to be more robust than the 'feasibility' aspect, for which no such intuition is apparent.

9.5 Conclusions

In this chapter I have steered a middle course between the Hicksian model in which a full ranking of tax policies can be made and models in which preferences are so frame dependent that nothing can be said about the impact of fiscal policies. We have seen that even when one makes only weak claims for the welfare relationship along the lines of Ex Ante Consent, it is possible to say something about optimal tax design. But of course, when preferences are non-Hicksian the set of feasible policy instruments expands to include all sorts of framing effects. It is this issue we tackle in the next chapter.

Appendix 1

I begin by repeating the relevant equation. Either the middle term or the right-hand term will be the basis for all that follows:

$$\frac{dW}{d\lambda_k} = \sum W_i\theta_i x_{ik} + \sum W_f f_l y_{lk} = \alpha^* \sum \left(W_i\theta_i - \frac{\alpha W_f}{\alpha^* U_f} \right) x_{ik} - \frac{\alpha W_f}{\alpha^* U_f} g_k y_k \tag{9.9}$$

Proof (Result 9.2) Consider the right hand term. Suppose (x, y) are such that $\pi_i = W_f \alpha g_i / U_f \alpha^*$, and $f_l/f_r = g_l\lambda_l/g_r\lambda_r$ l,r = 1,..., m, then clearly the result holds. We need to check that this is possible with the instruments available. Given strict convexity of preferences and monotonicity there exists a hyperplane which supports (x,y). Let that hyperplane define a consumer price vector, p. Given the hyperplane and strict convexity, (x,y) is uniquely optimal from the viewpoint of the consumer. Given strict convexity of production possibilities, there also exists a unique vector of producer prices (the g_is and $g_l\lambda_l$s) which supports that production. The difference between consumer prices and producer prices defines a tax vector, (t^x,t^y,t^f), where $t_i^x = p_i - g_i$ and $t_l^y = p_l - (1 + t^f)g_l\lambda_l$. (Note that T is then chosen to balance the budget.) We then have,

$$\frac{f_l U_f}{f_r U_f} = \frac{p_l}{p_r} = \frac{(1 + t_f)g_l\lambda_l + t_l^y}{(1 + t_f)g_l\lambda_l + t_r^y}$$

But then $f_l/f_r = g_l\lambda_l/g_r\lambda_r$ so $t_l^y = 0$, $l = 1, \ldots, m$ and the result follows.

Proof (Result 9.3) Suppose $x_{il} = 0$, all i, then technical efficiency is welfare desirable

Straightforward given the middle term in the expression for $dW/d\lambda_k$.

Proof (Result 9.4) If welfare has the form, $U(\theta^h(x), f(y))$, for some $\theta > 0$, then the equation for dW/dλ_k becomes,

$$\frac{dW}{d\lambda_k} = \alpha \left[\frac{W_l \theta}{U_l} \sum_i g_i x_{ik} - \frac{W_f(1+t_f)}{U_f} \sum_l g_l \lambda_l y_{lk} \right] \tag{9.10}$$

Now, if $1 + t_f = \frac{W_l \theta U_f}{U_l W_f}$ then the result follows.

Proof (Result 9.5) Suppose a. utility has the form $U = U(h(x), f(y))$, increasing in both h and f and b. welfare has the form U(θ^h(x),f(y)), for some $\theta > 0$. If following a decrease in technical efficiency, (a). the value of the change in x is negative and (b). the value of the change in y is negative when measured at market prices, then technical efficiency is welfare desirable.

This result follows from the expression for dW/dλ_k used to produce the previous result. If we consider the right hand terms in that expression, the terms summed represents the market value of changes in x and the market value of changes in y for a fall in technical efficiency. If both of these are negative the $dW/d\lambda_k$ as a whole is negative.

Appendix 2

The basic structure for this appendix draws on the work of Guesnerie (1995), Chapter 4. He defines the notion of a tax equilibrium, which for a single agent economy is a vector of consumer prices, a vector of production prices, a vector of public good production such that (i) the public good production vector is feasible and for the non-public goods, (ii) demand is utility maximizing, (iii) supply is profit-maximizing and (iv) excess demand is zero for all goods. The principle difference is that here is that we have no public goods and more importantly we drop feature (ii) and replace it with the assumption that (x,y) represents the choice at the relevant prices. More specifically, let $B = \{x : x \geq 0; y \geq 0; p^x x + (1 + t^f)p^y y - T \leq 0\}$ and use the notation $C(p, t^f, T)$ to mean $C(B)$.

I assume the following properties for the choice function:

1. $C(p, t^f, T) = \phi \rightarrow B = \phi$. Otherwise $C(p, t^f, T)$ is single-valued.
2. C is twice differentiable in (p, t^f, T).

Result 9.2A. Suppose that (p, t^f, T) is a tax equilibrium with all elements of x and y strictly positive. Then some rise in production efficiency is welfare desirable.

Proof Suppose an infinitesimal change in λ_k of $d\lambda_k < 0$. In order to show welfare desirability we need only find some resulting tax equilibrium such that both dx and dy are positive, with some elements strictly positive. Consider some accompanying changes in consumer parameters, (dp, dt^f, dT). Since C satisfies the assumptions of the implicit function theorem, there is a (dp, dt^f, dT) such that $(dx, dy) > 0$. Now, feasible changes in producer prices and output must only satisfy the condition,

$0 = w^x dx + w^y dy + w^k y^k d\lambda_k$. Since $d\lambda_k < 0$ then clearly $(dx, dy) > 0$ is feasible and since g satisfies the assumptions required for the implicit function theorem, then there exists a dw which supports such a change of production.

Note that since we have assumed continuity of C and a strictly convex production set with a twice differentiable surface, the existence of a tax equilibrium in this context is not a problem. On the other hand, there is nothing that guarantees that x and y are both strictly positive in any tax equilibrium.

Chapter 10
Framing Matters: Non-Standard Fiscal Policy

10.1 Introduction

In the last few chapters we have been considering the manipulation of the price and income terms in the generalised demand function *x(p,m,f)*. In this chapter the focus is on non-standard fiscal policy, meaning the manipulation of the f term in order to achieve government goals. For instance, Cullis et al. (2000), suggest that because individuals who pay for an item are more likely to consume it, because prescription charges may help individuals complete a course of treatment. The chapter is devoted to exploring the potential uses of this kind of non-standard policy in more detail. Some illustrations come from the theory of 'libertarian paternalism' proposed by Camerer et al. (2003). Much of the focus is on savings and pensions policies. For that reason I also include discussion of some relevant intertemporal decision-making anomalies.

No presumption is made about the goals of government: they might be to maximize welfare or to pursue some more limited objective (e.g. 'raise savings') which potentially has a welfarist justification, or they might arise from the aims of rent seeking groups etc. There is also no assumption that these policies can be used to manipulate individual behaviour at will. Policies based on framing must have their limits, just as standard fiscal policies do.

The sorts of instruments which might be tried are shown in Table 10.1. In the first column we have the class of anomaly or heuristic; in the second we have the policy lever and then the final column provides an example. So, for instance, given the tendency for individuals to stick with the status quo, setting the default as opt-in for pensions may lead to much higher savings levels compared to the case where individuals have to choose to opt in. As Camerer et al. (2003), note such policies have the merit that they do not penalise the rational agent, since typically the actual cost of switching is low. On the other hand they produce potentially large benefits for the individual who has difficulty making the commitment to save or become insured. Meanwhile, in his discussion of mental accounting, Thaler (1981), points out that individuals are much more sensitive to out-of-pocket expenses than they are to the opportunity costs. As a result, individuals may overuse cars relative to public transport, because the depreciation cost of using the former is not salient.

A. Munro, *Bounded Rationality and Public Policy*, The Economics of Non-Market Goods and Resources 12, DOI 10.1007/978-1-4020-9473-6_10,

© Springer Science+Business Media B.V. 2009

Table 10.1 Non-Hicksian Policies

Anomaly	Policy	Example
Reference Points and Anchoring	Manipulating the endowment point	Use of defaults in insurance and pension legislation.
	Using the choice set to manipulate reference points.	Providing dominated options
Mental Accounting	Offering commitment devices to reduce temptation	Penalties for early withdrawal from savings accounts
	Labelling to create accounts	Hypothecated taxes and expenditures.
	Using the non-equivalence of out-of-pocket and opportunity costs.	Educational attendance subsidies; information on depreciation costs of car use
	Exploiting the non-sunk nature of costs	Prescription charges to encourage consumption of full course of treatment.
Information Biases	Exploiting the availability and representativeness heuristics	'Tokenism' in visual advertising, restrictions on celebrity advertising

Conversely, individuals may undervalue education, if the opportunity costs of not attending college are hard to appreciate.

To explore further aspects of Table 10.1, there are four major elements of this chapter: first a discussion of labelling policies uses the UK's Winter Fuel Allowance as a motivating example; second we examine inconsistency in intertemporal decision-making which is then discussed further in the context of savings policy. The final substantive section is more speculative and focuses on framing effects in tax policy.

10.2 Creating Mental Accounts

As intimated, mental accounting is a potential motive for a variety of policies that when viewed through the lens of standard theory defy easy explanation. One of these is the UK's Winter Fuel Allowance, about which some background is in order.

Along with Ireland, the UK has the highest percentage of excess deaths during the winter months in Europe. Mortality rates regularly rise during periods of cold weather and deaths are concentrated in the over 65s. The exact link between temperature and excess death rates for the UK are unclear. Hypothermia accounts for a relatively small number of the deaths, most of which are attributed to respiratory or circulatory diseases (Aylin et al., 2001). Hypothesised and investigated causes of the deaths include poor quality housing stock, lack of central heating, pensioner behaviour and activity rates as well as poverty.

The Winter Fuel payment has been presented as a means of reducing excess mortality by allowing pensioners to spend more on their heating bills. Introduced as a token payment in 1997 and extended significantly four years later, the allowance is currently a lump-sum payment of £200 payable annually to those households claiming it, with an extra £100 for households where at least one person is aged 80 or over. Households who are not already in receipt of state benefits have to claim, but there is no income contingency and no requirement that the money be spent on fuel bills. Once a claim is paid in one year, the payment is automatically made in subsequent years unless the Department for Work and Pensions (DWP) is alerted to a material change in circumstances. Originally the allowance was available to men over 65 and women over 60 living in the UK, but following a ruling by the European Court of Human Rights, all individuals over 60 can claim. The UK government does not produce figures on the numbers potentially eligible. However, currently the total number of households receiving benefit is over 8.5 million.[1]

10.2.1 Theory

Within standard Hicksian theory the allowance is equivalent to a lump sum grant contingent solely on age (and the fact that the payment is claimed). As a result, the impact on heating should be minimal. As a guide, note that (own) estimates derived from the Family Expenditure Survey for 2001/2 suggest that the income elasticity of fuel expenditure is rather small at around 0.28. For an average two-adult retired household in the second lowest decile of the income distribution this translates into £2.78 per year from a £200 rise in income or about five pence per week. For households in the top fifth of the income distribution, the rise is only £1.50 per year.

More fundamentally the existence of the Minimum Income Guarantee (now the Pension credit) meant that when the Winter Fuel Allowance was introduced there was pre-existing vehicle for giving pensioners a lump-sum rise in income. The use of this vehicle would have avoided the significant costs of running the Winter Fuel programme. In other words, as instrument of policy, the allowance was dominated by adjustments to pre-existing programmes.

If the programme is hard to explain on welfarist or fuel poverty reduction grounds, it is equally difficult to produce a sensible political economy explanation that does not at some point invoke bounded rationality. The three main groups of actors relevant to the programme are pensioners, non-pensioners and elected politicians. Suppose there was a political outcome in which Δ was paid in cash to all pensioners at a net cost of $\Delta + c$ to non-pensioners. Compare this to an out-

[1] From 2003–04, the Food and Expenditure Survey (FES) – a nationally representative survey of UK households, asks questions about receipt of the benefit. From that it appears that above the age of 62, over 95% of eligible households are in receipt. Slightly lower uptake levels are recorded for households where the oldest person is 61. See Munro (2005).

come in which pensioners receive Δ at a cost $\Delta + c\prime$ where $c\prime < c$. Non-pensioners would prefer this second outcome provided their preferences satisfy non-satiation and provided the burden for each individual was lower under the second arrangement. Meanwhile pensioners would be indifferent or prefer the second arrangement if they were at all altruistic. Given this, politicians would also prefer the second arrangement, provided that their payoff was increasing in the utility of citizens.

One possible though not necessarily plausible explanation is based on asymmetric information along the lines of Coate and Morris (1995). For instance, non-pensioner citizens might be against a lump-sum subsidy to pensioners but in favour of a targeted heating subsidy.[2] If they are not able to distinguish between a heating subsidy proper and the Winter Fuel payment then there may exist a political equilibrium in which the Winter Fuel payment is made. One problem with this argument is the high level of publicity given to the programme: the government advertises it heavily, individual Labour party members of parliament produce localised figures on its uptake in their election material and newspapers carry reminders to pensioners to claim. More fundamentally, in a world where some actors can choose whether information is asymmetric or not, asymmetric information cannot be a part of any equilibrium if that is not compatible with incentives. In other words, if non-pensioners were unsure of the true nature of the payment, then there would be clear gains from 'political arbitrage' in the sense that there would be returns to an opposition politician or indeed newspapers from passing on the information. Yet the Winter Fuel payment remains popular, largely unchallenged in the U.K.

An alternative theory is that there is no citizens are actually bounded in their rationality, but that many voters believe that others are so. This is entirely possible, but then we are into the world of bounded rationality.

The most sensible explanation from within a bounded rationality framework is that the Winter Fuel payment creates a mental account out of which the income elasticity of expenditure is higher than in the standard model. The discussion of mental accounting in Chapter 2 ended with the observation that there was as yet no good formal model of it, although prospect theory could go some way towards explaining many of the features of mental accounts. So I begin there. We shall see though that there are problems with using the riskless version of prospect theory.[3]

Let us assume the following.

1. Individuals assess the utility of different options using Kahneman and Tversky style valuation functions. I shall take it that all valuation functions are twice differentiable, except at 0.
2. In particular w(x) is a concave function representing the utility gain from x (monetary) units of heating expenditure. Gains in income are valued by the

[2] In the final section of this chapter I argue that labels on taxes and subsidies may affect behaviour. One might therefore reasonably argue that a targeted heating subsidy can have a mental accounting effect along with the usual incentives to consume provided by a lower consumption price.

[3] I am grateful to Henrik Kleven for alerting me to this point.

function u(x) while v(x) is used to value losses in income or expenditures. These functions satisfy the usual properties of valuation functions in prospect theory. In particular we take $u(0) = v(0) = 0$ for all x, and $u(x) < v(x)$ and $u'(x) < v'(x)$, where $'$ indicates a first derivative.

3. Where cash grants, y, are not named or associated with heating, then heating expenditures and cash grants are not integrated into one mental account. As a result x is chosen to maximize: $w(x) + v(-x) + u(y)$.
4. Where cash grants are named or otherwise associated with heating, then heating expenditures and cash grants are integrated into a single mental account. As a result, x is chosen to maximize, V where:

$$V = w(x) + v(-x + y) \qquad \text{for } x > y$$
$$V = w(x) + u(y - x) \qquad \text{for } y > x.$$

The basic model is illustrated in Fig. 10.1 for the specific case where $y > x$. It can be seen that u(y)+v(−x) < u(y−x).

Where expenditure and grant are not integrated, the level of x is set by the condition.

$$w'(x) = v'(x) \tag{10.1}$$

Note that this is independent of y which is inaccurate but not greatly so given the evidence on the income elasticity of heating and given the scale of the fuel allowance. When expenditure and grant are integrated, then the relevant condition is either,

$$w'(x) = v'(-x + y) \tag{10.2}$$

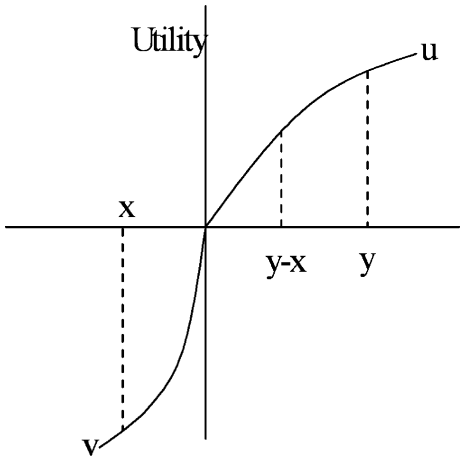


Fig. 10.1 Prospect Theory and the Winter Fuel Payment

or

$$w'(x) = u'(y - x) \tag{10.3}$$

The first case applies when heating expenditure exceeds the grant; the second case is relevant when the grant is greater than heating expenditure. Totally differentiating the first of these expression with respect to y yields,

$$\frac{dx}{dy} = \frac{v''}{w'' + v''} \tag{10.4}$$

Differentiating the second expression produces.

$$\frac{dx}{dy} = \frac{u''}{w'' + u''} \tag{10.5}$$

The crucial issue here is the shape of the valuation functions. In the standard version of prospect theory, the second derivative of u is negative and we have assumed a similar property for w. As a result, when the grant is higher than heating expenditure a rise in its level increases heating. This is what we would anticipate. On the other hand, with the standard formulation of prospect theory, (as in Fig. 10.1) v is convex. Thus the sign of 10.4 is ambiguous, but if in fact the first order conditions represent a maximum then the denominator of 10.4 is negative, implying that, with v'' positive, a rise in the heating allowance lowers expenditure on heating.

One way to 'fix' this problem is to drop the assumption that v is convex – a feature of the valuation function that makes some sense in a risky context, but perhaps less so in the absence of risk (see Chapter 2 or Munro and Sugden, 2003, for more detail). On the other hand, mental accounting can also be integrated into other theories and this might represent the more straightforward option. For instance, one simple format[4] is to write the utility function as $u = u(x_1 - \lambda_1 m_1/p_1, ..., x_n - \lambda_n m_n/p_n)$ where $m_i \geq 0$, $i = 1, ..., n$ represent labeled mental accounts, $p_i \geq 0$, $i = 1, ..., n$ are prices and the $\lambda_i \geq 0$, $i = 1, ..., n$ are coefficients representing the '*stickiness*' or effectiveness of the label.[5] In some cases $m_i = 0$ or $\lambda_i = 0$. A mental account of £1 for good i, is therefore akin to £λ_i of commited expenditure. Figure 10.2 illustrates the effect of the introduction of a mental account of m_1 in the context of a two good model with prices set equal to 1. E_0 is the consumer's choice in the absence of the account. Setting the account at m_1 means the consumer treats $\lambda_1 m_1$

[4] A more general format is to suppose that a vector of mental accounts, m, represents a reference point, not in goods but in expenditure. A rise in any element of m, then favours bundles with higher levels of the corresponding consumption good. Imposing continuity of preference in m and homogeneity in prices and m together as further reasonable assumption leads to a generalisation of consumer theory (Munro 2005).

[5] This simple formulation therefore implies that demand is homogeneous of degree zero in mental accounts and prices taken together.

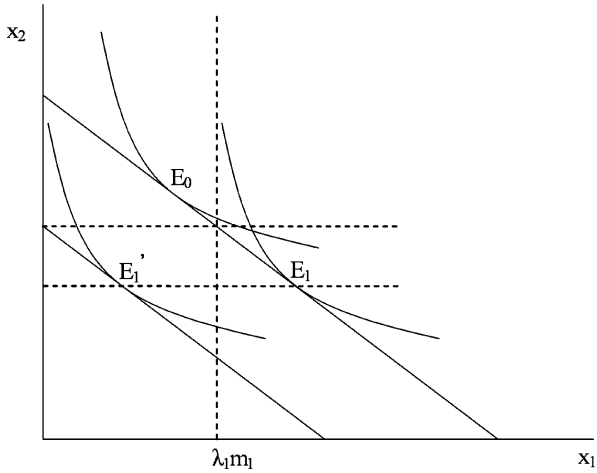


Fig. 10.2 The Effect of a Mental Account on Choice

as a committed expenditure that makes no contribution to utility. Thus consumption of x_2 at the new equilibrium, E_1 is equal to that which would obtain if there was no mental account but the individual had $\lambda_1 m_1$ less income (i.e. at E_1'). Meanwhile consumption of x_1 is equal to $\lambda_1 m_1$ plus the value of x_1 at E_1'. Letting $m = (m_1, ..., m_n)$ we can therefore write the demand function as,

$$x_i(p, M, m) = \frac{\lambda_i m_i}{p_i} + x_i^*\left(p, M - \sum \frac{\lambda_i m_i}{p_i}\right)$$

where x_i^* is a standard Marshallian demand function that depends only on prices, p, and an adjusted measure of income.[6]

The crucial feature which makes the policy more effective (in the sense of increasing the stickiness coefficient, λ_i) is the effective creation of the mental account. With the Winter Fuel payment, the name provides one stimulus to its creation. In addition, pensioners have to apply for the payment which may create some sense of obligation to spend the money received on heating. Finally it is heavily advertised and paid at a time of year (December) when fuel consumption is rising. As we saw with Prelec and Loewenstein (1998), results, timing the payment to match the period of highest consumption during the year, may lower the psychic costs of expenditure. More directly, Abeler and Marklein (2008), conduct a series of interesting experiments which illustrate the powerful effect of labels. In one field study, customers at a restaurant were randomly handed vouchers of equal face value as they entered the premises. For half the participants the voucher was for cash, while for the rest the voucher could be set against beverages purchased during the meal. The beverage voucher recipients spent on average 3.45 Euros more on beverages

[6] The formulation is reminiscent of the Stone-Geary or linear expenditure system or, more generally, the committed expenditures of the consumer demand model in Lewbel (1985).

than those receiving the cash voucher.[7] Meanwhile a study of US online grocery shopping (Milkman et al. 2007), found that individuals receiving a \$10 voucher spent around \$1.5 more on shopping than those who did not, which again is a much higher propensity than would be predicted given standard theories about the relationship between lifetime income and food expenditure.

Given the value of λ we can illustrate the effect of labelling a transfer through Fig. 10.3. A rise in income of Δm pushes out the budget constraint and increases heating expenditure from x_0 to x_1. If the same increase in income is labelled as heating expenditure and $\lambda_1 = 1$, then heating rises to x_2. For $0 < \lambda_1 < 1$, heating expenditure will lie between x_1 and x_2.

More formally, the responsiveness of expenditure to a marginal rise in the mental account is given by,

$$\frac{\lambda_i}{p_i}\left(1 - \frac{\partial x_i^*}{\partial M}\right) + \frac{\partial x_i^*}{\partial M}$$

So, as long as all other goods taken together are not inferior, an increase in income into a labelled account will therefore have a bigger increase in expenditure on the nominated good, compared to an unlabelled increase in income of the same size.

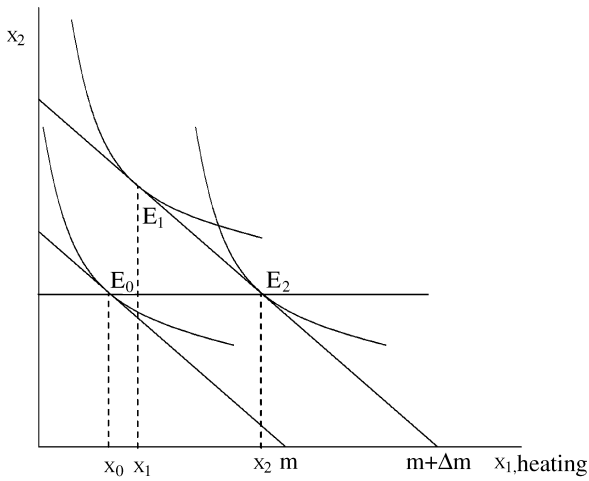


Fig. 10.3 The Effect of Labelling on Heating

[7] In other laboratory studies subjects first earned money then asked to allocate their income between two goods. Final payoffs were determined by an induced utility function of the two goods. For one treatment the earned money was given as a single budget; in the other treatment part of the total was given in the form of an intra-marginal in-kind subsidy. In other words the parameters of the experiment were set so that the individual who maximized utility treating all income as fungible, should not be affected by the difference in treatments. In fact, purchases of the labelled good were higher in the in-kind subsidy and departures from the standard solution were negatively correlated with mathematical ability. Abeler and Marklein speculate that mathematical ability may enable individuals to see more easily through the veil of the labelled budget.

Winter Fuel Payments are not the only labelled accounts. In the UK the Budget for 2003 introduced the Government's proposals for its Child Trust Fund. Children born after September 2002 receive an endowment at birth, means tested on parental income, but with a minimum value of 250 pounds and a maximum of 500 pounds. A top-up at age seven of £250 is also provided by the government. Until their 18th birthday, the child's grant is placed in a tax-free trust, but parents and others may add to it at a maximum rate of £1,200 per year. The impact and duration of the scheme remains to be seen and uptake has been patchy, but at least in standard theory, its effect is likely to be minimal: many families have at least some savings nominally put aside for matters such as college education, marriage expenses etc.

The impact of the Winter Fuel payment and Child Trust funds has not yet been clearly assessed. On the other hand, Child Benefit has been around a lot longer and also has large elements of the labelled mental account, particularly since in many countries where it is paid (e.g. the UK, Netherlands) there are no constraints on how the benefit should be used. Kooreman (2000), using data for the Netherlands concludes that the marginal propensity to spend on child clothing out of child benefit is 10 times that out of ordinary income. However, Edmonds (2002) (for Slovenia) and Blow et al. (2006), (for the UK) find no distinctive impact of child benefit on expenditure. Indeed the latter find only adult alcohol consumption to be distinctively sensitive to child benefit changes.[8] What makes the U.K. so different from the Netherlands is not clear, but it might lie in the way the benefit system labels and maintains the child benefit 'mental account'. In the UK payments are monthly while in the Netherlands, Child Benefit is paid quarterly making it larger and possibly more salient in the minds of parents.

In these examples accounts are created by naming and advertising as such, but other fiscal policies may create a mental account as a byproduct, with the nature of it sensitive to the framing of the policy. For instance, in the Hicksian model a subsidy for saving can also be implemented as a tax on current consumption accompanied by a lump-sum rebate. These two policies could have quite different effects on the creation of mental accounts and therefore on behaviour.

Promoting heating use by UK pensioners might be viewed as a sensible merit-wants style policy, but in theory governments can use mental accounts for other purposes. For instance in the theory of Leviathan, the government sets its tax policy so as maximize the feasible size of government. In doing so it faces the constraint of obtaining the consent of the electorate. Having many small taxes may help obscure the size of the government's income, but in addition, hypothecation may reduce the perceived deadweight loss from taxation, if citizens integrate the tax payment with the benefits from the expenditure side.

Overall, therefore, creating mental accounts by labels and by the appropriate timing of payments may be a means to raise the income effects associated with

[8] Lyssiotou (2005), uses a demand system rather than individual equations and, using the same data as Blow et al. (2005) does find a significant labelling effect for the UK – suggesting that differences in econometric methods might lie behind some of the disagreement.

some forms of government expenditure. However, as the empirical example of Child Benefit in the UK suggests, there is a need for some scepticism about the universal effectiveness of this kind of policy.

10.3 Intertemporal Issues: Pensions and Savings Policy

There are two different aspects to bounded rationality in the context of decisions about savings and pensions. The first concerns the intra-temporal allocation of invested funds; the second and the most prominent concerns the inter-temporal distribution of consumption. Rapid changes in our understanding of inter and intra-temporal choices have been fuelled by the emergence of behavioural finance as an alternative to the rational choice model of financial behaviour (Shleifer 2000). In part the growth of the behavioural approach has been prompted by some of the anomalies discussed in Chapter 2 and their persistence as anomalies despite many attempts to bring them within the scope of the rational choice model. Shefrin (1999), lists examples of regularities which seem to defy explanation by the rational choice model, including the equity-premium puzzle (whereby, over long periods of time the return on stocks exceeds the returns on bonds by an amount which is not consistent with standard values for individual risk aversion), and the tendency to hold on to losing stocks too long and sell rising stocks too quickly (Odean 1998).

The second aspect of bounded rationality concerns the inter-temporal allocation of consumption, where there is some evidence that consumers in many countries (particularly the Anglo-Saxon economies) undersave, relative to some desirable standard (see Bernheim, 1993 for the USA and Börsch-Supan and Brugiavini, 2001, for summary of the more heterogeneous European position). As a result many households face sharp drops in their standard of living upon retirement. For instance, Banks et al. (1998), find a significant fall in the consumption of non-durables by UK households around retirement age. After controlling for some possible explanations, including mortality risk labour market participation and job-related expenses, two thirds of the drop is explained. For the remainder they conclude that they are only the sudden arrival of a large (and negative) shock to wealth can reconcile the observed behaviour with the life-cycle model of consumption. Similar patterns have been found for Canada by Robb and Burbidge (1989), and for the USA by Bernheim et al. (2001) where consumption declines of up to 30% are found amongst the households with the lowest wealth.

Now, most individuals who save have multiple opportunities to invest and to observe the consequences of their previous choices. So if there is any instance where learning might supplant the need for government control, then perhaps intertemporal choice is it. Nevertheless a consensus that many individuals undersave suggests that the forces which create the apparent undersaving are quite powerful and distinct from some of the anomalies discussed earlier in the book. In the next sub-section I discuss some of the evidence on intertemporal preferences, together with theories which attempt to explain the evidence or to build models of investor behaviour on the back of it. Thereafter we return to the discussion of policy instruments.

10.3.1 Intertemporal Choice and Self-Control

The two prominent problems facing the rational choice model in the context of intertemporal choice concern, firstly, giving the wrong weight to future selves when planning consumption (i.e. a failure of strong rationality) and secondly, the issue of dynamic inconsistency (a failure of weak rationality), meaning that the optimal plans made in one period may not be optimal when viewed by the selves of other periods. To a significant degree, behavioural economics has joined these two concerns by picturing the typical consumer as someone who overweights the present because of intertemporally inconsistent preferences.

Suppose preferences are additively separable between time periods and can be represented by a utility function of the form,

$$U_t = \sum_{\tau=0}^{\tau=T} \delta_{t+\tau} u(x_{t+\tau}) \tag{10.6}$$

where $x_{t+\tau}$ is the consumption vector in period $t + \tau$ and $\delta_{t+\tau}$ is the discount factor attached to consumption in that period. Meanwhile suppose that individual welfare can be written as,

$$W_t = \sum_{\tau=0}^{\tau=T} \delta^*_{t+\tau} u(x_{t+\tau}) \tag{10.7}$$

Then the strong rationality issue can be defined as $\delta_{t+\tau} \neq \delta^*_{t+\tau}$ for some $\tau \geq 0$. Clearly, if it were for not for the time interpretation subscripts, then the wrong weights attached to the utility for individual periods would be seen as simply a specific case of merit wants. The time subscript is important because of the intuition it provides about the direction of bias: it is usually supposed that $\delta_t > \delta^*_t$, in other words the current period is overweighted in deliberations, compared to the interests of future selves. This view has a long history in economics as Peart (2000), points out. Stanley Jevons, for instance stated in the Theory of Political Economy that

> to secure a maximum benefit in life, all future pleasures or pains, should act upon us with the same force as if they were present, allowances being made for their uncertainty. (1871, p. 72).

He contrasted this ideal with the actual state of affairs, 'a future feeling is always less influential than a present one,' (1871, p. 72). Alfred Marshall and Irving Fisher expressed similar views, with the latter stressing that poverty made the distortions to preference that much greater. 'The smaller the income a man has, the higher is his preference for present over future income', (Fisher, 1930, p. 381). But the most well-known precedent for the view that current consumption is overweighted compared to future needs is associated with the name of Pigou, who argued that a failure to give equal weight to future consumption means only 'that our telescopic faculty is

defective and that we therefore see future pleasures, as it were, on a diminished scale' (Pigou, 1920, p. 25).

The second feature of bounded rationality in the context of intertemporal decision-making concerns the failure of future selves to necessarily follow the plan of consumption and saving laid out by the current self. Let X_t be the choice set for the individual at time t, a typical element of which is $x = (x_t, \ldots, x_T)$

Define a plan as,

$$x^*(t) = \underset{x_t, x_{t+1}, \ldots \in X_t}{\arg\max} \sum_{\tau=0}^{\tau=T} \delta_{t+\tau} u(x_{t+\tau}) \tag{10.8}$$

Write $x^*_{t+1}(t)$ to mean $x^*(t)$ excluding x_t, then dynamic consistency is the property that, for all t, $x^*_{t+1}(t) = x^*(t+1)$. In other words, future consumption planned at time t is also optimal at later times. As Strotz (1955), first pointed out, a sufficient condition for dynamic consistency is a constant rate of discounting, that is $\delta_{t+\tau} = \delta^\tau \delta_t$ for some $\delta \geq 0$ and all $t, \tau > 0$.[9]

The evidence that preferences do not have this simple feature is widespread, but not uncontested. A recent survey can be found in Frederick et al. (2002). In experimental work, discount rates are typically elicited by asking subjects to choose between x now and x+y t periods into the future. Alternatively subjects are asked for the value of y which makes them indifferent between the two options. The per-period discount rate, δ is then calculated as the value which solves the equation, $x = (x + y)\delta^t$. In an early, but fairly typical study, Thaler (1981), finds a value for δ of 0.22 (equivalent to a discount rate of 345%) when t represents one month, $\delta = 0.45$ for a one year interval and $\delta = 0.84$ when t is equal to 10 years. In other words, δ rises as t increases, rather than being constant as dynamic consistency would require. This pattern is usually referred to as 'hyperbolic discounting' on the grounds that a hyperbolic function often fits the data on δ_t fairly well, but in fact it seems more likely that the current period discount rate, δ_1 is low relative to discount rates in subsequent periods which thereafter remain fairly constant.[10] As a result, the following form for the evolution of δ has been proposed:

$$\begin{cases} 1 & t = 0 \\ \beta\delta^t & t > 0 \end{cases}$$

[9] It is also necessary for most of the budget sets commonly encountered in models of intertemporal choice.

[10] Read (2002), argues that discounting is sub-additive rather than hyperbolic. In other words, when time periods are sub-divided, the sum of the resulting discount rates exceeds that for the undivided period. This is similar to the part-whole bias found in the valuation of goods (see Bateman et al., 1997b for instance).

where $\beta < 1$ (see Laibson, 1997). More recently, Harrison et al. (2005), have used field and laboratory experiments to challenge much of the accepted wisdom on hyperbolic discounting. They note that typically subjects are given the choice between receiving payment now, in the laboratory or at a later date. Delaying payment introduces some uncertainty about whether payment will be received at all. Delaying all payment controls for this effect and, on the basis of their Danish evidence reduces the rise in δ over time, though without altogether eliminating the evidence for hyperbolic discounting in humans.

In addition to the evidence on hyperbolic discounting there is also some interesting experimental work (though much of it involving hypothetical choices only) which questions a number of the other assumptions of the basic intertemporal model. Prelec and Loewenstein (1998), for instance, observe that typically individuals have some preference for intertemporal variety, thus violating additivity. Faced with eating 100 meals of chicken tikka in 365 days, many individuals would prefer to spread them out evenly over the year, rather than eating them sequentially. They also show that individuals prefer an increasing sequence of payments (e.g. £1,000, £2,000,... £5,000) to the same sequence in reverse, even though the latter dominates the former given positive market interest rates.

An optimal consumption plan that is not dynamically consistent raises the question of what the actual plan will be. Two factors are important in this context:

- The agent's ability to commit themselves to a future plan of consumption;
- Their degree of awareness of the dynamic inconsistency.

The individual who is not clear-sighted may stumble from plan to plan. In each period she or he plans consumption for all future periods and, in each subsequent period plans only recently made are torn up and a fresh plan of lifetime consumption is laid out. This combination of a long planning horizon and myopia about the actions of their future selves lacks psychological credibility. At the other extreme, and possibly equally incredible, is the clear sighted agent who anticipates the decisions of future selves and hence is constrained to choose from only those plans which are dynamically consistent.

One important strand of the extensive literature on dynamic inconsistency has tried to say something about welfare. To a degree this means adopting the perspective of one of the life-time selves involved in the planning of consumption. For instance in the over-analysed instance of Odysseus and the Sirens, the optimality of his plan to lash himself to the mast is based on giving greater weight to the views of the pre- and post-Siren Odysseus compared to his preferences as he passes the rock. Most people will probably not object to this judgement in the case of Odysseus, but in other cases of dynamic inconsistency it is much harder to settle on a clear-cut view of the correct metric of welfare. As such, an optimal public policy may not be defined for dynamic inconsistency in general.

For some models of intertemporal choice, the welfare issues are clearer. An instance is provided by the theory of procrastination advanced by O'Donoghue and Rabin (2005). They consider an infinitely-lived individual who must decide whether and when to undertake a task which will cost c in the current period, but then yield

a stream of benefits v in all future periods. In this context procrastination is defined as never doing a task which has a positive net present value. The individual's future consumption is discounted at a rate of δ per period. It is supposed that the individual also underweights future consumption – placing a weight of $\beta \leq 1$ on all benefits or costs which accrue beyond the current period. O'Donoghue and Rabin distinguish between the preference for immediate gratification (as indicated by β) which leads to time-inconsistency and the subject's naivete about their future value of β. A subject who correctly anticipates their future value of β is described as sophisticated, while someone who believes that their future value of $\beta = 1$ is completely naïve, unless β is actually equal to one. I have already noted that complete naivety is not credible in a context where agents are making lifetime plans, but O'Donoghue and Rabin argue that models of dynamic inconsistency are not robust to the introduction of some naivety (i.e. a forecasted value of β somewhere between its true value and 1). O'Donoghue and Rabin advance the notion of a 'perception perfect strategy', a concept akin to that of sub-game perfection, in which (a) subjects choose optimally in each period, given their beliefs about the actions of their future self and (b) their beliefs about their future beliefs are correct. Note though that what an individual actually does in any given period may contradict what they believed they would do in previous periods.

An example illustrates their model and its main results. Suppose, in task A, $c = 5$, $v = 0.8$ and $\delta = 0.9$. The present value of doing the task in the first period is $-c + \delta v/(1 - \delta)$, which is 2.2 in this case. However, the individual actually appraises the value of the task as $-c + \beta\delta v/(1 - \delta)$, which for $\beta = 0.8$, makes the value of the task equal to 0.76. So the task is still worthwhile, in the sense that it has a perceived positive net present value. However, its net present value as seen by the agent, may be still higher if implementation is delayed by one or more periods, because this has a more marked impact on the value of the cost of the project, compared to the stream of future benefits. Implementation in the tth period yields a payoff of $\beta\delta^t[-c + \delta v/(1 - \delta)]$. Figure 10.4 shows the path of net present values for implementation of the task in periods one to ten. As can be seen, with sophisticated beliefs, the individual prefers any delay of up to eight periods, compared to implementation of the task immediately. However, doing the task in period one is preferred to leaving it until period nine or later. Suppose, in period 1, the individual believes that future selves will not implement the plan in periods 2 to 8, then he or she will prefer to do the task now rather than delay. Hence the longest possible delay is eight periods and all equilibria involve the agent executing the task in one of the first eight periods.

Now consider an individual who is naïve about the beliefs of future selves, thinking that future selves will weight future consumption by $\beta^* = 0.9$ rather than $\beta = 0.8$. The upper time path in Fig. 10.4 shows the path of net present values for a sophisticated person with $\beta = 0.9$. This is how the partially naïve person believes futures selves will behave. In other words, futures selves will not tolerate a delay of more than one period. The partially naïve person then compares doing the task immediately to leaving it, in which case, they believe, the maximum future delay will be one more period. Since the discounted value of executing the task in

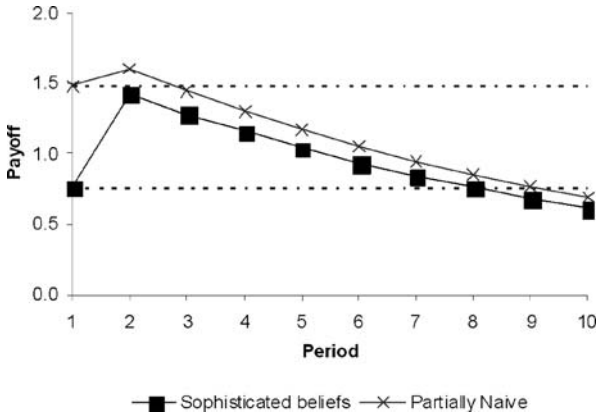


Fig. 10.4 The Value of Procrastination

period three is higher than its value in period one, they choose to postpone doing the task. Their perspective and analysis is repeated in the next period and so the task is never completed.

The critical factor here is that the maximum tolerable delay for the naï ve individual is strictly shorter than its equivalent for the sophisticated individual. Whenever this occurs the task will remain undone.

O'Donoghue and Rabin go on to show that procrastination may occur with a larger choice set when it does not occur on a smaller sub-set of choices. In other words, increasing the choices open to an individual can induce procrastination. For instance if the individual could only choose between doing nothing and implementing task B where $c = 1$ and $v = 0.3$, then for the same values of β and δ used above, the individual would implement the task immediately, whether they were naïve or not. Adding the option of doing task A used above means that if the individual were to execute a task then it would be A, since it has a higher net present value. However, as we have just seen, a sufficiently naïve individual would never do task A.

As the authors point out, the model has its limitations: the naïve individuals never learn from their faulty forecasts of future behaviour and neither they nor their sophisticated brethren employ any of the psychological and incentive devices that many individuals employ to avoid extensive delay of the kind the model produces. The model though does illuminate a number of issues. Clearly, complete procrastination of the kind found in the model is welfare-reducing for the individuals concerned. However, some delay often raises perceived utility, hence implementation in the first period may be against the perceived interest of some selves. As with Odysseus therefore commitment devices can have value.

The other feature of the model which is of interest is methodological: the agents mix quite sophisticated reasoning with, in some cases, extreme foolishness. The agents are assumed to be able to evaluate an infinite stream of payoffs and compare implementation of the task in different periods. They are able to posit the existence of future selves with different preferences. They select only plans which are consistent with their beliefs about the actions of these future selves. On the other

hand, naïve agents underestimate their own potential for future procrastination and they fail to learn from the failure of their earlier plans.

Many people seem to be aware of the procrastinating feature of their own decision-making and make commitments to try to overcome procrastination. Mayer and Russell (2005), offer the example of teachers in some districts of California many of whom choose to receive their income in 12 annual installments when the possibility of getting it in 10 payments is also available. For an older example, which uses a different language for intra-personal conflict, we turn to Charles Madge (1941), who in his investigative tour of wartime saving in English cities noted the large numbers of Christmas clubs and saving schemes:

> In psychological jargon, it is the function of the super-ego to make people save – not many are natural misers. Therefore, because of the known weakness of the super-ego in face of temptation, Lancashire women mill-workers are prepared to pay somebody to come round once a week and collect their sixpences or shillings. (p. 429–430).

Given these observations on intertemporal anomalies we turn to the issue of savings policy.

10.4 Regulating the Framing of Savings and Pensions Choices

In the UK in 1988, the then Conservative government liberalised the pensions market (Connolly and Munro, 1999). Previous to the reforms, individuals had been heavily constrained in their choice of options. All employees with earnings above a minimal level were forced to make contributions towards the first tier of the pension scheme – the Basic pension, the level of which was not income related. For the second tier of the pension system – the part related to income – employees could choose between the state scheme (SERPS) or could join an institutional scheme organised around their workplace. Typically, members of such schemes paid a proportion of their income each month into the scheme, as did their employers. As an inducement National Insurance contributions were reduced for members of such schemes.

The reforms left the structure of the first tier largely intact, but it added an extra element of choice for the second tier. Now, employees could opt for a pension scheme provided by a private supplier, provided that such a scheme satisfied certain minimal requirements. The advantage of the private schemes lay in their portability. Typically, in pension schemes organised through an employer, an individual who left before retirement faced reduced benefits compared to individuals who maintained a single employer throughout their working life. For individuals who switched jobs regularly or who had a pattern of interruptions to employment, private schemes could be more advantageous because they were attached to the individual rather than the company. On the other hand they did have some clear disadvantages for many potential purchasers (Blake, 1995). First, administrative costs were typically higher; secondly employers were typically not as generous in their contributions to private saving as they were to the collective pension schemes. Finally, almost all private schemes were defined contribution (DC) schemes, meaning that the value of the pension was determined by, for instance, the values of shares and stocks at the time

of an individual's retirement. Most employer-organised schemes, on the other hand, were defined benefit (DB) schemes, meaning that the generosity of the pension was some fixed fraction of an averaged measure of an individual's lifetime earnings. In short therefore, risks for workers were higher with DC schemes, compared to their more traditional options.

Despite these apparent deficiencies, private schemes were immediately popular. By the early 1990s several million individuals had taken out private pensions. In many instances, individuals switched out of employer-organised DC schemes into private pensions. What was startling about the up-take however, was not its level but the individuals who switched – hundreds of thousands of state-employees such as nurses and teachers opted out of their existing scheme and into the new pensions (Disney and Whitehouse, 1992). Nurses and teachers in the UK typically work for the state throughout their lifetime, their jobs are relatively secure and employer contributions to the pensions scheme are relatively generous. Hence, they were definitely not the people for whom the reforms were originally devised.

It is by no means certain that all these individual choices were irrational. For some people private pensions released them from a rigid pattern of savings which dictated that individuals made heavy contributions to pensions at times in their lives when, perhaps, they would have preferred to spend money in other ways. That said, it was clear that many people rapidly regretted their decisions and attributed them to heavy pressure from salespeople combined with misleading information about the consequences of their decisions. Above all there was the temptation to put off saving for the future and to enjoy a temporary rise in consumption. Much energy in the UK pensions system since the early 1990s has been devoted to sorting out the aftermath of the reforms, reversing hasty choices to opt-out and compensating individuals who claimed they were misled.

The example of pension policy reform suggests some of the forces which tend to limit saving for pensions:

- Having to opt-in to higher savings levels.
- Lack of clear and reliable information on the impact of savings rates on future consumption.
- Having a 'savings' account or policy rather than separate policies for different forms of saving.
- Having the possibility of raising saving in the future, but without commitment.
- Immediate consumption benefits from reductions in saving.

Not surprisingly therefore, Thaler (1994), argues that the ideal savings programme should have four features (see also Weiss, 1991):

1. it provides an immediate reward to saving
2. it is simple to use and understand
3. the money is perceive as 'off limits' to current spending
4. experts consider the programme a good deal (Thaler, 1994, p. 189).

Thaler and Benartzi (2004), offer an example of the theory at work, involving a US company which wished to raise the savings rates of 315 of its lower paid

employees. All were offered the chance to talk over their retirement planning with a consultant. Of those who accepted the offer, around one quarter then took the initial advice of the consultant . For the remainder, Thaler and Benartzi had devised a savings plan (SMarT) in which members would make no immediate rise in their contributions but would agree to commit 3 percentage points of their annual wage rises to their retirement savings plans. Subjects could, if they wished, change their minds at a later date and withdraw from their new commitment.

Figure 10.5 shows the results of the exercise. Thirty four percent of the group either refused the opportunity for a financial consultation or accepted the counsellor's suggested change of savings behaviour. About one in seven then refused to enter SMarT, but of the remaining 52% of individual, nearly all then stuck with the programme for three years of pay awards. For those who joined the scheme, average savings rates were 11.6% after four years, compared to the 3.1% prior to its introduction. In fact after three years, their savings rates exceeded those for the individuals who had taken the consultant's original advice. Meanwhile members of the workforce who had remained outside the scheme had savings rates which hovered between 6 and 7%.

The theory behind this scheme is closely related to the mental accounting literature discussed in Chapter 2 coupled with the problems of intertemporal optimisation discussed in the previous section. The source of the problem is viewed as procrastination on the part of the workforce. An immediate rise in savings rates, even when offered, is not taken because the pain of immediate loss of consumption is hard to bear given only the promise of the pleasure higher consumption in the distant future. But if as we saw, the pain of consumption foregone is lower than that from a reduction of actual consumption, individuals will be more willing to accept a scheme such as SMarT over an immediate rise in contributions.

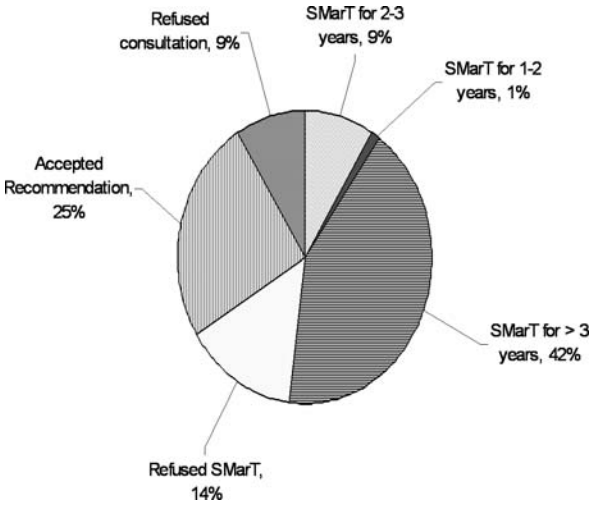


Fig. 10.5 Choices in Thaler and Benartzi (2004)

Although the positive side of SMarT-like policies is relatively clear, there still remains an important normative issue. The Thaler scheme passes the test of ex ante and ex post consent, because quite clearly participants can opt not to join or can opt out of the savings programme at a later date. As a result they also pass the related 'libertarian paternalism' test discussed in Chapter 5 – fully rational individuals would not find their path to optimisation hindered by the scheme, but boundedly rational agents appear to benefit.

Following the Thaler guidelines, Gorringe (1998), draws out the following general sorts of policy for governments which wish to see higher rates of saving for retirement.

- Legislate to make joining a (private) pension scheme a default option for everyone.
- Increase withholding tax rates
- Make the default option for tax refunds their payment into a pension scheme or other contractual savings account
- Provide information on saving for retirement
- Subsidies for saving

The last two elements are fairly standard devices. Each of the first three is designed to create inertia around the decision to save rather than consume. If taxes are withheld from income, then this limits consumption, but if the eventual refunds are then sent into savings accounts then this further reinforces the default option of saving rather than consumption.

10.4.1 Planning Ahead

One of the features of savings activity is its variability even after controlling for standard economic and demographic variables. In other words not everyone seems to suffer from problems of self-control . Across all ages and all income classes, some people save while others do not. For instance, for the U.K. Banks et al. (2002), show that around 10% of households in the top half of the income distribution have no financial wealth. Even in the richest 10% of the population, one in twelve households has £500 or less in financial assets. However, this diversity should not blind us to the fact that it is principally low income households that lack savings. Over half of the poorest 10% of households have no savings, while Banks et al. (2002), point out that for the poorest 20% of the population the median savings is just £50.

Psychologists (e.g. Warneryd, 1999) have examined the traits and attitudes that make some people save while others in similar economic circumstances do not. Not surprisingly, conscientiousness is an important factor associated[11] with saving (Webley et al., 2001) along with the personal discount rate, which tends to decrease as income and age rises. Meanwhile when Lunt and Livingston (1991), examined the lives of a group of Oxford residents, they found that against the positive influence

[11] The issue of causation rather than correlation has not been a central feature of this literature.

of education, income and age on saving, non-savers have greater preference for flexibility, tended to be more fatalistic about the future, were less likely to monitor their expenditure or to plan ahead and were more secretive about finances with friends and relations.

In the language of Ajzen (1991), much saving is planned behaviour which is most likely to occur when personal attitudes towards saving, beliefs about its efficacy and feelings of social pressure are aligned. Effective saving behaviour requires a person to plan ahead and to believe that setting aside money will make a difference. Not surprisingly therefore, 'planning ahead' is a key predictor of whether people save. In turn, planning ahead is correlated with a large number of factors which are summarised in Fig. 10.6. (The signs in this figure refer to whether the factor has a positive or negative correlation with planning.)[12]

An influence on planning ahead and a theme running through much of the work on savings psychology is the notion of 'locus of control', a concept from social psychology (e.g. Lefcourt, 1982) which refers to the extent to which individuals feel in control of what happens in their life. Individuals with an internal locus are people who feel in control; individuals with an external locus are more fatalistic, believing

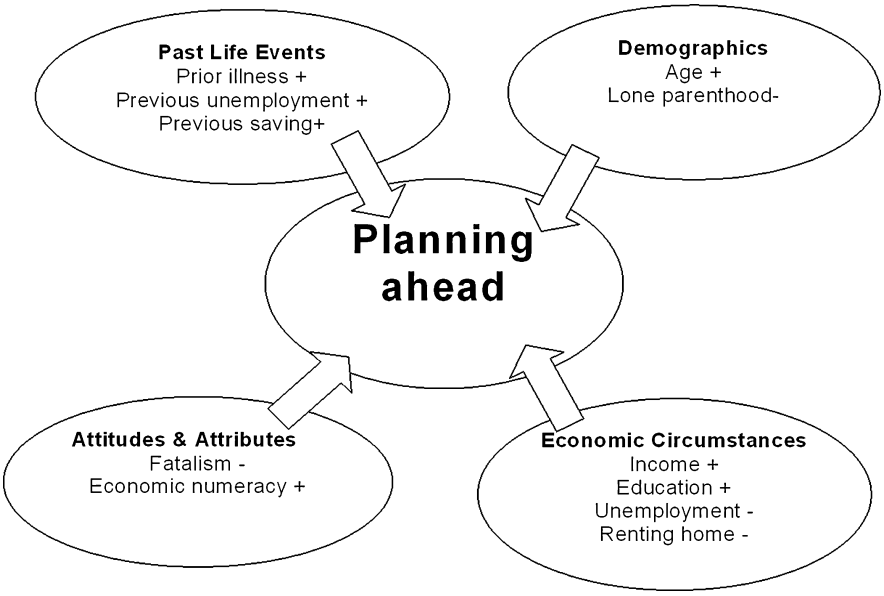


Fig. 10.6 Factors Affecting Planning Ahead

[12] Some sources: prior illness, previous employment (Atkinson et al., 2006), previous saving, age (Webley, et al., 2001), lone parenthood (Atkinson et al., 2006), fatalism (Lunt and Livingston, 1991; Eckel et al., 2003; Wu, 2005), economic numeracy (Banks and Oldfield, 2006), income (Lunt and Livingston, 1991; Atkinson et al., 2006), education (Lunt and Livingston, 1991), unemployment (Atkinson et al., 2006), renting home (Atkinson et al., 2006).

that events outside of them determine what goes on in their lives. Its relevance here is that non-saving behaviour and not planning ahead is associated with an external locus. Direct evidence of this link is provided by an important savings experiment financed by Human Resources Development Canada (Eckel et al., 2003). A wide range of subjects participated of whom, 72% had a family income of less than 120% of an official Canadian poverty line. In one treatment, participants were offered a real choice between 100 Canadian dollars in cash one week after the experiment or an investment of 400 dollars in a fund for education and training available for spending several months later. Subjects who had revealed high discount rates in another part of the experiment were more likely to opt for the immediate cash, as were subjects with an external locus of control.

In a similar vein Wu (2005), uses the 4,400 respondents to the 2001 US Survey of Consumer Finances to explore why people save. Out of the sample, 18% do not save at all. Those realising that they need to save more are more likely to be pessimistic about the future of the economy and more likely to agree with the view that they are unlucky with their finances. However this group is less likely to have done something about it and when they do save, they are less likely to shop around for financial products.

The ability to plan ahead is also affected by maths skills. Using the 2002 wave of the English Longitudinal Study of Ageing, Banks and Oldfield (2006), conclude that amongst the over 50s, a large proportion of the population have poor financial numeracy skills. The ability to solve relatively simple exercises in financial arithmetic is correlated with knowledge and understanding of financial products and also with levels of saving and retirement provision, even after other factors are controlled for.[13] The results match those of a wider study concerning the financial capability of the UK population, carried out on behalf of the Financial Services Authority, Atkinson et al. (2006), find a number of factors positively linked with planning ahead, including education, age, income and a number of factors negatively linked including renting, being unemployed, being a lone parent, and not having a current account. Those planning ahead were also more likely to have been active choosers of financial products, to keep informed about money and to have a high score in relatively simple exercise in financial arithmetic.

Inferring causation from some of these associations is difficult. With the exception of the Eckel et a1. (2003), study, the results indicate correlation rather than causation. Nevertheless, there have been some attempts to develop savings policy for low income earners on the assumption that financial training and heavily subsidised, small-scale saving can reduce fatalistic attitudes and encourage greater planning ahead. The most notable attempts to innovate can be seen in the case of anti-poverty policies designed to raise wealth rather than income.

The notion of asset-based welfare and particular, Individual Development Accounts (IDAs) has provided a radical challenge to the traditional format of the

[13] Bernheim and Garret (2003), report on the positive effects for saving of workplace-based financial education.

welfare state, given that historically, welfare minded governments have focused on providing support for incomes rather than savings. As conceived by Michael Sherraden (1991), IDAs were part of a proposal to reverse the emphasis and make asset based policies a central component of the welfare state. To that end IDA s consist of three main components: a matched subsidy to saving for low income groups; education in financial management skills for the recipients and restrictions on the use of the subsidized savings for uses such as homeownership or education. Though still experimental, IDAs have become a popular bipartisan policy in the US and to date over 40 states have approved them in some form or another, usually delivered with the active participation of the not-for-profit sector. With the introduction of the Child Trust fund scheme (mentioned above) and the development of the Savings Gateway project, asset-based welfare has spread to the UK.

Most IDAs are non-experimental: community groups and municipalities offer themselves as possible sponsors and then actively recruit potential participants. As such scientific assessment of their impact has been limited. The most compelling evaluation of IDAs is that conducted by Mills et al. (2004), reporting on an IDA in Minnesota, typical except for the fact that eligible applicants were randomly accepted or rejected for the programme. Fifty-three percent of people who signed up for the programme closed their account without ever qualifying for matching funds. Out of those who remained, average savings were $1,480 per participant, compared to $2,041 total financial assets (including retirement savings) at the start of the scheme. The largest long term impact of the scheme was on home ownership where after 48 months (1 year after the programme ended) those randomly accepted into the IDA had ownership rates 6 percentage points higher than those rejected. The wider impact of a group of 14 IDAs, linked under the heading of the American Dream Demonstration (ADD) is assessed in Moore et al. (2001). They report participants feeling higher confidence in the future (93% of their sample) and feeling more in control over their lives (85%).

The Savings Gateway programme has been the UK's pilot version of asset-based welfare. In the initial pilot (there is now a second) subjects were recruited non-randomly into five different schemes based in low-income urban districts. As with the U.S. schemes, savings were heavily subsidized and, with the exception of one variant, accompanied by some financial education. In their evaluation of the first group of pilots for the UK's Savings Gateway programme Kempson et al. (2005), find that average participants saved £361 out of the possible £375 over the lifetime of the projects. Perhaps more importantly, compared to a reference group of individuals from areas neighbouring the Savings Gateway pilots there is some evidence of a positive change in attitudes towards saving – 39% of the survey sample agreed or strongly agreed that they felt more in control of their life as a result of the scheme, while 60% felt more financially secure and a third stated that they were more likely to save for retirement. One of the schemes, in Hull, had no financial education and targeted potential recruits using benefit data, without the aid of community groups as intermediaries. As such, the Hull scheme represents a lower cost approach to asset-based welfare, compared to the US style of IDA. The Hull group saved as much as the groups where community groups were more heavily

involved. However, the large number of individuals who saved at or close to the maximum in all the pilot areas limit means that any underlying differences in the effectiveness of the schemes may have been masked by the low level of the cap on savings.

10.5 The Framing of Tax Policies

The idea that behaviour will depend on the framing of tax policies follows naturally from the arguments of this book, but it is important to distinguish two separate routes by which framing effects can be relevant for tax policy.

1. Since market behaviour is susceptible to a variety of framing effects, it would be no surprise if individuals were not also influenced by the framing of taxes and benefits.
2. Secondly, perceptions of tax fairness might well be susceptible to the manner in which the tax and benefit system is portrayed. As a result, a tax system might be politically acceptable in one frame, but unacceptable to citizens when presented in an alternative manner.

It is possible that these two routes are linked, but it is not a necessity. An individual with other regarding preferences could be immune from framing effects in their own choices, but nevertheless perceive that other citizens might not be so clear-headed in their assessment of government policies. As a result she or he would not be affected in their market choices by the framing of taxes, but could still have clear preferences over potential frames.

Schelling's 1981 example illustrates how framing may affect perceptions of the most acceptable tax system. Consider this slightly modified version of the choices (A versus B and C versus D in Table 10.2) faced by his students. In both pairs the choice is whether to target a higher subsidy for having children to higher income families or to families on low incomes. However in the two choices, the framing of the subsidy differs. Since every system with bonuses for some individuals is

Table 10.2 Framing Tax Systems

Option	For	High income	Low income
A	having children	£50,000 income families receive a £2,000 bonus	£15,000 income families receive £1,000
B	having children	£50,000 income families receive a £1,000 bonus	£15,000 income families receive £2,000
C	not having children	£52,000 income families receive a £2,000 penalty	£17,000 income families receive a penalty of £1,000
D	not having children	£52,000 income families receive a £1,000 penalty	£17,000 income families receive a penalty of £2,000

equivalent to one in which there are penalties for the complementary individuals, the only consistent answers to this pair of questions are the pairs {A,C} and {B,D}, but in fact the majority of Schelling's class opted for {B, C}. Such political preference reversals may occur either because the individual subjects were themselves confused by the framing devices or that they considered others to be fooled. There is no evidence on which interpretation is the more correct.

Experimental tests of Schelling's example have produced mixed results to date. Traub (1999), did not find support for Schelling's hypothesis in his data, but in dedicated experiments on the issue with 49 subjects, McCaffery and Baron (2004), discovered that subjects typically rated the fairness of bonuses to low income couples more highly than bonuses to high income couples but switched ratings when the choices were presented in the penalties format.

Traub (1999), used employees of 5 North German companies to examine wider issues of the dependence of perceptions of fairness on endowments. In one treatment, subjects were given information about a proposed tax schedule for a single, childless person then asked to propose a fair level of tax for a couple with 2 children and the same gross income. Some weeks later subjects were presented with information on a proposed tax schedule for the couple-with-children family in which the tax was at the level they themselves had proposed in the earlier questionnaire. They were then asked to propose a fair tax on a single, childless person with the same gross income. In other treatments the order was reversed. Traub found that in general tax fairness 'indifference curves' were non-reversible: the tax reduction for the couples with children in the first treatment was larger than the surcharge for a single person without children in the second treatment. In other words, people were generous in subsidizing the 'desirable' or 'deserving' but reluctantly to penalise heavily the relatively 'undeserving'.

In surveying the general issue of tax framing on behaviour, Krishna and Slemrod (2003) put forward a number of conjectures about the potential form of framing effects:.

1. Endowment effects – as with other choices, citizens may be averse to changes in the status quo.
2. Aggregation and splitting of gains and losses. According to Kahneman and Tversky (1979), (see also earlier in this chapter), the asymmetric shapes of value functions for gains and losses suggest that governments could raise the acceptability of expenditure by splitting it as finely as possible, while reducing the unacceptability of taxes by aggregating losses. Conversely, Krishna and Slemrod (2003), point out there may also be threshold effects – small gains and losses might not be noticed by individuals, raising the acceptability of many small taxes and reducing the attractiveness of splitting benefits into many fragments.
3. Timing effects. The timing of taxation and benefits might influence its acceptability. For instance, one reason for paying out the Working Tax Credit[14] on

[14] Working Tax Credit, formerly Working Family Tax Credit is the UK's system for subsidizing work for low income families. Payments are made weekly or monthly (at the choice of participants) directly into bank accounts.

a frequent basis is to reinforce the perceived gains from working. An annual lump-sum might be hard to manage for low income families, but might also lead to dislocation between the tax refund and the work that generated it also point to a marketing literature which concludes that many consumers feel more favourably towards purchases when prices are expressed in 'pennies per day' rather in a lump-sum format. Conversely benefits become more attractive when expressed in overall lump-sums rather compared to when they are broken down into daily amounts. Krishna and Slemrod (2003), suggest that politicians may seek to exploit such effects when describing the benefits and costs of their manifestors. Note though that the effect, if true, contradicts the aggregation argument given above. In fact it is only reconcilable in the presence of threshold effects.

4. Obfuscation. McCaffery and Baron (2005), argue that reducing the transparency of taxes can raise their public acceptability. Because individuals value more highly what is brought to their notice, losses and gains may be hidden by simply avoiding placing them in the spotlight of consumer attention. Indirect effects of taxes, such their consequences for prices may not therefore be salient. In some cases, this may not be simply an issue of consumers being mislead, but an active preference by citizens for hidden taxes. McCaffery and Baron (2005), cite the example of social security payments. If these are wholly framed as taxes from income then, they argue, workers will feel the pain of making the payments. However, if the deductions are defined as employer payments, then employees will not code the contributions as losses and therefore not suffer loss aversion.

10.5.1 Tax Evasion

Since Yitzhaki (1974), a number of researchers have pointed out the poor fit between a theory of behaviour based on expected utility on the one hand and the empirical regularities of tax evasion. In simple models of optimal evasion, taxpayers choose how much income to reveal to the authorities. If they fail to reveal all their income there is some probability of being caught and a consequential fine. Allingham and Sandmo (1972), show that if p is the probability of being caught and the fine (as a proportion of the evaded income) is γ, then some evasion is optimal provided that $1 - p - p\gamma$ is positive.[15] For many tax systems p is very small (2–3% according to Dhami and al-Nowaihi, 2004) while γ is typically less than 2. Hence $1 - p - p\gamma$ is close to 1 and evasion is optimal. In fact, Andreoni et al. (1998), report that around 30% of (US) citizens evade taxes.[16] In a similar vein expected utility theory underpredicts the level of tax evasion (i.e. how much tax is evaded) unless it is

[15] The smoothness of the utility function in income, means that the for infinitesimally small gambles the consumer is risk neutral. As a result, the consumer should evade some tax as long as the expected returns from doing so are positive. This yields the condition 1-p-pγ.

[16] One possible reason for this low percentage is that many salaried taxpayers have very little in the way of invisible income that could possibly be hidden.

assumed that individuals are risk averse to a degree that does not match behaviour elsewhere. Thirdly, expected utility is a function of total wealth or income, levels of tax evasion should not be affected by whether or not taxpayers must pay some or all taxes in advance. In reality, paying tax in advance seems to reduce evasion (Chang and Schultz, 1990, using US Internal Revenue Services data). Finally, if absolute risk aversion is declining in wealth, then higher tax rates should lead taxpayers to evade to a lesser degree. The opposite seems to be more realistic.

Bernasconi and Zanardi (2004) and Dhami and al-Nowaihi (2004), apply versions of prospect theory to these paradoxes.[17] In the latter, the authors demonstrate that cumulative prospect theory used with typical parameters, when allied to an assumption that the reference point is the post-tax, no evasion income and coupled with a stigma attached to evasion, can explain the four regularities. Leaving aside the introduction of a stigma term, the key assumption here is the choice of reference point. If the reference point was gross income or post-tax with evasion, then their results would not follow because all states of the world would be coded as losses. So, a critical but as yet unresolved question for policy framing is how reference points are established and maintained.

10.5.2 Labour Supply and Tax Framing

Much of the debate on personal taxation over the last twenty years has been focused on its impact on incentives to work. We end this chapter with a theoretical discussion of the possible impact of framing effects on labour supply choices. As with much of the previous sections, the 'alternative' theory used to predict the possible impact of framing is very much guided by theories of reference dependent preferences.

Figure 10.7 shows a standard income-leisure trade-off diagram. The budget constraint has four important components:

- A minimum income level, b, which provides basic support even in the absence of paid employment;
- A terms of trade for converting leisure into consumption via paid work;
- A supplementary element of benefit, paid up to income y^l, above which
- There is a taper, at a rate of a, through which the supplementary element is gradually withdrawn, up to an upper level of income y^u.

This budget constraint mimics some aspects of tax/benefit systems such as those operating in the USA and UK, where working tax credits (WTCs) provide support for families and individuals with low incomes from the labour. In many cases the basic support for individuals out of work differs from the support level given to individuals in work, but for simplicity I have not made this distinction in the diagram.

[17] Using hypothetical payoffs in a sequence of evasion dilemmas, Schepanski and Kelsey (1990), find supporting evidence of an endowment effect in tax evasion.

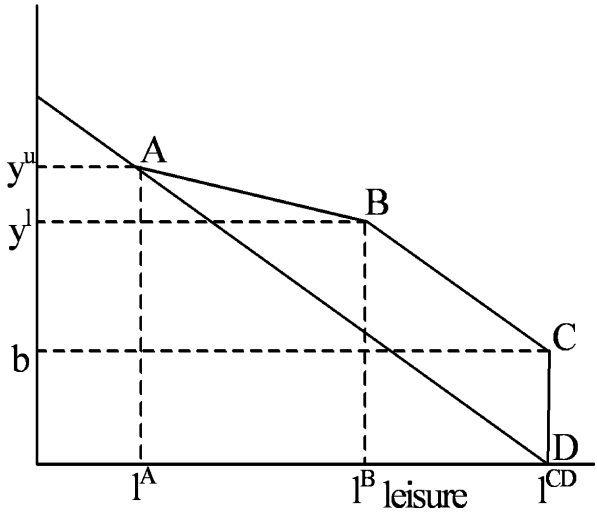


Fig. 10.7 Income-Leisure Trade-off

Consider the four points, A–D, evaluated with regard to their suitability as potential reference points in the framing of policy. In doing so I restrict attention to frames which are truthful.[18] The table below shows the view of each of them, as depicted as gains and losses from the other points. So, as viewed from B, point A is a gain in income, coupled with a loss in leisure. In the final column we also have a summary of the option in a positive language or spin.

What is clear from the discussion on reference point effects earlier in the book is that A appears to be the reference point maximizes the attractiveness of working, since all increases in leisure from A are accompanied by a change in consumption that is presented as a loss.

There are a number of important ways in which the actual policy context differs from this simple analysis of Table 10.3 and that make the choice of optimal reference point potentially complex. Chief amongst them is the fact that for many individuals more hours of work are not necessarily a bad thing, even leaving aside the additional income they can generate. In other words it is not straightforwardly true that increased leisure is the good. Work has often been supposed to bring psychological benefits and to be pleasurable for many individuals. Consequently, it is not unusual to see 'U' shaped indifference curves in labour economics texts.[19] For the framer

[18] Some of the benefits of misleading boundedly rational individuals have been explored in Chapter 7. Here, I am concerning with how the same information can be represented in a manner likely to change labour market behaviour. Such a constraint is potentially important because with the high effective marginal rates of taxation typical with working tax credit schemes all options involving higher labour supply can appear unattractive, no matter what frame is employed.

[19] Of course, in the absence of poverty and unemployment traps, the upward sloping section of the indifference curve will not typically be observable in the labour market. The point is that in drawing preferences in such a manner, the authors are recognising the inherently dual nature of the psychology of work.

Table 10.3 Framing Income-Leisure Choices

From	A	B	C	D	Positive description
To					
A	–	Loss (y^u–y^l)	Loss (y^u–b)	Loss (y^u)	Work, income and independence from the State
	–	Gain (l^B–l^A)	Gain (l^{CD}–l^A)	Gain (l^{CD} –l^A)	
B	Gain (y^u–y^l)	–	Loss (l^B–l^A)	Loss (y^u)	Work, income and extra benefit
	Loss (y^l–b)	–	Gain (l^{CD}–l^B)	Gain (l^{CD}–l^B)	
C	Gain (y^u–b)	Loss (l^{CD}–l^A)	–	Loss (b)	Benefit Income b and maximum leisure (no work)
	Loss (l^{CD}–l^B)	Gain (y^B–b)	–	–	
D	Gain (y^u)	Gain (y^l)	Gain (b)	–	No income & maximum leisure (no work)
	Loss (l^{CD}–l^A)	Loss (l^{CD}–l^B)	–	–	

aiming to avoid the usual reductions in labour supply associated with income maintenance schemes this has only benefits, since it means that, for instance, the choice of A versus D can be framed as a pair of gains ('you get to work more and you get income') rather than as a trade-off between one gain and one loss.

In additional to the multi-faceted nature of work another important dimension of the policy task is that typically wages will differ the amongst the target population. To see why this makes a difference, consider one possible way of framing the consumer's choices:

> If you work 1 hour less, you will have £x less to spend each week.

For an individual, for a given job, it is straightforward to calculate what x is. If he or she is on the section of the budget constraint between A and B, $x = (1 - t)(1 - a)w$, where w is the wage rate, t is the tax rate on labour market income and a is the 'taper rate' – the rate at which benefit is withdrawn for every post-tax £1 earned. When w varies across the population, then x will also vary. Now public documents and other messages framing benefit rules cannot typically be targeted to particular individuals, but rather must aim at some average potential beneficiary. This makes frames such as the one in the example either vague or irrelevant to a proportion of the target audience. Vagueness would tend to undermine the credibility of the framing effect and would therefore be more likely to leave the individual's prior frame intact.

Some frames may be less affected by the restriction that the message be suitable to a wide range of individuals. For instance 'if you work more than 16 hours you may be entitled to up to £100 of extra help each week', does not require mention of a specific wage rate to be truthful. However, it does not say anything about the taper. To be wage independent, that statement could be of the form:

as you work more and earn more, the support you receive will be withdrawn, at a rate of £ a for every £1 extra you receive in your pay packet.

On the other hand, its converse: if you work less, then for every £1 less you receive in your pay packet you will receive an extra £ a in support, sounds far more complicated and therefore less convincing as a frame.

In both these last two examples, independence from mentioning the wage rate is achieved by separating out the benefit and the post-tax income in the description of the choice set. Doing so opens up the possibility of manipulating the set of mental accounts as well as the perception of choices as gains or losses. We have seen how mental accounts can alter choices. It appeared that simple rules such as separate out gains, integrate small losses into large gains, integrate losses and (possibly) separate out large losses and small gains could alter the perceived utility to be had from choices. Here, the receipt of benefit is potentially separable from the receipt of income from work. If we were to follow the simple prospect theory rules, then to maximize the favourably of A over C or D would require either integrating the loss of benefit associated with a gain in wage income (if C or D is the reference point), and also, possibly, integrating the gain in benefit associated with a loss of wage income when A is the reference point.

More could be said on this issue. In the UK for instance, the Labour government is separating out the support given to members of working but low income families into the basic working tax credit (WTC) available to all low income households supplying more than 4 hours a week of paid labour and a child tax credit (CTC) paid as a supplement to wage income for low income families with one or more children. Consider the hypothetical example in Table 10.4, where an individual in a household with children must choose between the status quo and working one extra hour per week.

Table 10.4 Framing WTC Incentives

	Work	Receive	Lose	Lose
Old	1 extra hour	£5 extra wage income	£2.75 working family tax credit	–
New	1 extra hour	£5 extra wage income	£1.85 working tax credit	£0.90 child tax credit

Financially these two presentations are equivalent. In both of them the subject has £2.25 available from working one extra hour. On the other hand, what was once one gain of income and one loss of income is now one gain of income and two losses. If the choice is framed in this way, then prospect theory would predict that working longer hours would be less attractive under the new frame compared with the old. On top of that, some of the lost income is specifically linked to children, which may evoke a quite different sense of loss.

Framing effects may also impact on the decision to work. Consider the following two options, A and B.

- Option A: Unemployed; current unemployment benefit of £ x.
- Option B: Full-time Employed; a wage of £ y.

Suppose someone was asked, which would you choose, option A or option B? Compare the result to the question C:

- C: Suppose you are currently unemployed and receive £ x of unemployment benefit per week. You are offered full-time employment. What is the minimum wage you would be willing to accept for that employment?

The evidence on preference reversal (e.g. see Cubitt et al., 2004), would suggest that many individuals would give contradictory answers to these two types of questions. In particular, if the employed/unemployed dimension of the problem is viewed as more prominent than the income dimension, then there would be a tendency for individuals to set a reservation wage in their answer to question C which would be higher than values of y for which they would accept option B. The evidence on loss aversion would also suggest a tendency for individuals to choose option A when endowed with option A and to choose option B when endowed with option B. In both cases, therefore behaviour in the labour market would be sensitive to the framing of the choices available to the individual concerned.

There is one final point to be emphasised. It was noted above that marginal tax rates can be very close to 100% with tax credit schemes. In some cases rates exceed 90% (in the UK for instance – see Connolly and Munro, 1999). Consider two frames for the individual facing a marginal rate of 95% for instance, with an hourly wage rate of £ 5:

- If you work one hour more you will gain an £ 0.25 extra income
- If you work one hour less you will lose £0.25.

Does either frame present the choice between fewer and more hours in a more positive light? Probably not. Consequently, it is worth reminding ourselves that even the most sophisticated of framing strategies may have little impact on behaviour if one choice is so clearly better than another.

10.6 Conclusion

In this chapter I have explored some non-standard fiscal policies. Government often seem to use them unwittingly (e.g. the Winter Fuel allowance) and there has been little in the way of an analytical approach to their employment within economics. The apparent success of Thaler and Benartzi's savings scheme does suggest that insight from the behavioural and experimental economics literature can be used to design and understand framing effects.

I have also applied some of the ideas from the early sections of the book to the issues of pensions and savings. In some ways it is easier to do this with behaviour in financial markets than in other contexts. Partly this is because anomalies are well-established in financial markets and this in turn is due to the fact that the standards of rationality are somewhat clearer in situations where risk and return are the prime factors determining choice in the rational choice model.

As in other chapters, the emphasis has been very much on the individual and the framing of individual choice, but it is worth ending with the reminder that much behaviour takes place in a social setting and that the government often has the ability to frame the social context of decisions as well as the individual frame. The positive impact on recycling of using transparent bags for household rubbish is predicated on the notion that individuals will wish to avoid shame of public exposure. Other deterrent policies, from tax evasion to littering often seek to exploit the same sorts of emotions. There is therefore much more to be learnt on the effects of policy framing.

Chapter 11
Stated Preference and Non-Market Decisions

11.1 Introduction

There are many different means by which non-market decisions can be formalised. Some of the approaches regularly employed such as contingent valuation and the travel cost method share a common underlying philosophy (e.g. welfarism), but some do not. Advocates of Citizens' Juries (e.g. Crosby,1991) or consensus conferences, attach as much value to the process by which decisions are made as to whether the conclusions are reflective of preferences. On the other hand, typical cost-benefit techniques such as hedonic pricing or contingent valuation are almost exclusively concerned with the outcome (preference satisfaction) rather than the process.

There has been little discussion of why one might choose a method that viewed from the underlying normative principles appears dominated by another method. For instance, within non-market valuation it is generally only stated preference techniques such as contingent valuation which can produce estimates for total values, including the values of non-users. In theory therefore, revealed preference methods, such as the travel cost approach or dose response methods are inferior because they cannot estimate benefits to non-users. Nevertheless these techniques are alive and well and frequently chosen in preference to contingent valuation, raising the question of why one might choose such methods, or more generally what is the optimal valuation technique.

The aim of this chapter is to use some of the concepts from earlier in the book and apply them to stated preference methods. It starts from the observation that most decision processes place some weight on individual preferences. The by now familiar problem is that expressions of preference are potentially susceptible to framing effects.

What makes the problem of framing worse with stated preference is that typically the reliability of data becomes weaker the further decisions are removed from the realm of actual and repeated choice (e.g. List, 2002).[1] In choosing a process for

[1] 'An obvious hypothesis, for which there is some support, is that the more ambiguity in one's preferences... the more one's expressed preferences will be subject to procedural and descriptive influences' (Schkade and Payne, 1994, p. 105).

A. Munro, *Bounded Rationality and Public Policy*, The Economics of Non-Market Goods and Resources 12, DOI 10.1007/978-1-4020-9473-6_11,

© Springer Science+Business Media B.V. 2009

producing information on preferences, the decision-maker therefore faces choices between methods which are more prone to framing effects and methods, which might be less prone to anomalies, but which capture only part of the value placed by the population on a potential benefit. For instance, regular users of a threatened wilderness may have much more robust and precise preferences towards its conservation, compared to non-users, but nevertheless non-users may still place some value on the wilderness. One way to conceptualise the dilemma is in the form of a trade-off between the variance of an estimate of value and its bias: for instance regular wilderness users may provide lower variance estimates of value, compared to values obtained from all citizens, but by design an estimate based only one the values of one section of the population will be biased. It is this trade-off that is analysed here.

The trade-offs between closely related methods may be relatively straightforward to analyse because of the shared normative principles, but as I have already remarked many alternative decision-making frameworks differ quite fundamentally in their underlying rationale. One way to deal with conflicting views on the appropriate decision process is to try to judge processes according to a common set of criteria, even if the original motives for the processes differ. If, for instance, using Citizens' Juries leads to higher preference satisfaction than say contingent valuation (which rests on a goal of preference satisfaction), then on purely instrumental grounds one can judge the former as superior to contingent valuation without having to debate the relative merits of the underlying philosophies. Thus we may be able to produce at least an incomplete ordering of decision processes even in the absence of agreement about deeper philosophies about how decisions should be made.

Figure 11.1 summarises many of the alternative means by which decisions can be made in an environmental context. As can be seen, methods vary to the extent that they are base themselves on consumer preferences and on the extent to which an attempt is made to estimate the a total figure for valuation. Strong sustainability for instance, gives little weight to preference. Meanwhile, contingent valuation does attempt to estimate a total value, but many other methods such as the travel cost technique, dose response and hedonic pricing are partial valuation methods, in that by design they omit some aspects of valuation, such as non-use value. In what follows I compare different processes according to the degree that they elicit a reliable figure for the Hicksian consumer surplus. Other criteria could be used, but given consumer surplus is at the heart of the cost-benefit approach and its associated controversies, it makes sense to use it as the vehicle for comparing decision processes. In order to motivate the comparison further, I also use three processes that have widespread use and advocacy.

1. Contingent valuation.
2. Partial valuation.
3. Citizen's Jury, an example of deliberative methods.

Contingent valuation is perhaps the most widely used non-market valuation method and possibly the most controversial (e.g. Diamond and Hausman, 1994). Meanwhile, partial valuation methods (see Fig. 11.1) are processes that by intent

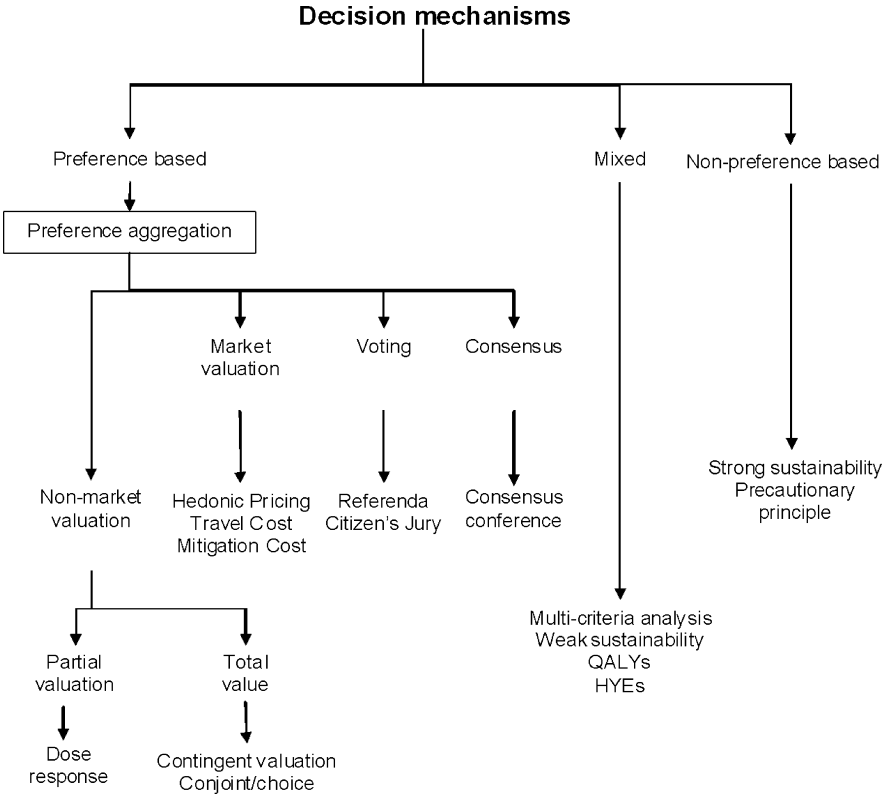


Fig. 11.1 Decision Methods

do not seek to capture all of consumer surplus. They might do this in one of three ways:

1. Restrictions on subjects. The researcher may for instance only gather preference data on actual users of a resource. This is typically the case with techniques such as the travel cost method.
2. Restrictions on aspects of valuation. The user/non-user distinction is fairly obvious, but a process might also seek to identify only some part of use value and ignore other contributions to welfare. For instance, the damage done by pollutants might be estimated purely from data on mortality, ignoring morbidity, the damage to plant life animals and the built environment, as well as possibly harder to measure costs such as the reduced amenity value from poor visibility.
3. Restrictions on preference variability. One argument for such an approach is that, by imposing this restriction on elicitation procedures, researchers simplify the cognitive requirements for respondents.

As an example of the last kind of partial valuation method, consider Quality adjusted life years, or QALYS – a widely used tool in health economics, especially

within cost effectiveness analysis or cost utility analysis (see Torrance, 1985). According to the QALY method, the benefits of different treatments are measured in terms of the utility of the resulting health status. More specifically, consider a profile q = (q_1, q_2,..., q_n) where a typical element q_i represents the quality of an individual's health in year i. For instance, q_i may include pain and discomfort caused by hip joint problems or recurring bouts of angina. An individual is called a discounted QALY maximizer if in choosing between different profiles, he or she acts as if maximizing, $V(q) = \sum v(q_i)r_i$ where the r_i represent discount rates for the ith period. In the QALY method the values of different treatments are compared: the one with the highest QALY per monetary unit is then chosen. QALYs are estimated using either the time method or the standard gamble technique. In the latter, the subject is asked to state a probability p such that he or she is indifferent between the status quo health state, q_s, for n years and the alternative of risky treatment which may result in total cure, with probability p or may cause death, with probability (1–p). For simplicity the value of perfect health v(h) is usually normalised to 1, while the value of death v(0) is usually set equal to zero, in which case p solves,

$$V(q) = \sum_i v(q_i)r_i = v(q_s)\sum_i r_i = p\sum_i v(h)r_i + (1-p)\sum_i v(0)r_i = p\sum_i r_i \tag{11.1}$$

Hence p = v(q_s), giving the value of the current health state. In the alternative, time method, the subject is asked to state an equivalence between an anticipated n years of the status quo health state and T years of perfect health. If the individual is a QALY maximizer, in the sense that r_i = 1 for all i, the time approach yields,

$$V(q) = \sum_i v(q_s) = nv(q_s) = Tv(h) = T \tag{11.2}$$

So v(q_s) = T/n.

As Culyer (1990), points out the use of QALYs may have normative justifications outside the standard welfarist framework. However, from within the welfarist viewpoint, they can be seen as an approximate means of measuring welfare in particular cases, a pragmatic compromise between practicality and theoretical validity.[2] Only if the subject does not discount will the time method produce a figure for v(q_s) that is theoretically consistent with rational choice theory. And the two QALY methods will only produce the same answer if the agent does not discount and if he or she is risk neutral in v. Each of these assumptions is heroic, as is the broader assumption that preferences are additively separable in health and other goods. Without this assumption, values would depend on income and preferences towards other goods.

[2] Of course to be part of a full cost-benefit analysis an exchange rate between money and QALYs must also be established (see Chilton et al., 2003 for a recent exploration).

Possibly, the most fundamental problem with QALYSs as an approximate guide to welfare is that the methods used to elicit them are the same methods which, in other contexts, have been shown to produce significant framing effects. The standard gamble and the healthy years equivalent methods are both forms of matching, in which the subject adjusts the level of one attribute until equivalence is produced between the two options on offer. Using the standard gamble technique produces many of the anomalies (e.g. the common ratio effect) familiar from the study of risky choice. Meanwhile, the time method may produce time inconsistencies and comparison of the two methods yields preference reversals (Wakker et al., 1997).

Finally in our list of alternative decisions processes, a Citizens' Jury is an example of a deliberative method (Bohman and Rehg, 1997), which is a process where citizens are actively and publicly involved in decision-making in a manner that encourages reason and argument. The jury method involves placing a small group of ordinary people together over a period of hours or days and asking them to reflect upon a policy issue. Typically, the jurists receive written and oral information from experts and get the chance to interrogate witnesses about the basis of their evidence.

Citizens' Juries originate in the USA (Crosby, 1991), but have since spread to many parts of the globe (Coote and Lenaghan, 1997) and are used in environmental decision-making (e.g. Aldred and Jacobs). Employed by public agencies as well as NGOs, much of the justification for their use arises from the notion that in a functioning democracy a good decision is defined by the process by which the decision is made as much as by the match between means and ends created by the final choice. However, at least one potentially important justification of Citizens' Juries is that they enable information to be transmitted and understood in such a way that judgement biases are eliminated. Blamey et al. (2002), for instance argue strongly that juries are less likely to provoke yea-saying, strategic behaviour, protest voting and biases created by social desirability. It follows therefore the Citizens' Juries are at least potential alternatives to methods such as contingent valuation even when viewed through the lens of welfarist principles.

The plan for the remainder of the chapter is as follows. In the next section I outline the basic methodology of comparison and illustrate the problem of comparing decision processes using the example of elicitation effects in contingent valuation. Following that some simple formulae are derived for comparing the performance of different processes. These formulae are then used to carry out some illustrative calculations. In the final section, a summary is offered, along with some thoughts on implications.

11.2 A Formal Approach to Optimal Decisions

Consider a decision-maker who must estimate an unknown parameter, μ, using some estimator $\widehat{\mu}$. Acting on $\widehat{\mu}$ rather than μ produces a loss, L, for the decision maker in the sense that a less than optimal decision may be made. The loss is zero if $\mu = \widehat{\mu}$, but otherwise positive and increasing in the difference between $\widehat{\mu}$ and μ. A conventional model for the loss function is that the costs of a mistaken decision are

proportional to the square of the difference between the true value and the estimate. That is,

$$Loss = aE\left[(\widehat{\mu} - \mu)^2\right] = a.var\,(\widehat{\mu}) + a\,(bias\,(\widehat{\mu}))^2 = aMSE\,(\widehat{\mu}) \qquad (11.3)$$

where $var(.)$ is the variance, $bias(\widehat{\mu}) = E\,[(\widehat{\mu} - \mu)]$ and MSE is the mean squared error. In practice loss functions may not be symmetric, convex or continuous in the manner implied by the quadratic formulation, but if costs do have this form, then it follows that if the optimal estimator minimizes the expected loss from the decision, then it also minimizes the mean square error. Consequently, the optimal estimator is not necessarily unbiased. Figure 11.2 illustrates the issue, showing two probability distributions for estimators: one estimator (shown by the solid line) is unbiased, but because it has a larger variance, the probability that is closer to its true value is higher with the alternative estimator (shown by the broken line), despite the fact that the latter is biased.

When the parameter estimate is an aggregate measure of preferences, estimated from data obtained from a sample of the population affected by a project, variance in can arise in two ways: first, through a combination of sampling and variability of the underlying preference parameter in the population. Secondly, variance can arise through *framing variation*.

As defined here, we can think of frame variation arising from two distinct sources. On the one hand the experimenter typically does not have the time or the resources to examine all aspects of his or her elicitation method. Many features of the design will be pre-tested through piloting and focus groups and the past experience of other researchers provides a stock of guidance, but there may also be aspects of the design (or the interaction of the design and the environment in which it is used) which can elude the researcher. The variables which determine values are almost countless: particular payment vehicles may be more or less acceptable; data and questions early in the survey might provide anchors for later valuation questions; the attitude and dress of interviewers can also influence subjects. Thus

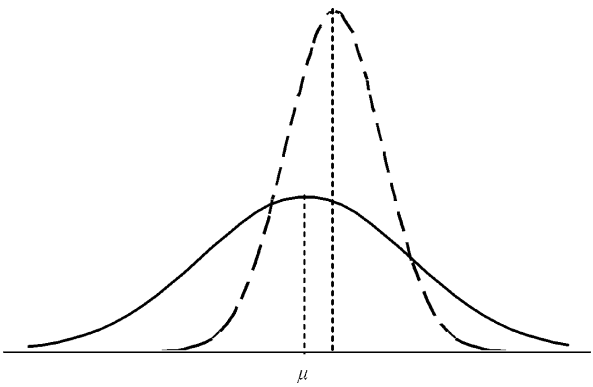


Fig. 11.2 Optimal Estimators

even in the best of surveys, there will be some residual uncertainty about the exact frame elicited by the researcher.

Secondly, recall that the researcher is attempting to identify a true valuation of the object. In order to do so, there must be a model of the relationship between the frame and the true valuation of the resource. Running a revealed preference study in parallel with the hypothetical exercise can give insight into the correct frame, but the nature of non-market valuation is that this is impractical for many goods. Without revealed preference behaviour therefore, the best frame represents something of a guess (albeit calculated) on the part of the researcher and there may be conflicting arguments what frame should be used. For instance, suppose that a proposed project involves environmental degradation of some kind, so that viewed from a property rights perspective willingness to accept compensation (WTA) is the appropriate measure of lost value. But as is widely noted in the contingent valuation literature (e.g. Mitchell and Carson, 1989), WTA appears more prone to producing unreliable estimates of value, compared to willingness to pay (WTP) and in cases where through repetition and learning the WTA-WTP gap is closed, it appears that most of the change occurs in the WTA estimate (Shogren et al., 1994). This suggests that the figures obtained from a one-shot WTP measure may be a better estimate of true WTA than a one-shot WTA figure. With arguments on both sides the researcher may consider both WTA and WTP acceptable frames for eliciting preferences over the environmental degradation.

Most of the possible source of framing effects will be familiar from earlier chapters. In fact, many of the anomalies already discussed in this book were first discovered or explored within the context of cost-benefit analysis, particularly where researchers have sought to explore the robustness of the many methods used for valuing goods outside the market-place. The most notorious example is the gap between WTA and WTP examined in Chapter 2, but it is still worthwhile mentioning two more areas of controversy, both of which have lead to widespread queries about the usefulness of a cost-benefit analysis that is based on stated preference.

11.2.1 Part-Whole and Scope Effects

A part-whole effect is said to occur when the valuation placed on the parts of a good sum to more than the valuation of the whole. A prime example of this is shown by the experiment conducted by Bateman et al. (1997b), in which vouchers for parts of meal at a Norwich pizza restaurant were bought from and sold to student subjects using a Becker-de Groot Marschak mechanism. Some of the WTP results are summarised in Fig. 11.3. Vouchers were offered for both the main course and for dessert plus coffee. Subjects were asked their willingess to pay to acquire both vouchers (i.e. the whole meal). Alternatively they were asked to value one voucher and the other voucher conditional on owning the first. Subjects endowed with no vouchers, for instance, were willing to pay on average £3.00 to acquire both vouchers. Meanwhile the mean wtp for a main course voucher was £2.18 and the

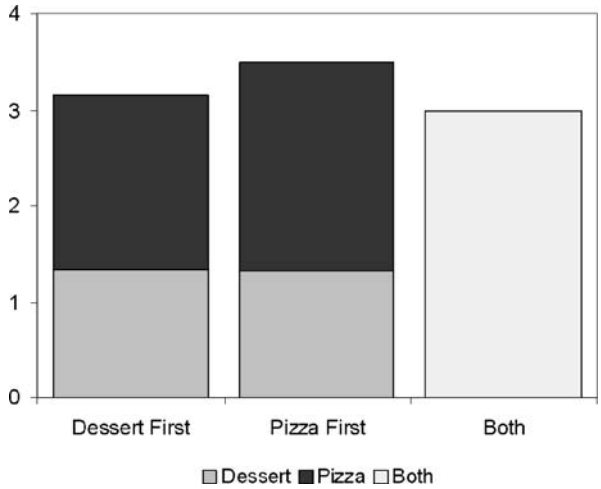


Fig. 11.3 Part-Whole Effects in Pizza Valuation

mean WTP for a dessert plus coffee voucher given the subject already held a pizza course voucher was £1.32 – meaning that on average subjects were willing to pay £0.50 more for the parts compared to the whole. This value is different from zero at a significance level of less than 0.1%.[3] The part-whole effect found in the pizza restaurant study may be an example of a wider phenomenon. Studies of decision in a variety of situations have found attribute-splitting and event-splitting effects. The former effect is a tendency for the valuation of a multi-attribute good to increase if an attribute on which it rates highly is re-described as two sub-attributes, e.g. if the 'performance' attribute of a high-performance car is split into 'acceleration' and 'top speed' (Weber et al., 1988). The latter effect is a tendency for a lottery which gives a prize with a particular probability to be valued more highly if it is described so that the prize is won in each of two low-probability events than if it is won in a single event (Starmer and Sugden, 1993).

Part-whole effects are one possible cause of scope effects – the lack of sensitivity of values elicited in contingent valuation to changes in the size or scope of the good in question. A classic, though controversial example of a scope effect is provided by Kahneman and Knetsch (1992) telephone survey of values for the preservation of Ontario lakes, where the experiments found no difference between the average values elicited for 2 and 2,000 lakes. While part-whole effects imply the rejection of Hicksian preferences, scope effects are more of an example of an extreme value anomaly. In other words, it is perfectly compatible with rationality for an individual to be (nearly) satiated with a good and therefore to show only a small increase in the valuation following a large increment to the good. Kahneman and Knetsch (1992), proposed that the lack of sensitivity to scope was due to the 'warm glow'

[3] The second difference shown in the figure is not significant, but the two WTA discrepancies were significant. It is worth noting that to compensate for potential income effects, subjects endowed with a voucher for a part were endowed with less money than those given no vouchers.

produced in the minds of respondents when asked to make a contribution to a public good. The absence of sensitivity to scope has been seen as a mark of a weak design. As Carson et al. (2001, p. 183) say 'poorly executed survey design and administration procedures appear to be a primary cause of problems in studies not exhibiting sensitivity to scope.' Nevertheless, as the part-whole paper illustrates, a lack of coherence in underlying preferences may also be a source of the apparent anomaly.

11.2.2 Elicitation Effects

Elicitation effects occur when responses gathered from subjects are sensitive to the method of elicitation in a manner inconsistent with standard, Hicksian consumer theory. Two widely reported elicitation effects are starting point effects and yea-saying. Starting point effects (SPE) occur when reported valuations are correlated with some initial valuation cue, such as the bid value in dichotomous choice (DC) questions. Yea-saying describes the phenomenon of a subject agreeing to a proposal in the form of a direct question that she or he would reject under other conditions. For instance, a subject may agree to a bid price in a dichotomous choice format but then provide a lower stated valuation in a subsequent valuation exercise. A key difference between the two elicitation effects is that yea-saying is a unidirectional phenomenon, i.e., it raises willingness to pay or reduces willingness to accept whereas starting point bias can work in either direction depending on the value of the cue.

Starting point effects are an example of anchoring discussed in Chapter 3 where it was presented as an example of an information processing anomaly. In the context of valuation, anchoring occurs when an individual's reported or revealed valuation is correlated with some prior numerical cue. Since its preliminary identification by Slovic and Lichtenstein (1971) manifestations of anchoring have been identified in numerous and diverse settings including the guessing of answers to multiplication problems and estimating the number of African countries in the United Nations (Kahneman et al., 1982). And as we saw, a particularly stark example of anchoring can be found in the recent work of Ariely et al. (2003), who asked subjects for the final two digits of their US social security number and found that it was closely correlated with individuals' subsequent valuations of a variety of goods.

The iterative bidding format of CVM gives subjects a sequence of opportunities to give information about their valuation of a good. For instance, the experimenter may ask the subject if they are willing to pay x (the 'starting point') for an improvement in a public good and then, in a follow-up question, ask the subject to state a maximum willingness to pay. Intended as an improvement upon the open-ended format (OE), the iterative bidding CVM revealed that final open-ended valuations were often correlated with the initial value of x – hence the term, starting point effect (Rowe et al., 1980; Brookshire et al., 1982; Boyle et al., 1985; Kealy et al., 1988; Boyle and Bishop, 1988b; Cameron and Quiggin, 1994; Bateman et al., 1995; Herriges and Shogren, 1996).

Anchoring is one possible reason for SPE, with the anchor provided by the initial value of x offered to subjects. The initial bid value might act as a clue or hint towards the good's value, especially when respondents are confused or unfamiliar with the good concerned (Brookshire et al., 1982; Bishop et al., 1983; Kealy et al., 1988; McFadden, 1999; Brown et al., 1996). Since the domain of CVM often involves the valuation of unfamiliar, non-marketed goods, this starting point problem has become recognised as a potentially serious flaw inherent in iterative bidding techniques (Boyle et al., 1985, p. 193).

The DC protocol offers an alternative to bidding games. It has been suggested that this approach might simulate a more market like setting, since it involves a simple accept-reject bid decision. The increased simplicity might help subjects to feel more comfortable answering a CVM valuation question (Sellar et al., 1985). Furthermore, CVM surveys are typically limited by time, information, interaction, consultation and market experience all of which may cause subjects to become rather uncertain about their responses. If subjects only have to think whether their true value is above or below some suggested amount, the whole valuation process may be simplified (Bishop et al., 1983). Comparisons between OE and DC formats have shown that OE mean and median estimates (of WTP) are consistently lower than the DC estimated means (Sellar et al., 1985; Bishop et al., 1983; Cummings et al., 1995; Brown et al., 1996; Ready et al., 1996; Boyle et al., 1996; Bateman et al., 1995). One view is that these results demonstrate that OE questioning is typically subject to strategic behaviour. A second explanation is that just as in iterative bidding processes, anchoring may occur in the DC format.

Yea-saying provides an alternative explanation of the results. This phenomena is documented in the existing psychology literature (Arndt and Crane, 1975; Crouch and Keniston, 1960), and has also been proposed in the economics CVM literature (Kriström, 1993; Kanninen, 1995; Brown et al., 1996) as a possible influence on DC responses.

Brown et al. (1996), argued that the simplicity of the take-it-or-leave-it choice might generate a conflicting objective in respondents. Torn between answering truthfully and showing a positive preference, if a DC bid is above her/his maximum WTP, a subject may still respond positively because s/he would like to demonstrate a positive preference for the good in question. In addition to this, we might also include the notion of the good respondent (Orne, 1962). Orne described how subjects, when faced with officialdom, might respond positively to questions, only because they wrongly believe that such a response is exactly what the interviewer (in a position of perceived authority) wishes to hear.

Although some results can be interpreted as evidence for yea-saying some caution is warranted. Point estimates from DC data which are used to compare with moments from open-ended data, have been found to be rather dependent upon the original specification of the bid function and the Hanemann (1984), specifications have not been wholly supported (Cooper and Loomis, 1992). Given that many of the previous studies cited above compare OE with DC estimates based upon these functional specifications, it would be unwise to place too much emphasis on these studies as evidence of yea-saying.

Another approach, thereby avoiding the functional form problem entirely, is to test using synthetic data sets or implicit preferences. The synthetic DC responses are constructed by allocating a 'yes' if an OE valuation is greater than or equal to those bids used in the actual DC questioning. Differences between actual and implicit responses can then be tested. Studies which have previously used these 'synthetic' data sets have revealed evidence to support the yea-saying tendency (Kriström, 1993; Holmes and Kramer, 1995; Bateman et al., 1995; Boyle et al., 1996; Ready et al., 1996). The studies cited produce evidence based on stated preference exercises with hypothetical goods. It is not clear whether the results can be replicated with individuals making real choices about real goods, hence the value of a controlled experiment.

Frykblom and Shogren (2000), use real choices and a split-sample design to value an environmental economics text using 108 Swedish university students. One treatment undergoes a Vickrey auction while the others face dichotomous choice (DC) questions set at a variety of bid levels. The authors argue that both yea-saying and anchoring will increase the acceptance of the proposal at high bid levels, while the two effects work in opposite directions for low bid levels. Hence it is possible to test between the impact of these two effects by comparing the distribution of values derived from the auction with the upper and lower parts of the distribution derived from the DC exercise. On the basis of their results, they conclude that yea-saying is present, but not anchoring. Alternatively, one could design an experiment in which for some subjects, after the DC questions there is an incentive compatible open-ended (OE) valuation question. If only anchoring occurs the values derived from open-ended questions should be consistent with the values from the DC questions, but if only yea-saying is present then the distribution of values derived from the OE questions should be independent of the bid level in the DC question and equal to the distribution obtained from subjects who face an open-ended question without a prior DC question. This provides a clear cut means of distinguishing between anchoring and yea-saying. Bateman et al. (2006), use this method, with students as subjects and teabags and wine as the commodities As the Fig. 11.5 suggests, they conclude in favour of yea-saying as the cause of the results. This diagram shows none of the properties of the stylised anchoring effect shown in Fig. 11.4. Instead, the cumulative percentages bidding at each level (in pounds) appear highly similar for all three treatments. Such results are at variance with the work of Ariely et al. (2003), who admittedly use less familiar goods. On the other hand they are consistent with Frykblom and Shogren (2000) and with the recent evidence on US consumption presented by Van Soest and Hurd (2003). Partly the difference may lie in the goods involved. In our experiment and in the last two studies cited, familiar goods were the objects of valuation, whereas anchoring effects seemed to have been found most clearly when subjects were facing novel or unfamiliar valuation tasks, which is often the case in environmental valuation.

The deeper conclusion from this literature is that even in an incentive compatible environment with well-trained and experienced subjects, real goods and simple choices, different frames produce different valuations. Thus the researcher cannot be entirely certain about which method of elicitation is the correct one.

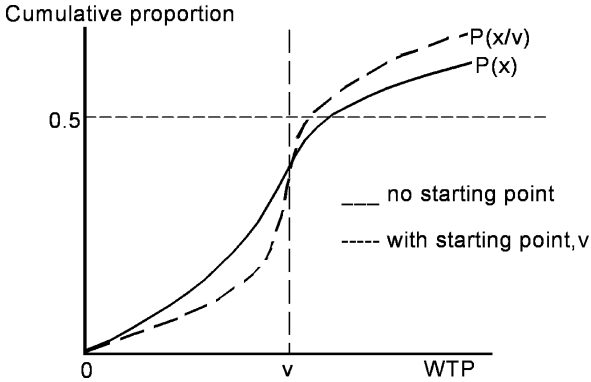


Fig. 11.4 The Effect of a Starting Point on Valuations

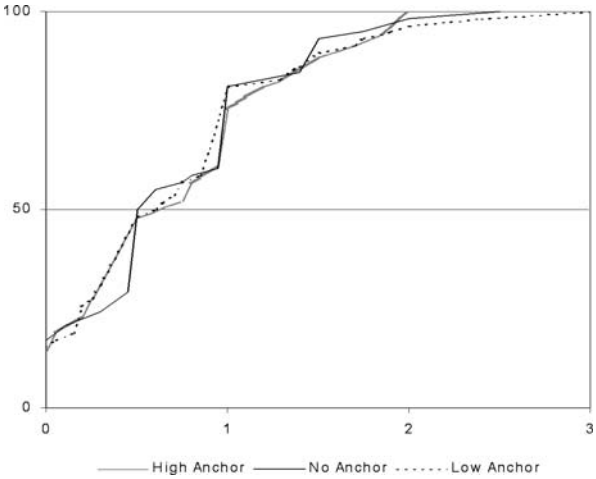


Fig. 11.5 WTP for Teabags

11.2.3 Responding to Anomalies

We saw in Chapter 5 that there are a number of possible responses to anomalies. It is worth revisiting some of the options. One approach I argued against was switching horses in mid-race i.e. to invoke some alternative value system such as maximizing opportunity or, to take an example from Fig. 11.1, strong sustainability. After all, the practical difficulties of implementing the normatively optimal framework do not imply that previously rejected alternatives become morally superior. However, the alternative approaches may have superior instrumental value an idea pursued below. The second approach discussed was the managed retreat. When Diamond and Hausman (1994), asked 'is some number better than no number' they responded that very often with stated preference it was better to have no number. The key questions for the managed retreat is proponent 'when is no number better?' and 'how then should decisions be made'. A third response was to agree a golden standard for preference elicitation. Anomalies are dealt with by rejecting methods which produce

contradictory evidence on preferences. Since this latter approach has been influential within non-market valuation, I shall devote more space to it here.

11.2.3.1 Gold Standard Frames

Researchers exploring new methods for valuing non-market goods have sought to establish ex ante and ex post tests of their acceptability (Mitchell and Carson, 1989; Bishop, 2003), which, within environmental economics, has produced three notions of validity: content, construct and criterion.

- Content validity refers to whether the procedures used are appropriate to the valuation task.
- Construct validity is typically split into two aspects: theoretical validity refers to whether the results of the valuation exercise match prior hypotheses derived from economicy theory – for instance, is valuation increasing in income or is more apparently preferred to less. Convergent validity on the other hand, holds if the results concur with those obtained by other approaches. So if a hedonic approach and a travel cost method yield similar figures for the value of a seaside beach, there would be convergent validity.
- The final element is criterion validity which is a test of whether the results match those obtained by a method already certified as acceptable. Comparisons of hypothetical wtp with figures obtained from real second-price auctions represent examples of criterion validity, insofar as real choices represent a widely accepted test of value. Meanwhile the examples of part-whole bias and starting-point effects represent failures of theoretical validity if one accepts that consumer behaviour should not exhibit such properties.

These notions of validity represent standards for the way the valuation exercise is framed. With content validity the emphasis is on ex ante tests of whether for instance the good to be valued is described accurately and the information clear to respondents. With the other validity criteria, the judgement on the frame is ex post – its suitability is scored according to whether it delivers valuation figures in accordance with theory, other methods and those produced by widely accepted techniques. A gold standard or optimal frame is therefore one where the manner in which preference information is elicited is specified precisely and which passes all tests of validity. Methods which meet the gold standard produce acceptable information, but methods which do not meet the standard require the rejection of information or at least a significant reduction in its value. For economists, especially those involved in environmental valuation, guidance on what constitutes an optimal or gold standard frame is available in large quantities. Within the contingent valuation approach, Mitchell and Carson's (1989) comprehensive book is probably still the dominant source of framing advice, supplemented by the guidelines produced by the NOAA panel under the chairmanship of Kenneth Arrow (Arrow et al., 1993) which produced what it called 'a fairly complete set of guidelines compliance with which

would define an ideal CV survey'.[4] More recently some further advice has been put forward by Carson (1997), amongst others and there are a number of environmental valuation manuals being produced or available for government economists (e.g. Bateman et al., 2002).

Since they have been so influential in shaping the recent development of contingent valuation it is worth considering some of the details of the NOAA panel recommendations (see Table 11.1). Some of them are clearly particular to environmental economics and some of them reflect areas where, in the view of the panel, more research was required. In addition to the specific recommendations in the table, there is a general requirement that,

> In the absence of a set of reliable reference surveys, the burden of proof shall remain on the researchers. A survey should be judged unreliable to the extent it exhibited the following defects:
>
> - a high non-response rate
> - inadequate responsiveness to scope
> - lack of understanding of the task by respondents
> - lack of credibility of the restoration scenario" (Arrow et al., 1993, p. 37, bullets added).

Carson (1997), is a sensible overview of some general principles that emerged in the wake of the NOAA panel. For him, fundamentally a well-conducted CV exercise must have three attributes: the good to be valued must be well-defined; there must be plausible means of provision and there must be a plausible means of making trade-offs. A good that is vaguely defined by the surveyors may be understood in widely differing ways by subjects in the survey. To aid the credibility of the exercise with regard to the payment vehicle, Carson recommends coercion where this is credible. Ideally, the vehicle should also appear reasonably fair, as well as plausible and understandable to the respondents in the survey. In addition to these features of the survey frame, the sampling process should be carefully designed so that population estimates of value can be derived; statistical methods should allow for the possibility of zero willingness to pay values and the whole procedure should be transparent.

Transparency is also a feature advocated by the practitioners who argue that preferences are not always in place ready for the survey analyst, but must be constructed by the individual respondent. In other ways the recommended methodology is fundamentally different. Gregory and Slovic (1997), for instance argue that

> the analyst therefore functions as an architect, helping respondents build their values from simpler pieces rather than (following the economics model) as an archaeologist whose task is to uncover values presumed to exist. (Gregory and Slovic, 1997, p. 177).

Typically, options differ along many dimensions. For some dimensions, such as environmental damage or risks to health, subjects may be unused to expressing values in terms of a single metric such as money. They may also be cognitively

---

[4] It is worth noting in passing that the guidelines prompted a major research agenda for environmental valuation and current practices depart from the guidelines in a number of ways. For instance, choice experiments are probably more commonly proposed now, rather than referenda.

Table 11.1 The Optimal Frame – NOAA Panel Recommendations

General conduct
Probability sampling.
Make efforts to minimise non-responses.
Generally use face to face interviews.
Pre-test for interview effects.
Transparent reporting of the study.
Carefully pre-test the survey instrument and provide evidence of understanding from the final respondents.
The Specifics of contingent valuation
Conservative design – one which is more likely to produce under-stated WTP rather than over-stated figures.
Use the WTP figure (because it is more conservative) rather than WTA.
Use a referendum format.
Describe the policy accurately.
Pre-test photographs.
Remind participants about substitutes.
Check for the temporal stability of the responses.
Provide a 'no-answer' option, with a diagnostic follow-up to elicit the reasons for the refusal.
Also follow-up questions on the yes or no responses, with questions about motives.
Ask questions on potential co-variates, such as income, prior knowledge and interest, attitudes towards the environment, distance from the site, understanding, credibility of the scenario.
Check for understanding.
Remind subjects about alternatives for spending income.

Source: Arrow et al., 1993

overloaded if faced with choices between two complex options which differ in many important ways. Consequently, they argue that subjects should be first asked to compare alternatives along single dimensions, using valuation scales natural to the dimension. Only later should subjects be asked to make trade-offs between different dimensions and invited to attach weights. This they argue, limits the cognitive overload that might otherwise hinder subject's ability to choose between alternatives.

Unlike the typical CV instrument which seeks to gain values from a representative sample of the population, Gregory and Slovic's constructive approach uses only small numbers of key stakeholders. They recommend a five-step approach to valuation:

1. First the problem should be structured. Subjects should be given all relevant information about the options and the key values or objectives should be identified, possibly by the stakeholders themselves. These objectives should not just reflect the immediate ends of the issue being considered, but fundamental objectives too.
2. Once the objectives have been identified, achievement of them should be measured in some ways, either in general available scales, such as profit or mortality or through specially constructed scales.

3. Next the stakeholders should be asked to consider trade-offs between objectives. Subjects should consider appropriate weights for the scales.
4. In the fourth step, stakeholders should compare the options on offer using the weighting system devised in the previous step. Gregory and Slovic also recommend that, at this stage new alternatives can be offered to increase the acceptability of the 'best' option to the widest range of stakeholders.
5. This leads into the final step in which sensitivity analysis is carried out and the facilitators search for consensus amongst the participating stakeholder groups and suggest areas for further research.

It is clear that this approach differs in a number of important ways from the standard CV method, but which ever way it is pursued, there are two fundamental problems with the golden frame perspective on valuation – as we saw in Chapter 5. First the guidance may not be exact enough, in the sense that the researcher however well-funded is not in a position to control exactly the perception of the survey and its purpose in all respondents. The impact of changes in the way information is presented visually; or changes in elicitation formats, or in the order in which questions are asked may all be tested by the researcher, but some aspect of the survey design or its implementation is bound to escape the attention of the researcher or may interact with the target sample's experience in an unanticipated manner. In other words there is always some residual variance in the frame elicited in the mind of the subject and this is one source of the frame variance discussed above.

Secondly, in drawing up the definition of the gold frame, there is often no definitive reason for favouring a frame with aspect x rather than aspect y. 'Each of the three main response formats has strengths and weaknesses' states Boyle (2003, p. 137), referring to the choice between open-ended, dichotomous choice and payment cards in a recent guide to the contingent valuation method. The dichotomous choice elicitation method is often recommended (e.g. by the NOAA panel, 1993), largely on the grounds of its perceived incentive compatibility. However, the bid levels provided to subjects in DC may provoke starting point effects and therefore responses at variance to individuals' true preferences (if they have them). Dichotomous choice is also inefficient in that it produces very little information per subject. In response to the anchoring problem, one might use payment cards – which is where a subject sees a card or screen with a set of bid levels and is asked for his or her preferred option. In response to the inefficiency of the DC method, there has been a recent trend towards conjoint methods, in which subjects face a sequence of choices, often between more than two options. Both the alternative elicitation formats lack the clear incentives for truth-telling provided by the DC method, but they clearly have other advantages and it is not obvious where the balance lies. This then is the second source of frame variance: the set of acceptable frames contains more than one element.

11.3 Formal Comparison

Processes are compared according to their mean squared error , that is the sum of the variance and the square of the bias, (Munro, 2006). Suppose that there are F acceptable frames (with F > 1) from within which one frame is selected at random. The estimator is a random variable, $\widehat{\mu}$, which in frame j, we label as $\widehat{\mu}_j$ j=1,. . .,F, with an expected value of $\overline{\widehat{\mu}}_j$ within frame j and an expected value across frames of $\overline{\widehat{\mu}}$. In other words,

$$\overline{\widehat{\mu}} = \frac{1}{F}\sum_{j=1}^{j=F}\overline{\widehat{\mu}}_j = \frac{1}{F}\sum_{j=1}^{j=F} E\left[\mu_j\right] \tag{11.4}$$

The *within frame* variance of $\widehat{\mu}_j$ is $\sigma_j^2 = E\left[\left(\mu_j - \overline{\widehat{\mu}}_j\right)^2\right]$, with the expected value of it across frames defined as $\sigma^2 = \frac{1}{F}\sum_{j=1}^{j=F}\sigma_j^2$. Meanwhile, the *frame variance,* denoted by σ_f^2 is the variance across the frames of the mean estimates. That is,

$$\sigma_f^2 = \frac{1}{F}\sum_{j=1}^{j=F}\left(\overline{\widehat{\mu}}_j - \overline{\widehat{\mu}}\right)^2 \tag{11.5}$$

We let n be the sample size, with individual observations indexed by i = 1,. . ., n. Each random variable within this sample is assumed to be i.i.d. We let $\widehat{\mu}_{ji}$ be the ith element of the sample when the frame is j. The total variance of the process, $var\left(\widehat{\mu}\right)$, is then

$$\begin{aligned}
var\left(\widehat{\mu}\right) &= E\left[\left(\sum_i\left(\frac{\widehat{\mu}_i}{n}\right) - \overline{\widehat{\mu}}\right)^2\right] = E\left[\left(\sum_i\left(\frac{\widehat{\mu}_i - \overline{\widehat{\mu}}}{n}\right)\right)^2\right] \qquad (11.6)\\
&= E\left[\frac{1}{F}\sum_{j=1}^{j=F}\left(\sum_{i=1}^{i=n}\left(\frac{\widehat{\mu}_{ji} - \overline{\widehat{\mu}}}{n}\right)\right)^2\right]\\
&= \frac{1}{F}\sum_{j=1}^{j=F} E\left[\left(\sum_i\left(\frac{\widehat{\mu}_{ji} - \overline{\widehat{\mu}}_j + \overline{\widehat{\mu}}_j - \overline{\widehat{\mu}}}{n}\right)\right)^2\right]\\
&= \frac{1}{F}\sum_{j=1}^{j=F}\left(\frac{\sigma_j^2}{n} + \left(\overline{\widehat{\mu}}_j - \overline{\widehat{\mu}}\right)^2\right)\\
&= \frac{1}{F}\sum_{j=1}^{j=F}\left(\frac{\sigma_j^2}{n} + \left(\overline{\widehat{\mu}}_j - \overline{\widehat{\mu}}\right)^2\right)\\
&= \frac{\sigma^2}{n} + \sigma_f^2
\end{aligned}$$

The penultimate line of this expression arises from the i.i.d. nature of the sample. The formula for MSE is therefore:

$$MSE = \frac{\sigma^2}{n} + \sigma_f^2 + \left(\widehat{\mu} - \mu\right)^2 \tag{11.7}$$

Comparing processes is then a matter of comparing the MSE for different decision-making approaches. In theory, MSE may be lowered by splitting the sample across frames. Such a strategy may be feasible when it is the researcher who has full control over the frame. However when there is residual uncertainty over how subjects will interpret the valuation process, then it may not be in the power of the researcher to choose the exact frame.

Each of the processes selected for the comparison has its strengths and weaknesses which we need to formalise. Most jury exercises begin by recruiting a 'representative' sample of individuals. To be manageable, Citizens' Jury have to be small, so there is a potential problem of high variance due to the small sample of individuals involved. For instance, the Jefferson Institute argue in favour of a jury of 18 people. Unless preference variability across the population is extremely small, sampling variability with deliberative processes is likely to be large. In the guidelines for the original US juries, one of the essential elements was random sampling for members.[5] Let n be the size of the Citizens' Jury. When the jury convenes, n individuals arrive at a collective value for the mean value. Suppose that, by conversation, they arrive at the correct frame. I shall call this an *ideal* Citizens' Jury (ICJ) and further suppose that the mean value produced by the jury is the mean of their individual values. If the n members are chosen randomly, then the expected value obtained from this process is also μ.

With contingent valuation on the other hand judgement biases are not necessarily eliminated within any given frame, but samples are potentially large enough to limit sampling variability and the set of acceptable frames is ideally chosen so that there is no ex ante bias.

For the PV method the benefits can be decomposed. For the purposes of the exercise it does not matter what these components are, only that they can be conceptually separated in what follows. Let $\mu = \mu_e + \mu_n$ where the labels of the components are chosen to indicate 'estimated' and 'non-estimated' values respectively. For the partial valuation method only μ_e is estimated. A partial valuation method may be chosen because the part of the benefits estimated are not subject to significant framing effects[6] – in other words the frame variance is zero, but by definition it will omit the values of some users or uses or as in the case of QALYS, force data on preferences into a very specific functional form. As a result we can

[5] The process of juror recruitment has often been rather less formal (Aldred and Jacob, 2000). In some cases volunteers are recruited by word of mouth or advertisement.

[6] This is the underlying claim about the superior value of revealed preference methods.

expect systematic bias. However, as with contingent valuation, the sample may be chosen to be sufficiently large and random to all-but eliminate sampling variation.

Formally therefore, I define each of the methods to be compared as follows:

Definition 11.1 Contingent valuation (CV): $n = \infty, \overline{\widehat{\mu}} - \mu = 0$.

Definition 11.2 Ideal Citizens' Jury (ICJ): $\overline{\widehat{\mu}} - \mu = 0;\ \sigma_f^2 = 0$

Definition 11.3 Partial valuation (PV): $n = \infty,\ \overline{\widehat{\mu}} - \mu = -\overline{\mu}_n < 0;\ \sigma_f^2 = 0$

With these definitions we have:

$$\begin{aligned} MSE_{CV} &= \sigma_f^2 \\ MSE_{ICJ} &= \frac{\sigma^2}{n} \\ MSE_{PV} &= \left(\overline{\widehat{\mu}} - \overline{\mu}\right)^2 = \bar{\mu}_n^2. \end{aligned} \tag{11.8}$$

The optimal decision process is then the one with the lowest MSE.

Define $\alpha = \sigma_f/\mu$ as the *coefficient of frame variation*, let $\beta = \sigma/\sqrt{n}\mu$ be the *coefficient of sample variation*with $B = \sigma/\mu$ as the *coefficient of population variation* and let s $= \bar{\mu}_n/\mu$ be the share of total benefits omitted by the partial valuation method. Then, CV is as least as good as PV when,

$$\alpha \leq s \tag{11.9}$$

Meanwhile CV is at least as good as ICJ provided,

$$\alpha \leq \beta \tag{11.10}$$

Finally, PV is as least as good as ICJ, when

$$s \leq \beta \tag{11.11}$$

Figure 11.6 summarises the circumstances under which each decision rule is superior to the other two. We have three regions. As the proportion of total benefits omitted by the PV estimation technique becomes smaller, the PV method raises its advantage compared to the other two rules. Once the coefficient of frame variation is larger than s, then PV produces a lower MSE than contingent valuation. Similarly if the coefficient of sample variation is large compared to s, then PV is superior to the ideal Citizens' Jury. For the comparison between ICJ and CV, s is irrelevant – what matters is the relative size of the coefficients of variation. If the frame coefficient is larger then ICJ is superior to CV, but if the sample coefficient is larger then CV is the superior elicitation method. Thus for the comparison between CV and ICJ three things matter: the size of the citizen's jury sample; the variance of tastes within the population and the sensitivity of expressions of preferences to changes in the frame.

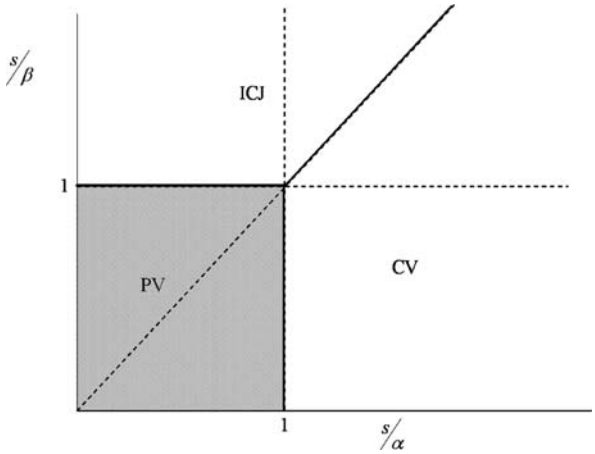


Fig. 11.6 A Comparison of Three Methods

The particularly simple formulae and diagram represent the result of some strong simplifying assumptions, particularly that ICJ and partial valuation methods do not suffer from framing variation. Adding in these elements would raise the relative advantage of the contingent valuation process. On the other hand, allowing for sampling variation in contingent valuation and partial valuation would lower their advantages relative to the ICJ.

11.3.1 Example 1: WTA Versus WTP

To illustrate the formulae we consider two examples. The first example supposes that there are only two possible frames, namely buying and selling and therefore two possible values: willingness to pay (WTP) and willingness to accept (WTA). There is doubt over the correct frame for valuation purposes, not because of uncertainty over property rights and the researcher believes that WTA and WTP are equally likely to be the correct way of eliciting valuation. Let WTA = $(1+\gamma)wtp$ then the mean value across frames is $(1+0.5\gamma)wtp$. The frame variance is $\gamma^2 wtp^2/4$.

The first table shows critical values of n (the size of the jury) when the CV approach and ICJ have equal MSE. In other words,

$$n = \frac{(2+\gamma)^2 B^2}{\gamma^2} \tag{11.12}$$

To determine the range of sensible values for the coefficients in this expression, recall the experimental evidence: the ratio of mean WTA to WTP typically exceeds 1 by a significant amount and ratios of four or more are not uncommon (Chapter 2 and Horowitz and McDonnell, 2002, or Sayman and Onculer, 2005).

Unfortunately, there is no systematic evidence on the coefficient of population variation from the CV literature, but a sample of CV studies is assembled in

Table 11.2 Evidence on the Coefficient of Population Variation, B

Authors	Method	Measure	Good	B
Halvorsen and Sœlensminde (1998)	OE	WTP	Reducing premature deaths	0.9
Adamowicz et al. (1993)	OE	WTA	Movie tickets	0.92
Bateman et al. (1995)	OE	WTP	Flood defences	1.70
Whittington et al. (1990)	OE	WTP	Water, public standpipe	0.54
Whittington et al. (1990)	OE	WTP	Water, private supply	1.32
Boyle et al. (1996)	OE	WTP	Oil spill clear up	0.34
Silberman et al. (1992)	OE	WTP	Beach quality (users)	1.37
Silberman et al. (1992)	OE	WTP	Beach quality (never will use)	1.83
Kontoleon and Swanson (2003)	OE	WTP	Panda reserves	1.05
Thayer (1981)	Bidding game	WTP	Landscape preservation	1.08
Amigues et al. (2002)	OE	WTA	Riparian habitat conservation	2.80

Notes: OE = open ended. Sample standard deviations used.

Table 11.2. Although this sample is not random or systematic and none of the studies claim to randomly sample members of the population, it suggests that a range for B from 0.5 to 2.5 is reasonable.

As can be seen in Table 11.3 the critical size for the Citizens' Jury is sensitive to the value of B. Proponents of juries suggest 16–25 members as a sensible range of numbers for an effective process, which would imply that despite the small numbers involved, the ICJ may outperform the CV exercise when it is not clear whether the optimal frame should present the policy changes as a loss or a gain and the variance of preferences across the population is small.

The final column of the table illustrates a different comparison: between contingent valuation and partial valuation for different values of the WTA/WTP ratio. In the column, s^* is the critical share of total benefits not estimated by the partial valuation method such that $\alpha = s^*$. In the example $s^* = \gamma/(\gamma + 2)$. From the column we can see that when WTA is only 50% bigger than WTP then the partial valuation method must estimate 75% of total benefits to be of equal value to the contingent valuation method. However, by the time the ratio of WTA/WTP is equal to 4, the partial valuation method can be superior even if it omits 2/3 of the benefits.

Estimates of s are also hard to come by. Typically when contingent valuation is compared to the travel cost method, for example, both procedures are partial valuations, since it is only visitors to the site who are sampled (Carson et al., 1996). One

Table 11.3 Critical Values of n for Equivalence of CV and ICJ

	Coefficient of population variation					
WTA/WTP ratio	0.5	1	1.5	2	2.5	Critical s*
1.5	6	25	56	100	156	0.20
1.75	3	13	30	54	84	0.25
2	2	9	20	36	56	0.33
2.5	1	5	12	22	34	0.43
3	1	4	9	16	25	0.50
4	1	3	6	11	17	0.60

source is the pioneering study on hunting permits by Bishop and Heberlein (1979), which included a travel cost exercise. For a zero cost of time the method produced estimates of \$159,000 for total benefits, which should be compared to \$880,000 obtained from the real purchase of permits off the hunters (i.e. a WTA figure). This suggests a rough estimate of $s = 0.82$, if we ignore possible differences between WTA and WTP. Meanwhile, when the zero cost of time figures are compared to hypothetical WTP elicited from a comparison group of hunters, $s = 0.46$. In the case of non-use values, Greenley et al. (1981), estimate 19–27% of benefits for improvements to water quality in a south Colorado, USA, river system represent existence value.[7] If non-use benefits were the ones omitted from a partial valuation (i.e.. $s = 0.19 - 0.27$), this would suggest that partial valuation outperforms a full valuation for figures within the standard range for the WTA/WTP ratio. In the context of healthcare, Clarke (2002), compares travel cost[8] and contingent valuation estimates for rural pregnancy services in Australia using a mail-based survey. He finds that estimates from the travel cost method are approximately 56% of those obtained via contingent valuation (i.e. $s = 0.44$). Meanwhile, Kennedy (2002) compares revealed preference and contingent valuation measures for radon protection and obtains a figure for the former which is 67% of the latter (i.e. $s = 0.33$). Again, making the heroic assumption that this difference is purely due to benefits omitted from the revealed preference methods it would suggest that partial valuation outperforms contingent valuation for many realistic values of the WTA/WTP ratio.[9] In the Bishop and Heberlein (1979), article referred to above, the WTA/WTP ratio for hypothetical valuation was 4.8, giving a critical value of 66% for s. The intermediate cost-of time travel cost estimate would therefore dominate hypothetical valuation, but an ICJ would easily dominate both given a value of B $\approx$ 1 (see Li et al., 1996 for this last figure).

Figure 11.7 offers a simultaneous comparison of all three approaches. For this illustration, I assume that $s = 0.4$ and that a jury of up to 25 is possible, which is fairly generous to the ICJ method. This is reflected in the figure, where we can see that none of the approaches has a decisive advantage over the other methods for typical parameter combinations. Comparing all three approaches (CV, ICJ and PV), contingent valuation has its clearest advantage when the WTA/WTP ratio is low (i.e.

---

[7] Cummings and Harrison (1995) argue forcefully that it is not possible to separate the values held by one individual into existence or non-use and use values.

[8] It is well-known that travel cost methods produce estimates for the Marshallian consumer surplus. For normal goods this underestimates the Hicksian compensating surplus. In theory if income and price elasticities are identified, the Hicksian surplus can be estimated. The point is this figure would still omit the non-use value for non-users and hence even the adjusted figure from travel cost studies will be a partial valuation measure.

[9] A more general expression for comparing PV with the other measures can be obtained if we suppose that the variance of the partial valuation method is some fraction δ of the CV variance. We then obtain $\alpha\sqrt{1-\delta} \leq s$ for the comparison between CV and partial valuation. To put this into some kind of perspective: if the frame variance for partial valuation is half the figure for CV, then the critical figures for s in the final column of Table 11.3 would be reduced by about 30%.

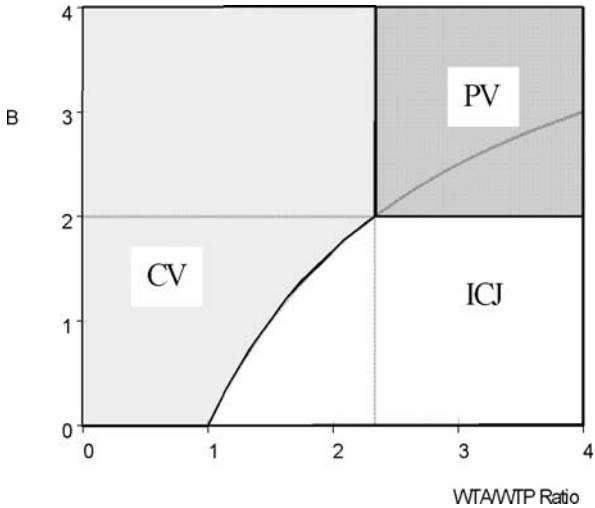


Fig. 11.7 Optimal Decisions for $s = 0.4$ and $n = 25$

close to 1), preference variation within the population is high and, at the same time partial valuation methods capture only a small percentage of the total value.

11.3.2 Example 2: Elicitation Methods

Table 11.4 provides an example where all the elicitation methods used produce a figure for WTP. In Bateman et al. (1995), two versions of a CV exercise are conducted on an aspect of the Norfolk Broads (a national park located near to Norwich , UK). Some respondents face a straightforward open ended (OE) question. Others face a dichotomous choice, with a follow-up open-ended question. In the table IB OE refers to this question; c^{***} refers to the estimates of WTP from the dichotomous choice answers based on the assumption that any values between zero and infinity are possible, while c^{**} uses the same data, but assumes that no values beyond the highest bid level (£500) are credible. The mean figures from the four estimation methods form a pattern that might be anticipated: OE produces the lowest estimate, the dichotomous choice estimates are somewhat higher and the follow-up OE question produces an estimate which is lower than that from the dichotomous choice questions but possibly because of anchoring it is still higher than the basic OE estimate.

Within this set of alternatives, let us consider different groups of acceptable frames. For the first acceptable set, the prior is that all frames appear equally credible; in the remaining four cases only two of the elicitation methods are possible candidates for being the elicitors of the true value. One group consists of just the open-ended WTP figures; one eliminates the datasets where there is some truncation by the researchers of the acceptable values; a fourth group consists of just the dichotomous choice variants, while the final group uses two sources of data

Table 11.4 Mean WTP Values and Critical Jury Sizes for Different Acceptable Sets

	Acceptable frames				
	All	Open ended	Un-truncated	All DC	OE After DC
OE	67	67	67	–	–
IB OE	74.9	74.9	–	–	74.9
C**	112	–	–	112	112
C***	144	–	144	144	144
Mean Value Across Frames	99.5	71.0	105.5	128.0	109.5
Coefficient of Frame Variation	0.28	0.05	0.30	0.10	0.26
Critical n, B = 0.5	3	121	3	24	4
Critical n, B = 1	13	484	11	96	15
Critical n, B = 1.5	29	1089	25	216	34

(OE and dichotomous choice) drawn from the same individuals. In the table therefore five estimates of the critical n are produced for each possible B which is varied between three values from 0.5 to 1.5.

The critical value of n varies enormously. When the choice is across OE or DC-based methods of elicitation, an ideal CJ consisting of 16–25 members might outperform CV for some low values of B. On the other hand, if the set of possible frames lies within OE variants or within DC variants then ICJ is inferior. In the actual study, the value for B from the raw open-ended WTP was 1.7 which would rule out ICJ for any set of acceptable frames. Meanwhile the highest value for frame variation is 0.28, meaning that any partial valuation would have to capture at least 72% of the total valuation in its estimates for it to dominate the mix of contingent valuation methods used in the research.

11.4 Summing Up

This final chapter has examined one means by which non-market valuation methods can be compared. In a sense it provides an answer to Diamond and Hausman's when is no number better than any number question. Concurring with my argument in Chapter 5, generally the answer is 'never', given a choice between decision processes. At the same time we encountered clear reasons why the figures obtained from a contingent valuation exercise may be less useful from those obtained from a method that appears a priori inferior.

I stressed the possibility that partial valuation methods and approaches such as the Citizens' Jury may, under specific circumstances be superior to contingent valuation, particularly when WTA is included in the set of acceptable frames, but it is worth reiterating the point made earlier, that whereas contingent valuation and its associated techniques have been exhaustively examined for anomalies and inadequacies many of the alternatives have not received similar scrutiny. It is particularly worth emphasising that the potential advantage of the partial valuation methods lies in the absence or reduction of framing variance when such alternatives are used. To

the extent that a partial valuation method is subject to the same kinds of framing variance that plagues fully specified contingent valuation, then the partial valuation method loses its superiority. The same applies to Citizens' Juries and consensus methods which have not thus far been tested extensively for framing invariance.Indeed the evidence from group decision-making (see for example the survey in Kerr et al., 1996) is far from comforting on the issue. That literature suggests that group dynamics can differ radically from the judgement and choices made by individuals. In particular there can be shift to the extremes, rather than the process of averaging that might be supposed.

One interpretation of the examples is that the frame variance drives the results. In other words, without any restriction on the set of acceptable frames partial valuation and ICJ methods dominate contingent valuation. To the extent that some frames are excludable from the set of acceptable frames, for instance by using the criteria set out by the NOAA panel, then the advantage of alternative decision-processes over contingent valuation diminishes.

A final point worth making, both about this chapter and more generally is the lack of appropriate data for policy. The comparative model of this chapter suggests that estimates of the variation in tastes within the sample are often not reported from valuation exercises, but would provide a useful guide to the choice of technique. Similarly, a database comparing values from users to total values and more exercises in which frame variance was compared amongst different groups such as users versus non-users would be a step towards the optimal choice of decision processes.

References

Abeler J, Marklein F (2008) Fungibility, labelling and consumption. University of Bonn and IZA Bonn

Adamowicz WL, Bhardwaj V, McNab B (1993) Experiments on the difference between willingness to pay and willingness to accept. Land Econ 69:416–427

Adamowicz WL, Boxall P, Williams M, Louviere J (1998) Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. Am J Agric Econ 80(1):64–75

Ahmad E, Stern N (1991) The theory and practice of tax reform in developing countries. Cambridge University Press, Cambridge.

Ajzen I (1991) The theory of planned behaviour. Organ Behav Human Decis Proc 50:179–211

Alberini A, Kanninen B, Carson R (1997) Modeling response incentive effects in dichotomous choice contingent valuation data. Land Econ 73(3):309–324

Aldred J, Jacobs M (2000) Citizens and wetlands evaluating the Ely citizens' jury. Ecol Econ 34:217–232

Allingham MG, Sandmo A (1972) Income tax evasion a theoretical analysis. J Public Econ 1:323–338

Amigues JP, Boulatoff C, Desaigues B, Gauthier C, Keith JE (2002) The benefits and costs of riparian analysis habitat preservation a willingness to accept/willingness to pay contingent valuation approach. Ecol Econ 43(1):17–31

Anderlini L, Canning D (2001) Structural stability implies robustness to bounded rationality. J Econ Theory 101(2):395–422

Anderson C, Lepper M, Ross L (1980) Perseverance of social theories: the role of explanation in the persistence of discredited information. J Pers Soc Psychol 39(6):1037–1049

Anderson M (1974) Family structure in 19th century Lancashire. Cambridge University Press, Cambridge

Andolfatto D (2002) A theory of inalienable property rights. J Polit Econ 110(2):382–393

Andreoni J (1995) Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. Quart J Econ 110(1):1–21

Andreoni J, Erard B, Feinstein J (1998) Tax compliance. J Econ Lit 36:818–860

Angeletos GM, Laibson D, Repetto A, Tobacman J, Weinberg S (2001) The hyperbolic consumption model: calibration, simulation, and empirical evaluation. J Econ Perspect 15(3):47–68

Arcelus FJ, Levine AL (1986) Merit goods and public choice – the case of higher-education. Public Finance-Finances Publiques 41(3):303–315

Arcelus FJ, Levine AL (1988) Merit goods and public choice - the case of higher-education - reply. Public Finance-Finances Publiques 43(3):471–472

Ariely D, Huber J, Wertenbroch K (2005) When do losses loom larger than gains? J Mark Res 42(2):134–138

Ariely D, Loewenstein G, Prelec D (2003) Coherent arbitrariness: stable demand curves without stable preferences. Quart J Econ 118:73–105

Arkes HR, Blumer C (1985) The psychology of sunk cost. Organ Behav Hum Decis Process 35:24–40

Armour A (1995) The Citizens' Jury model of public participation: a critical evaluation. In: Renn O, Webler T, Widemann P (eds) Fairness and competence in citizen participation. Kluwer Academic Publishers, Dordrecht

Arndt J, Crane E (1975) Response bias, yea-saying and the double negative. J Mark Res 12:218–220

Arrow K (1977) Extended sympathy and the possibility of social choice. Am Econ Rev: Papers Proc 67:219–225

Arrow K (1982) Risk perception in psychology and economics. Econ Inq 20:1–9

Arrow K, Solow R, Portney PR, Leamer EE, Radner R, Schuman H (1993) Report of the NOAA panel on contingent valuation. Federal Register 58(10):4602–4614

Atkinson A, McKay S, Kempson E, Collard S (2006) Levels of financial capability in the UK: results of a baseline survey. Prepared for the Financial Services Authority by Personal Finance Research Centre, University of Bristol

Auld D, Bing P (1971) Merit wants: a further analysis. Finanzarchiv 30(2):257–265

Ayers WM Jr (1973) Efficiency and equity in provision for merit wants: the case of federal food programs for the low-income population. PhD thesis, Cornell University, Aug 1973

Aylin P, Morris S, Wakefield J, Grossinho A, Jarup L, Elliott P (2001) Temperature, housing, deprivation and their relationship to excess winter mortality in Great Britain, 1986–1996. Int J Epidemiol 30:1100–1108

Baigent N (1981) Social choice and merit goods. Econ Lett 7(4):301–305

Baker M, Nagel S, Wurgler J (2005) The effects of dividends on consumption. NBER Working Paper, 12288

Banford NJ, Knetsch JL, Mauser GA (1979) Feasibility judgements and alternative measures of benefits and costs. J Business Administration 11:25–35

Banks J, Blundell R, Tanner S (1998) Is there a retirement-savings puzzle? Am Econ Rev 88(4):769–788

Banks J, Emmerson C (2005) Prepared for retirement? The adequacy and distribution of retirement resources in England. Institute for Fiscal Studies, IFS Reports

Banks J, Oldfield Z (2006) Understanding pensions: cognitive function, numerical ability and retirement saving. Institute for Fiscal Studies, IFS Reports

Banks J, Smith Z, Wakefield M (2002) The distribution of financial wealth in the UK: evidence from 2000 BHPS data. The Institute For Fiscal Studies, WP02/21

Barberis N, Huang M (2001) Mental accounting, loss aversion, and individual stock returns. J Financ 56(4):1247–1292

Barberis N, Huang M, Santos T (2001) Prospect theory and asset prices. Quart J Econ 116(1):1–53

Barberis N, Huang M, Thaler RH (2006) Individual preferences, monetary gambles, and stock market participation: a case for narrow framing. Am Econ Rev 96(4):1069–1090

Bardsley N, Moffatt P (2007) The experimetrics of public goods: inferring motivations from contributions. Theor Decis 62(2):161–193

Barkan R, Busemeyer JR (2003) Modeling dynamic inconsistency with a changing reference point. J Behav Decis Making 16(4):235–255

Barnes J Jr (1984) Cognitive biases and their impact on strategic planning. Strategic Manage J 5(2):129–137

Barnes M (1999) Building a deliberative democracy an evaluation of two Citizens' Juries. Institute for Public Policy Research, London

Baron J, Ritov I (1994) Reference points and omission bias. Organ Behav Human Decis Proce 59(3):475–498

Bartle I, Vass P (2005) Self-regulation and the regulatory state – A survey of policy and practice. Centre for the Study of Regulated Industries, London

Barzel Y (1997) Economic analysis of property rights. Cambridge University Press, Cambridge

Basu K (1976) Retrospective choice and merit goods. Finanzarchiv 34(2):220–225

Bateman IJ, Carson RT, Day B, Hanemann M, Hanley N, Hett T, Jones-Lee M, Loomes G, Mourato S, Ozdemiroglu E, Pearce DW, Sugden R, Swanson T (2002) Economic valuation with stated preference techniques: a manual. Edward Elgar, Cheltenham

Bateman IJ, Kahneman D, Munro A, Starmer C, Sugden R (2005) Testing competing models of loss aversion: an adversarial collaboration. J Public Econ 89(8):1561–1580

Bateman IJ, Langford IH, Munro A, Starmer C, Sugden R (2000) Estimating four Hicksian welfare measures for a public good: a contingent valuation investigation. Land Econ 76(3): 355–373

Bateman IJ, Langford IH, Rasbash J (1999) Willingness-to-pay question format effects in contingent valuation studies. In: Bateman IJ, Willis KG (eds) Valuing environmental preferences. Oxford University Press, Oxford, pp 511–539

Bateman IJ, Langford IH, Turner RK, Willis KG, Garrod GD (1995) Elicitation and truncation effects in contingent valuation studies. Ecol Econ 12(2):161–179

Bateman IJ, Munro A (2003) Non-cooperative decision-making and measures of household surplus. Centre for Social and Economic Research on the Global Environment, EDM-2003-12

Bateman IJ, Munro A (2004) Testing economic models of the household an experiment. Centre for Social and Economic Research on the Global Environment, EDM-2004-01

Bateman IJ, Munro A (2005) An experiment on risky choice amongst households. Econ J 115(502):C176

Bateman IJ, Munro A, Rhodes B, Starmer C, Sugden R (1997a) Does part-whole bias exist? an experimental investigation. Econ J 107(441):322–332

Bateman IJ, Munro A, Rhodes B, Starmer C, Sugden R (1997b) A test of the theory of reference-dependent preferences. Quart J Econ 112(2):479–505

Bateman IJ, Munro A, Rhodes B, Starmer C, Sugden R (2006) Anchoring and yea-saying with private goods: an experiment. In: List J (ed) Using experimental methods in environmental and resource economics. Edward A. Elgar, Cheltenham

Bateman IJ, Willis KG (1999) Valuing environmental preferences: theory and practice of the contingent valuation method in the US, EU, and developing countries. Oxford University Press, Oxford

Baumeister R (2003) The psychology of irrationality why people make foolish, self defeating choices. In: Brocas I, Carillo J (eds) The psychology of economic decisions, vol 1: rationality and well-Being. Oxford University Press, Oxford, chap 1, pp 3–16

Baumol WJ, Oates WE (1975) The theory of environmental policy: externalities, public outlays, and the quality of life. Prentice-Hall, Harlow

Becker G (1981) Treatise on the family. Harvard University Press, Cambridge.

Becker G, DeGroot M, Marschak J (1963) Stochastic models of choice behavior. Behav Sci 8(1):41–55

Becker GM, DeGroot MH, Marschak J (1964) Measuring utility by a single-response sequential method. Behav Sci 9:226–232

Becker SW, Brownson F (1964) What price ambiguity? Or the role of ambiguity in decision-making. J Polit Econ 72:62–73

Bell DR, Lattin JM (2000) Looking for loss aversion in scanner panel data: the confounding effects of price response heterogeneity. Mark Sci 19(2):185–201

Benartzi S, Thaler RH (1995) Myopic loss aversion and the equity premium puzzle. Quart J Econ 110(1):73–92

Benartzi S, Thaler RH (2001) Naive diversification strategies in defined contribution saving plans. Am Econ Rev 91(1):79–98

Benbassat J, Pilpel D, Tidhar M (1988) Patients' preferences for participation in clinical decision making: A review of published surveys. Behav Med 24(2):81–88

Bendor J, Diermeier D, Ting M (2003) A behavioral model of turnout. Am Polit Sci Rev 97(2):261–280

Benjamin DK, Dougan WR (1997) Individuals' estimates of the risks of death part I- a reassessment of the previous evidence. J Risk Uncertain 15:115–133

Benjamin DK, Dougan WR (2001) Individuals' estimates of the risks of death part II – new evidence. J Risk Uncertain 22:35–57
Bergmann BR (2006) Reducing inequality – merit goods vs. income grants. Dissent 53(1):67–71
Bernasconi M, Loomes G (1992) Failures of the reduction principle in an Ellsberg-type problem. Theory Decis 32:77–100.
Bernasconi M, Zanardi A (2004) Tax evasion tax rates and reference dependence. FinanzArchiv 60:422–445
Bernheim BD (1993) Is the baby boom generation preparing adequately for retirement? Summary Report. Merrill Lynch, Princeton and NJ
Bernheim BD (1996) Rethinking saving incentives. Stanford Economics Working Paper, 96-009
Bernheim BD, Garrett DM (2003) The effects of financial education in the workplace: evidence from a survey of households. J Public Econ 87:1487–1519
Bernheim BD, Garrett DM, Maki DM (2001) Education and saving: the long-term effects of high school financial curriculum mandates. J Public Econ 80(3):435–465
Berti AE, Bombi AS (1988) The child's construction of economics. Cambridge University Press Paris: Editions de la Maison des sciences de l'homme, Cambridge [Cambridgeshire], New York
Besley T (1988) A simple model for merit good arguments. J Public Econ 35(3):371–383
Besley T (1989) Ex ante evaluation of health states and the provision for ill health. Econ J 99(394):146–166
Besley T, Coate S (1997) A model of representative democracy? Quart J Econ 112:85–114
Beverly S, McBride A, Schreiner M (2003) A framework of asset-accumulation stages and strategies. J Fam Econ Issues 24(2):143–156
Biel A, Johansson-Stenman O, Nilsson A (2006) Emotions, morality and public goods: the WTA-WTP disparity revisited. Working papers in Economics no. 193, Department of Economics, School of Business, Economics and Law
Binmore K (1999) Why experiment in economics? Econ J 109(453):16–24
Bishop RC (2003) Where to from here? In: Champ PA, Boyle KJ, Champ TC (eds) A primer on non-market valuation. Springer, Dordrecht, pp 537–566
Bishop RC, Heberlein TA (1979) Measuring values of extramarket goods: are indirect measures biased? Am J Agric Econ 61(5):926–930
Bishop RC, Heberlein TA (1986) Does contingent valuation work? In: Cummings RG, Brookshire DS, Schulze W (eds) Valuing environmental goods. An assessment of the contingent valuation method. Rowman and Allen, Totowa, New Jersey, pp 123–147
Bishop RC, Heberlein TA, Kealy MJ (1983) Contingent valuation of environmental assets: comparisons with a simulated market. Nat Resour J 23:619–633
Blackorby C, Donaldson D (1988) Cash versus kind, self-selection, and efficient transfers. Am Econ Rev 78(4):691–700
Blackorby C, Donaldson D, Weymark JA (1984) Social choice with interpersonal utility comparisons a diagrammatic introduction. Int Econ Rev 25(2):327–356
Blake D (1995) Pension schemes and pension funds in the United Kingdom. Oxford, Oxford University Press.
Blamey R, Smith RJR, Niemeyer S (2002) Citizens' juries and environmental value assessment. Australian National University, Research School of Social Sciences
Blavatskyy P (2006) Violations of betweenness or random errors? Econ Lett 91:34–38
Bleichrodt H (2002) A new explanation for the difference between time trade-off utilities and standard gamble utilities. Health Econ 11(5):447–456
Bleichrodt H, Pinto JL (2002) Loss aversion and scale compatibility in two-attribute trade-offs. J Math Psychol 46(3):315–337
Bleichrodt H, Pinto JL, Wakker PP (2001) Making descriptive use of prospect theory to improve the prescriptive use of expected utility. Manage Sci 47(11):1498–1514
Block-Lieb S, Janger EJ (2006) The myth of the rational borrower: rationality, behavioralism, and the misguided "reform" of bankruptcy law. Texas Law Rev 84(6):1481–1565
Blomquist S, Micheletto L (2006) Optimal redistributive taxation when government's and agents' preferences differ. J Public Econ 90(6–7):1215–1233

Blow L, Walker I, Zhu Y (2006) Who benefits from child benefit? Institute for Fiscal Studies Research Reports, No. 749

Blumenschein K, Johannesson M (1998) An experimental test of question framing in health state utility assessment. Health Policy 45(3):187–193

Boadway R, Bruce N (1984) Welfare economics. Blackwell, Oxford

Bohm P, Lind H (1993) Preference reversal, real-world lotteries, and lottery-interested subjects. J Econ Behav Organ 22(3):327–348

Bohman J, Rehg W (eds) (1997) Deliberative democracy: essays on reason and politics. MIT Press, Cambridge, USA

Bone J, Hey J, Suckling J (2004) A simple risk-sharing experiment. J Risk Uncertai 28(1):23–38

Bonini N, Rumiati R (1996) Mental accounting and acceptance of a price discount. Acta Psychol 93(1–3):149–160

Börsch-Supan A, Brugiavini A (2001) Savings: the policy debate in Europe. Oxford Rev Econ Pol 17:116–143

Boshara R (2001) Building assets: a report on the asset-development and IDA field. Corporation for Enterprise Development, USA

Boven LV, Loewenstein G, Dunning D (2003) Mispredicting the endowment effect: underestimation of owners' selling prices by buyer's agents. J Econ Behav Organ 51(3):351–365

Bowker J, Stoll J (1988) Use of dichotomous choice nonmarket methods to value the whooping crane resource. Am J Agric Econ 70:372–381

Bowman D, Minehart D, Rabin M (1999) Loss aversion in a consumption-savings model. J Econ Behav Organ 38(2):155–178

Boyce RR, Brown TC, McClelland GH, Peterson GL, Schulze WD (1992) An experimental examination of intrinsic values as a source of the WTA-WTP disparity. Am Econ Rev 82(5):1366–1373

Boyle KJ (1989) Commodity specification and the framing of contingent-valuation questions. Land Econ 65(1):57–63

Boyle KJ (2003) Contingent valuation in practice. In: Champ PA, Boyle KJ, Brown TC (eds) A primer on non-market valuation. Kluwer Academic Press, Dordrecht, pp 111–170

Boyle KJ, Bergstrom JC (1992) Benefit transfer studies: myths, pragmatism, and idealism. Water Resour Res 28(3):657–663

Boyle KJ, Bishop RC (1988a) Valuing wildlife in benefit-cost analysis a case study involving endangered species. Water Res Res 23:943–950

Boyle KJ, Bishop RC (1988b) Welfare measurements using contingent valuation: a comparison of techniques. Am J Agric Econ 70(1):20–28

Boyle KJ, Bishop RC, Welsh MP (1985) Starting point bias in contingent valuation bidding games. Land Econ 61(2):188–194

Boyle KJ, Johnson FR, McCollum DW, Desvouges WH, Dunford RW, Hudson SP (1996) Valuing public goods: discrete versus continuous contingent-valuation responses. Land Econ 72(3):381–396

Boyle KJ, MacDonald HF, Cheng H, McCollum DW (1998) Bid design and yea saying in single-bounded, dichotomous-choice questions. Land Econ 74(1):49–64

Braga J, Humphrey SJ, Starmer C (2006) Market experience eliminates some anomalies – and creates new ones. University of Nottingham, CEDEX Discussion Paper in Economics

Braga J, Starmer C (2005) Preference anomalies, preference elicitation and the discovered preference hypothesis. Environ Resour Econ 32(1):55–89

Breeze J (1998) Can paternalism be justified in mental health care? J Adv Nurs 28(2):260–265

Brendl CM, Markman AB, Higgins ET (1998) Mental accounting as self-regulation: representativeness to goal-derived categories. Zeitschrift Fur Sozialpsychologie 29(2):89–104

Brennan G, Lomasky LE (1983) Institutional aspects of 'merit goods' analysis. Finanzarchiv 41(2):183–206

Brennan G, Walsh C (eds) (1990) Rationality, individualism and public policy. Centre for Research on Federal Financial Relations, Australian National University, Canberra

Brooks P, Zank H (2005) Loss averse behavior. J Risk Uncertai 31(3):301–325
Brookshire DS, Coursey DL (1987) Measuring the value of a public good: an empirical comparison of elicitation procedures. Am Econ Rev 77:554–566
Brookshire DS, Randall A, Stoll JR (1980) Valuing increments and decrements in natural resource service flows. Am J Agric Econ 62:478–488
Brookshire DS, Schulze WD, D'Arge RC (1982) Valuing public goods: a comparison of survey and hedonic aroaches. Am Econ Rev 72:165–177
Broome J (1999) Ethics out of economics. Cambridge University Press, Cambridge.
Brouwer R, Langford IH, Bateman I, Turner RK (1999) A meta-analysis of wetland contingent valuation studies. Reg Environ Change 1(1):47–57
Brown TC (2005) Loss aversion without the endowment effect, and other explanations for the WTA-WTP disparity. J Econ Behav Organ 57(3):367–379
Brown TC, Champ P, Bishop R, McCollum DW (1996) Which response format reveals the truth about donations to a public goods? Land Econ 72:152–166
Browning EK (1981) A theory of paternalistic in kind transfers. Econ Inq 19(4):579–597
Browning M, Chiappori P (1998) Efficient intra-household allocations: a general characterization and empirical tests. Econometrica 66(6):1241–1278
Buchanan J (1960) The theory of public finance. South Econ J 26(3):234–238
Buchanan J (1990) The domain of constitutional economics. Constitutional Polit Econ 1(1):1–18
Buchanan J, Musgrave R (1999) Public finance and public choice: two contrasting visions of the state. MIT Press, Cambridge, USA
Burrows P (1993) Patronising paternalism. Oxford Econ Papers 45:542–572
Buschena D, Zilberman D (2000) Generalized expected utility, heteroscedastic error, and path dependence in risky choice. J Risk Uncertain 20(1):67–88
Cachon G, Camerer C (1996) Loss-avoidance and forward induction in experimental coordination games. Quart J Econ 111(1):165–194
Caillaud B, Rey P (1995) Strategic aspects of vertical delegation. Europ Econ Rev 39(3):421–431
Camerer CF (1987) Do biases in probability judgment matter in markets? Experimental evidence. Am Econ Rev 77(5):981–997
Camerer CF (1988) Gifts as economic signals and social symbols. Am J Sociol 94:180–214
Camerer CF (1995) Individual decision-making. In: Roth A, Kagel J (eds) Handbook of experimental economics. Princeton University Press, Princeton, pp 349–443.
Camerer CF (1998) Bounded rationality in individual decision making. Exp Econ 1(2):163–183
Camerer CF (2000) Prospect theory in the wild: evidence from the field. In: Kahnemann D (ed) Choices, values and frames. Oxford University Press, Oxford, pp 288–300
Camerer CF (2005) Three cheers – psychological, theoretical, empirical – for loss aversion. J Mark Res 42(2):129–133
Camerer CF, Babcock L, Loewenstein G, Thaler RH (1997) Labor supply of New York city cabdrivers: one day at a time. Quart J Econ 112(2):407–441
Camerer CF, Fehr E (2002) Measuring social norms and preferences using experimental games: a guide for social scientists. Institute for Empirical Research in Economics, University of Zurich
Camerer CF, Hogarth RM (1999) The effects of financial incentives in experiments: a review and capital labour production framework. J Risk Uncertai 19(1):7–42
Camerer CF, Issacharoff S, Loewenstein G, O'Donoghue T, Rabin M (2003) Regulation for conservatives: behavioral economics and the case for "asymmetric paternalism". University PA Law Rev 151(3):1211–1254
Camerer CF, Kunreuther H (1989) Decision processes for low probability events: policy implications. J Policy Anal Manage 8(4):565–592
Camerer CF, Loewenstein G (2004) Behavioral economics: past, present, future. In: Camerer CF, Loewenstein G, Rabin M (eds) Advances in behavioral economics. Princeton University Press, Princeton, NJ, pp 3–51
Camerer CF, Thaler RH (1995) Anomalies: Ultimatums, dictators and manners. J Econ Perspect 9(2):209–219

Camerer CF, Weber M (1992) Recent developments in modeling preferences: uncertainty and ambiguity. J Risk Uncertain 5(4):325–370
Cameron TA, Quiggin J (1994) Estimation using contingent valuation data from a dichotomous choice with follow-up questionnaire. J Environ Econ Manage 27:218–234
Capéau B, Ooghe E (2003) Merit goods and phantom agents. Econ Bulletin 8(8):1–5
Caplan B (2001) Rational ignorance versus rational irrationality. Kyklos 54(1):3–26
Carbone E (1998) Investigation of stochastic preference theory using experimental data. Econ Lett 57:305–312
Carbone E, Hey J (1994) Estimation of expected utility and non-expected utility preference functionals using complete ranking data. Models and Experiments on Risk and Rationality. Kluwer, Dordrecht
Carmichael HL, Macleod WB (2006) Welfare economics with intransitive revealed preferences: a theory of the endowment effect. J Public Econ Theory 8(2):193–218
Carson RT (1994) Contingent valuation and lost passive use: damages from the Exxon Valdez. Resources for the Future, Washington
Carson RT (1995) Contingent valuation surveys and tests of insensitivity to scope. University of California, San Diego, Department of Economics
Carson RT (1997) Contingent valuation: theoretical advances and empirical tests since the NOAA panel. Am J Agric Econ 79(5):1501–1507
Carson RT (1998) Valuation of tropical rainforests: philosophical and practical issues in the use of contingent valuation. Ecol Econ 24:15–29
Carson RT (2000) Contingent valuation: a user's guide. Environ Sci Technol 34(8):1413–1418
Carson RT, Flores NE, Martin KM, Wright JL (1996) Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods. Land Econ 72(1):80–99
Carson RT, Flores NE, Meade NF (2001) Contingent valuation: controversies and evidence. Environ Resour Econ 19(2):173–210
Carson RT, Groves T, Machina MJ (1999) Incentive and informational properties of preference questions. Plenary Address, European Association of Environmental and Resource Economists, Oslo, Norway
Carson RT, Hanemann WM, Kopp RJ, Krosnick JA, Mitchell RC, Presser S, Rudd PA, Smith VK, Conaway M, Martin K (1997) Temporal reliability of estimates from contingent valuation. Land Econ 73(2):151–163
Carson RT, Hanemann WM, Kopp RJ, Krosnick JA, Mitchell RC, Presser S, Ruud PA, Smith VK, Conaway M, Martin K (1998) Referendum design and contingent valuation: the NOAA panel's no-vote recommendation. Rev Econ Stat 80(2):335–338
Carson RT, Jeon Y (2000) On overcoming inherent informational deficiencies in estimating the distribution of willingness to pay. UCSD Economics Discussion Paper, Feb 2000
Carson RT, Mitchell RC (1993) The issue of scope in contingent valuation studies. Am J Agric Econ 75(5):1263–1267
Carson RT, Mitchell RC, Hanemann WM, Kopp RJ, Presser S, Ruud PA (1992) A contingent valuation study of lost passive use values resulting from the Exxon Valdez oil spill. A report to the Attorney General of the State of Alaska
Carter R (1977) Justifying paternalism. Can J Philos 7(1):139–142
Chakravarti D, Krish R, Paul P, Srivastava J (2002) Partitioned presentation of multicomponent bundle prices: evaluation, choice and underlying processing effects. J Consum Psychol 12(3):215–229
Chami R (1998) Private income transfers and market incentives. Economica 65(260):557–580
Champ PA, Boyle KJ, Brown TC (2004) A primer on nonmarket valuation. Springer, Dordrecht
Chang HF (2000) A liberal theory of social welfare fairness, utility and the Pareto principle. Yale Law J 110:173–235
Chang O, Schultz J Jr (1990) The income tax withholding phenomenon evidence from TCMP data. J Am Taxation Assoc 12(1):88–93
Charles S, Westaway T (1981) Ignorance and merit wants. Finanzarchiv 39(1):74–78

Charness G, Levin D (2005) When optimal choices feel wrong: a laboratory study of Bayesian updating, complexity, and affect. Am Econ Rev 95(4):1300–1309
Cheema A, Soman D (2006) Malleable mental accounting: the effect of flexibility on the justification of attractive spending and consumption decisions. J Consum Psychol 16(1):33–44
Chen MK, Lakshminarayanan V, Santos LR (2006) How basic are behavioral biases? evidence from Capuchin monkey trading behavior. J Polit Econ 114(3):517–537
Cherry T, Frykblom P, Shogren J, List J, Sullivan M (2004) Laboratory testbeds and non-market valuation: the case of bidding behavior in a second-price auction with an outside option. Environ Resour Econ 29(3):285–294
Cherry TL (2001) Mental accounting and other-regarding behavior: evidence from the lab. J Econ Psychol 22(5):605–615
Cherry TL, Crocker TD, Shogren JF (2003) Rationality spillovers. J Environ Econ Manage 45(1):63–84
Chilton S, Baker R, Donaldson C (2003) Determining the societal value of a QALY by surveying the public in England and Wales: a research protocol. NCCRM Publications, Birmingham
Choi JJ, Laibson D (2003) For better or for worse default effects and 401(k) savings behaviour. In: Wise DA (ed) Perspectives on the economics of aging. Chicago University Press, Chicago (for NBER)
Choi JJ, Laibson D, Madrian B, Metrick A (2002) Defined contribution pensions: plan rules, participant decisions, and the path of least resistance. In: Poterba JM (ed) Tax policy and the economy 16. MIT Press, Cambridge, USA, pp 67–113
Clancy M, Grinstein-Weiss M (2001) Financial education and savings outcomes in individual development accounts. Centre for Social Development Working Paper, 01-2
Clarke P (2002) Testing the convergent validity of the contingent valuation and travel cost methods in valuing the benefits of health care. Health Econo 11:117–127
Coate S (1995) Altruism, the samaritan's dilemma and government transfer policy. Am Econ Rev 85:46–57
Coate S, Morris S (1995) On the form of transfers to special interests. J Polit Econ 103(6):1210–1235
Colell AM, Whinston MD, Green JR (1995) Microeconomic theory. Oxford University Press, New York
Conlisk J (1996) Why bounded rationality? J Econ Lit 34(2):669–700
Connolly S, Munro A (1999) Economics of the public Sector. Prentice Hall, Harlow
Cooper J, Loomis J (1992) Sensitivity of willingness-to-pay estimates to bid design in dichotomous choice contingent valuation models. Land Econ 68:211–224
Coote A, Lenaghan J (1997) Citizens' Juries: from theory to practice. IPPR, London
Cosmides L, Tooby J (1996) Are humans good intuitive statisticians after all? rethinking some conclusions from the literature on judgment under uncertainty. Cognition 58(1):1–73
Coursey DL, Hovis JL, Schulze WD (1987) The disparity between willingness to accept and willingness to pay measures of value. Quart J Econ 102(3):679–690
Cox AD, Cox D, Zimet G (2006) Understanding consumer responses to product risk information. J Mark 70(1):79–91
Cox JC, and Epstein S (1989) Preference reversals without the independence axiom. Am Econ Rev 79:408–26.
Cox JC, Grether DM (1996) The preference reversal phenomenon: response mode, markets and incentives. Econ Theory 7(3):381–405
Critchfield TS, Kollins SH (2001) Temporal discounting basic research and the analysis of socially important behaviour. J Allied Behav Anal 34(1):101–122
Crosby N (1991) Citizens' Juries as a basic democratic reform. Jefferson Centre, Minneapolis
Crouch A, Keniston K (1960) Yea-sayer and nay-sayers: agreeing response set as a personal variable. J Abnormal Soc Psychol 60:151–174
Cubitt RP, Munro A, Starmer C (2004) Testing explanations of preference reversal. Econ J 114(497):709–726

Cubitt RP, Starmer C, Sugden R (1998) On the validity of the random lottery incentive system. Exp Econ 1(2):115–131

Cubitt RP, Sugden R (2001) Dynamic decision-making under uncertainty: an experimental investigation of choices between accumulator gambles. J Risk Uncertain 22(2):103–128

Cullis J, Jones P, Lewis A (2006) Tax framing, instrumentality and individual differences: are there two different cultures? J Econ Psychol 27(2):304–320

Cullis J, Jones P, Propper C (2000) Waiting lists and medical treatment: analysis and policies. Handbook Health Econ 1:1201–1249

Culyer AJ (1971) Merit goods and the welfare economics of coercion. Public Finan 26(4):546–572

Culyer AJ (1990) Commodities, characteristics of commodities, characteristics of people, utility and quality of life. In: Baldwin S, Godfrey C, Propper C (eds) Quality of life: perspectives and policy. Routledge, London, chap 1, pp 9–27

Cummings RG, Harrison GW (1995) The measurement and decomposition of nonuse values: a critical review. Environ Resour Econ 5(3):225–247

Cummings RG, Harrison GW, Rutstrom EE (1995) Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive-compatible? Am Econ Rev 85(1):260–266

Danziger K (1958) Children's earliest conceptions of economic relationships. J Soc Psychol 47:231–240

Darley JM, Gross PH (1983) A hypothesis-confirming bias in labeling effects. J Person Soc Psychol 44:20–33

Davis R (1963) Recreation planning as an economic problem. Nat Resour J 3:239–249

Debreu G (1970) Economies with a finite set of equilibria. Econometrica 38:387–392

Desvousges WH, Johnson FR, Dunford R, Boyle K, Hudson S, Wilson K (1993) Measuring natural resource damages with contingent valuation: tests of validity and reliability. In: Diamond P, Hausman JA (eds) Contingent valuation: a critical assessment. North Holland, Amsterdam, pp 91–164

Dhar R (1996) New developments in mental accounting: implications for consumer choice. Adv Consum Res 23:210–235

Dholakia UM, Simonson I (2005) The effect of explicit reference points on consumer choice and online bidding behavior. Mark Sci 24(2):206–217

Diamond PA, Hausman JA (1994) Contingent valuation: is some number better than no number? J Econ Perspect 8:45–64

Diamond PA, Mirrlees JA (1971) Optimal taxation and public production: I–production efficiency. Am Econ Rev 61:8–27

Diecidue E, Wakker PP (2001) On the intuition of rank-dependent utility. J Risk Uncertain 23(3):281–298

van Dijk E, van Knippenberg D (1998) Trading wine: on the endowment effect, loss aversion, and the comparability of consumer goods. J Econ Psychol 19(4):485–495

Dhami S, al-Nowaihi A (2004) Why do people pay taxes: expected utility versus prospect theory', Leicester Discussion Paper 03/18, University of Leicester.

Disney R, Whitehouse E (1992) The personal pension stampede. Institute for Fiscal Studies, London

Dohmen T, Falk A, Huffman D, Sunde U, Shu J, Wagner GG (2005) Individual risk attitudes: new Evidence from a large, representative experimentally-validated survey. IZA Bonn Discussion Paper No. 1730

Downs A (1957) An economic theory of democracy. Harper, New York

Doyle JR, O'Connor DJ, Reynolds GM, Bottomley PA (1999) The robustness of the asymmetrically dominated effect buying frames, phantom alternatives and in-store purchases. Psychol Mark 16(3):225–243

Dubourg WR, Jones-Lee MW, Loomes G (1994) Imprecise preferences and the WTP-WTA disparity. J Risk Uncertain 9:115–133

Dubourg WR, Jones-Lee MW, Loomes G (1997) Imprecise preferences and survey design in contingent valuation. Economica 64(256):681–702

Dupont DY (2003) Do children matter? an examination of gender differences in environmental valuation. CSERGE, University of East Anglia

Dupont DY, Lee GS (2002) The endowment effect, status quo bias and loss aversion: rational alternative explanation. J Risk Uncertain 25(1):87–101

Duxbury D, Keasey K, Zhang H, Chow SL (2005) Mental accounting and decision making: evidence under reverse conditions where money is spent for time saved. J Econ Psychol 26(4):567–580

Dworkin G (1971) Paternalism. In: Wasserstrom RA (ed) Morality and the law. Wadsworth, Belmont, California

Earl P (2005) Behavioural economics and the economics of regulation. Briefing paper prepared for the New Zealand Ministry of Economic Development, New Zealand

Eckel C, Johnson C, Montmarquette C (2003) Human capital investment by the poor: calibrating policy with laboratory experiments. CIRANO, University of Montreal

Eden L, Shoup CS (eds) (1991) Retrospectives on public finance. Duke University Press, Durham, USA

Edmonds E (2002, August) Reconsidering the labeling effect for child benefits: evidence from a transition economy. Econ Lett 76(3): 303–309

Eecke WV (2001) The concept of "merit good" or the epistemological necessity of an ethical concept in economic science. Laval Theologique Et Philosophique 57(1):23–40

Eecke WV (2003) Adam Smith and Musgrave's concept of merit good. J Soc Econ 31:701–720

Einhorn HJ, Hogarth RM (1981) Behavioral decision theory: processes of judgment and choice. J Account Res 19(1):1–31

Eisenberger R, Weber M (1995) Willingness to pay and willingness to accept for risky and ambiguous lotteries. J Risk Uncertain 10:223–233

El-Safty AE (1976) Adaptive behavior, demand and preferences. J Econ Theory 13(2):298–318

Ellemann-Jensen P (1991) The social costs of smoking revisited. British Journal of Addiction 86:957–966

Ellsberg D (1961) Risk, ambiguity, and the savage axioms. Q J Econ 75(4):643–669.

Elster J (1985) Sour grapes. Cambridge University Press, Cambridge

Elster J, Roemer JE (1991) Interpersonal comparisons of well-being. Cambridge University Press, Cambridge

Emmerson C, Wakefield M (2001) The saving gateway and the child trust fund: is asset-based welfare 'well fair'? The Institute For Fiscal Studies, Commentary, 85

Epley N, Gilovich T (2001) Putting adjustment back in the anchoring and adjustment heuristic differential processing of self-generated and experimenter-provided anchors. Psychol Sci 12:391–396

Erev I, Barron G (2005) On adaptation, maximization, and reinforcement learning among cognitive strategies. Psychol Rev 112(4):912–931

Feehan JP (1990) A simple-model for merit good arguments – comment. J Public Econ 43(1):127–129

Fehr E, Tyran JR (2005) Individual irrationality and aggregate outcomes. J Econ Perspect 19(4):43–66

Ferris SP, Haugen RA, Makhija AK (1988) Predicting contemporary volume with historic volume at differential price levels: evidence supporting the disposition effect. J Finan 43(3):677–697

Fiorito R, Kollintzas T (2004) Public goods, merit goods, and the relation between private and government consumption. Eur Econ Rev 48(6):1367–1398

Fischer GW, Carmon Z, Ariely D, Zauberman G (1999) Goal-based construction of preferences: task goals and the prominence effect. Manage Sci 45:1057–1075

Fischhoff B (2000) Informed consent for eliciting environmental values. Environ Sci Technol 34(8):1439–1444

Fischoff B (1991) Value elicitation: is there anything in there? Am Psychol 468:835–847

Fischoff B, Slovic P, Lichtenstein S (1982) Lay foibles and expert fables in judgments about risk. Am Stat 36(3):240–255

Fisher FM (1987) Household equivalence scales and interpersonal comparison. Rev Econ Stud 54:519–524.
Fisher AH (2000) Introduction to special issue on irreversibility. Resour Energy Econ 22:189–196
Fisher I (1930) The theory of interest. Macmillan, New York
Folkes VS (1988) The availability heuristic and perceived risk. J Consum Res 15(1):13–23
Fortin B, Lacroix G (1997) A test of the unitary and collective models of household labour supply. Econ J 107(443):933–955
Foster V, Bateman IJ, Harley D (1997) Real and hypothetical willingness to pay for environmental preservation: a non-experimental comparison. J Agric Econ 48(2):123–138
Fox CR, Tversky A (1995) Ambiguity aversion and comparative ignorance. Quart J Econ 110(3):585–603
Fox JA, Shogren JF, Hayes DJ, Kliebenstein JB (1998) CVM-X: calibrating contingent values with experimental auction markets. Am J Agric Econ 80(3):455–465
Franciosi R, Kujal P, Michelitsch R, Smith VL, Deng G (1996) Experimental tests of the endowment effect. J Econ Behav Organ 30:213–226
Frank RH (1987) If homo-economicus could choose his own utility function, would he want one with a conscience. Am Econ Rev 77(4):593–604.
Fraser CD (1995) Misperceived job hazards and welfare. J Public Econ 56(1):97–123
Fraser CD (1996) On tort as an implicit insurance system with state dependent utility the case of child mortality risk. Int Rev Law Econ 16(4):449–459
Frederick S, Loewenstein G, O'Donoghue T (2002) Time discounting and time preference: a critical review. J Econ Lit 40(2):351–401
Frey B, Eichenberger R (1994) Economic incentives transform psychological anomalies. J Econ Behav Organ 23(2):215–234
Frey B, Stutzer A (2004) Does the political process mitigate or accentuate individual biases due to mispredicting future utility? In: Behavioral public finance: toward a new agenda conference, University of Michigan, Ann Arbor, Michigan
Friedman D (1998) Monty hall's three doors: construction and deconstruction of a choice anomaly. Am Econ Rev 88(4):933–946
Frisch D, Baron J (1988) Ambiguity and rationality. J Behav Decis Making 1:149–157
Frykblom P, Shogren JF (2000) An experimental testing of anchoring effects in discrete choice questions. Environ Resour Econ 16(3):329–341
Fuchs G, Guesnerie R (1983) Structure of tax equilibria. Econometrica 51(2):403–434.
Gale D, Mas-Colell A (1975) An equilibrium existence theorem for a general model without ordered preferences. J Math Econ 2(1):9–15
Gasparini L (1997) Equality of opportunity and optimal cash and in-kind policies. Instituto y Universidad Torcuato Di Tella
Genesove D, Mayer C (2001) Loss aversion and seller behavior: evidence from the housing market. Quart J Econ 116(4):1233–1260
Georgescu-Roegen N (1936) The pure theory of consumer's behavior. Quart J Econ 50(4):545–593
Georgiou S, Langford IH, Bateman IJ, Turner RK (1998) Determinants of individuals' willingness to pay for perceived reductions in environmental health risks: a case study of bathing water quality. Environ Plann A 30:577–594
Gilovich T, Griffin DW, Kahneman D (2002) Heuristics and biases: the psychology of intuitive judgement. Cambridge University Press, Cambridge
Glaeser EL, Laibson D, Sacerdote B (2000) The economic approach to social capital. National Bureau of Economic Research, Cambridge, Mass, USA
Glaeser EL, Luttmer EFP (2003) The misallocation of housing under rent control. Am Econ Rev 93(4):1027–1046
Gneezy U, Rustichini A (2000) Pay enough or don't pay at all. Q J Econ 115(3):791–810.
Gneezy U, Potters J (1997) An experiment on risk taking and evaluation periods. Quart J Econ 112(2):631–645
Godwin RK (1991) Charges for merit goods: third world family planning. J Public Policy 11(4):414–429

Goldstein D, Gigerenzer G (2002) Models of ecological rationality: the recognition heuristic. Psychol Rev 109(1):75–90
Gorringe P (1998) Background to the policy implications of limited rationality. In: Grimes A, Jones A, Procter R, Scobie G (eds) Economics for policy: expanding the boundaries. Institute of Policy Studies, Wellington.
Gorter J (2000) Melioration, matching and rational choice: a study on the interface between economics and psychology. PhD Thesis, University of East Anglia
Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford CA
Gourville JT, Soman D (1998) Payment depreciation: the behavioral effects of temporally separating payments from consumption. J Consum Res 25:160–74.
Green D, Jacowitz K, Kahneman D, McFadden D (1998) Referendum contingent valuation, anchoring, and willingness to pay for public goods. Resour Energy Econ 20(2):85–116
Green L, Fry A (1994) Discounting of delayed rewards – a lifespan comparison. Psychol Sci 5:33–36
Green L, Myerson J, Lichtman D, Rosen S, Fry A (1996) Temporal discounting in choice between delayed rewards: the role of age and income. Psychol Aging 11:79–84
Greenley D, Walsh RG, Young RA (1981) Option value: empirical evidence from a case study of recreation and water quality. Quart J Econ 96(4):657–673
Greenwald A (1980) The totalitarian ego. Am Psychol 35(7):603–618
Gregory R, Wellman K (2001) Bringing stakeholder values into environmental policy choices: a community-based estuary case study. Ecol Econ 39(1):37–52
Gregory RS (2000) Valuing environmental policy options: a case study comparison of multiattribute and contingent valuation survey methods. Land Econ 76:151–173
Gregory RS, Slovic P (1997) Commentary a constructive approach to environmental valuation. Ecol Econ 21:175–181
Grether DM (1980) Bayes' rule as a descriptive model: the representativeness heuristic. Quart J Econ 95:537–557
Grether DM, Plott CR (1979) Economic theory of choice and the preference reversal phenomenon. Am Econ Rev 69(4):623–638
Grether DM, Plott CR (1982) Economic theory of choice and the preference reversal phenomenon: reply. Am Econ Rev 72:575
Greven PJ (1970) Four generations: population, land, and family in colonial andover. Cornell University Press, New York
Grinblatt M, Han B (2005) Prospect theory, mental accounting, and momentum. J Finan Econ 78(2):311–339
Guala F (2000) Artefacts in experimental economics: preference reversals and the Becker-DeGroot-Marschak mechanism. Econ Philos 16(1):47–75
Guesnerie R (1995) The theory of taxation. Econometric society monographs. Cambridge University Press, Cambridge.
Guter B, Priddat BP (1992) [On the economy of community needs: new attempts towards an ethical basis for the theory of merit goods]. Zur okonomie der gemeinschaftsbedurfnisse neuere versuche einer ethischen begrundung der theorie meritorischer. Zeitschrift fur Wirtschafts und Sozialwissenschaften 112(2):239–259
Haigh MS, List JA (2005) Do professional traders exhibit myopic loss aversion? An experimental analysis. J Finan 60(1):523–534
Hakes J, Viscusi W (1997) Mortality risks: a Bayesian reassessment. J Risk Uncertain 15:135–150
Halvorsen B, Soelensminde K (1998) Differences between willingness-to-pay estimates from open-ended and discrete-choice contingent valuation methods the effects of heteroscedasticity. Land Econ 74(2):262–282
Hammack J, Brown GM (1974) Waterfowl and wetlands towards bio-economic analysis. John Hopkins University Press, Baltimore
Hammitt JK (2002) QALYs versus WTP. Risk Anal 22:985–1001

Hammond PJ (1976) Endogenous tastes and stable long-run choice. J Econ Theory 13(2):329–340
Hammond PJ (1981) Ex-ante and ex-post welfare optimality under uncertainty. Economica 48(191):235–250
Hammond PJ (1988) Consequentialist foundations for expected utility. Theory Decis 25(1):25–78
Hanemann WM (1984) Welfare evaluations in contingent valuation experiments with discrete responses. Am J Agric Econ 66(3):332–341
Hanemann WM (1991) Willingness to pay and willingness to accept: how much can they differ? Am Econ Rev 81(3):635–647
Hanley ND, Macmillan DC, Philip L, Alvarez-Farizo B (2002) Valuing the non-market benefits of wild goose conservation: a comparison of interview and group based aroaches. Ecol Econ 43(1):49–59
Hanley ND, Shogren JF (2005) Is cost-benefit analysis anomaly-proof? Environ Resour Econ 32(1):13–24
Harbaugh WT, Krause K, Berry TR (2002) GARP for kids on the development of rational choice behaviour. Am Econ Rev 91(5):1539–1545
Harbaugh WT, Krause K, Vesterlund L (2001) Are adults better behaved than children? Age, experience, and the endowment effect. Econ Lett 70(2):175–181
Hardie BGS, Johnson EJ, Fader PS (1993) Modeling loss aversion and reference dependence effects on brand choice. Mark Sci 12(4):378–394
Harris J, Blair EA (2006) Consumer preference for product bundles: the role of reduced search costs. J Acad Mark Sci 34(4):506–513
Harrison GW (1994) Expected utility theory and the experiments. Empir Econ 19(2):223–253
Harrison GW, Johnson E, McInnes MM, Rutström EE (2004) Controls for risk aversion in the laboratory. University of Central Florida, Department of Economics
Harrison GW, Lau MI, Rutström EE (2005) Dynamic consistency in Denmark: a field experiment. Working Paper 5-02, University of Central Florida, Department of Economics
Harrison GW, Lau MI, Williams MB (2002) Estimating individual discount rates in denmark: a field experiment. Am Econ Rev 92(5):1606–1617
Harrison GW, List JA (2004) Field experiments. J Econ Lit 42(4):1013–1059
Harsanyi JC (1955) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. J Polit Econ 63(4):309–321
Hartman RS, Doane MJ, Woo CK (1991) Status quo bias in the measurement of value in service. Resour Energy 12:197–214
Hartog J, Ferrer-i-Carbonell A, Jonker N (2002) Linking measured risk aversion to individual characteristics. Kyklos 55(1):3–26
Hayes DJ, Shogren JF, Shin SY, Kliebenstein JB (1995) Valuing food safety in experimental auction markets. Am J Agric Econ 77(1):40–53
Head JG (1966) On merit goods. Finanzarchiv 25(1):1–29.
Head JG (1990) On merit wants: reflections on the evolution, normative status and policy relevance of a controversial public finance concept. In: Brennan G, Walsh C (eds) Rationality, individuals and public policy. Australian National University, Canberra, pp 211–244
Head JG (1991) Merit wants: analysis and taxonomy. In: Eden L (ed) Retrospectives on public finance. Duke University Press, Durham, NC
Head JG, Shoup CS (1969) Public goods, private goods and ambiguous goods. Econ J 79(315):567–572
Heath C (1995) Escalation and de-escalation of commitment in response to sunk costs – the role of budgeting in mental accounting. Organ Behav Human Decis Proc 62(1):38–54
Heath C, Ocurry S (1994) Mental accounting and consumer spending. Adv Consum Res 21:119–134
Heath TB, Chatterjee S, France KR (1995) Mental accounting and changes in price – the frame dependence of reference dependence. J Consum Res 22(1):90–97
Heath C, Tversky A (1991) Preference and belief: ambiguity and competence in choice under uncertainty. J Risk Uncertain 4:5–28

Herne K (1998) Testing the reference-dependent model: an experiment on asymmetrically dominated reference points. J Behav Decis Making 11(3):181–192
Herriges JA, Shogren JF (1996) Starting point bias in dichotomous choice valuation with follow-up questioning. J Environ Econ Manage 30(1):112–131
Herrnstein RJ, Rachlin H, Laibson DI (1997) The matching law: papers in psychology and economics. Blackwell Synergy, Oxford
Hertwig R, Fanselow C, Hoffrage U (2003) Hindsight bias: how knowledge and heuristics affect our reconstruction of the past. Memory 11(4):357–377
Hey JD (2001) Does repetition improve consistency? Exp Econ 4(1):5–54
Hey JD (2005) Why we should not be silent about noise. Exp Econ 8(4):325–345
Highhouse S, Johnson MA (1996) Gain/loss asymmetry and riskless choice: loss aversion in choices among job finalists. Organ Behav Human Decis Proc 68(3):225–233
Hillman AL (1980) Notions of merit want. Public Finance-Finances Publiques 35(2):213–226
Hines JR, Thaler RH (1995) Anomalies – the flypaper effect. J Econ Perspect 9(4):217–226
Ho TH, Lim N, Camerer CF (2006) Modeling the psychology of consumer and firm behavior with behavioral economics. J Mark Res 43(3):307–331
Hochman HM, Rodgers JD (1969) Pareto optimal redistribution. Am Econ Rev 59(4):542–557
Hochman HM, Rodgers JD (1970) Pareto optimal redistribution: reply. Am Econ Rev 60(5):997–1002
Hoehn J, Randall A (1987) A satisfactory benefit cost indicator from contingent valuation. J Environ Econ Manage 14:226–247
Hoffrage U, Hertwig R, Gigerenzer G (2000) Hindsight bias: a by-product of knowledge updating. J Exp Psychol: Learn Mem Cogn 26(3):566–581
Holmes T, Kramer RA (1995) An independent sample test of yea-saying and starting point bias in dichotomous-choice contingent valuation. J Environ Econ Manage 29:121–132
Holt CA (1986) Preference reversals and the independence axiom. Am Econ Rev 76:508–515
Holt CA (1995) Industrial organization: a survey of laboratory research. In: Roth A, Kagel J (eds) Handbook of experimental economics. Princeton University Press, Princeton, pp 349–443
Horowitz JK, McConnell KE (2002) A review of WTA/WTP studies. J Environ Econ Manage 44(3):426–447
Horowitz JK, McConnell KE (2003) Willingness to accept, willingness to pay, and the income effect. J Econ Behav Organ 51(4):537–545
Hsee CK (1996) The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. Organ Behav Human Decis Proc 67(3):247–257
Hsee CK, Leclerc F (1998) Will products look more attractive when presented separately or together? J Consum Res 25(2):175–186
Hsee CK, Loewenstein GF, Blount S, Bazerman MH (1999) Preference reversals between joint and separate evaluations of options: a review and theoretical analysis. Psychol Bull 125(5):576–590
Hsee CK, Weber EU (1999) Cross-national differences in risk preference and lay predictions. J Behav Decis Making 12(2):165–179
Hsiao WC (1995) Abnormal economics in the health sector. Health Policy 32(1–3):125–139
Hubin DC (1994) The moral justification of benefit/cost analysis. Econ Philos 10:169–194
Huck S, Oechssler J, Kirchsteiger G (1997) Learning to like what you have: explaining the endowment effect. Wirtschaftswissenschaftliche Fakultät, Humboldt-Universität zu Berlin Wirtschaftswissenschaftliche Fakultät
Huffman D, Barenstein M (2004) Riches to rags every month? The fall in consumption expenditures between paydays. IZA Discussion Paper 1430, IZA Bonn
Humphrey SJ (2006) Does learning diminish violations of independence, coalescing and monotonicity? Theory Decis 61(2):93–128
Humphrey SJ, Verschoor A (2004) Decision-making under risk among small farmers in East Uganda. J Afr Econ 13(1):44–101
Hurd M, McFadden D, Chand H, Gan L, Merrill A, Roberts M (1997) Consumption and savings balances of the elderly: experimental evidence on survey response bias. National Bureau of Economic Research Working Paper, Cambridge, USA

Huysentruyt M, Read D (2005) Insurance decisions of the very poor: a field study. mimeo, London School of Economics
Ireland NJ (1994) On limiting the market for status signals. J Public Econ 53(1):91–110
Irwin JR, Slovic P, Lichtenstein S, McClelland GH (1993) Preference reversals and the measurement of environmental values. J Risk Uncertain 6:5–18
Isenberg DJ (1986) Group polarisation: a critical review and meta-analysis. J Pers Soc Psychol 50(6):1141–1151
Itaoka K, Saito A, Krupnick A, Adamowicz W, Taniguchi T (2006) The effect of risk characteristics on the willingness to pay for mortality risk reductions from electric power generation. Environ Resour Econ 33(3):371–398
Itoh H (1994) Job design, delegation and cooperation a principal agent analysis. Eur Econ Rev 38:691–700
Jackson SB, Shoemaker PA, Barrick JA, Burton FG (2005) Taxpayers' prepayment positions and tax return preparation fees. Contemp Account Res 22(2):409–447
Johannesson M, Liljas JB, O'Conor RM (1997) Hypothetical versus real willingness to pay some experimental results. Appl Econ Lett 4(3):149–159
Johnson EJ, Hershey J (1993) Framing, probability distortions and insurance decisions. J Risk Uncertain 7:35–51
Jolls C, Sunstein CR, Thaler R (1998) A behavioral approach to law and economics. Stanford Law Rev 50(5):1471–1550
Jones M, Sugden R (2001) Positive confirmation bias in the acquisition of information. Theory Decis 50:59–99
Jones PR, Cullis J (2000) 'Individual failure' and the analytics of social policy. J Soc Policy 29:73–93
Jones PR, Cullis J (2002) Merit want status and motivation: the knight meets the self-loving butcher, brewer, and baker. Public Finan Rev 30(2):83
Jones PR, Cullis J (2003) Key parameters in policy design: the case of intrinsic motivation. J Soc Policy 32(04):527–547
Jones PR, Cullis JG, Lewis A (1998) Public versus private provision of altruism: can fiscal policy make individuals 'better' people? Kyklos 51(1):3–24
Jones-Lee MW (1992) Paternalistic altruism and the value of statistical life. Econ J 102(410):80–90
Jones-Lee MW, Hammerton M, Philips PR (1985) The value of safety results of a national sample survey. Econ J 95:49–72
Joyce EJ, Biddle GC (1981) Anchoring and adjustment in probabilistic inference in auditing. J Account Res 19(1):120–145
Kachelmeier SJ, Shehata M (1992) Culture and competition: a laboratory market comparison between China and the West. J Econ Behav Organ XIV:145–168.
Kagel JH (1995) Cross-game learning: experimental evidence from first-price and english common value auctions. Econ Lett 49:163–170
Kahneman D (2000a) Evaluation by moments past and future. In: Kahneman D, Tversky A (eds) Choices, values and frames. Cambridge University Press, Cambridge, chap 37, pp 693–708
Kahneman D (2000b) Experienced utility and objective happiness a moment-based approach. In: Kahneman D, Tversky A (eds) Choices, values and frames. Cambridge University Press, Cambridge, chap 37, pp 673–692
Kahneman D (2003a) Maps of bounded rationality: psychology for behavioral economics. Am Econ Rev 93(5):1449–1475
Kahneman D (2003b) A psychological perspective on economics. Am Econ Rev 93(2):162–168
Kahneman D (2006) Maps of bounded rationality: psychology for behavioural economics. Psikhologicheskii Zhurnal 27(2):5–28
Kahneman D, Knetsch JL (1992) Valuing public goods: the purchase of moral satisfaction. J Environ Econ Manage 22:57–70.
Kahneman D, Knetsch JL, Thaler RH (1986a) Fairness and the assumptions of economics. J Bus 59(4):285–300

Kahneman D, Knetsch JL, Thaler RH (1986b) Fairness as a constraint on profit seeking: entitlements in the market. Am Econ Rev 76(4):728–741
Kahneman D, Knetsch J, Thaler R (1990) Experimental tests of the endowment effect and the Coase theorem. J Polit Econ 98(6):1325–1348
Kahneman D, Knetsch JL, Thaler RH (1991) Anomalies – the endowment effect, loss aversion, and status-quo bias. J Econ Perspect 5(1):193–206
Kahneman D, Ritov I (1994) Determinants of stated willingness to pay for public goods: a study in the headline method. J Risk Uncertain 9(1):5–37
Kahneman D, Slovic P, Tversky A (1982) Judgment under uncertainty: heuristics and biases. Cambridge University Press, Cambridge
Kahneman D, Sugden R (2005) Experienced utility as a standard of policy evaluation. Environ Resour Econ 32(1):161–181
Kahneman D, Thaler R (2006) Utility maximization and experienced utility. J Econ Perspect 20(1):221–234
Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47(2):263–291.
Kahneman D, Tversky A (eds) (2000) Choices, values, and frames. Cambridge University Press, Cambridge
Kanbur R, Pirttilä J, Tuomala M (2004) Moral hazard, income taxation, and prospect theory. University of Tampere, mimeo
Kanbur R, Pirttilä J, Tuomala M (2006) Non-welfarist optimal taxation and behavioral public economics. J Econ Surv 20(5):849–868
Kanninen BJ (1995) Bias in discrete response contingent valuation. J Environ Econ Manage 28:114–125
Kaplow L, Shavell S (2001, April) Any non-welfarist method of policy assessment violates the pareto principle. J Polit Econ 109:281–286
Karni E, Safra Z (1987) Preference reversal' and observability of preferences by experimental methods. Econometrica 55:675–685
Kealy MJ, Dovido JF, Rockel M (1988) Accuracy in valuation is a matter of degree. Land Econ 64:158–171
Kemp MC, Ng YK (1976) On the existence of social welfare functions, social orderings and social decision functions. Economica 43(169):59–66
Kempson E, McKay S, Collard S (2005) Incentives to save: encouraging saving among low-income households. Final report on the Saving Gateway pilot project. University of Bristol, Personal Finance Research Centre, March
Kennedy CA (2002) Revealed preference valuation compared to contingent valuation: radon-induced lung cancer prevention. Health Econ 117:585–598
Keren G, Gerritsen LEM (1999) On the robustness and possible accounts of ambiguity aversion. Acta Psychologica 103(1):149–172
Kerr NL, MacCoun RJ, Kramer GP (1996) Bias in judgment: comparing individuals and groups. Psychol Rev 1034:687–719
Kermer DA, Driver-Linn E, Wilson TD, Gilbert DT (2006) Loss aversion is an affective forecasting error. Psychol Sci 17(8):649–653
Keynes JM (1921) A treatise on probability. Macmillan, London
Kilka M, Weber M (2001) What determines the shape of the probability weighting function under uncertainty? Manage Sci 47(12):1712–1726
Kind P, Rosser RM, Williams A (1982) Valuation of quality of life: some psychometric evidence. In: Jones-Lee M (ed) The value of life and safety. North-Holland Publishing Company, Amsterdam
Kleinig J (1984) Paternalism. Rowman & Allanheld, Totowa, NJ
Knetsch JL (1989) The endowment effect and evidence of nonreversible indifference curves. Am Econ Rev 79:1277–1284.
Knetsch JL, Sinden JA (1984) Willingness to pay and compensation demanded: experimental evidence of an unexpected disparity in measures of value. Quart J Econ 99(3):507–521

Knetsch JL, Tang FF, Thaler RH (2001) The endowment effect and repeated market trials: is the Vickrey auction demand revealing? Exp Econ 4(3):257–269
Knez P, Smith VL, Williams AW (1985) Individual rationality, market rationality, and value estimation. Am Econ Rev 75(2):397–402
Kobberling V, Wakker PP (2005) An index of loss aversion. J Econ Theory 122(1):119–131
Kocher MG, Sutter M (2005) The decision maker matters individual versus group behaviour in experimental beauty contest games. Econ J 115(1):200–223
Kollman K, Miller J, Page S (1992) Adaptive parties in spatial elections. Am Polit Sci Rev 86(4):929–937
Kolstad CD, Guzman RM (1995) Information and the divergence between willingness-to-accept and willingness-to-pay. University of California, Santa Barbara, Department of Economics
Kontogianni A, Skourtos M, Langford I, Bateman I, Georgiou S (2001) Integrating stakeholder analysis in non-market valuation of environmental assets. Ecol Econ 37(1):123–138
Kontoleon A, Swanson T (2003) The willingness to pay for property rights for the giant panda: can a charismatic species be an instrument for nature conservation? Land Econ 79(4):483–499
Kooreman P, (2000) The labelling effect of a child benefit system. Am Econ Rev 90:571–583
Korobkin R (2003) The endowment effect and legal analysis. Northwest Univ Law Rev 97(3):1227–1293
Kovalchik S, Camerer CF, Grether DM, Plott CR, Allman JM (2004) Aging and decision making: a broad comparative study of decision behavior in neurologically healthy elderly and young individuals. J Econ Behav Organ 58(1):79–94
Krishna A, Slemrod J (2003) Behavioral public finance: tax design as price presentation. Int Tax Public Finan 10:189–203
Kriström B (1993) Comparing continuous and discrete contingent valuation questions. Environ Resour Econ 31:63–71
Krupnick AJ, Cropper ML (1992) The effect of information on health risk valuations. J Risk Uncertain 5(1):29–48
Kuhberger A, Schulte-Mecklenbeck M, Perner J (2002) Framing decisions: hypothetical and real. Organ Behav Human Decis Proc 89(2):1162–1175
Laibson D (1996) Hyperbolic discount functions, undersaving, and savings policy. NBER Working Paper, 5635
Laibson DI (1997) Golden eggs and hyperbolic discounting. Quart J Econ 112(2):443–477
Laibson DI, Repetto A, Tobacman J, Hall R, Gale W, Akerlof G (1998) Self-control and saving for retirement. Brookings Pap Econ Act 1998(1):91–196
Lake IR, Bateman IJ, Parfitt JP (1996) Assessing a kerbside recycling scheme: a quantitative and willingness to pay case study. J Environ Manage 46(3):239–254
Langer T, Weber M (2001) Prospect theory, mental accounting, and differences in aggregated and segregated evaluation of lottery portfolios. Manage Sci 47(5):716–733
Langer T, Weber M (2005) Myopic prospect theory vs. myopic loss aversion: how general is the phenomenon? J Econ Behav Organ 56(1):25–38
Langford IH, Bateman IJ, Jones AP, Langford HD, Georgiou S (1998) Improved estimation of willingness to pay in dichotomous choice contingent valuation studies. Land Econ 74(1):65–75
Langford IH, Georgiou S, Day RJ, Bateman IJ (1999) Comparing perceptions of risk and quality with willingness to pay: a mixed methodological study of public preferences for reducing health risks from polluted coastal bathing waters. Risk Decis Policy 4(3):201–220
Layard R (2006) Happiness: Lessons from a New Science. Penguin: London, UK.
Lazo JK, McClelland GH, Schulze WD (1997) Economic theory and psychology of non use values. Land Econ 73(3):358–371
Lefcourt HM (1982) Locus of control: current trends in theory and research. Erlbaum, Hillsdale, NJ.
Levy M, Levy H (2003) Prospect theory: much ado about nothing? Manage Sci 48(10):1334–1349
Lewbel A (1985) A unified approach to incorporating demographic or other effects into demand systems. Rev Econ Stud 52(1):1–18.

Lewbel A (2000) Identification of the binary choice model with misclassification. Econometric Theory 16(04):603–609
Li CZ, Löfgren KG, Hanemann WM (1996) Real versus hypothetical willingness to accept: the Bishop and Heberlein model revisited. CUDARE Working Paper, 793
Li S (2006) Preference reversal: a new look at an old problem. Psychol Rec 56(3):411–428
Lichtenstein S, Slovic P (1971) Reversals of preference between bids and choices in gambling decisions. J Exp Psychol 89:46–55
Lim SS (2006) Do investors integrate losses and segregate gains? Mental accounting and investor trading decisions. J Bus 79(5):2539–2573
Lindman H (1971) Inconsistent preferences among gambles. J Exp Psychol 80:590–597
List JA (2001) Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards. Am Econ Rev 91(5):1498–1507
List JA (2002) Preference reversals of a different kind: the "more is less" phenomenon. Am Econ Rev 92(5):1636–1643
List JA (2003) Does market experience eliminate market anomalies? Quart J Econ 118(1):41–71
List JA (2004) Neoclassical theory versus prospect theory: evidence from the marketplace. Econometrica 72(2):615–625
List JA, Gallet CA (2001) What experimental protocol influence disparities between actual and hypothetical stated values? Environ Resour Econ 20(3):241–254
List JA, Lucking-Reiley D (2000) Demand reduction in multiunit auctions: evidence from a sportscard field experiment. Am Econ Rev 90(4):961–972
List JA, Shogren JF (1998) Calibration of the difference between actual and hypothetical valuations in a field experiment. J Econ Behav Organ 37(2):193–205
List JA, Shogren JF (1999) Price information and bidding behavior in repeated second-price auctions. Am J Agric Econ 81(4):942–949
List JA, Shogren JF (2002) Calibration of willingness-to-accept. J Environ Econ Manage 43(2):219–233
Loewenstein G (1999) Experimental economics from the vantage-point of behavioural economics. Econ J 109(453):25–34
Loewenstein G, Adler D (1995) A bias in the prediction of tastes. Econ J 105:929–937.
Loewenstein G, O'Donoghue T (2004) Animal spirits: affective and deliberative processes in economic behavior. Soc Sci Res Network Available at SSRN: http://ssrn.com/abstract=539843
Loewenstein G, Weber E, Hsee CK, Welch N (2001) Risk as feelings. Psychol Bull 127(2):267–286
Loewenstein G, O'Donoghue T, Rabin M (2002) Projection bias in predicting future utility. University of California-Berkeley, mimeo
Loomes G (2005) Modelling the stochastic component of behaviour in experiments: some issues for the interpretation of data. Exp Econ 8(4):301–323
Loomes G, Moffatt P, Sugden R (2002) A microeconometric test of alternative stochastic theories of risky choice. J Risk Uncertain 24(2):103–130
Loomes G, Starmer C, Sugden R (1989) Preference reversal information processing effect or rational non-transitive choice? Econ J 99:140–151
Loomes G, Starmer C, Sugden R (1991) Observing violations of transitivity by experimental methods. Econometrica 59(2):425–439
Loomes G, Sugden R (1982) Regret theory: an alternative theory of rational choice under uncertainty. Econ J 92(368):805–824
Loomes G, Sugden R (1983) A rationale for preference reversal. Am Econ Rev 73(3):428–432
Loomes G, Sugden R (1986) Disappointment and dynamic consistency in choice under uncertainty. Rev Econ Stud 53(2):271–282
Loomes G, Sugden R (1995) Incorporating a stochastic element into decision theories. Eur Econ Rev 39(3):641–648
Loomes G, Sugden R (1998) Testing different stochastic specifications of risky choice. Economica 65(260):581–598
Lopes LL (1994) Psychology and economics: perspectives on risk, cooperation, and the marketplace. Ann Rev Psychol 45:197–227

Lord CG, Ross L, Leer MR (1979) Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. J Pers Soc Psychol 37:2098–2109
Lovallo D, Kahneman D (2000) Living with uncertainty: attractiveness and resolution timing. J Behav Decis Making 13(2):179–190
Lunt PK, Livingstone SM (1991) Psychological social and economic determinants of savings: comparing recurrent and total savings. J Econ Psychol 12:621–641
Lyssiotou P (2005) Reconsidering the labeling effect of a child benefit system. Unpublished paper, University of Cyprus.
MaClean A (2000) Now you see it, now you don't: consent and the legal protection of autonomy. J Appl Philosophy 17(3)
Macmillan D, Hanley ND, Lienhoop N (2005) Contingent valuation: environmental polling or preference engine? University of Stirling, Department of Economics, Scotland
Madge C (1941) War economics and war potential: the propensity to save in Blackburn and Bristol. Econ J 12:621–641
Malinowski (1961) Argonauts of the West Pacific. Dutton, New York
Mandler M (2004) Status quo maintenance reconsidered: changing or incomplete preferences? Econ J 114(499):F518
Mankiw N (1985) Small menu costs and large business cycles: a macroeconomic model of monopoly. Quart J Econ 100(2):529–537
Mann S (2003) Why organic food in Germany is a merit good. Food Policy 28(5–6):459–469
Mann S (2006) Merit goods in a utilitarian framework. Rev Polit Econ 18(4):509–520
del Mar Racionero M (2001) Optimal tax mix with merit goods. Oxford Economic Papers 53:628–641
Marschak TA (1978) On the study of taste changing policies. Am Econ Rev 68(2):386–391
Marshall A (1890) Principles of economics. Macmillan and Co., London
Masatlioglu Y, Ok E (2005) Rational choice with status-quo bias. J Econ Theory 121:1–29
Mayer T, Russell T (2005) Income smoothing and self-control: the case of schoolteachers. Econ Inq 43(4):823
McCaffery EJ, Baron J (2003) The Humpty Dumpty blues: disaggregation bias in the evaluation of tax systems. Organ Behav Human Decis Proc 91(2):230–242
McCaffery EJ, Baron J (2004) Framing and taxation: evaluation of tax policies involving household composition. J Econ Psychol 25:679–670
McCaffery EJ, Baron J (2005) Isolation effects and the neglect of indirect effects of fiscal policies. mimeo and University of Southern California Law School
McClure CE Jr (1968) Merit wants: a normatively empty box. Finanzarchiv 27(3):474–483
McClure CE Jr (1990) Merit wants. In: Brennan G, Welsh CE (eds) Rationality, individualism and public policy. Centre for Research on Federal Financial Relations, ANU, Canberra, pp 178–185
McDaniels TL (1992) Reference points, loss aversion, and contingent values for auto safety. J Risk Uncertain 5(2):187–200
McDaniels TL, Gregory RA, Arvai J, Cheunpagdee R (2003) Decision structuring to alleviate embedding in environmental valuation. Ecol Econ 46:33–46
McFadden D (1999) Rationality for economists? J Risk Uncertain 19(1):73–105
McGraw AP, Tetlock PE, Kristel OV (2003) The limits of fungibility: relational schemata and the value of things. J Consum Res 30(2):219–229
McLean IS, List C, Fishkin JS, Luskin RC (2000) Does deliberation induce preference structuration? Evidence from Deliberative Opinion Polls. University of Oxford
McLennen EF (1990) Rationality and dynamic choice: foundational explorations. Cambridge University Press, Cambridge
Mehre R, Prescott EC (1985) The equity premium: a puzzle. J Monet Econ 15(2):145–61.
Melumad ND, Mookherjee D, Reichelstein S (1995) Hierarchical decentralization of incentive contracts. Rand J Econ 26:654–672
Milkman KL, Beshears J, Rogers T, Bazerman MH (2007) Mental accounting and small windfalls: evidence from an online Grocer. Harvard NOM Working Paper No. 08-024

Mill JS (1859) On liberty J. W. Parker, London.
Mill JS (1891) Principles of political economy: with some of their applications to social philosophy. Routledge, London
Mills G, Patterson R, Orr L, DeMarco D (2004) Evaluation of the American Dream demonstration: final evaluation report. American Dream
Milne F, Shefrin HM (1988) Ex post efficiency and ex post welfare some fundamental considerations. Economica 55(217):63–79
Mitchell RC, Carson RT (1989) Using surveys to value public goods: the contingent valuation method. Resources for the Future, Washington
Monga A, Zhu R (2004) What makes you happier – a nonloss or a gain? The moderating role of regulatory focus. Adv Consum Res 31:33–34
Moore A, Beverly S, Schreiner M, Sherraden M, Lombe M, Cho EYN, Johnson L, Vonderlack R (2001) Saving, IDA programs, and effects of IDAs: a survey of participants. Center for Social Development, Washington University, St Louis, Missouri
Morrison GC (1997) Resolving differences in willingness to pay and willingness to accept: comment. Am Econ Rev 87(1):236–240.
Morrison, GC (1998) Understanding the disparity between WTP and WTA: endowment effect, substitutability, or imprecise preferences? Econ Lett 59(2):189–194.
Morrison GC (2000) The endowment effect and expected utility. Scot J Polit Econ 47(2):183–197
Mulligan C, Philipson T (2000) Merit motives and government intervention: public finance in reverse. University of Chicago - Population Research Center Working Paper 2000–03, Chicago - Population Research Center.
Munnich K (2006) An axiomatic characterization of value judgments relative to a reference point. Math Soc Sci 51(1):58–69
Munro A (1989) Merit wants, rationing and in kind distribution. Department of Economics, University of Stirling
Munro A (1992) In-kind distribution, uncertainty, and merit wants: a simple model. Public Finan Rev 20(2):175
Munro A (1997) Loss aversion and welfare economics. Working Paper, Economics Research Centre, University of East Anglia.
Munro A (2005) The economics of the winter fuel allowance. Royal Holloway, University of London
Munro A (2006) Optimal decision processes: on the choice between environmental valuation methods. Royal Holloway, University of London
Munro A, Ferreira de Souza Y (2007) Truck, barter and exchange versus the endowment effect: a virtual field experiment. Royal Holloway, University of London
Munro A, Sugden R (1997) A theory of general equilibrium with reference- dependent preferences. Mimeo, University of East Anglia.
Munro A, Hanley N (1999) Information, uncertainty and contingent valuation. In: Bateman I, Willis K (eds) Valuing environmental preferences. Oxford University Press, Oxford.
Munro A, Sugden R (2003) On the theory of reference-dependent preferences. J Econ Behav Organ 50(4):407–428
Musgrave RA (1986) Public finance in a democratic society: collected papers of Richard A. Musgrave. New York University Press, New York
Musgrave RA (1987) Merit goods. In: The new Palgrave: a dictionary of economics. Macmillan, London
Mussweiler T, Strack F (2000) Numeric judgments under uncertainty: the role of knowledge in anchoring. J Exp Soc Psychol 36:495–518
Myagkov M, Plott CR (1997) Exchange economies and loss exposure: experiments exploring prospect theory and competitive equilibria in market environments. Am Econ Rev 87:801–828.
Narasimhan C, He C, Anderson ET, Brenner L, Desai P, Kuksov D, Messinger P, Moorthy S, Nunes J, Rottenstreich Y, Staelin R, Wu G, Zhang ZJ (2005) Incorporating behavioral anomalies in strategic models. Mark Lett 16(3–4):361–373

Neill H, Cummings R, Ganderton P, Harrison G, McGuckin T (1994) Hypothetical surveys and real economic commitments. Land Econ 70(2):145–154
Neilson WS (2002) Comparative risk sensitivity with reference-dependent preferences. J Risk Uncertain 24(2):131–142
Ng YK (1981) Welfarism: a defence against Sen's attack. Econ J 91(362):527–530
Ng YK (1992) Utilitarianism and interpersonal comparison. Soc Choice Welfare 9(1):1–15
Ng YK (1997) A case for happiness, cardinalism, and interpersonal comparability. Econ J 107(445):1848–1858
Nichols D, Smolensky E, Tideman T (1971) Discrimination by waiting time in merit goods. Am Econ Rev 61(3):312–323
Nisbett R, Ross L (1980) Human inference: strategies and shortcomings of human judgment. Prentice Hall, Englewood Cliffs, NJ
Novemsky N, Kahneman D (2005a) The boundaries of loss aversion. J Mark Res 42(2):119–128
Novemsky N, Kahneman D (2005b) How do intentions affect loss aversion? J Mark Res 42(2):139–140
Nunes JC (2001) Mental accounting: flexible accounts, order effects and incommensurable entries. Adv Consum Res 28:70–70
Nyborg K (2000) Homo economicus and homo politicus interpretation and aggregation of environmental values. J Econ Behav Organ 42:305–322
Odean T (1998) Are investors reluctant to realize their losses? J Financ 53(5):1775–1798
O'Donoghue T, Rabin M (2005) Optimal taxes for sin goods. Swedish Econ Policy Rev 12:7–39
Okada EM (2001) Trade-ins, mental accounting, and product replacement decisions. J Consum Res 27(4):433–446
Olsen JA, Smith RD (2001) Theory versus practice: a review of 'willingness-to-pay' in health and health care. Health Econ 10(1):39–52
Orne M (1962) On the social psychology of the psychological experiment. Am Psychol 17:776–783
Osborne M, Slivinski A (1996) A model of political competition with citizen-candidates. Quart J Econ 111(1):65–96
Ott RL, Andrus DM (2000) The effect of personal property taxes on consumer vehicle-purchasing decisions: a partitioned price/mental accounting theory analysis. Public Financ Rev 28(2):134–152
Palfrey T, Rosenthal H (1985) Voter participation and strategic uncertainty. Am Polit Sci Rev 79(1):62–78
Parkhurst GM, Shogren JF, Dickinson DL (2004) Negative values in Vickrey auctions. Am J Agric Econ 86(1):222–235
Pauly M (1967) Mixed public and private financing of education: efficiency and feasibility. Am Econ Rev 57(1):120–130
Pazner E (1972) Merit wants and the theory of taxation. Public Financ 27:460–472
Peart SJ (2000) Irrationality and intertemporal choice in early neoclassical thought. Can J Econ 33(1):175–189.
Perry CJ, Rock MT, Seckler DW (1997) Water as an economic good: a solution, or a problem? International Irrigation Management Institute, USA
Persson T, Tabellini G (2000) Political economics: explaining economic policy. MIT Press, Cambridge, USA
Peters HE, Unur AS, Clark J, Schulze WD (2004) Free-riding and the provision of public goods in the family: a laboratory experiment. Int Econ Rev 45(1):283–299
Pettibone JC, Wedell DH (2000) Examining models of nondominated decoy effects across judgment and choice. Organ Behav Human Decis Proc 81(2):300–328
Philip L, Hanley ND, Macmillan DC, Alvarez-Farizo B (2002) Valuing the non-market benefits of wild goose conservation: a comparison of interview and group based approaches. Ecol Econ 43(1):49–59
Pigou AC (1920) The economics of welfare. Macmillan, London.

Pinto P (1997) Is the person trade-off a valid method for allocating health care resources? Health Econ 6:71–81
Plott CR (1991) Will economics become an experimental science? South Econ J 57(4):901–919
Plott CR (1996) Rational individual behavior in markets and social choice processes: the discovered preference hypothesis. In: Arrow K, Colombatto E, Perleman M, Schmidt C (eds) Rational foundations of economic behaviour. Macmillan, London, pp 225–250
Plott CR, Zeiler K (2005) The willingness to pay-willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. Am Econ Rev 95(3):530–545
Plous S (1989) Thinking the unthinkable the effects of anchoring on likelihood estimates of nuclear war. J Appl Soc Psychol 19:67–91
Pollak RA (1970) Habit formation and dynamic demand functions. J Polit Econ 78(4):745–763
Pollak RA (1988) Tied transfers and paternalistic preferences. Am Econ Rev 78(2):240–244
Poterba J, Steven M, Venti F, Wise DA (1998) Lump sum distributions from retirement savings plans receipt and utilization. In: Wise DA (ed) Inquiries in the economics of aging. Chicago University Press, Chicago, pp 85–105
Prelec D (1998) The probability weighting function. Econometrica 60:497–528.
Prelec D, Loewenstein G (1998) The red and the black: mental accounting of savings and debt. Mark Sci 17(1):4–28
Prelec D, Simester D (2001) Always leave home without it: a further investigation of the credit-card effect on willingness to pay. Mark Lett 12(1):5–12
Raban DR, Rafaeli S (2003) Subjective value of information: the endowment effect. University of Haifa, Israel
Rabin M (1998) Psychology and economics. J Econ Lit 36(1):11–46
Rabin M (2000) Risk aversion and expected-utility theory: a calibration theorem. Econometrica 68(5):1281–1292
Rabin M, Schrag J (1999) First impressions matter: a model of confirmatory bias. Quart J Econ 114(1):37–82
Racionero MDM (2000) Optimal redistribution with unobservable preferences for an observable merit good. Int Tax Public Financ 7(4–5):479–501
Randall A, Stoll J (1980) Consumer's surplus in commodity space. Am Econ Rev 70(3):449–455
Ranyard R, Hinkley L, Williamson J, McHugh S (2006) The role of mental accounting in consumer credit decision processes. J Econ Psychol 27(4):571–588
Rawls J (1971) A theory of justice. Oxford University Press, Oxford
Read D (2002) Is time-discounting hyperbolic or subadditive? J Risk Uncertain 23(1):5–32.
Ready RC, Buzby J, Hu D (1996) Differences between continuous and discrete contingent value estimates. Land Econ 73:397–411
Reiling S, Boyle K, Phillips M, Anderson M (1990) Temporal reliability of contingent values. Land Econ 66(2):128–134
Rhodes B (1997) Experiments on loss aversion. PhD thesis, University of East Anglia
Richmond H (1998) Paternalism and consent some educational problems. New Zealand J Philos Educ 32(2):251
Richter WF, Weimann J (1991) Merit goods and social marginal utility of income. Jahrbuch Fur Sozialwissenschaft 42(1):118–130
Ritov I, Baron J (1992) Status-quo and omission biases. J Risk Uncertain 5(1):49–61
Robb AL, Burbidge JB (1989) Consumption, income, and retirement. Can J Econ 22(3):522–542
Rockenbach B (2004) The behavioral relevance of mental accounting for the pricing of financial options. J Econ Behav Organ 53(4):513–527
Rose R, Falconer (1992) Individual taste or collective decision public policy on school meals. J Soc Policy 21:349–373
Rosenzweig MR (1988) Risk, implicit contracts and the family in rural areas of low-income countries. Econ J 98(393):1148–1170
Ross TW (1991) On the relative efficiency of cash transfers and subsidies. Econ Enquiry 29:485–496.

Roth AE (1988) Laboratory experimentation in economics – a methodological overview. Econ J 98(393):974–1031
Rowe RD, d'Arge RC, Brookshire DS (1980) An experiment on the economic value of visibility. J Environ Econ Manage 7:1–19
Ryan A (1991) Merit goods and benefits in kind: paternalism and liberalism in action. In: Wilson T (ed) The state and social Welfare: the objectives of policy. Longman, Harlow, pp 89–104
Sagi JS (2006) Anchored preference relations. J Econ Theory 130(1):283–295
Salanie F, Treich N (2005) Regulation in Happyville. University of Toulouse, Working Paper
Salanie F, Treich N (2006) Regulating an agent with different beliefs. In: Boyer M, Hiriart Y, Martimort D (eds) Frontiers on regulation and liability. Ashgate Publishing Ltd., Sussex
Salkeld G, Ryan M, Short L (2000) The veil of experience: do consumers prefer what they know best? Health Econ 9(3):267–270
Samuelson P (1952) Probability, utility, and the independence axiom. Econometrica 20(4):670–678
Samuelson W, Zeckhauser R (1988) Status quo bias in decision making. J Risk Uncertain 1(1):7–59
Sandmo A (1983) Ex post welfare economics and the theory of merit goods. Economica 50(197):19–33
Savage LJ (1954) The foundations of statistics. Wiley, New York.
Sayman S, Onculer A (2005) Effects of study design characteristics on the WTA-WTP disparity: a meta-analytical framework. J Econ Psychol 26:289–312
Schelling TC (1981) Economic reasoning and the ethics of policy. Public Interest 63:37–61
Schepanski A, Kelsey D (1990) Testing for framing effects in taxpayer compliance decisions. J Am Taxation Assoc Fall:60–77.
Schkade D, Payne J (1994) How people respond to contingent valuation questions: a verbal protocol analysis of willingness to pay for an environmental regulation. J Environ Econ Manage 26(1):88–109
Schmidt K (1988) [More on merit goods: critical notes and alternatives to the theory of public goods]. Mehr zur meritorik kritisches und alternatives zu der lehre von den offentlichen gutern. Zeitschrift fur Wirtschafts und Sozialwissenschaften 108:383–403
Schmidt U, Hey JD (2004) Are preference reversals errors? An experimental investigation. J Risk Uncertain 29(3):207–218
Schmidt U, Traub S (2002) An experimental test of loss aversion. J Risk Uncertain 25(3):233–249
Schmidt U, Zank H (2005) What is loss aversion? J Risk Uncertain 30(2):157–167
Schneider S, Lopes L (1986) Reflection in preferences under risk: who and when may suggest why. J Exp Psychol: Human Percept Perform 12(4):535–548
Schreiner M (2005) Match rates, individual development accounts, and saving by the poor. J Income Distrib 13(3):112–129
Schreiner M, Sherraden M, Clancy M, Johnson L, Curley J, Grinstein-Weiss M, Zhan M, Beverly S (2001) Savings and asset accumulation in individual development accounts. Center for Social Development, George Warren Brown School of Social Work, Washington University, St Louis, MO
Schroyen F (2005) An alternative way to model merit good arguments. J Public Econ 89(5–6): 957–966
Schwarze J, Wagner G (1990) Preference research of merit goods – the case of old-age security in the Federal Republic of Germany. Jahrbucher Fur Nationalokonomie Und Statistik 207(5):464–481
Schweitzer ME, DeChurch LA, Gibson DE (2005) Conflict frames and the use of deception: are competitive negotiators less ethical? J Appl Soc Psychol 35(10):2123–2149
Schwert GW (2002) Anomalies and market efficiency. Simon School of Business Working Paper No. FR 02-13
Scott C (2003) Regulation in the age of governance: the rise of the post-regulatory state. In Jordana J, Levi-Faur D (eds) The politics of regulation in the age of governance. Edward Elgar, Cheltenham, pp 145–74.
Segal U (1988) Does the preference reversal phenomenon necessarily contradict the independence axiom? Am Econ Rev 78:233–235

Seidl C (2002) Preference reversal. J Econ Surv 16(5):621–655
Selden TM (1993) Should the government provide catastrophic insurance. J Public Econ 51(2):241–247
Sellar C, Stoll JR, , Chavas JP (1985) Validation of empirical measures of welfare change a comparison of non-market techniques'. Land Econ 61:156–175
Sen AK (1970) Collective choice and social welfare. Holden-Day, San Francisco
Sen AK (1977) Rational fools: a critique of the behavioral foundations of economic theory. Philos Public Aff 6(4):317–344
Sen AK (1986) Social choice theory. In: Arrow KJ, Intriligator MD (eds) Handbook of mathematical economics, vol 3. Elsevier, Amsterdam, pp 1073–1181
Shafer W, Sonnenschein H (1975) Equilibrium in abstract economies without ordered preferences. J Math Econ 2:345–348
Shafir E, Diamond P, Tversky A (1997) Money illusion. Quart J Econ 112(2):341–374
Shafir E, Thaler RH (2006) Invest now, drink later, spend never: on the mental accounting of delayed consumption. J Econ Psychol 27(5):694–712
Shaked A, Sutton J (1982) Relaxing price competition through product differentiation. Rev Econ Stud 49:3–13.
Shalev J (2002) Loss aversion and bargaining. Theory Decis 52(3):201–232
Shaughnessy J (1977) Misconceptions of probability: an experiment with a small-group, activity-based, model building approach to introductory probability at the college level. Educ Stud Math 8(3):295–316
Shefrin H (1999) Beyond greed and fear. Harvard Business School Press, Boston, MA
Shefrin H, Statman M (1985) The disposition to sell winners too early and ride losers too long: theory and evidence. J Financ 40(3):777–790
Shefrin H, Statman M (1993) Behavioral-aspects of the design and marketing of financial products. Financ Manage 22(2):123–134
Shefrin H, Statman M (2000) Behavioral portfolio theory. J Financ Quant Anal 35(2):127–151
Shefrin H, Thaler R (1988) The behavioral life-cycle hypothesis. Econ Inq 26(4):609–643
Sherraden M, (1991) Assets and the poor. ME Sharpe, Armonk, NY.
Shibata H (1988) Merit goods and public choice – the case of higher-education – comment. Public Finance-Finances Publiques 43(3):466–470
Shleifer A (2000) Inefficient markets: an introduction to behavioral finance. Oxford University Press, Oxford
Shogren JF, Margolis M, Koo C, List JA (2001) A random nth-price auction. J Econ Behav Organ 46(4):409–421
Shogren JF, Shin SY, Hayes DJ, Kliebenstein JB (1994) Resolving differences in willingness to pay and willingness to accept. Am Econ Rev 84(1):255–270
Siegel JJ, Thaler RH (1997) Anomalies – the equity premium puzzle. J Econ Perspect 11(1):191–200
Silberman J, Gerlowski DA, Williams NA (1992) Estimating existence value for users and nonusers of New Jersey beaches. Land Econ 68(2):225–236
Sileo P (1995) Intransitivity of indifference, strong monotonicity, and the endowment effect. J Econ Theory 66:198–223
Simon H (1955) A behavioral model of rational choice. Q J Econ 69(1):99–118.
Simon H (1985) Human nature in politics: the dialogue of psychology with political science. Am Polit Sci Rev 79(2):293–304
Simonsohn U, Loewenstein G (2006) Mistake #37: the effect of previously encountered prices on current housing demand. Econ J 116(508):175–199
Simonson I (1990) The effect of purchase quantity and timing on variety seeking behavior. J Mark Res 32:150–162
Sinclair WS (1976) The economic and social impact of Kemano II Hydroelectric Project on British Columbia's Fisheries Resources. Department of the Environment, Canada
Slembeck T, Tyran JR (2004) Do institutions promote rationality? An experimental study of the three-door anomaly. J Econ Behav Organ 54:337–350

Slovic P (1995) The construction of preference. Am Psychol 50(5):364–371
Slovic P, Griffin D, Tversky A (1990) Compatibility effects in judgment and choice. In: Hogarth R (ed) Insights in decision-making. University of Chicago Press, Chicago, chap 1, pp 5–27
Slovic P, Lichtenstein S (1971) Comparison of Bayesian and regression approaches to the study of information processing in judgement. Organ Behav Human Perform 6:649–744
Smith VL (1962) An experimental study of competitive market behavior. J Polit Econ 70:111–137.
Smith VL (2003) Constructivist and ecological rationality in economics. Am Econ Rev 93(3):465–508
Smith V, Walker JM (1993) Monetary rewards and decision cost in experimental economics. Econ Inq 31(2): 245–261.
Solow JL (1994) Paternalistic preferences, interpersonal transfer and reciprocity. South Econ J 61(2):379–386
Soman D (1997) Contextual effects of payment mechanism on purchase intention: check or charge. Unpublished Working Paper, University of Colorado
Soman D (2001) The mental accounting of sunk time costs: why time is not like money. J Behav Decis Making 14(3):169–185
Soman D, Cheema A (2001) The effect of windfall gains on the sunk-cost effect. Mark Lett 12(1):51–62
Spash CL (2000) Multiple value expression in contingent valuation economics and ethics. Environ Sci Technol 34(8):1433–1438
Starmer C (1999) Experimental economics: hard science or wasteful tinkering? Econ J 109(453):5–15
Starmer C (2000) Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. J Econ Lit 38(2):332–382
Starmer C, Sugden R (1991) Does the random-lottery incentive system elicit true preferences? An experimental investigation. Am Econ Rev 81(4):971–978
Starmer C, Sugden R (1993) Testing for juxtaposition and event-splitting effects. J Risk Uncertain 6(3):235–254
Statman M (2003) Behavioural portfolios: hope for riches and protection from poverty. PRC WP 2003-9, Pension Research Council
Stephens M Jr (2003) '3rd of Tha Month' do social security recipients smooth consumption between checks? NBER Working Paper, 9135, Cambridge, USA
Stephens M Jr (2004) Paycheck receipt and the timing of consumption. NBER Working Paper, 9356, Cambridge, USA
Strahilevitz MA, Loewenstein G (1998) The effect of ownership history on the valuation of objects. J Consum Res 25(3):276–289
Strauss A (1952) The development and transformation of monetary meanings in the child. Am Sociol Rev 17(3):275–286
Strotz RH (1955) Myopia and inconsistency in dynamic utility maximization. Rev Econ Stud 23(3):165–180.
Sugden R (1984) Reciprocity: the supply of public goods through voluntary contributions. Econ J 94(376):772–787
Sugden R (1986) The economics of rights, co-operation and welfare. Blackwell, Oxford
Sugden R (1991) Rational choice: a survey of contributions from economics and philosophy. Econ J 101(407):751–785
Sugden R (2003a) Coping with preference anomalies in cost-benefit analysis. CSERGE Working Paper, EDM 03-03
Sugden R (2003b) Reference-dependent subjective expected utility. J Econ Theory 111(2):172–191
Sugden R (2004) The opportunity criterion: consumer sovereignty without the assumption of coherent preferences. Am Econ Rev 94(4):1014–1033
Sugden R (2005) Coping with preference anomalies in cost-benefit analysis: a market-simulation approach. Environ Resour Econ 32(1):129–160
Sugden R, Williams A (1978) The principles of practical cost-benefit analysis. Oxford University Press, Oxford

Sunstein CR, Thaler RH (2003) Libertarian paternalism is not an oxymoron. Univ Chicago Law Rev 70(4):1159–1202
Teal G (2000) Effects of gender and parental status on the economic valuation of increasing wetlands, reducing wildlife contamination and increasing salmon populations. Soc Nat Resour 13(1):1–14
Thaler RH (1980) Toward a positive theory of consumer choice. J Econ Behav Organ 1(1):39–60
Thaler RH (1981) Some empirical evidence on dynamic inconsistency. Econ Lett 8(3):201–207
Thaler RH (1985) Mental accounting and consumer choice. Mark Sci 4(3):199–214
Thaler RH (1988) Anomalies: the ultimatum game. J Econ Perspect 2(4):195–206
Thaler RH (1990a) Anomalies – saving, fungibility, and mental accounts. J Econ Perspect 4(1):193–205
Thaler RH (1990b) Saving, fungibility and mental accounts. J Econ Perspect 4(1):193–205
Thaler RH (1994) Psychology and savings policies. Am Econ Rev 84(2):186–192
Thaler RH (1999) Mental accounting matters. J Behav Decis Making 12(3):183–206
Thaler RH, Benartzi S (2004) Save more tomorrow^{tm}: using behavioral economics to increase employee saving. J Polit Econ 112(1):S164–S187
Thaler RH, Gould W (1982) Public-policy toward life saving – should consumer preferences rule. J Policy Analy Manage 1(2):223–242
Thaler RH, Kahneman D, Knetsch J (1991) The endowment effect, loss aversion, and status quo bias: Anomalies. J Econ Perspect 5(1):193–206
Thaler RH, Sunstein CR (2003) Libertarian paternalism. Am Econ Rev 93(2):175–179
Thaler RH, Tversky A, Kahneman D, Schwartz A (1997) The effect of myopia and loss aversion on risk taking: an experimental test. Quart J Econ 112(2):647–661
Thayer M (1981) Contingent valuation techniques for assessing environmental impacts: further evidence. J Environ Econ Manage 8(1):27–44
Thurow L (1976) Government expenditures: cash or in-kind aid? Philos Public Aff 5(4):361–381
Tietz R (1992) An endowment effect in market experiments. In: Lea SE, Webley P, Young BM (eds) New directions in economic psychology: theory, experiment and applications. Edward Elgar, Aldershot, pp 99–121
Tobin J (1970) On limiting the domain of inequality. J Law Econ 13(2):263–277
Todd PM, Goodie AS (2002) Testing the ecological rationality of base rate neglect. Proceedings of the seventh international conference on simulation of adaptive behavior
Torrance GW (1985) Measurement of health state utilities for economic appraisal. J Health Econ 5:1–30
Traub S (1999) Framing effects in taxation. Springer-Verlag, Berlin Heidelberg. MIT Press, Cambridge MA USA, pp 215–223
Treas J (1993) Money in the bank: transaction costs and the economic organization of marriage. Am Sociol Rev 58(5):723–734
Tversky A, Kahneman D (1971) Belief in the law of small numbers. Psychol Bull 76(2):105–110
Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185:1124–1131.
Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. Science 211:453–458.
Tversky A, Kahneman D (1986) Rational choice and the framing of decisions. J Bus 59(4):251–278
Tversky A, Kahneman D (1991) Loss aversion in riskless choice: a reference-dependent model. Quart J Econ 106(4):1039–1061
Tversky A, Kahneman D (1992) Advances in prospect-theory – cumulative representation of uncertainty. J Risk Uncertain 5(4):297–323
Tversky A, Slovic P, Kahneman D (1990) The causes of preference reversal. Am Econ Rev 80(1):204–217
Vagliasindi PA (1994) The pure theory of impure public-expenditures reconsidered – a note on the public and merit characteristics approach. Public Finance-Finances Publiques 49(2):257–281
Van De Veer D (1986) Paternalistic intervention: the moral bounds of beneficence. Princeton University Press, Princeton

Van Dijk E, Van Knippenberg D (1996) Buying and selling exchange goods: loss aversion and the endowment effect. J Econ Psychol 17(4):517–524
Van Soest A, Hurd M (2003) A test for anchoring and yea-saying in experimental consumption data. RAND Corporation
Varian HR (1978) Microeconomic analysis. WW Norton and Company, USA
Villadsen B (1995) Communication and delegation in collusive agencies. J Account Econ 19(2):315–344
Viscusi WK, Magat WA, Huber J (1987) An investigation of the rationality of consumer valuations of multiple health risks. RAND J Econ 18:465–479
Wakker PP (2003) The data of Levy and Levy (2002) "Prospect theory: much ado about nothing?" actually support prospect theory. Manage Sci 49(7):979
Wakker PP, Thaler R, Tversky A (1997) Probabilistic insurance. J Risk Uncertain 15(1):7–28
Wakker PP, Tversky A (1993) An axiomatization of cumulative prospect-theory. J Risk Uncertain 7(2):147–175
Warneryd KE (1999) The psychology of saving: a study on economic psychology. Edward Elgar, Cheltenham
Weale A (1978) Paternalism and social policy. J Soc Policy 7(2):157–172
Weber M, Eisenfuhr F, von Winterfeldt D (1988) The effects of splitting attributes on weights in multiattribute utility measurement. Manage Sci 34:431–445.
Weber M, Keppe HJ, Meyer-Delius G (2000) The impact of endowment framing on market prices – an experimental analysis. J Econ Behav Organ 41:159–176
Webley P (2005) Children's understanding of economics. In: Barrett M, Buchanan-Barrow E (eds) Children's understanding of society. Psychology Press, Hove, chap 3, pp 43–67
Webley P, Burgoyne CB (2001) The economic psychology of everyday life. Psychology Press, Hove
Webley P, Lewis A, Mackenzie C (2001) Commitment among ethical investors: an experimental approach. J Econ Psychol 22(1):27–42
Weiss DM (1991) Paternalistic pension policy: psychological evidence and economic theory. Univ Chicago Law Rev 58(4):1275–1319
Wenzel HD, Wiegard W (1981) Merit goods and 2nd best taxation. Public Finance-Finances Publiques 36(1):125–139
West EG, McKee M (1983) De gustibus est disputandum: the phenomenon of "merit wants" revisited. Am Econ Rev 73(5):1110–1121
Whittington D, Briscoe J, Mu X, Barron W (1990) Estimating the willingness to pay for water services in developing countries: a case study of the use of contingent valuation surveys in Southern Haiti. Econ Dev Cult Change 38(2):293–311
Wiles R, Higgins J (1996) Doctor patient relationships in the private sector: patients' perceptions. Sociol Health Illn 18(3):341–356
Winnett A, Lewis A (1995) Household accounts, mental accounts, and savings behavior – some old economics rediscovered. J Econ Psychol 16(3):431–448
Wirl F (1999) Paternalistic principals. J Econ Behav Organ 38(4):403–419
Wittman DA (1995) The myth of democratic failure: why political institutions are efficient. University of Chicago Press, Chicago
Wu S (2005) Fatalistic tendencies: an explanation of why people don't save. Contrib Econ Anal Policy 4(1)
Yan Y (2005) The gift and gift economy. In: Carrier JG (ed) A handbook of economic anthropology. Edward Elgar, Cheltenham, chap 15, pp 246–261
Yates JF, Zukowski LG (1976) Characterization of ambiguity in decision making. Behav Sci 21:19–25
Yitzhaki S (1974) A note on income tax evasion: a theoretical analysis. J Public Econ 3:201–202
Zhao J, Kling C (2001) A new explanation for the WTP/WTA disparity. Econ Lett 73:293–300.
Zhan M, Sherraden M, Schreiner M (2004) Welfare recipiency and savings outcomes in individual development accounts. Soc Work Res 28(3):165–182

Zhang Y, Fishbach A (2005) The role of anticipated emotions in the endowment effect. J Consum Psychol 15(4):316–324

Zhou RR, Pham MT (2004) Promotion and prevention across mental accounts: when financial products dictate consumers' investment goals. J Consum Res 31(1):125–135

Index

A

Abeler, Johannes, 233
Absolute Risk Aversion, 252
Acyclicity, 39, 86, 218, 223
Adulthood
  Regulation of, 204
Agency, 134, 136, 139, 187
Ahmed, Ehtisham, 220
Allais paradox, 10
Allingham, Michael, 251
Altruism, 4, 150, 188, 195, 197–199, 201
  Regulation and, 189
Altruistic Agent, 195
Ambiguity Aversion, 56, 57
Anchoring, 51–53, 63, 80, 268, 269
  Role of Expertise, 53
Anderlini, Luca, 140
Anomalies, viii, 17, 75, 100
  Anchoring, 52
  Availability, 51
  Confirmation Bias, 54
  Decoy Effects, 80
  Description Invariance, 17
  Elimination by Repetition, 61, 63, 73
  Ellsberg Paradox, 57
  Extreme Event, 17
    Reflection Effect, 18
  Failure to Predict, 33, 72
  Field Evidence, 19
  Implicit, 17
  Information Processing, 50–52, 54
  Mental Accounting, 39
  Preference Reversal, 44
  Reference Dependent Preferences, 22
  Representativeness, 50
Ariely, Dan, 1, 63, 267, 269
Arrow, Kenneth, 17, 104, 108, 111, 271
Asian Disease Example, 7
Asset Based Welfare, 247, 248
Asymmetric Information, 15, 116, 134, 135, 139, 187, 193–195, 206
Australia, 280
Autarchy, *see* Paternalism
Availability Heuristic, 51
Aware, 135

B

Baigent, Nick, 171
Baked Beans, 7
Banks, James, 245, 247
Baron, Jonathan, 250, 251
Barzel, Yoram, 151
Bateman, Ian, 30, 217, 265, 269, 281
Bayes' Theorem, 10, 49, 50, 55, 58, 71, 73, 112, 166, 185
Becker de Groot Marschak (BDM), 27, 63
Becker, Gary, 203
Becker, Selwyn, 57
Benartzi, Shlomo, 243
Bernasconi, Michele, 21, 252
Bernheim, B. Douglas, 236
Berti, Anna Emilia, 190, 191
Besley, Timothy, 135, 156, 170, 180
Bias, *see* Anomalies
Binmore, Ken, 49
Bishop, Richard, 22
Blow, Laura, 235
Boadway, Robin, 95
Bombi, Anna Silvia, 190, 191
Boundaries of the State, 142
Bounded Rationality, 2
Boyle, Kevin, 53, 268, 274
Braga, Jacinto, 61
Broome, John, 97

C
Camerer, Colin, 9, 17, 19, 20, 22, 55, 57, 64, 227
Cameron, Trudy-Ann, 267
Canada, 236
Canning, David, 140
Caplan, Brian, 21, 137, 142, 147
Carmichael, Lorne, 39
Carson, Richard, 267, 271, 272, 279
Carter, Rosemary, 118
Charness, Gary, 71
Cheating, 191
Chicken Tikka, 239
Child Benefit, 235
Child Trust, 248
Child Trust Fund, 235
Children, 166, 188–191, 202, 205
  Understanding of Money, 191
Chocolate, 33, 61, 62
  Belgian, 29
  Swiss, 22
Citizen Candidate, 156
Citizens' Jury, 16, 259, 260, 263, 277, 279, 282
  Ideal, 276, 282
Coate, Stephen, 150, 156, 230
Coding Gains and Losses, 43
Cognitive Limitations, 137
Coherent Arbitrariness, 63
Collective Trading, 89
Common Ratio Effect, 10, 37
Community Preferences, 169
Compatibility, 46
Competence Theory, 57
Complexity of Decisions, 122, 123
Compulsory State Provision, 197
Confirmation Bias, 54, 55, 152, 157, 159
Conlisk, John, 9, 20, 66
Connolly, Sara, 242, 256
Conscientiousness, 245
Consensus Rule, 110
Consent, 107, *see* Paternalism, 120
  Contractarian, 121, 128
  Ex Ante, 118, 126
  Ex Post, 118, 124, 126
  Hypothetical Rational, 119, 127
  Prior, 118
Consequentialist, 105
Consumer Sovereignty, 15, *see* Paternalism, 163, 166, 169, 182, 187, 189
Contingent Valuation, 2, 102, 104, 259, 260, 263, 265, 266, 274, 277, 280
  Open-Ended, 267
Contractarian Consent, *see* Consent, Contractarian
Contraction Property, 13
Cost minimization, 211
Cost-Benefit Analysis, 260
Coursey, Donald, 33, 61
Cox, James, 61
Crocker, Thomas, 62
Crosby, Ned, 259, 263
Cubitt, Robin, 47, 256
Cullis, John, 142, 227
Cycles, 78

D
Decoy, 80, 81, 83
Decoy Effects, 80, 82
Delegation, 136
  Strategic, 136
Department for Work and Pensions, 229
Dependency Relationships, 189
Description Invariance, 17
Desirability, 210, 224
  Ex Ante, 221
  Ex Post, 215
Desirable, 215
Dhami, Sanjit, 251
Diamond, Peter, 102, 210, 213, 270
Dichotomous Choice (DC), 53, 267, 274, 282
Diminishing Sensitivity, 91, 216
Discounting, 237
Disney, Richard, 243
Disposition Effect, 31, 236
Dutch Book, 61
Dworkin, Gerald, 114, 119
Dynamic Inconsistency, 237–240

E
EAC, *see* Ex Ante Consent
Earl, Peter, 149
Eckel, Catherine, 247
Editing, 34
Edmonds, Edward, 235
Education, 179
Efficiency
  Technical, 211
Egalitarianism, Specific, 108
Elicitation Effects, 267
  Yea-Saying, 268
Ellsberg Paradox, 56, 57
Elster, Jon, 6, 111
Endogenous Irrationality, 147
Endowment Effect, 22, 32
Enforcement, 200

EPC
Ex Post Consent, 124
Equity Premium Puzzle, 21, 236
Equivalent Gain, 30
Equivalent Loss, 30, 217
European Union, 149
Tobacco Advertising n, 149
Ex Ante Consent, 124, 126, 219, 224
Ex Post Consent, 124
Excess Deaths in Winter, 228
Expected Utility Theory, 50, 56
Experiments, 18, 47
Field, 25
Hypothetical, 56
Incentives in, 19, 55
Limitations, 20
Markets, 58
Experts, Valuation by, 103
Extended Sympathy, 111
Externalities, 15, 138, 259

F
Familiarity and Markets, 106
Fatalism, 246
Feasible, 215
Feedback, 63
Ferreira de Souza, Yannick, 31
Field Experiments, 31
Financial Education, 248
Financial Services Authority, 247
Fischoff, Baruch, 55
Fisher, Irving, 237
Flat Pack Furniture, Japan, 149
Flypaper Effects, 134
Fox, Craig, 57
Frame Neutrality, 8, 14, 97
Frame Variation, 264, 283
Frames, 6–8, 12, 17, 49, 63, 95, 110, 123, 130, 133
Acceptable, 265, 282
Asian disease, 6
Gold Standard, 103, 271
Market, 49
Optimal, 49, 100, 103, 105
Optimal (see also Gold Standard), 100, 103
Unacceptable, 103
Frey, Bruno, 80
Friedman, Daniel, 63
Frykblom, Per, 53, 269
Fundamental Theorems of Welfare Economics, 92, 138, 220
Fungibility, 234
Fungible, 39

G
General Equilibrium, 83, 87
Acyclicity, 86
Reference dependent preferences and, 83
Glaeser, Edward L., 60
Gneezy, Uri, 19
Goodhart's Law, 16
Gorrige, Peter, 245
Gorter, Joeri, 69, 70
Government
Central, 134
Local, 134
Gregory, Robin, 272
Grether, David, 44, 55, 61
Guesnerie, Roger, 220, 223, 225
Guilt Subsidy, 193

H
Habit Formation, 83
Hammond, Peter, 11
Hanemann, Michael, 22
Hanley, Nick, 165
Happiness, 101
Objective, 102
Harrison, Glenn, 239
Harsanyi, John, 172
Head, John, 14, 163–167
Health, 53, 151, 179, 208
Efficiency in, 211
Hedonic Pricing, 260
Herding Effects, 63
Herne, Kaisa, 33, 217
Herriges, Joseph, 53
Herrnstein, Richard, 69
Heuristics, 46, 50, 51
Hines, John, 134
Holt, Charles, 45, 59
HRC, *see* Hypothetical Rational Consent
Hsee, Christopher, 47
Humphreys, Stephen, 71
Hurd, Michael, 53
Hypothecation, 235

I
IDA, *see* Individual Development Accounts
In-Kind Transfer, 179
Incentive Compatibility, 16, 175, 177, 181, 183, 194
Incentives, 20
Market-based, 15
Income Tax, 175
Incompleteness, 93
Increasing Returns, 138

Independence, axiom of Expected Utility Theory, 45
Individual Development Accounts, 247
Induced Preference, 60
Information, 112, 130, 181, 182
  Aggregation, 100, 108, 111, 130
  Extra-Preference, 111
  Misinformation, 183
  Non-Preference, 14, 112
  Processing, 10, 49
  Provision of, 180, 182
  Value of, 180
  Wrong, 164, 179, 180
Integrity, 121
Ireland, Norman, 169
Irrationality, 9, 142
  Rationality of, 137

J
Jevons, Stanley, 237
Johnson, Catherine, 247
Jones, Martin, 55
Jones, Philip, 142

K
Kagel, John, 62
Kahneman, Daniel, 6, 17, 18, 30–32, 34, 36, 40, 46, 48, 50, 52, 72, 85, 103, 230, 250, 266
  and Happiness, 102
Keren, Gideon, 57
Keynes, John Maynard, 56
Kleinig, John, 115, 117, 119, 122
Knetsch, Jack, 22
Kolstad, Charles, 67
Kooreman, Peter, 235

L
LA, *see* Loss Aversion
Labelling, *see* Mental Accounts
Labour supply, 252
Laibson, David, 239
Layard, Richard, 101
Learning, 49, 58, 59, 61–63, 65–67, 69, 71, 108, 129, 146
  Agency, 195
  Bayesian, 71
  Melioration, 70
  Universally Beneficial, 145
Levin, Dan, 71
Lichtenstein, Sarah, 44, 55, 267
List, John, 31, 33, 47, 259
Locus of Control, 246
Loewenstein, George, 30, 33, 43, 72
Loomes, Graham, 21, 45, 63
Lord, Charles G., 54
Loss Aversion, 21, 38, 213, 216–218, 221
Lottery, 62, 63, 122

M
MacLean, Alistair, 122
Macleod, Bentley, 39
Madge, Charles, 242
Marginal Social Value of Information, 181
Market Experiments, 59
Market Failure, 116, 133, 138–141
  Bounded Rationality and, 140, 141, 146, 159
Market Frame, 106
Markets, 75, 105, 107
  Learning in, 58, 61, 64, 65
  Rationality and, 60
Marklein, Felix, 233
Marshall, Alfred, 18, 237
McCaffery, Edward, 251
McClure Jr., Charles, 165, 166, 169
McFadden, Daniel, 6, 17, 80
McLennen, Edward, 13
Mean Squared Error, 264, 275, 277
Mehra, Rajnish, 21
Melioration, 69, 70, 72
Mental Accounting, 18, 39–42, 44, 134, 255
  and Health Club Use, 40
  Consumption Expenditure, 43
  Saving, 43
Mental Accounts, 230, 232, 233, 235
  Creating, 228
  Fungibility, 235
Menu Costs, 141
Merit Good, 169
Merit Wants, 3, 4, 11, 12, 14, 15, 94, 113, 135, 142, 163, 170, 171, 179, 185, 187, 194, 207, 235
  Agency, 191
  Community Preferences, 169
  Information processing and, 165
  Optimal taxation, 173
  Risk, 172
  Spurious Categories, 168
Merit Worthy, 11, 163
Messages, Potentially Credible, 184
Mickey Mouse, 31
Mill, John Stuart, 115–117, 122, 165, 189
Minimum Standards of Care, 197, 198
Mirrlees, James, 210, 213
Misinformation, 183

Mitchell, Robert, 265, 271
Money Pump, 61, 62, 66
Monotonicity, 112, 207
Monty Hall Problem, 62
Moral Hazard, 142
Morrison, Gwendolyn, 29
MSE, *see* Mean Squared Error
Mugs, 22
Musgrave, Richard, 14, 164, 166, 169
Myopia, 33

N
Naive, 183
Nanny state, 148
National Monuments as Merit Goods, 170
Near-Optimality, 139
Negative Income Tax, 186
Netherlands, 235
Neutrality, 14
Neutrality, Strong, 95, 97
Ng, Yew-Kwang, 101
No Reverses, 125
NOAA Panel, 104, 271, 272, 283
Non-Hicksian Policies, 228
Non-linear pricing, *see* Optimal taxation
Non-Market, 146, 259
    Goods, 211
    Valuation, 265
Non-Sensitivity Axiom, 216–218
Norwich, 28, 265, 281
NR, *see* No Reverses
Numeracy, 247

O
O'Donoghue, Ted, 239
Odysseus, 239
Okada, Erica Mina, 40
Optimal Frames, 130
Optimal Policy
    Children, 204
Optimal Taxation, 221, 223
    Merit wants, 173
Optimization, 19
Orne, Martin, 268
Osborne, Martin, 135

P
Palfrey, Thomas, 159
Paretian Welfare, 1, 8, 95
    Strong Pareto Principle, 97
Pareto, 91, 108, 140
    Pareto Improvement, 150
Pareto Efficiency, 138
    ex ante, 220
    ex post, 220
Pareto Optimum, 90, 220
ParetoF, 97
Part-Whole Bias, 265, 271
Partial Valuation, 260, 276, 280
Paternalism, 94, 114, 115, 119, 123
    Consent, 123, 124, 128
        Hypothetical Rational, 127
    Libertarian, 227, 245
    Strong, 122
Payment Cards, 274
Pensions, 236, 242–245
Personal Integrity, 119
Pigou, A. C., 164, 237
Planned Behaviour, 246
Planning Ahead, 245, 246
    Numeracy skills and, 247
Plott, Charles, 19, 33, 44, 65
Policy, Convincing, 184
Political Economy, 135, 151
Political Failure, 151
Poverty
    Poverty and Irrationality, 237
Preference
    Constructed, 272
    Discovered, 19, 66
    Ex Ante, 219
    Hicksian, 14, 22, 38, 86, 127
    Measurement Problem, 93
Preference Reversal, 17, 44–47, 61
Prelec, Drazen, 36, 42, 233, 239
Prisoners' Dilemma, 138
Probability
    Distorted, 36
Procrastination, 43, 240–242
Product Differentiation, 80
Production Efficiency, 210
Prominence, 46, 47
Prospect Theory, 33, 36–38, 230
    and Tax Evasion, 252
    Cumulative, 34
    Diminishing Sensitivity, 35
Provision of Information, 181
Public Goods, 15, 138, 259

Q
QALYS, 100, 261, 276
Quiggin, John, 267

R
Rabin, Matthew, 152, 239, 241
Racionero, Maria del Mar, 163, 175, 180
Rationality, 135
    Convergence to, 61–64

Ecological, 21
Endogenous, 142
Social, 60
Strong, 9
Weak, 8
Rationality Spillovers, 62
Rawls, John, 121, 128
RDPS, *see* Reference Dependent Preferences
Reference Dependent Preferences, 22, 28, 38, 39, 75, 76, 83, 85, 91, 123, 213, 215–217, 220, 221, 223
Tax, 207
Reference Point, 32, 33, 48, 61, 64, 69, 75, 79, 86, 134, 214, 218, 223
Effects in Housing, 30
Speed of Adjustment, 33, 84
Referendum, 105
Reflexive Equilibrium, 88
Reflexive Optimum, 87, 88, 91
Reframing, 96, 97, 99
Regret, 70
Regulating Care, 195
Regulation, 149
Relative Closeness Effect, 217
Relief from Duty, 192
Rent Seeking, 151
Representativeness, 50, 55
Returns to Scale, 187
Risk, 10, 37
Riskless Choice, 38, 216
Roemer, John, 111
Rustichini, Aldo, 19

S

Safety Standards, 149, 150
Salop, Stephen, 80
Samaritans Model, 150
Samuelson, Paul, 10
Sandmo, Agnar, 172, 179
Savage, Leonard, 10
Savings, 43, 236, 242, 245
SMarT, 244
Savings Gateway, 248
Schelling, Thomas, 250
Scope Effects, 265
Scope Sensitivity, 267
Second Price Auction, *see* Vickrey Auction
Self-Control, 136, 149, 237, 245
Sen, Amartya, 1, 93, 95, 109, 125, 208
Separability, 102
Shaked, Avner, 81
Shefrin, Hersh, 2, 19, 236
Shleifer, Andrei, 236
Shogren, Jason, 28, 33, 53, 61, 265, 269
Sileo, Patrick, 39
Simon, Herbert, 9, 136
Simonson, Itamar, 72
Slemrod, Joel, 250
Slovenia, 235
Slovic, Paul, 44, 46, 272
SMA, *see* Strong Market Acceptability
Small Print Protection, 149
Smith, Vernon, 59
Social Rationality, 60
Social State, 6
Specific Egalitarianism, 168, 171
Stakeholders, in non-market valuation, 273
Starmer, Chris, vii, 17, 47, 55, 266
Stated Preference, 14, 53, 102, 259
Status Quo Bias (see also Endowment Effect), 30
Status symbols, 169
Stern, Nicholas, 220
Stickiness of Labels, 233
Stonehenge, 170
Strack, Fritz, 53
Strategic Delegation, *see* Delegation
Strong Market Acceptability (SMA), 107
Strong Neutrality, 97, 169
Sugden, Robert, 9, 30, 39, 49, 55, 75, 78, 85, 86, 90, 91, 124, 215, 217
Sunk Costs, 40, 134
Sure Thing Principle (STP), 10

T

Tax, 173, 174, 179, 180, 183, 186, 207, 208, 210, 214, 215, 220, 221, 223, 224
ad Valorem, 208
and Loss Aversion, 213
Commodity, 173, 221
Evasion, 16, 251
Framing, 249, 250, 252
Income, 175, 179, 186
Marginal Rate, 256
Merit Wants, 174
Negative Income Tax, 186
Non-linear, 175, 179, 180
Obfuscation, 251
Optimal, 181, 224
Reform, 214, 220, 221
Ex-post Preferred, 215
Tax Equilibrium, 225
Tax Policy, 207
Technical Efficiency, 207, 208, 210, 211, 213
Term Limits, 155
Thaler, Richard, 2, 22, 30, 31, 39, 41–43, 227, 238, 243, 256

Time Preference, 43
Tobin, James, 108, 168
Transitivity, 67
Traub, Stefan, 22, 186, 250
Travel Cost, 260
Tversky, Amos, 6, 15, 18, 32, 34, 36, 40, 46, 50, 52, 57, 85, 91, 104, 216, 230, 250, 267
Tyran, Jean-Robert, 62

U
Undersaving, 236
Unemployment Benefit, 255
United Kingdom, 134, 149, 166, 235, 243, 248, 255
  Chilly Beaches of, 6
USA, 30, 236, 280

V
Validity
  Construct, 271
  Content, 271
  External, 73
  Theoretical, 271
Valuation Function, 34
Veil of Ignorance, 128
Vickrey Auction, 27, 29, 269
Voting, Paradox of, 159

W
Wakker, Peter, 37, 263
Walker, Ian, 235
Warm Glow, 266
Wason Test, *see* Confirmation Bias
Weak Rationality, 66, 84
Weber, Martin, 57
Webley, Paul, 245
Weighting Function, *see* Prospect Theory
Welfare, viii, 8, 75, 93, 95, 100–102, 105, 114, 124, 127, 128, 130, 131, 218
  Children, 203
  Fundamental Theorem of, 92
  Individual, 9
  Monotonicity, 207
  Non-individualistic, 169
  Social Welfare Function, 164
Welfare Economics, 98
Welfare Frame Neutrality, 99, 103
Welfarism, 95, 182, 259
WFN, *see* Welfare Frame Neutrality
Willingness to Accept (WTA), 22, 67, 79, 265, 278
Willingness to Pay (WTP), 22, 39, 53, 58, 268, 281
Winter Fuel Payment, 228–230
Working Tax Credit, 250, 255
WTA-WTP Gap, 18, 29, 31–33, 37, 61, 68, 265
  and Substitutability, 28
Wu, Stephen, 247

Y
Yates, Frank, 57
Yea-saying, *see* Elicitation Effects., 269

Z
Zeckhauser, Richard, 22
Zhao, Yu, 69
Zhu, Yu, 235
Zukowski, Lisa, 57