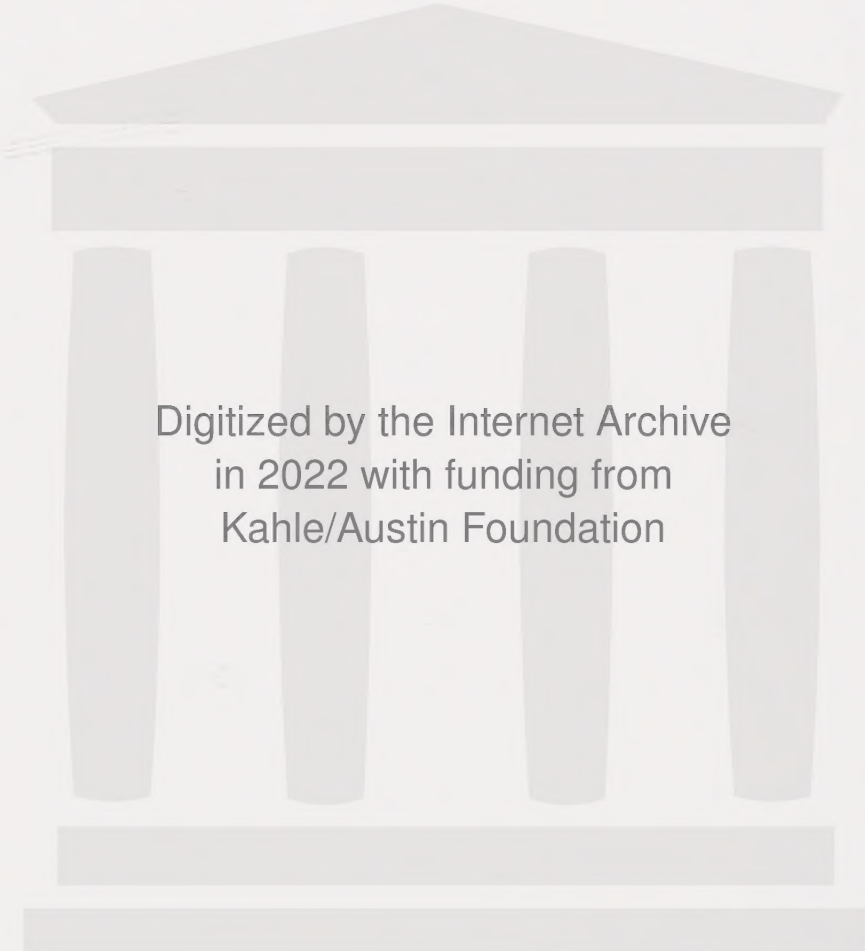


Introduction to Transport Policy

A PUBLIC POLICY VIEW

Peter Stopher • John Stanley



Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

WITHDRAWN
FROM THE LIBRARY OF
UNIVERSITY OF ULSTER

100588771



Introduction to Transport Policy

Introduction to Transport Policy

A Public Policy View

Peter Stopher

and

John Stanley

Institute of Transport and Logistics Studies, University of Sydney, Australia

Edward Elgar

Cheltenham, UK • Northampton, MA, USA

© Peter Stopher and John Stanley 2014

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical or photocopying, recording, or otherwise without the prior permission of the publisher.

Published by
Edward Elgar Publishing Limited
The Lypiatts
15 Lansdown Road
Cheltenham
Glos GL50 2JA
UK

Edward Elgar Publishing, Inc.
William Pratt House
9 Dewey Court
Northampton
Massachusetts 01060
USA

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

Library of Congress Control Number: 2013949815



ISBN 978 1 78195 244 3 (cased)
ISBN 978 1 78195 246 7 (paperback)
ISBN 978 1 78195 245 0 (eBook)

Typeset by Servis Filmsetting Ltd, Stockport, Cheshire
Printed and bound in Great Britain by T.J. International Ltd, Padstow

Contents in brief

<i>Preface</i>	xi
<i>Acknowledgements</i>	xiii
<i>List of abbreviations</i>	xiv
1 Introduction	1
2 History of transport with policy implications	9
3 Policy needs and policy processes	21
4 Economic and sustainability foundations	60
5 Traffic theory and transport planning foundations	92
6 Social exclusion	116
7 Tackling the externalities – environment	132
8 Tackling the externalities – health and safety	154
9 Tackling the externalities – congestion	175
10 Tackling the externalities – fuels and technology	195
11 Agglomeration and other wider economic benefits	206
12 Road user charges	221
13 Potential solutions – public transport investment and technology	243
14 Potential solutions – TSM, TDM and VTBC	265
15 Goods movement	291
16 An integrated land use/transport policy	309
<i>Index</i>	333

Contents in full

<i>Preface</i>	xi
<i>Acknowledgements</i>	xiii
<i>List of abbreviations</i>	xiv
1 Introduction	1
Chapter overview	1
1.1 Context	1
1.2 Structure	2
1.3 Approach	5
1.4 Postscript	8
References	8
2 History of transport with policy implications	9
Chapter overview	9
2.1 A brief history of transport technologies	9
2.2 Transport in the early twenty-first century	12
2.3 Transport and urban form	13
2.4 Travel-time budgets	16
2.5 Principal policy directions	18
References	20
3 Policy needs and policy processes	21
Chapter overview	21
3.1 Rationale for policy	21
3.2 Market failures and transport policy	22
3.3 How do we know if society is better off?	29
3.4 Policy approaches	35
3.5 Generic approach to a policy cycle	48
3.6 Generating alternative solutions	55
3.7 Participation	57
3.8 Conclusions	58
References	59
4 Economic and sustainability foundations	60
Chapter overview	60

4.1	Scope	60
4.2	Economic evaluation using cost–benefit analysis	61
4.3	Preferences and a wider context for money and other values	77
4.4	Total economic value	80
4.5	Cost concepts	81
4.6	Discounting	82
4.7	Conclusions	89
	References	90
5	Traffic theory and transport planning foundations	92
	Chapter overview	92
5.1	Rationale for traffic theory	92
5.2	Space and time	93
5.3	Flow, density and speed	95
5.4	System performance	102
5.5	Travel-demand forecasting	107
5.6	From speed–flow to congestion costs	111
5.7	Implications for policy	114
	References	115
6	Social exclusion	116
	Chapter overview	116
6.1	Some context	116
6.2	Some definitions	119
6.3	Social policy goals	120
6.4	Transport disadvantage	121
6.5	Mobility, social inclusion and well-being	122
6.6	Minimum service levels	125
6.7	Regional mobility policy: a social enterprise model	126
6.8	Some other ‘social’ matters	129
6.9	Policy directions	130
	References	130
7	Tackling the externalities – environment	132
	Chapter overview	132
7.1	Historical background	132
7.2	Environmental impacts	136
7.3	Valuing environmental externalities	151
7.4	Conclusions	152
	References	153

8	Tackling the externalities – health and safety	154
	Chapter overview	154
8.1	Health and safety externalities	154
8.2	Transport accidents	154
8.3	Transport and health	163
8.4	Valuing safety policies and programmes	172
8.5	Conclusions	173
	References	173
9	Tackling the externalities – congestion	175
	Chapter overview	175
9.1	Defining congestion	175
9.2	Historical background to recurring congestion	181
9.3	Responses to traffic congestion	182
9.4	Policies for mitigating congestion	189
9.5	Some concluding comments on congestion	192
	References	193
10	Tackling the externalities – fuels and technology	195
	Chapter overview	195
10.1	Introduction	195
10.2	Oil supply	196
10.3	Policy implications	200
10.4	Policy directions	203
	References	204
11	Agglomeration and other wider economic benefits	206
	Chapter overview	206
11.1	The relevance of agglomeration economies to transport policy	206
11.2	How do we measure agglomeration benefits in land transport?	210
11.3	Other wider economic benefits	212
11.4	Significance of agglomeration benefits and other wider economic benefits	213
11.5	High-speed rail in Australia	217
11.6	Conclusions on agglomeration and wider economic benefits	218
	References	219
12	Road user charges	221
	Chapter overview	221

12.1	Context and some principles	221
12.2	Road cost recovery including external costs	225
12.3	Fuel taxes as charges for road use by cars	227
12.4	Heavy vehicle charging	231
12.5	Congestion pricing schemes	232
12.6	Land value capture for land transport funding	235
12.7	Increased borrowings	239
12.8	Public–private partnerships	239
12.9	Asset sales	240
12.10	Conclusion on sustainable pricing and funding	240
	References	241
13	Potential solutions – public transport investment and technology	243
	Chapter overview	243
13.1	Land transport policy context	243
13.2	The need for public transport	247
13.3	Providing public transport	257
13.4	The future role that public transport might play	260
13.5	Conclusions on public transport	262
	References	263
14	Potential solutions – TSM, TDM and VTBC	265
	Chapter overview	265
14.1	Introduction	265
14.2	Transport system management	266
14.3	Travel demand management	275
14.4	Non-coercive behaviour change strategies	286
	References	290
15	Goods movement	291
	Chapter overview	291
15.1	Context	291
15.2	Congestion	295
15.3	Network access	297
15.4	Encroachment	299
15.5	Energy use/security	301
15.6	Coordination	301
15.7	Safety	302
15.8	Environmental outcomes	303
15.9	The regulatory framework	304
15.10	Freight in land transport plans	305

15.11	Concluding comments	307
	References	307
16	An integrated land use/transport policy	309
	Chapter overview	309
16.1	Context	309
16.2	The need for integration	311
16.3	National policy directions for land transport	314
16.4	Some land use/transport research findings	318
16.5	Land use/transport integration in Melbourne	322
16.6	New York Sustainable Streets Plan	328
16.7	Sustainable funding	330
16.8	Regulation for the environment and safety	330
16.9	Conclusions	331
	References	332
	<i>Index</i>	333

Preface

For several years, we have taught postgraduate students at the Institute of Transport and Logistics Studies at the University of Sydney about Transport Policy. During that time, we have been very aware that there was no book that covered the material that we taught, and this has been the genesis of this book. It is our hope that the material in this book will assist many current and future students of both public policy and transport policy, and will provide a foundation for the advice and analysis that need to lie behind good policy. In this book, we have not assumed prior knowledge on the part of students and others who may wish to learn from this book, but have tried to provide within it the necessary foundations for a study of public policy, especially in the area of urban transport. We have also then illustrated the development of policy by choosing a number of currently topical issue areas and discussing these within the context of setting policy directions.

Of course, over time, the issues that are most significant in the policy arena undergo change. Today's burning issues may become of relatively little concern tomorrow, when new issues will arise to take their place. Thus, the issue areas that we have chosen to focus on in this book will, over time, become less important and new issues will take their place. Nevertheless, we hope that the approaches we have taken will be useful for new issues that we cannot necessarily anticipate at this time.

The authors of this book, themselves, come from two quite different backgrounds – one in engineering and planning, and the other in economics. These different backgrounds mean that we approach some issues differently and do not necessarily see things the same way. We believe this is particularly advantageous for this book, because our differences in approach and opinion, with neither one of us right and neither one of us wrong, may illuminate the issues, and particularly the formation of policy direction, better than if we came from the same backgrounds and were in full agreement throughout.

For us, it has been both interesting and informative to work together on writing this book. Each of us has learnt from the other, and each of us has been able to contribute unique viewpoints and experiences to the writing of this book. We hope that we have been able, within the confines of this book

and from this diversity, to provide an enriching experience for any reader. We have also had the good fortune to work in a number of different parts of the world and hope that our views reflect some of this variety of experiences. Again, not only are there no absolutes in public policy, but what may be best for one nation or region at one time may be far from best for another nation or region at this time. For this reason, we have frequently not tried to deduce specific policy from the issues we have discussed, but have rather attempted to set up a sound basis for the development of good policy.

One final point that we would like to make in this Preface is that it is our strong belief that planning and policy must always go hand in hand. Development of policy without planning often leads to a waste of public monies and political embarrassment, when a policy direction that has been pursued is found to be completely inappropriate. Similarly, planning undertaken without policy is barren, in that it is policy that provides direction for planning the future. Without clearly enunciated policies, planning becomes a futile exercise that will not affect the future of our nations, regions and cities. Policy and planning, then, must be done together, with policy directions informing the planning, and planning providing useful inputs as to the likely consequences of adopting particular policies. Undertaken together, planning and policy provide the needed direction and justification for investment and implementation in the future.



Acknowledgements

As noted in the Preface, the genesis of this book has been our joint teaching of a postgraduate class in Transport Policy over the past several years. We have benefited enormously from the students in these classes over the years, from many of whom we have gained new insights and ideas about this topic. These various students have also contributed through their widely varying backgrounds and the variety of countries from which they have come. We acknowledge these contributions and are grateful for their inspiration.

We would especially like to acknowledge the inspiration of the late Professor David Pearce OBE, great teacher and advocate for the environment and the role of values in economics.

We also wish to acknowledge the generous help and support provided by Professor David Hensher, Director of the Institute of Transport and Logistics Studies at the University of Sydney. David has provided encouragement to us in this project and has also provided us the opportunities to be able to accomplish the writing of this book.

We acknowledge the support and assistance of the University of Sydney, especially for permitting one of us to take a six-month sabbatical leave to devote to much of the writing of this book.

We would like to acknowledge our publisher and those assisting in the publishing process for their encouragement in this project and for providing us the opportunity to publish this as a book. We are most grateful also for the help we have received through the publisher in the editing of our writing and the preparation of this book.

Finally, we would like to acknowledge the support provided by our wives – Carmen Stopher and Janet Stanley – during the writing of this book. They have supported and encouraged us, and provided us with inspiration and strength to complete this. Without their support, it is unlikely this book would have been written.

Abbreviations

ABS	anti-locking braking system
APC	average private costs
AVR	average vehicle ridership
BART	Bay Area Rapid Transit
BCR	benefit/(net) cost ratio
BITRE	(Australian) Bureau of Infrastructure, Transport and Regional Economics
BMI	body mass index
BPR	Bureau of Public Roads
BRT	bus rapid transit
BUZ	bus upgrade zone
CAAA	Clean Air Act Amendments
CAFE	Corporate Average Fuel Economy
CBA	cost-benefit analysis
CBD	central business district (downtown)
CCTV	closed circuit television
CEA	cost-effectiveness analysis
CFCs	chlorofluorocarbons
CGE	computable general equilibrium
CNG	compressed natural gas
CO	carbon monoxide
CO ₂	carbon dioxide
COAG	Council of Australian Governments
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CT	community transport services
dB	decibel
dBA	decibel measured on the A scale
DfT	Department for Transport
DOT	Department of Transport(ation)
EA	environmental assessment/analysis
EIS	environmental impact statement
EPNL	effective perceived noise level
EU	European Union
FHWA (US)	Federal Highway Administration

FONSI	finding of no significant impact
GBE	government business enterprise
g/dl	grams per decilitre
GDP	gross domestic product
GHG	greenhouse gas
g/kWh	grams per kilowatt-hour
GPS	global positioning system
HC	hydrocarbon
HFCs	hydrofluorocarbons
HOV	high-occupancy vehicle
HSR	high-speed rail
HVCCLT	Hunter Valley Coal Chain Logistics Team
ITS	intelligent transport systems
km/h	kilometres per hour
LGA	local government area
LPA	locally preferred alternative
LPG	liquefied petroleum gas
MAC	ministerial advisory council
MCC	marginal congestion costs
MDL	mass/distance/location
MOU	memorandum of understanding
MOVES	Motor Vehicle Emissions Simulator
MPB	marginal private benefits
mph	miles per hour
MSC	marginal social costs
NA	negative affect
NEPA	National Environmental Policy Act
NO	nitric oxide
NO ₂	nitrogen dioxide
NOI	notice of intent
NO _x	oxides of nitrogen, especially nitrogen dioxide
NPBC	negotiated performance-based contract
NPV	net present value
NRTC	National Road Transport Commission (Australia)
NSTIFC	National Surface Transportation Infrastructure Financing Commission
NSW	New South Wales (Australia)
O ₃	ozone
OECD	Organisation for Economic Co-operation and Development
PA	positive affect
pce	passenger car equivalent
pc/h/ln	passenger cars per hour per lane

pc/km/ln	passenger cars per kilometre per lane
PDF	probability density function
PER	public environment report
PESTEL	political, economic, social, technological, environmental and legal
PFCs	perfluorocarbons
pkm	passenger kilometre
PKT	passenger (or person) kilometres of travel
PM	particulate matter
PM _{2.5}	particulate matter of less than 2.5 microns
PM ₁₀	particulate matter of less than 10 microns
PMT	passenger miles of travel
PPM	policy and procedure memorandum
ppm	parts per million
PPP	public–private partnership
PT	public transport
PV	present value
PWI	Personal Wellbeing Index
RISs	regulatory impact statements
SCATS	Sydney Coordinated Adaptive Traffic System
SE	social exclusion
SEU	Social Exclusion Unit
SGEM	Sydney general equilibrium model
SO ₂	sulphur dioxide
SOC	social opportunity cost
SO _x	oxides of sulphur
SSP	Sustainable Streets Plan
STPR	social time preference rate
TDM	travel demand management
TfL	Transport for London
TIF	tax increment financing
tkm	tonne kilometre
TOD	transit-oriented development
TRESIS	Transport, Environment and Social Impact Simulator
TSM	transport system management
TUB	total user benefits
UITP	International Association of Public Transport
UK	United Kingdom
UN	United Nations
US EPA	United States Environmental Protection Agency
US/USA	United States/United States of America
VCR	volume/capacity ratio

vkm	vehicle kilometre
VKT	vehicle kilometres of travel
VMT	vehicle miles travelled
VOC	volatile organic compound, i.e. unburnt hydrocarbons
VTBC	voluntary travel behaviour change
WEB	wider economic benefits
WHO	World Health Organization
WTA	willingness to accept
WTP	willingness to pay

Introduction



CHAPTER OVERVIEW

Despite transport problems being a topic of frequent dinner table conversation, comprehensive assessment of policy directions for transport has been the subject of remarkably little academic analysis. This chapter introduces the scope of the book, which is intended to help redress this shortcoming. The primary focus is on urban transport policy, with the emphasis being on policy analysis rather than analysis of the policy process. Importantly, the chapter sets out some key propositions that have been important in shaping the authors' approach to the particular matters that are considered in subsequent chapters.

1.1 Context

Have you ever sat in your car on a crowded freeway, or been squashed in a peak-hour train, and wondered 'Why don't they do something about this?' Most people who live in cities have – all too frequently. Urban transport is, in consequence, a matter that is of significant personal and political interest and a common barbecue or dinner table conversation topic. We are all experts in various ways, by virtue of experience, skills or both!

The way urban transport and land use systems are planned, and the policies that shape this planning and its delivery affect the quality of people's lives in very real ways. They affect the productivity of the cities in which people live, congestion levels, greenhouse gas emissions, air pollution, the opportunities available to people living in different parts of a city and the health and safety of residents and visitors. Some of these factors are reasons why people care strongly about transport policies and programmes. Others are reasons why they should care strongly about the quantity and quality of urban transport infrastructure and services.

This book is the result of many years teaching Transport Policy to post-graduate students at the Institute of Transport and Logistics Studies at the University of Sydney. Our student cohort is international, from mixed

undergraduate disciplines and highly engaged with urban transport issues. In some cases this engagement is because they are employed in the field. In others it is because of personal interests. In all cases, personal experiences provide a great stepping stone to thinking about transport policy, with our role being to provide frameworks which can assist students to think about structuring their ideas about the subject matter.

The book is mainly written for students who may wish to pursue a career in transport policy, particularly urban transport policy, and for those who are already practising professionals in the field. It will also be of value to those who might need to understand how transport policy analysts and advisers are likely to think and for those who simply want a few good arguments to help their barbecue or dinner table conversations.

We have chosen to concentrate on current major urban transport policy issues, for several connected reasons. Cities are now home to over half the world's population, and this proportion is on the rise. Transport problems seem to be endemic to cities and to be increasing in their significance. While some countries are making progress on some indicators of national transport system performance, it is more common to see such indicators in retreat at the city level. Congestion levels are increasing. Greenhouse gas emissions are on the rise. Massive numbers of road deaths and serious injuries continue to be recorded annually. Social exclusion continues to be exacerbated by poor mobility opportunities. Button (2010) observes that regulation of private transport, which sits behind many of these issues, has not achieved the same progress as other areas of regulatory reform. Even small advances in the way urban transport systems and services are planned and function will potentially be of value to very large numbers of people. Such advances are most likely to be realisable if urban transport policy is well structured and targeted.

In choosing this focus, we have decided to stay away from many of the significant continuing land transport policy debates, such as public versus private ownership of transport infrastructure and services, and also from sea and air transport policy issues. This is not because we see these as of no significance but simply because we do not believe it is possible to do justice to a wider canvas at the same time.

1.2 Structure

The approach we have taken to compiling the book is to start with a little historical context (Chapter 2) and then present some basic frameworks and analytical tools that we have found helpful in our own endeavours in transport

policy, and supportive transport planning, over many years (Chapters 3 and 4). Our experience suggests that it is relatively unusual for policy analysts and policy makers to have a solid foundation in policy thinking. These two chapters seek to provide some key foundational material, with a welfare economic bias. In taking this particular slant, we acknowledge that there is no uniquely correct way to undertake policy analysis but recognise the widespread acceptance of a welfare economic approach. The presentation is intended to help policy analysts and policy makers understand the implications of such an approach, while being open to alternative perspectives. The chapters include many references to complementary material that will enable readers to go deeper or wider as they wish. The foundational material in Chapters 3 and 4 is primarily intended for subsequent application to urban transport policy issues but has been written in such a way that it should also be of use to those working in other sectoral policy fields (particularly land use policy and social policy).

Good urban transport policy depends significantly on the policy analyst having an understanding of some basic transport and traffic planning concepts. While the transport policy analyst or decision maker can seek detailed advice from a suitable transport planning expert, a grasp of some basic transport planning concepts is essential to asking the right questions and understanding the likely strengths and weaknesses of the answers. Chapter 5 presents some key concepts, with a focus on road use by motor vehicles. Similar concepts are relevant to other urban means of travel.

Chapters 6 to 10 discuss the major urban transport policy problems, which are usually called ‘externalities’ (Chapter 3 fully explains the meaning of this term). The chapters encompass the familiar problems of: social exclusion (Chapter 6), one of the most recent of the urban transport problem areas; the environment (Chapter 7), where we consider air pollution, noise and greenhouse gas emissions; health and safety (Chapter 8), with the discussion encompassing well-known problems such as transport accidents, together with the more recent policy concerns of obesity and road rage; traffic congestion (Chapter 9), the favourite topic for most of our students, the most common dinner table urban transport topic, and probably the most difficult urban transport problem for policy analysts and policy makers; and fuels and technology (Chapter 10), which underpin some of the other externalities (such as greenhouse gas emissions) and raise policy issues of energy security, which are very important concerns for some countries.

The discussion and analysis in Chapters 6 to 10 deal with what might be thought of as the ‘bad effects’ of urban transport, effects that frequently

require policy interventions to rectify problems. In contrast, Chapter 11 looks at the role urban transport systems can play in promoting the productivity of cities. This is sometimes called a positive externality (a 'good'). This is a relatively new area of research, which is enabling a more complete understanding of the importance of effective urban transport systems.

Chapters 12 to 16 move from the diagnosis of the preceding chapters to policy prognosis. Chapter 12 tackles the complex problems associated with charging for road use, a policy measure that can impact positively on each of the negative externalities discussed in Chapters 6 to 10 and can also affect the positive productivity externalities of urban areas. Pricing reform has promised much over 50 years but delivered little, making it an interesting subject for review from a policy perspective.

Chapter 13 considers the role of investment in public transport in tackling the transport problems and opportunities of urban areas. Like pricing, public transport service improvements hold the prospect of having positive benefits in terms of reducing the impact levels of all the problem areas discussed in Chapters 6 to 10. Rail, in particular, is also fundamental to achievement of the potential productivity benefits discussed in Chapter 12. Chapter 13 seeks to identify where and how public transport can add value in urban areas.

Pricing is an important way to raise revenue to improve urban transport systems and services, as discussed in Chapter 12. It is also a way to change traveller behaviour in directions that policy makers may require. In recent years, however, there has been strong interest in many cities in other ways of influencing and changing traveller behaviour, towards a greater travel role for transport modes with low impacts on the environment and on people, as well as attempting to get more out of the huge investments already made in urban transport infrastructure. Chapter 14 discusses a range of such measures, from both supply and demand sides.

The discussion through Chapters 6 to 14 primarily focuses on the movement of people in urban areas, with references to freight or goods movement. Chapter 15 consolidates a range of important policy issues associated with urban goods movement, exploring problems, opportunities and some potential solutions. Goods movement is of increasing policy concern in many cities, partly because of the tight macroeconomic environment and associated need to lift economic productivity but also because of the tendency for goods traffic to grow strongly in many cities, with associated external costs imposed on urban communities.

The essence of modern urban transport policy is the idea of integration: looking at problems and solutions in a ‘joined-up’ way rather than as a series of separate issues. Chapter 16 takes the opportunity to illustrate how this might be approached, taking both national and city-level perspectives, with land use/transport policy integration being a particularly important focus. Goods movement and funding perspectives are also included.

The book has not dealt in much detail with cycling and walking as urban travel modes, but their importance is recognised and acknowledged (e.g. in the integration chapter, Chapter 16). The tools we present in the book will enable interested policy analysts and decision makers to think through problems and opportunities that involve cycling and walking.

1.3 Approach

Parsons’s (1995) very valuable text on public policy reviews the evolution of policy analysis and how dominant paradigms have changed during this evolution. Parsons distinguishes between *analysis of the policy process* and *analysis in and for the policy process*. The current book is firmly embedded in the latter space, as it relates to urban transport, but inevitably requires some consideration of the policy process itself, since it influences the nature and scope of the analysis that is undertaken and the outcomes that are delivered. However, largely for reasons of space constraints, our focus on the policy process is minimal, being mainly concerned with development of integrated land use/transport planning and policy processes (Chapter 16) and with aspects of the implementation of freight policy (Chapter 15). For example, we devote very little attention to questions of institutional design, to processes of policy implementation or delivery or to post hoc evaluation of the outcomes of policy delivery. Our focus, instead, is mainly on policy problems, possible solutions and analysis.

Institutional design is an important issue in urban transport policy, particularly in relation to matters such as the development and delivery of integrated approaches, public–private partnerships (e.g. in motorway or rail construction and operation) and the ownership and regulation of public transport. Specialist forums such as the biennial international Thredbo Conference series, in which one of us is closely involved, deal with the latter topic in great detail, playing an important agenda-setting role. Interested readers can pursue this subject through the Thredbo Conference website.¹

Parsons (1995) emphasises the importance of frameworks in public policy and discusses a range of diverse approaches that have been used in this regard.

He suggests that 'a primary task for the student of public policy is to understand and clarify the discourse or frameworks which structure the analysis of policy problems, content and processes' (Parsons 1995, p. 57).

With our own professional backgrounds in engineering and economics respectively, we plead guilty to the following framework beliefs in terms of how urban transport policy should be approached:

- Market failure is the fundamental basis for policy intervention, with applied welfare economics (substantially grounded in utilitarianism) providing an integrating framework for approaching policy development to tackle such market failure. Chapter 3 discusses many important issues on this theme and sets out some of the shortcomings we see with a welfare economics-based approach. It also suggests how some of these shortcomings might be handled. For example, distributional considerations incline us towards Rawlsian primary social goods (Rawls 1971) or a Sen (2009) capabilities-based approach to questions of social exclusion as this is affected by mobility. Chapter 6 discusses this particular question. With respect to major environmental questions such as greenhouse gas emissions and climate change, uncertainties and the potential scale of adverse impacts suggest the need for a risk management framework, which is likely to involve setting emission targets that provide a framework to condition the operation of markets. Chapter 7 considers such matters.
- Governments have a vital role in tackling market failures, but there are limitations on the capacity of governments to solve urban transport problems, for reasons such as the inherent complexity of these problems, inefficiencies in some governmental approaches and resource constraints. The public management approach to improving the efficiency and effectiveness of government (e.g. by using private sector approaches) has much to commend it in this regard, but we believe that this approach has gone too far in some cases, with a growing list of failed public–private partnerships (for example) as evidence of this position.
- What Parsons (1995) calls a 'stagist' (or rational decision-making policy cycle) approach to policy formation,² implementation and evaluation is useful, although we recognise limitations with the stagist approach. The central argument in favour of this approach is the need to reduce complexity to a manageable level (Parsons 1995). However, Parsons (1995, p. 81) argues that a stagist framework 'lends itself to a "managerialist", "top down" approach to the policy process and to a view of the role of policy analysis as a form of elite resource, rather than as involving wider social enlightenment'. The usual stages in urban transport policy development, implementation and evaluation are discussed in Chapter 3.

- Information is important for good policy, including information on facts, values and beliefs that are relevant to the policy issues at hand. We place particular importance on analysts understanding the value basis of policy analysis and the importance of being sensitive to how different value positions can influence outcomes of the analysis process. Many following chapters highlight values issues.
- The emphasis on values and information draws us to policy analytical approaches that include extensive open consultative processes. Our views on the constraints that limit the capacities of governments and markets to resolve all the problems of market failures also support an open consultative approach, in part because it prompts a search for what Etzioni (1994) calls 'communitarian' solutions to some problems. Etzioni argues that policy should seek to promote institutions that stand between the individual and the state (e.g. family, voluntary organisations, schools, churches, neighbourhoods and communities), an idea that resonates with that of the Big Society (see, for example, Blond 2010). Several arguments support the search for communitarian solutions to some transport problems. Chapter 6 discusses a relevant example one of us has been involved with in rural/regional areas.
- The recognition of what Simon (1957) calls 'bounded rationality', with the allied 'satisficing' decision-making model, is vital for shaping how policy should be approached, leading the policy analyst away from the rational comprehensive approach that was widely favoured in urban transport planning in the 1960s, towards less ambitious but more realistic analytical endeavours. The 'managerial approach' that characterises the rational comprehensive approach is still relevant but should (we believe) incorporate an open and consultative orientation, which is often not part of a managerial approach. Chapter 3 presents some discussion on these matters and on other decision-making frameworks.
- While most policy is inevitably incremental in nature (Lindblom 1959), the scale and trends in urban transport problems in many cities suggest that transformational approaches are also needed if fundamental issues are to be resolved better. The mixed scanning approach proposed by Etzioni (1967) is an effective way to approach policy in this regard, since it seeks to blend use of policy-making processes to pursue basic strategic directions, as well as the more usual incremental issues. This approach underlines the importance of integrated approaches. Chapters 3 and 16 are particularly relevant to these issues.

These points, in effect, summarise key elements of the framework within which the current authors approach urban transport policy. Those who adopt different frameworks are likely to reach different policy conclusions on

many occasions. However, our emphasis on the importance of information and on recognising and respecting pluralistic values should mean that the process of inquiry involved in policy development and implementation that we propose will enable those with different frameworks or values to reach many of their own conclusions from the information that our approach collects along the way.

We have been through a similar process ourselves in writing this book, particularly because one of us has a background in engineering and planning, while the other has a background in economics and policy. This means that we will come to certain issues from different viewpoints, and each viewpoint informs the other in important ways, being a microcosm of what planners and policy makers, pursuing integrated transport policy and planning, should expect. We also feel it important to stress that, in our opinions, there are no absolutes in policy debates and that the process of discussion and reasoning is really the essence of the process.

1.4 Postscript

Oh, and one final thing in starting: this book may not manage to remove the problems of urban transport congestion that we bemoaned at the start. We do hope, however, that it enables the reader to understand why not!

NOTES

- 1 www.thredbo-conference-series.org.
- 2 A stagist approach involves a policy life cycle of: problem, problem definition, identifying alternative responses/solutions, selection of the preferred policy option, implementation, evaluation and so on back through the cycle (Parsons 1995, p. 77).



REFERENCES

- Blond, P. (2010), *Red Tory: How Left and Right Have Broken Britain and How We Can Fix It*, London: Faber and Faber.
- Button, K. (2010), *Transport Economics*, 3rd edn, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Etzioni, A. (1967), 'Mixed scanning: a "third" approach to decision-making', *Public Administration Review*, **27**, 385–92.
- Etzioni, A. (1994), *Rights and the Common Good: The Communitarian Perspective*, Belmont, CA: Wadsworth.
- Lindblom, C. (1959), 'The science of muddling through', *Public Administration Review*, **19**, 78–88.
- Parsons, W. (1995), *Public Policy: An Introduction to the Theory and Practice of Policy Analysis*, Aldershot, UK and Brookfield, VT, USA: Edward Elgar Publishing.
- Rawls, J. (1971), *A Theory of Justice*, Cambridge, MA: Harvard University Press.
- Sen, A. (2009), *The Idea of Justice*, London: Penguin Books.
- Simon, H. (1957), *Administrative Behaviour*, 2nd edn, New York: Macmillan.

History of transport with policy implications



CHAPTER OVERVIEW

As a foundational building block for formulating transport policy in the twenty-first century, it is important to understand how modern-day transport has come to be and how it relates to land use. In this chapter, a brief history of transport is provided that shows how transport remained largely unchanged for nearly 5000 years, but has changed enormously in the most recent 200 or so years. Furthermore, a brief exploration is provided of the suspected relationships between urban form and transport. The concept of a human travel-time budget is also discussed briefly. The chapter concludes with a summary of recent policy directions in transport.

2.1 A brief history of transport technologies

As far as we can tell, prior to somewhere around 3000 BC, humans walked and were largely restricted to transporting only what they could carry over distances that they could walk. The most significant first invention in transport was that of the wheel, which appears to have occurred in Mesopotamia around 3200 BC (Anthony 2007). However, the wheel was not initially used for transport, but rather was invented as a potter's wheel. As a means of transport, the wheel appears likely to have been invented around 5000 years ago (Anthony 2007). Initially, wheeled vehicles depended on human power to move. Somewhere in the following 1000 years, animals began to be harnessed to wheeled vehicles to provide a means to move heavier commodities. They were also used without wheeled vehicles as beasts of burden, with loads placed directly on the animals. Around 2000 BC, spoked wheels made their appearance and, at about this time, horses appear as a means of transport, either to pull wheeled vehicles or to ride on.

Apart from coracles (small, lightweight boats powered by paddles) and then the evolution from them to a variety of sailing and rowed vessels for transport over water, developments over the next 3000 years were largely in terms

of improving the construction of wheels and developing improved wheeled vehicles for animals and people to pull or push. People started living in towns about 7000 BC. The evolution of wheeled transport affected the development of towns and their related agricultural hinterlands. This can still be seen in many countries, where the spread of the farming land seems to be governed by the travel-time budget of a human with a bullock cart.

In the late 1400s, Leonardo da Vinci theorised about flying machines, but nothing was actually built from his theories. The next significant transport invention was the omnibus, invented in France by the philosopher and mathematician Blaise Pascal in 1662. The name 'omnibus' is from the Latin, meaning 'for all', and subsequently was shortened to the word 'bus', which signifies an on-road vehicle 'for all'. Pascal's omnibus was a horse-drawn vehicle, and he operated five routes in Paris with fixed routes and fixed fares. His omnibus was designed to provide transport for the poor of Paris and operated for 15 years before the company went out of business as a result of bankruptcy. It was another 150 years before the next omnibus operation was started.

The next point in time when significant transport inventions appeared was 1763, which saw the invention of the first steam-powered boat and also the hot-air balloon. The bicycle made its first appearance in 1790, thus representing the first major change in land transport for a little less than 4000 years. In 1801, Richard Trevithick invented the first steam locomotive, which ran on roads, not rails. His invention was followed, two years later, by improved road building under the guidance of Thomas Telford. In 1807, Robert Fulton introduced the first steamboat with regular passenger service. These steamboats spread rapidly across the world and, within a few years, would be found traversing some of the major rivers in the eastern United States. One problem with them, however, was that they quite frequently blew up, especially if two captains of steamboats tried to race each other to the next boarding point – a not uncommon occurrence. George Stephenson introduced the first steam locomotive to run on rails in 1814. From a slow beginning, in which a man with a red flag had to precede the locomotive, the steam locomotive began to spread across the world, and rails, or iron roads, began to criss-cross various nations.

In 1815, John McAdam invented tarmacadam for surfacing roads, a product that later became known as tarmac. There is, as it were, a short pause in the development of means of transport at this point, with the next major invention appearing almost 50 years later in the form of the first underground railway, built in London in 1863 from King's Cross to Edgware Road. This

was, of course, still a steam railway and was built using cut-and-cover techniques, so that the tunnels were only a little below the surface of the ground. In 1867, the first motorcycle was invented, using steam power. In 1871, the first cable car was invented. Karl Benz built the first internal combustion engine in 1885 in an automobile, although automobiles so powered did not immediately become publicly available. Electric-powered motors now began to appear as well, and the first electric underground trains began running in 1890 between Stockwell and King William Street in London, a part of what is now the Northern Line. This was also the first deep-tunnel underground train, made possible by new tunnelling capabilities.

At the end of the nineteenth century, horse-drawn trams were operating in many cities. These trams were different from buses, in that the vehicles had flanged wheels and ran on steel rails, with the vehicles pulled by a team of horses. In 1893, Rudolf Diesel invented the diesel engine, which was not used in a land vehicle until about 1922, although it began to be used in ships and submarines in the preceding decade. Also, before the end of the nineteenth century, the first monorail was opened in Wuppertal in Germany. This monorail began operation in 1896 and is still operating today. Early in the next century, electricity gradually replaced the horses for trams, using an overhead wire or wires. The first aeroplane with an engine was invented and flown by the Wright brothers in 1903, beginning the age of powered flight. The internal combustion engine, in the meantime, was undergoing significant developments, so that in 1908 Henry Ford invented the assembly line for cars and the first Model T Fords became available at a price that many people could afford. Prior to this, cars had generally been for the rich alone and required a knowledgeable chauffeur to operate the vehicles. Ford wished to produce a car that a worker could afford, and did so by 1908 with the Model T, after having produced his first car, the Model A, in 1903.

Over the next decades, motorbuses and trolley buses began to replace trams, because these vehicles had more manoeuvrability than the tram. By the 1930s, most trams had been replaced, although some continue to operate in various cities around the world, such as those in Melbourne, Australia. Until the early 1900s, road vehicles used solid rubber tyres. The first pneumatic tyre was invented for the bicycle by John Dunlop, while the first pneumatic tyre for cars was invented by André Michelin, although it was not successful. It was not until 1911 that the first rubber tyre with a pneumatic inner tube was invented by Philip Strauss. This began a revolution for road-based transport, with cars, lorries or trucks, motorbuses and various other vehicles being invented or developed over the ensuing years. However, from this

point in the early twentieth century, ground transport saw relatively little major development, and attention moved instead to air and water. The first liquid-fuelled rocket was launched in 1926. It took the Second World War to see the development of significant advances in rocket propulsion, which was the forerunner of the space programmes in the USA and USSR in the 1960s. The modern helicopter was invented in 1940, and the first supersonic flight occurred in 1947. Jet aircraft, which made an appearance late in the Second World War, entered into passenger service in 1949.

In the 1950s, trolley buses were rapidly phased out of use in most cities in the Western world, being replaced almost entirely by buses. The hovercraft was invented in 1956, and the bullet train appeared in 1964. The first manned moon mission took place in 1969, and the first jumbo jet entered service in 1970. The first space shuttle was launched in 1981. The first self-balancing scooter – the Segway – was invented in 2001, and the first magnetically levitated train was put into service in Shanghai in 2004.

As this brief history shows, for thousands of years humankind was satisfied with animal-drawn vehicles, and animals as beasts of burden and also as a form of transport to be ridden. However, in the years from the mid-1700s until the mid-1900s, ground transport underwent massive transformations. While high-speed trains (bullet and maglev) were introduced in the latter part of the twentieth century and into the beginning of the twenty-first century, there has been little major advance in ground transport, especially for the urban area, apart from modifications to the motive power of the car, such as hybrid engines and fuel cell technology. Most recent technological development has taken place in air and space travel.

2.2 Transport in the early twenty-first century

This brief history of the evolution of transport shows a rather startling fact. Today, humankind travels around urban areas and also between urban areas in most parts of the world, using methods of travel that are generally over 100 years old. Travel in the urban area is by:

- walking (invented too long ago to know exactly when);
- bicycles (invented in the nineteenth century);
- trains (invented in the early to mid-nineteenth century);
- buses (invented in their present form in the early twentieth century);
- cars (invented in the early twentieth century in a form available to most people); and
- underground railways (invented in the mid-nineteenth century).

In many fields of endeavour, the last 40 to 50 years have seen major advances and changes. Fifty years ago, the person on the street did not know what a computer was. Advances in computing, medicine, space exploration and many other fields have been phenomenal in the past 50 years. Urban transport in particular, while benefiting from largely unseen offshoots of this technological progress, still remains rooted in the nineteenth and early twentieth centuries, albeit that performance levels (e.g. speed, safety levels and environmental performance) have improved. Interestingly, in the second third of the twentieth century, there were a number of inventions of such things as driverless transport vehicles and possible new rail-based transit technologies, many of which have found their way into intra-airport circulation systems, but almost none of these have appeared outside these specialist locations, and none have influenced the urban transport picture significantly.

From a policy perspective, this means that urban transport in particular is ripe for new technologies to be introduced. Most modern transport operates on fossil fuels and runs over roads and rails that use materials that are not sustainable, for reasons that are elaborated in various chapters in this book. The ever-increasing population of the world demands new solutions to mobility within and between urban areas, solutions that are not yet apparent. Transport policy needs to be open to the development of revolutionary new methods of moving people within and between urban areas.

2.3 Transport and urban form

While looking at the history of the evolution of ground-based transport, it is also appropriate to consider briefly the effects of transport on urban form and vice versa. A study of the growth of urban areas in Western countries, particularly, is quite instructive. Over the 4500 years from the invention of the wheel until around the early 1800s, conurbations were generally quite small throughout Western civilisation. In medieval times, the size of the city was often limited either by the ability of sound to travel or by the speed with which a horse could travel from one side of the city to the other. Sound had to do with the distance over which a human shout could be heard, so that, if enemies were seen approaching the walls of the town, the watchman could shout to warn defenders within the city to assemble at the appropriate point to repel the possible enemy. Alternatively, the city was limited to the travel time it would take to get from one side of the city to the other to provide the same warning, in time enough that defenders could assemble where needed. The remnants of walled towns throughout much of Europe attest to this

limitation in size and indicate that most towns and cities remained quite small until well into the eighteenth century.

The first major change in city size grew out of the Industrial Revolution, with its accompanying revolutions in the means of transport available. With the earliest forms of new transport becoming available being railways, many small towns grew up around railway stations along railway lines that fanned outward from the major towns and cities. As trams, buses and other road-based vehicles developed, the major cities began to expand outwards, often along the railway lines, and the newer small towns also grew along the rail lines, filling in development between towns and cities. The advent of the motorcar brought with it much greater mobility. The car and bus together brought further development that filled in between the rail lines, so that many small towns and cities that had grown up initially around train stations now were swallowed up into the urban fabric of the original major town. Evidence of this is clear in major cities around the world, such as London, Manchester, Birmingham, Glasgow, Edinburgh, New York, Chicago, Philadelphia, Sydney, Melbourne and so on. In all of these and many other cities, one now finds other business districts or just suburbs that carry the names of what were once individual small towns, located around the original city. In Sydney, for example, suburbs such as Parramatta, Chatswood, Epping, Ryde and so forth were once individual towns surrounding Sydney, but are now part of a continuous urbanised area that stretches well over 40 kilometres from the city centre.

Thus, transport has played a major role in shaping urban areas, making it possible for people to reside at increasing distances from their workplaces, and also allowing commercial and industrial developments to grow up at increasing distances from ports and markets. The increased mobility provided by evolving means of transport has provided accessibility of businesses and industry to workforces, workers to residences at increased distances from work and, in modern times, the increasing scale of retail outlets that provide economies of scale, but can draw patrons from increasing distances from their locations. The authors of this book grew up in the immediate post-war years in the UK and Australia. At that time, most retail outlets were small speciality stores located along the main roads of each small village, town or city. The word 'supermarket' did not exist in the vocabulary, and neither did the concept of a 'shopping centre'. Only with the burgeoning growth of car travel did supermarkets, shopping centres, megamalls and the like appear and the small speciality shops gradually disappeared. Without the car, it is doubtful that this change in urban structure could have taken place. Thus, one can consider that transport was the driving force behind the development of the

modern city, with its residential suburbs, industrial and commercial parks, major shopping centres and so on.

However, once the city has developed to a large conurbation, the roles of transport and land use seem to reverse themselves, to a significant extent. In the developed city, pressures to develop or densify land uses and the urban structure bring with them pressures to expand the transport system to serve the newly developed or more densely developed areas of the city. This has tended to produce the dilemma of urban and transport planners as to whether it is transport or development that leads. Unfortunately, also, because of the way in which government is often organised, land use and transport are often located in different portfolios of government, and these separate portfolios often do not communicate effectively on those aspects of land use planning and transport planning that would benefit from integration and extensive interaction. This issue is dealt with later in this book in Chapter 16.

Another significant issue in the area of transport and land use concerns what makes public transport, as we know it in the early twenty-first century, viable. Rail systems are efficient at moving large numbers of people from many origins to a few destinations, and then distributing the large numbers of people from few origins back to many destinations. Because of this, rail works best in cities that have a strong central area, with extensive employment, and that have grown in a more or less radial pattern. Rail also requires quite high densities of suburban development, to place a sufficient number of potential riders within easy reach of stations. Ideally, a modern train system needs stations to be spaced 3 to 5 kilometres apart in the suburban areas, although it can serve station spacings as close as 1 kilometre in the city centre. Low-density sprawl is not conducive to rail service, because there is a lack of sufficient patrons to make rail service viable. Also, rail service tends not to perform well in decentralised cities, where the demand patterns are from many origins to many destinations.

Bus systems can move more modest volumes of passengers and are more readily adaptable to changing land use patterns, because buses travel over the road system, rather than requiring substantial investment in a fixed right of way. Buses can serve a strongly radial-orientated city, but can also serve reasonably well in decentralised cities, as long as there are still concentrations of workplaces and moderately high-density residential developments. However, as density falls and workplaces become more dispersed, bus service will also tend to become less and less viable. Bus service does, however, have the ability to change more rapidly with changing land use patterns than rail service. A hybrid of rail and bus – bus rapid transit – loses some, but not all,

of the flexibility of normal bus service, and offers a higher speed that can be more competitive with the car. At the same time, the levels of investment required are generally significantly lower for bus rapid transit than for rail. Bus rapid transit may use dedicated lanes within existing road rights of way, or may operate over separate rights of way. Especially in the latter case, bus rapid transit may be a precursor to rail, where the bus rapid transit is able to build patronage that will eventually shift to rail, when volumes become high enough to warrant the investment in rail.

It should be clear, then, that low-density residential development and highly dispersed work locations are both problems for serving transport needs by public transport as we know it in the early twenty-first century. Such land use patterns encourage the use of the car and discourage use of and investment in public transport. On the other hand, in many countries, the goal and desire of many households is to be able to live in a single-family dwelling on a sufficiently large parcel of land to offer some isolation from immediate neighbours. Such aspirations tend to operate against the success of public transport and also to raise issues of long-term sustainability.

2.4 Travel-time budgets

In the context of previous discussion on transport and urban form, it is useful, before moving on to other issues, to examine a concept that was put forward about 50 years ago. This concept is that of a travel-time budget. The proponents of this idea observed from numerous cities around the world that the average amount of time spent in travelling per day seemed to be somewhere around 1 hour and 15 minutes, and was not found to vary greatly from city to city, nor from country to country (Szalai 1972; Zahavi 1973; Newman and Kenworthy 1989). Curiously, this observation was found to hold regardless of whether a country was considered to be a developed or a developing nation. From these observations, it was postulated that people have a constant amount of time out of the day that they are willing to spend in travel, and that this constant amount of time, while showing some day-to-day variation, will actually average out to about the same amount for every person. This is the travel-time budget, and has been suggested to be around 75 minutes plus or minus some small amount. While this idea has proved to be rather controversial, with some professionals arguing that they cannot see evidence to support it (Kitamura et al. 1992; Purvis 1994; Levinson and Kumar 1995; Levinson and Wu 2005), while others provide substantial evidence for it (Szalai 1972; Zahavi 1973, 1974; Zahavi and Ryan 1980; Zahavi and Talvitie 1980; Schaefer and Victor 1997; Schaefer 2000), if it is, in fact, potentially true, it has enormous policy implications.

First, it can be argued that any investment or policy that reduces the amount of time required for certain travel that people undertake on a daily, weekly or even less frequent basis will not diminish the amount of total travel time on the system. Rather, if people still wish to spend their budget of daily travel time, then new travel will be undertaken, to fill up the time saved, so that the total daily amount of travel time expended remains about the same. Second, if people are rational in their decisions (and we assume that most people are relatively rational), then it also follows (Zahavi 1973) that people are likely to attempt to maximise the total distance that they can travel within their constant travel-time budget, because this will probably maximise their access to various urban activity opportunities. If the car is the fastest means of travel, this would mean, first of all, that those people who do not have a car are disadvantaged, because they can travel lesser distances only, using slower means of travel. This could, at least in part, explain the desire that people exhibit to increase car ownership, especially in less developed countries, where car ownership is low. In addition, it would suggest that people will generally choose the fastest mode of travel from among available choices, because such means of travel will allow them to cover greater distances than by slower means.

From these observations, it follows that policies and investments that shorten travel times will not succeed in reducing the amount of travel as measured by person kilometres of travel (PKT), but will rather result in people travelling longer distances than before, or adding new travel, so that the total time spent travelling will remain approximately the same, but PKT will be likely therefore to increase. Second, efforts aimed at getting people to use their cars less and travel by walking, bicycling or public transport are unlikely to have much overall success. For long-term success, it would seem necessary that the alternative means of travel to the car must offer faster overall travel than the car. In the past, most policies that were designed to increase public transport use failed to look at the potential to speed up public transport and make it competitive in speed with the car. However, if travel-time budgets really exist, then such policies and investments are essential. Further, policies that aim at consciously reducing car speeds, especially as a means to encourage people to abandon their cars in favour of public transport, are also likely to be met with considerable disapproval by the public and are not likely to remain in position long.

Clearly, it is important to understand if such a thing as a daily travel-time budget really does exist. It is hoped that research currently being pursued at the University of Sydney, Australia will lead to clarification as to whether or not this concept has a scientific and observable foundation. If it does, it requires some considerable rethinking of policy directions in transport.

2.5 Principal policy directions

Throughout the world, both in countries that are well developed and those that are still developing, the current focus of urban transport policy is primarily on meeting the access requirements of growing urban populations in a sustainable manner, against a background trend towards ever greater urbanisation of the population, and also a trend of increasing motorisation, or near-stability of motorisation in those countries that have already reached very high levels of car ownership. In addition, as national economies around the world become more and more globalised and import and export of goods becomes a larger and larger part of these economies, handling increasing volumes of freight transport, especially by road, is another facet of the issues facing most governments around the world. Perhaps not surprisingly, therefore, many policies from transport focus on attempting to reduce congestion, with four primary directions considered:

- Building more roads and investing in road transport, often with accompanying urban sprawl, lower densities of development, and decentralisation of activities. While the 1960s and 1970s, in particular, saw many countries engaged in major highway development programmes, to cater for urban growth and rising car ownership, the period since has been more about managing road traffic growth and the associated congestion.
- Restraining the use of cars through tolls, road user charges, or other procedures that limit the availability of the car.
- Investing in public transport, walking and bicycling, in the hopes of diverting people from travelling by car to travelling by these alternative modes (such policies also generally require densification of residential development and concentration of employment locations, contrary to what road building encourages).
- Investing in more compact cities, to reduce the need to travel and the demand for travel.

The idea of managing access requirements in a sustainable manner means that policy concerns must extend well beyond issues of congestion, to include wider economic considerations, together with environmental and social policy questions. On the economic side, for example, current urban transport policy concerns encompass the idea of maximising the productivity of the urban system as it is impacted by transport, the productivity of the transport system itself (the network and the productivity of those using the network) and questions of how to achieve the most efficient and effective policy interventions.

The environmental performance of transport systems has an important modern policy history that reaches back to the late 1960s, in particular, encompassing such matters as the environmental impacts of facility construction (e.g. freeway construction) and the environmental performance of modal operations. Regulation of vehicular emissions, for example, has become the single most significant way of controlling noxious emissions from motor vehicles, through mechanisms such as the ever-tightening Euro standards.

The rapid increase in motorisation that took place in highly developed economies following the Second World War saw a correspondingly significant and unacceptable jump in the road toll. Many newly developing cities are currently experiencing similar problems. Safety regulation, therefore, also has a long history in terms of transport policy, from the red flag carrier mentioned at the beginning of this chapter, to more recent measures such as driver licensing, compulsory seat belt wearing, drug and alcohol testing, vehicle design standards and suchlike. Safety measures typically focus on the driver, the vehicle and the environment. Regulatory harmonisation is a major current policy focus, to improve the efficiency of the regulatory process (e.g. reducing compliance costs).

Another important strand in the evolution of current transport policy relates to concerns about the social or distributional dimension of transport. In the 1960s and 1970s, this concern was particularly focused on the social impacts of freeway construction, where neighbourhood destruction and social severance drew many people to the barriers. More recently the policy concern has shifted more towards recognising the link between transport and a person's capacity to be socially engaged and, in consequence, the person's well-being. This is a question about access to a society's resources. It includes matters related to disability access but also ranges far more broadly, to encompass access to urban opportunities by all urbanites and visitors.

These are just a few illustrations of how the transport policy agenda has emerged over the past few decades. It is interesting to note how these issues reflect and have been influenced by the broader emergence of public policy thinking over the past 50 years, through the public administration model (e.g. the 1960s and 1970s) and subsequent public management model (the 1980s, 1990s and 2000s) to the public value management model (the 2000s to the present).

Important questions to be raised, which we hope are answered in the balance of this book, are how to choose suitable policy directions, whether such

policy directions will work in the long run and whether this will solve our present transport problems.



REFERENCES

- Anthony, D.A. (2007), *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*, Princeton, NJ: Princeton University Press.
- Kitamura, R., J. Robinson, T.F. Golob, M. Bradley, J. Leonard and T. van der Hoorn (1992), 'A comparative analysis of time use data in the Netherlands and California', Research Report UCD-ITS-RR-92-9, Institute of Transportation Studies, University of California, Davis.
- Levinson, D. and A. Kumar (1995), 'Activity, travel, and the allocation of time', *Journal of the American Planning Association*, **61** (4), 458–70.
- Levinson, D. and Y. Wu (2005), 'The rational locator re-examined: are travel times still stable?', paper presented to the 84th Annual Meeting of the Transportation Research Board, Washington, DC, January.
- Newman, P. and J. Kenworthy (1989), *Cities and Automobile Dependence: An International Sourcebook*, Aldershot, UK: Gower Publishing.
- Purvis, C. (1994), 'Changes in regional travel characteristics and travel time expenditures in the San Francisco Bay Area', *Transportation Research Record*, **1466**, 99–109.
- Schaefer, A. (2000), 'Regularities in travel demand: an international perspective', *Journal of Transportation Statistics*, **3** (3), 1–31.
- Schaefer, A. and V.G. Victor (1997), 'The past and future of global mobility', *Scientific American*, **227** (4), 36–9.
- Szalai, A. (1972), *The Use of Time: Daily Activities of Urban and Suburban Populations in Twelve Countries*, The Hague, Netherlands: Mouton Publications.
- Zahavi, Y. (1973), 'The TT-relationship: a unified approach to transportation planning', *Traffic Engineering and Control*, **15** (5), 205–12.
- Zahavi, Y. (1974), 'The "UMOT" project', report prepared for the US Department of Transportation and the Ministry of Transport of the Federal Republic of Germany.
- Zahavi, Y. and J.M. Ryan (1980), 'Stability of travel components over time', *Transportation Research Record*, **750**, 19–26.
- Zahavi, Y. and A.P. Talvitie (1980), 'Regularities in travel time and money expenditures', *Transportation Research Record*, **750**, 13–19.

3

Policy needs and policy processes



CHAPTER OVERVIEW

This chapter provides foundational material for the remaining chapters. It takes a broad public policy perspective, which has relevance well beyond the transport sector. Policy interventions are usually a response to some form of market failure. The chapter considers common forms of market failure, with transport examples. It then asks how a policy analyst or decision maker might know if particular policy responses are likely to make society better off. This is followed by an overview of a range of alternative decision-taking approaches that are often used in policy formulation, together with an overview of the policy cycle. Some possible outcome indicators for transport policy are considered, and there is a brief discussion of the role of participation.

3.1 Rationale for policy

Markets are the primary means of allocating resources in most economies. They largely determine (for example) what is consumed, what is produced, the selling prices of goods and services, wage rates, interest rates and so on. Markets are usually regarded as an efficient way of determining resource allocation. The idea of an efficient outcome is often linked to the concept of a *Pareto optimum*, defined as a situation where it is not possible, through reallocating resources, to make someone better off without making someone else worse off.

The way markets reflect preferences of buyers and sellers is fundamental in them being regarded as efficient at allocating resources, as demonstrated in much of the foundational literature of welfare economics. However, there are a number of important situations where free markets fail in terms of achieving an efficient allocation of resources. These situations include:

- public goods (e.g. defence, law and order);
- merit or quasi-public goods (e.g. local roads);

- externalities (e.g. noise, air pollution, congestion);
- natural monopoly (e.g. rail);
- limited extent of markets (e.g. not all things that are valued pass through markets);
- lack of information; and
- distributional considerations.

They are explored in more detail below.

Market failures are the major rationale for policy interventions led by governments, on the presumption that such intervention will deliver better outcomes for society than a market that is failing. There is an active debate about the efficiency and effectiveness of government intervention that needs to be acknowledged in arguing a case for a governmental policy role, which usually also involves a related programme and/or project role. Much of the history of transport policy over the past 80 years has reflected changing attitudes about the effectiveness of government intervention (see, for example, Button 2010).

We use the terminology ‘led by government’ deliberately. For example, the private sector is playing an increasingly important and independent role in public transport service provision in many countries where service provision has been deregulated or made the subject of competitively tendered or negotiated contracts with government. The private sector role also extends, for example, to many toll roads. Such changes in institutional arrangements for transport service provision and operation mean that governments increasingly need to take policy decisions in a collaborative way, a trend further reinforced by citizens who expect to be involved in decisions that affect their well-being.

3.2 Market failures and transport policy

Public goods

A public good has two key characteristics: it is *non-rivalrous* and *non-excludable*. The former characteristic means that consumption of the good by one person does not reduce the amount available for consumption by others. The latter means that, once it is provided, it is available to all (no one can be excluded). Common examples are defence and law and order. In the transport context, policing of road safety has the characteristics of a public good.

Understanding people’s preferences for provision of public goods is difficult, for well-known reasons such as the ‘free-rider’ problem. If a public good is

provided for one, it is provided for all, so there are incentives for people not to reveal the true value of the good to them, in the hope that others may pay the cost. Such issues suggest that markets are not efficient in determining the optimum provision of public goods – undersupply is usually expected. This suggests survey techniques (for example) will struggle to identify demand for public goods. Varian (2006) discusses various approaches that have been proposed to reveal preferences for public goods. Some form of government intervention is usual in provision of goods. Public discussion about suitable levels of provision of particular public goods is common in electioneering.

Introducing the idea of a public good, with underlying qualities of non-rivalry and non-excludability, invites presentation of a simple matrix that illustrates the broad possibilities defined by these two qualities. Table 3.1 shows the two-by-two possibilities, with some relevant transport examples.

Table 3.1 Typology of market access

Market access	Excludable	Non-excludable
Rivalrous	Private goods: e.g. motor vehicles, food, clothing, holidays	Common goods or common property resources ('tragedy of the commons'): e.g. road space (traffic congestion)
Non-rivalrous	Club goods: e.g. theatres, private game parks, intellectual property, local roads in gated communities	Public goods: e.g. road policing; local roads that are open-access.

The category of rivalrous/non-excludable is particularly interesting for transport policy. This category includes those cases where people are not excluded and their (rivalrous) competition for use of the good reduces the value that others derive from using it. Use of unpriced urban roads, where congestion is a common occurrence, is a well-known example. The troposphere is another, which is a sink for the accumulation of greenhouse gases, implicated in climate change. This category of cases was the subject of Hardin's (1968) 'tragedy of the commons'. The lack of defined property rights to common property resources is central to problems that arise in this context, and policy solutions may involve seeking to assign such property rights, such as tradable emissions permits for greenhouse gas emissions or congestion pricing for peak congested roads.

Club goods are goods where exclusion is possible and non-rivalrous. Local roads have been included as an example of a public good, on the basis that

access is usually universally available. However, local roads that limit access by outsiders might be seen as club goods (e.g. gated communities).

Merit goods

A merit good (or service) is one which society, through its political processes, has decided should be provided on the basis of considerations of need rather than ability and willingness to pay. The good is provided in the private marketplace, but there is a social decision to ensure some base level is available, irrespective of individual preferences or circumstances. There is an implication that some people will undervalue the worth of the merit good, if left to their own devices, and that both society in the broad and some (possibly myopic) individuals will be better off if a base level of availability is assured. In short, positive externalities are expected from the provision of a base level of the good (see the next section on externalities), which may be benefits to the wider society and/or benefits to some individuals who would otherwise have under-consumed the merit good. Education is a common example, with private schools often being available, but a publicly funded system is usually provided that is available to all, to ensure that each person 'consumes' at least a minimum socially determined amount, for personal and wider societal reasons. As with public goods, free markets are expected to under-provide merit goods.

The concept of merit goods might be seen as linked to Rawls's (1971) notion of the provision of *primary social goods* (e.g. as a 'just' outcome arising from a social contract that is intended to look after the interests of the most disadvantaged) and Sen's (2009) concept of *capabilities* (morally significant freedoms that are needed to achieve a certain sort of functioning), work that was taken further by Nussbaum (2001).

In transport policy, the concept of merit goods is perhaps most relevant to the problem of *social exclusion*. An important transport policy concept in some jurisdictions is that of providing a minimum service level for public transport in a particular geographic area, by way of a social safety net. Chapter 6 discusses this matter in detail. A regional equivalent may be the provision of all-weather road access, to help assure reasonable car-based mobility in the absence of alternatives and to support operation of buses and of cycling.

Externalities

Perhaps the most important concept for modern urban transport policy is that of an *externality*. Pearce and Nash (1981) define an externality as

unpriced goods or bads that accrue to third parties. Distinctions can be drawn between *consumption externalities* and *production externalities*. A consumption externality arises when a person's enjoyment of some good or service is affected by another person's production or consumption behaviour and that effect is not priced. A *negative consumption externality* arises where that impact is to reduce the benefit that is derived from the person's use of the good or service; the converse applies for a positive consumption externality. Examples of consumption externalities that are of interest for transport policy are mostly 'bads', such as noise (see Chapter 7), air pollution (Chapter 7) and traffic congestion (Chapters 9 and 12).

A positive consumption externality that is transport policy relevant might be the benefits that urban residents receive from the wide range of services (e.g. theatres, libraries, events, etc.) that are available to them because of the agglomeration benefits made possible by urban rail systems. We are not aware of any quantification of such a possible impact, although there is evidence of consumption externalities (benefits) of cities (Borck 2007).

Production externalities arise when the production possibilities of a firm are influenced by the unpriced activities of another firm or by a consumer. These may also be positive or negative. The agglomeration benefits that accrue to businesses in large cities, which are likely to depend in part at least on urban heavy rail or metro capacity, are a relevant example. Agglomeration effects are discussed in Chapter 11. If unpriced production externalities are present, free markets will oversupply goods or services that are characterised by negative externalities and undersupply those with positive externalities.

An important requirement for the existence of an externality is the lack of suitable pricing of the effect that gives rise to the externality (i.e. pricing that internalises the relevant externalities into price systems). Pricing solutions are, therefore, a commonly proposed policy response from economists to deal with problems of externalities, linked to concepts such as 'the polluter pays'. This is seen as a way of dealing with the market failure associated with an externality. Regulatory solutions are more common (e.g. emissions requirements on new vehicles to reduce problems of air pollution).

Coase (1960) argued that assigning property rights to one or other party in a case of 'externalities' would allow the market to reach an efficient outcome via a negotiated process. It does not matter to the achievement of an efficient outcome in which party the property rights are vested, although one may have a view on the fairness of allocating rights to one or other party. An efficient solution is possible through negotiation provided transaction costs are

not so large as to make negotiation infeasible. In the latter cases, governmental pricing or regulatory solutions are likely to be preferred, if the external costs or benefits are sufficiently large to warrant action, provided these solutions are efficiently delivered. The lack of negotiated solutions to external costs in transport policy reflects the pervasive problem of transaction costs.

Hausman and McPherson (2006) make the important point that to tax (or price) rather than fine (or regulate) pollution is to treat pollution as socially and morally acceptable. Appropriate policy solutions to problems of externalities thus involve decisions on the fairness of the relevant approaches.

Monopolies

A monopoly describes a market that has only one seller. This situation may arise, for example, if scale economies exist in production of a good or service (i.e. unit costs decline as the volume or scale of output increases), such that one seller is able to capture the market, given the size of that market relative to the minimum efficient scale of production.

Profit maximisation in a monopoly market is usually expected to involve the monopolist taking advantage of its market power to lift prices above the level at which the net benefits from consumption of the good are maximised. In general, prices will be higher and output lower in a market where the supplier behaves as a monopolist rather than competitively. This forgone potential benefit is sometimes called the *deadweight loss* of a monopoly. Where monopoly is present, then, free markets will undersupply (and overprice) the good or service in question, relative to an efficient solution. Governments usually dislike monopolies because of the associated welfare loss.

A situation sometimes known as a 'natural monopoly' is one where a firm cannot operate at an efficient level of output (i.e. where the price of its product is sold at the marginal cost of production) without losing money (i.e. average costs exceed marginal costs at the output level at which price equals marginal cost). Any standard economic textbook usually contains a demonstration of this situation (see, for example, Varian 2006).

The track component of major transport infrastructure services, where lumpiness and lack of alternative use opportunities (immobility) are common characteristics, is a transport area where the question of monopoly is often encountered. Railways are a relevant example, frequently leading to ownership by government to help ensure supply at an efficient price/output combination. Some countries regard local bus services as natural monopolies (e.g.

Australian urban areas, and London, where there is what is usually called 'competition *for* the market'), but others look to the market to provide such services under a regime of open competition (e.g. the UK outside of London, where there is 'competition *in* the market').

Limited extent of markets

Not all the things that people value can be bought and sold in markets. For example, the enjoyment of a beautiful view, personal friendships, and knowledge of the existence of blue whales may be highly valued by some people and affect their well-being, but are seldom the subject of market transactions. If a major new transport project, such as a road or heavy rail line, severs a cohesive community, how does this particular impact get reflected in markets? Can changes in property prices be a guide, even though they are likely to reflect far more than such social severance outcomes?

Chapter 4 discusses some ways in which policy analysts might seek to value such matters, and considerable progress has been made over the past 50 years in imputing money values to some benefits and costs of transport interventions that do not have market prices. Limits to valuation are also discussed. The techniques of cost-benefit analysis, summarised in Chapter 4, are intended to incorporate such situations, but the difficulties of so doing (i.e. the difficulties of valuing all the potential policy/project impacts that do not pass through markets) are a criticism of this tool.

It is perhaps worth making the point here that not all countries use quantitative and comprehensive money-based evaluation tools, such as cost-benefit analysis, which relies on imputing monetary values for the externalities in order to find a satisfactory way forward. The USA largely abandoned these approaches in the 1960s, and replaced them with the cost-effectiveness approach. In this approach, the internal costs of alternative policies are determined, an array of externalities is identified, and the magnitude of those externalities is identified. However, no attempt is made to place monetary values on the externalities in the analysis process. Instead, it is left to the political bargaining approach to trade off different magnitudes of different externalities, one against another. The result is, of course, subjective, in the sense that two different political bodies, faced with exactly the same policy with the same costs and same magnitude externalities, may arrive at different decisions as to which policy to pursue. On the other hand, the cost-benefit approach can readily generate quite different results simply by using different monetary values of the externalities. In the end, either approach is somewhat subjective, a necessary fact of life in analysis and evaluation for policy, and

neither approach will necessarily lead to consistent decisions. Transparency is therefore vital in assessing impacts, whatever approach is used.

Lack of information

Standard economic arguments about competitive markets being efficient allocators of resources depend partly on an assumption about perfect information being available to buyers and sellers. Often this is not the case. For example, people may have false beliefs about the consequences of particular actions for their well-being, such as smoking. Alternatively, the consequences of some activities may be uncertain but with the possibility of being highly disadvantageous in some situations (e.g. the possibility of 'fat tails' in the probability distribution of temperature rise associated with climate change, as discussed in Chapter 7). In such circumstances, markets are likely to be poorly equipped to make efficient resource allocation decisions. Our view is that people should have the best possible information available to them to exercise reasoned judgement about matters of public policy. If there are doubts about the quality of such information in particular decision contexts, then every effort should be made to improve information quality, consistent with the scale of issue at hand, to assist informed policy decision taking.

Distributional considerations

Paretian welfare economics outlines the conditions under which competitive markets will produce efficient resource allocation outcomes, in the sense that it is not possible by reallocating resources to make somebody better off without someone else being made worse off. However, what is an efficient allocation in any context will depend on how the ownership of goods and services and factors of production is distributed. An infinite number of efficient allocations are possible, depending on the distribution of goods or services and factors of production between the members of society. If market prices are being used to reflect people's preferences for particular goods or services in a project evaluation, it needs to be recognised that these prices depend in part on the existing distribution of wealth and incomes, a matter to which we return in Chapter 4. Governments frequently act, through mechanisms such as taxation, income transfers and the direct provision of particular goods and services (such as merit goods as outlined above), to influence the distribution of resources (in the widest sense) between people in a society, recognising that markets alone are not sufficient to achieve socially just distributional outcomes.

So far as transport policy is concerned, distributional considerations are important in two major ways. First, particular transport policies and

associated programmes may be implemented specifically to meet needs of particular groups, to improve their well-being. Arguments for justice might mean that people who have some particular characteristics that indicate social disadvantage (e.g. unemployment) might be provided with fare concessions on public transport, for example. People living in areas of concentrated disadvantage might have high-quality public transport provided to improve their access to jobs and other opportunities. Second, in assessing the merits of particular transport policies, programmes or projects, the incidence of benefits and costs between different groups in the society is an important consideration that needs to be included in the assessment. As indicators of value, market values are loaded dice as a means of quantifying value. Chapter 4 explores this issue in greater detail, suggesting some ways this issue might be approached in transport policy evaluation.

3.3 How do we know if society is better off?

Social welfare functions

The various examples of market failures outlined above are endemic in the transport sector and underline why transport policy is such an active and wide-ranging field of endeavour. Market failures of central relevance to transport policy encompass such important issues as (for example):

- traffic congestion (typically costed at over 1 per cent of gross domestic product in developed economies);
- air pollution (less costly than congestion but still costly and a major killer);
- noise (a serious source of disturbance);
- greenhouse gas emissions (transport accounts for about one-quarter of international emissions);
- road deaths, which amount annually to about 1.2 million internationally; and
- social exclusion, where a lack of transport opportunities has been shown to be an important contributor.

If some form of policy (and allied programme and/or project) response to such issues is to be put together, on what basis might the policy analyst proceed to determine suitable policy advice?

Policy is largely about guiding action to improve the well-being of a society, where markets cannot be relied upon to achieve the desired result. Economists typically associate well-being with a person's preferences, but the

preceding discussion has suggested that there are also questions of justice, rights and freedom that may provide a basis for policy interventions.

Good policy is not likely to result from a series of ad hoc responses to particular societal problems of the kind outlined. It is more likely, in our view, to flow from a deliberative approach that consciously seeks ways to improve the well-being of a society, examining a range of alternatives for achieving this, cognisant of the scale of the issue(s) being tackled, with a clear view on ways in which the well-being of society might be judged as having improved. This implies a policy process that can be followed to maximise the likelihood of a successful outcome. Fundamentally, it requires some idea of how to judge whether a society is better off in one set of circumstances than another. The concept of a social welfare function is sometimes used to assist this process.

A social welfare function is simply a way to 'add up' the well-being of different people in a society, which is easily said, but not so easily done! It provides a way to bring together a range of impacts (benefits and costs) of policies (and programmes or projects) and to rank different distributions of well-being (often called 'utility' in welfare economics) among people. To do this, it needs a way to consider such matters as:

- What is to be counted or valued?
- How is it to be counted or valued?
- How are trade-offs to be made between counted or valued matters?
- How should different (groups of) people be treated?

Welfare economics provides the foundation for what is probably the most commonly used approach to tackling such questions. The fundamental *value judgement* behind welfare economics and cost-benefit analysis is that 'individual preferences should count' (Pearce and Nash 1981). Nash et al. (1975a, 1975b) discuss a range of issues that arise with the application of this value judgement, including circumstances in which an analyst might choose to restrict its application, a matter to which we return in Chapter 4 in relation to cost-benefit analysis of transport policies. Hausman and McPherson (2006) present a comprehensive assessment of the role of preferences in informing policy and the moral foundations of preference-based approaches.

A key point about a value judgement is that it is not verifiable or falsifiable in the way that an empirical statement is, but there is still an important role for moral reasoning, argument and judgement: 'some answers to moral questions are better than others and . . . rational argument can help one to judge which answers are better' (Hausman and McPherson 2006, p. 8). In

practical terms, this quickly leads to the importance of transport (and other) policy analysts achieving two things. First, at all times, transport and policy analysts must be clear about the value judgements that underlie their assessments, so decision makers and other interested stakeholders can be clear about those positions and, should they so wish, be able to test the implications of adopting different value positions. Second, policy propositions must be subjected to value sensitivity analysis by analysts, which means examining the implications of alternative value judgements for evaluation outcomes. Policy evaluations are usually subjected to sensitivity testing of various technical input assumptions (e.g. traffic and construction cost forecasts for a major road project). They should also be subject to value sensitivity analysis.

Voting

Majority rule through *voting* is one way for a society to allow individual preferences to count in ranking alternative possible social states (or transport policy options). The voting paradox shows a well-known problem with this ranking method. A desirable quality of a social welfare function or social ordering is that it be transitive: if society prefers X to Y and Y to Z, then it should prefer X to Z. The voting paradox shows that this cannot be assured, as illustrated in Table 3.2. In that table X, Y and Z denote different choices that can be taken, and A, B and C are the members of the society seeking to rank these alternatives. Two people prefer X to Y, two prefer Y to Z and two prefer Z to X: the transitivity requirement is violated.

Table 3.2 Voting and intransitive preferences

Person A	Person B	Person C
X	Y	Z
Y	Z	X
Z	X	Y

Arrow (1963) took this idea further and showed that there is no perfect way to add together individual ranked preferences to make one social preference. His 'impossibility theorem' set five 'reasonable conditions' that a democratic decision rule (or welfare function) should possess and showed that these cannot all be met at the same time. Decision rules that incorporate strength of preference are one way to avoid Arrow's outcome, although they do violate one of what he thought were reasonable conditions (dictatorship).

In terms of transport policy implications, we note three points about the preceding discussion:

1. Irrespective of Arrow's impossibility theorem, voting as a means of suggesting social ranking of particular transport policy options is likely to be an inefficient process, because of the transaction costs involved.
2. Indicators of a society's welfare that incorporate an element of individual strength of preference are preferred, in our view (a value judgement), to those that simply rank order options (e.g. voting). For example, an initiative that harms one person greatly and delivers a very small benefit to two people, of comparable circumstances (to avoid distributional considerations), should not generally be judged favourably by a two-to-one vote in our view.
3. Sampling of opinion on policies and/or their impacts is a useful input to the policy process, including in those situations where the issues at hand are difficult to quantify.

Some examples of social welfare functions

The idea of a *Pareto improvement* was introduced earlier in this chapter, with its origins in preference satisfaction as the way to judge the well-being of society as between alternative social states. A Pareto improvement implies a very restrictive social welfare function, namely that society is better off if at least one person can be made better off, with no one being made worse off. This is a most unusual circumstance.

The most widely used social welfare function, in terms of transport policy (although it is unusual to see it referred to by name), is the *classical utilitarian* (or *Benthamite*) social welfare function. This seeks to measure the 'utility' of each member of society (u_i) and aggregate across all n members of the society, as shown in equation 3.1:

$$W(u_1, \dots, u_n) = \sum u_i \quad (3.1)$$

As formulated, this makes no distinction between members of the society in terms of their degree of deservingness or need. A critical component of the approach is that it requires interpersonal comparisons of utility to be made, an unavoidable part of efforts to judge the relative performance of policy options. As discussed in Chapter 4, cost-benefit analysis (CBA) can be seen as an attempt to operationalise this formula, using money as the measure of utility (more accurately, CBA usually seeks to measure *changes* between different options).

A generalised version of the classical utilitarian welfare function introduces distributional weights to reflect a judgement about the relative deservingness of, or relative value derived by (or cost imposed on), different groups, in what Varian (2006) calls a *weighted sum-of-utilities welfare function*, shown in equation 3.2:

$$W(u_1, \dots, u_n) = \sum a_i u_i \quad (3.2)$$

In this formulation, the a_i weights measure how important each person's utility or welfare is judged to be in the overall social welfare function. Chapter 4 discusses some approaches to determining such weights. It is important to note, however, that the failure to specify particular weights (values of a_i) is *not* a value-free approach to policy. It is simply an approach that uses the (usually implied) value judgement that all groups or people should be weighted equally, irrespective of their personal circumstances (e.g. income, physical capacities, etc.).

Philosopher John Rawls has proposed an alternative form of social welfare function, focusing on the idea of justice in what is sometimes called a substantive theory of welfare (which says what things are intrinsically good). Rawls (1971) used the concept of a 'veil of ignorance' to ask what social choice rules a society might adopt. Under this veil of ignorance a member of a society does not know what position he or she will occupy in that society. Rawls suggests that, under this veil, society is likely to agree that social states should only be judged by how they affect the welfare of the worst off and that this should be judged in terms of an index of primary social goods.

Sen (2009) took a different approach and argued that welfare should be assessed in terms of a person's capabilities. How the person then used those capabilities was a matter for the individual. Capabilities are a basis for freedom, broadly understood (Hausman and McPherson 2006). Nussbaum (2001) set out ten capabilities that she thought were relevant cross-culturally as a basis for well-being and freedom. We return to her list in the discussion of transport policy and social exclusion in Chapter 6.

Different forms of social welfare function (e.g. Paretian, majority rule, utilitarian or Rawlsian) can give different rankings among choice alternatives. Table 3.3 shows two alternative choice possibilities for a community of two people and a measure of the welfare impacts of each alternative for each person. The Pareto rule would rank Y as preferred to X, since starting at X it is possible to make at least one person (A) better off by moving to Y, without making person B worse off. On the Rawls criterion, society would be

indifferent between X and Y, since the welfare of the worse-off person (B) is no different between X and Y (with the units understood as primary social goods in the Rawls case), even though person A is much worse off in X than Y, a consequence that we sometimes find disturbing from application of the Rawls approach. Utilitarianism would rank choice Y ahead of X ($21 > 11$). Majority rule would also rank Y ahead of X, because person A is better off (and presumably in favour of Y), whereas person B has the same welfare level for both X and Y (and is, therefore, presumably indifferent between them). If person A could find a way to switch some of her gain from being in state Y, rather than in state X, to person B, then a felicitous outcome might perhaps be achieved!

Table 3.3 Choice possibilities and welfare outcomes (1)

Choice	Person A welfare	Person B welfare
X	10	1
Y	20	1

Table 3.4 complicates comparisons somewhat by adding a third person. On Rawls's criterion, Y would be preferred to X, because the worst-off person, C, is better off in Y than X. However, person B is decidedly worse off in Y than X, a situation that may be distasteful to some as a consequence of the Rawlsian decision rule. Majority rule would yield indifference between X and Y, one person favouring each and the third person being indifferent. Utilitarianism would rank X ahead of Y. The Pareto criterion would not provide guidance, because we are not able to make one person better off without making someone else worse off, unless compensation can be arranged for the loser (along the lines suggested in relation to Table 3.2). This idea of compensation has some historical importance in the development of the theory of cost–benefit analysis, as outlined in Chapter 4.

Table 3.4 Choice possibilities and welfare outcomes (2)

Choice	Person A welfare	Person B welfare	Person C welfare
X	100	80	60
Y	100	61	61

These two simple examples serve to show that different social welfare functions can give different social rankings of alternative choice possibilities. They underline the importance of policy analysts being clear about the decision rules or (implicit) social welfare function they are using. They also show

why it is important for analysts to undertake value sensitivity testing in relation to how impacts on particular groups or individuals are put together to form a view on policy (programme or project) desirability. The examples emphasise for us (1) the importance of seeking measures of strength of preference when seeking to rank alternative possible choices (our value judgement) and, related to this, (2) the potential value from using redistributive measures as part of a policy package, to broaden the range of potential winners and associated likelihood that social welfare will indeed be improved if a particular course of action is taken.

3.4 Policy approaches

Understanding political decision making

One of the major errors that was made in the early days of transport planning and policy development was a failure to understand how political decisions are made. Transport planning and regional policy making largely had their origins in the immediate post-Second World War years, especially as computers began to be developed for non-military uses and it was thought that the size of the perceived urban transport problems of the time could potentially be handled by computers.

Much of the development of initial transport planning activity, leading to policy formulation, was undertaken by professionals who were originally trained as traffic engineers and who brought to the problem the attitudes and training of engineers. This led to a mainly engineering view of the world, which consists of rational decision making. It took around 20 to 30 years for transport planners and engineers to realise that one of the major reasons that many of their recommended solutions were not implemented was because they did not understand the way in which politicians make decisions. Such a failure to understand the decision-making process is bound to lead to a situation in which the solutions recommended by the technical experts or policy analysts do not provide the answers sought by those charged with the responsibility to make decisions. Indeed, it is also important to recognise that the technical experts do not make the decisions; rather the technical experts or policy analysts provide recommendations to decision makers and also answer their questions about the consequences of different decisions.

There is a wide variety of different models of political decision making that might be followed, and the policy analysis process needs to be integrated into this process. However, the set of decision-making models that we use here to illustrate the issues is primarily that originally suggested in a transport

context by Meyer and Miller (2001), who draw on the classical public policy decision-making models of Simon (1957), Lindblom (1959) and others, with the addition of one further model attributed to Etzioni (1967). This set of decision-making models comprises six different models, where an individual may use any one or may use a combination of models, depending on the context and the types of decisions that need to be reached, their value position, and the political environment in which decisions are being taken. The six models are termed: the rational actor approach, the satisficing approach, the incrementalist approach, the organisational process approach, the political bargaining approach, and the mixed scanning approach.

The rational actor approach

The rational actor approach is the decision-making approach that most closely aligns with the engineering approach to decision making. It is a decision-making process that is driven by the systems analytic approach to problem solving (Hitch and McKean 1967). In this approach, it is assumed that the process begins with the enunciation of basic values, from which a set of goals is determined for the particular issue at hand. Recognising that goals are rather abstract and often represent idealised end states that may not be achievable in reality, the next step is to develop a set of objectives from the goals. The objectives are measurable and achievable expressions of what is contained within the goals. So that the achievement of these objectives can be assessed, the next step in the process is to define criteria, which constitute the measurement of outcomes from the objectives. A useful illustration of this process to this point might be focused on a single value applied to a metropolitan area. The value is that people should be able to live in a good environment. One of the goals from that value might be that the metropolitan area should have clean air. The objectives from that goal might be that each of oxides of nitrogen, oxides of sulphur, volatile organic compounds, particulate matter and carbon monoxide in the atmosphere should be reduced from current levels. The criteria by which attainment of these objectives might be measured would specify particular target levels of each of the air pollutants that should not be exceeded in a specified time period within a specified frequency. The approach then continues with identification of the problems to be tackled, formulation of alternatives, and evaluation of the alternatives against the criteria, which leads to the selection of the optimal solution.

The planners and engineers who were initially involved in transport planning and policy at the regional level assumed that this was the model of decision making that would be used by all persons involved in the process of policy formulation and selection of transport solutions. This leads to defining the

decision process and formulation of policy through a goal-setting approach. With goals, objectives and criteria defined, it is necessary first to measure the current situation and then to identify the problems, where these are defined as those aspects of the system that are failing to meet the criteria. An exhaustive search is then undertaken of all the alternatives that could be undertaken to correct these failures in the transport system. Each alternative must then be assessed against all of the criteria, to determine which alternative solution will produce the best outcome. If the goals and objectives are complete, and all possible alternatives are identified and tested against the criteria, then the result will be the selection of the optimal or best solution.

The rational actor decision-making model starts out by assuming that decision makers are completely informed and that they wish to maximise achievement of a set of goals and objectives. It is a normative model of decision making. It requires a highly structured and data-intensive planning process to support it, because the planning process needs to be able to identify all feasible alternatives and to be able to evaluate all of these alternatives against a complete set of evaluation criteria, and then needs to be able to rank alternatives against goal achievement. This requires the most structured and data-intensive planning process of any of the decision-making models, and requires extensive analysis to be able to satisfy the demands of the process.

In reality, the rational actor cannot function, because the demands of this model are beyond the capacities of decision makers and planners alike. It is unachievable, because an exhaustive set of alternatives cannot be identified, all relevant objectives and criteria cannot be specified and, even if either or both of these could be done, it is beyond human capacity to be able to comprehend all of the alternatives and come out with an optimal solution. The approach is also flawed, in that, as Etzioni (1967) points out, values tend to be fluid and both affect and are affected by the decisions that are made; information about consequences is usually only partial at best; and neither the time nor the resources exist to be able to assess all of the information required to support this decision-making model. Indeed, as Etzioni points out, the rational actor decision maker 'will become frustrated, exhaust his resources without coming to a decision, and remain without an effective decision-making model to guide him' (Etzioni 1967, p. 386).

The satisficing approach

Recognising the impracticality of the rational actor approach, the satisficing model attempts to retain many aspects of the rational actor approach, but to do so more realistically. The satisficing approach, which is based on Simon's

(1957) bounded rationality approach, is still goal-driven, but is based on an assumption that it is not possible or feasible to seek for an optimal solution, but is sufficient to seek for a solution that satisfies certain minimum criteria. Typical of this approach is to define a limited set of goals and objectives, and hence fewer criteria. It is recognised at the outset that all alternatives cannot be specified, so that a process of discovery of alternatives is undertaken, flowing from the definition of the limited set of objectives and criteria. The criteria are also specified differently, in that they are now usually defined in terms of a minimum level that must be satisfied. In relation to the rational actor model, the satisficing model recognises that all alternatives cannot be defined, nor can all possible consequences be assessed, and that the 'best' solution is not likely to be found. Within the satisficing model, it is also possible for criteria to be redefined during the search process, particularly if a number of feasible alternatives have been found, none of which are able to satisfy one or more specific criteria.

The satisficing model is characterised by a sequential discovery process of alternatives and their consequences. For example, a particular alternative may be identified and assessed against the limited set of criteria, in which process it may be discovered that this particular alternative does not satisfy certain criteria, but by altering the alternative in certain ways – generating effectively a new alternative – those criteria can be satisfied. Only a restricted range of situations and consequences will be examined, although, again, these are subject to modification as the process unfolds. This model also may result in the definition of a set of actions that can be implemented in recurrent situations.

To support this model of decision making, the planner must assist in identifying acceptable levels of performance and assist in developing a limited list of alternatives, but be open to identifying or testing additional alternatives that may be generated as the decision-making process evolves. Determination of the impacts and consequences of the limited set of alternatives will also be limited, although, as with the list of alternatives, must be open to considering additional impacts as the decision-making process unfolds. While the evaluation criteria will be limited, the process is still driven by the attainment of specific goals and objectives. This model may be considered to be the pragmatic version of the rational actor approach.

The incrementalist approach

The incrementalist approach, which owes its origins to Lindblom (1959), represents a complete break from the goal-driven rational actor and satisficing

approaches. Where those approaches are proactive, being based on setting goals and objectives, the incrementalist approach is reactive. The incrementalist approach starts from a definition of a problem, and proceeds to seek out possible alternatives to solve the problem. In a transport policy nexus, this is typical of the approach often taken by road authorities, which may identify, for example, an accident 'blackspot' or a failing traffic signalised intersection, and seeks to find a solution to that problem. The approach is remedial in nature, that is, seeking to determine alternatives that will solve the immediate problem. The focus of this decision-making approach, as its name implies, is generally on marginal or incremental differences in consequences. In the incrementalist approach, only a small number of alternatives will usually be considered, along with a small set of 'important' consequences. The problem may also be continually redefined to make it more manageable. The incrementalist approach also will not lead to the 'right' decision, or even a single decision, but may result in a never-ending series of attacks on the problem, each one being a reaction to the latest situation, arising from the latest incremental solution implementation.

Planning to support this decision-making approach is quite limited. It is restricted to identifying a few incremental alternatives (i.e. marginal changes to the existing situation) and assessing a limited number of consequences from these alternatives. The emphasis is on the differences between alternatives, and there is no attempt to find an optimal or even a satisfactory alternative. The decision making itself is usually completed in limited time, with limited information and expertise as inputs. The alternatives considered usually vary only slightly from existing policies and programmes.

An example of this type of decision making, which is quite often found in transport policy, is provided by a certain southern US city. Similar to many cities in the US, this particular city was originally laid out with a complete grid system of streets. Over the years, however, many street segments were closed to permit large-scale developments to take place, which required the use of the right of way for a one- or two-block length. This resulted by the 1990s in an incomplete grid system, which would require almost any reasonably lengthy north-south or east-west movement to negotiate several 'dog's legs' in the grid to reach its destination. However, there also existed in this city one diagonal street. This diagonal street had, for many years, been one of the major congested streets within the urban area. Successive efforts (incremental) had been made to widen this roadway, prohibit parking along it, grade-separate certain intersections, and take various other actions to increase capacity. All of this was to no avail, in that congestion soon returned to the highway, so that new solutions needed to be sought. However, in a

traffic modelling exercise, it was found that, if the majority of the broken links in the grid were restored as roadways, the congestion on the diagonal highway all but disappeared. Clearly, most of the traffic on this highway had no business on the highway, but used it as a way to circumvent the discontinuities in the grid system. The incrementalist approach would not find this solution, because it would always fail to identify the actual reasons for the congestion. Rather, incrementalist solutions would continue to focus on reducing the congestion on that highway in a reactive approach, rather than a proactive one of seeking to understand why the congestion was arising and tackling that problem.

In summary, the incrementalist approach is remedial in nature (taking a 'band-aid' or 'sticking plaster' approach to solving problems), does not seek after future goals, is reactive not proactive, has to keep coming back to fix the same or related problems, and considers only marginal or incremental changes from the current situation. It is politically popular, because incrementalism does not 'rock the boat'. Also, the failure actually to solve the problem may be relegated to the next term of office of the politicians, so is not an immediate concern to the incumbent politicians.

The organisational process approach

Again, this is an approach to decision making that departs completely from any of the previous three models and is one that is often found in government departments, among other places. It is based on the recognition that decision making is influenced by the organisations to which decision makers belong or for which they have responsibility. It is based on the formal and informal organisational structures that usually exist in public agencies, on the channels of communication that exist within these agencies, and on standard operating procedures. Governmental action is the output of organisations. Decisions that are made by government leaders usually trigger organisational routines within the bureaucracy. These organisational routines define the range of effective choice. For example, a highway authority will not usually look at public transport or land use solutions to what it perceives as a 'highway problem'.

This approach to decision making recognises also that policies and programmes can be successful only to the extent that responsible organisations have the capability to carry out their responsibility and implement the policies or programmes. The approach is especially appropriate in organisations that may have design standards or other standard operating procedures that are mandated to be used, provided the basis of these standards takes

cognisance of wider interrelationships in the transport and land use systems. For example, in transport policy, this approach would be relevant in designing a new highway, where there will exist a range of standards specifying such things as lane widths, maximum gradients, safe horizontal curves, required sight distances and so on, once the political decision has been taken that a highway is the best solution.

The role of planning to support this model of decision making is quite different. In this case, the major concern will be the impacts of the alternatives on the programme or project development process itself, on budgets and on feasible implementation. Because standard operating procedures exist, it will often be considered unnecessary to assess the consequences of alternatives on the world at large (these should be taken into account in setting the standards), and the planning process will therefore not be required to take such things into account. Implementation is usually the key to the decision making in this case, so that the planning process also needs to be focused on implementation.

In summary, the organisational process approach to decision making examines only a restricted range of alternatives, which are based on the organisation's capabilities and portfolio, tends to be driven by standards and standard operating procedures, looks more at the impacts of alternatives on the organisation itself, its programming and its budgeting process, and focuses on implementation. It is very evident in government, where agencies often exist in 'silos' with little or no cross-communication, so that the range of options and alternatives that can be considered is restricted within each governmental agency, and solutions that would require multiple agency input are rarely identified. Such agencies also tend to operate in an environment of 'This is the way we do this', reinforced by the publication of volumes of standard operating procedures.

The political bargaining approach

The fifth model of decision making is the political bargaining approach. This is an approach that recognises that the decision process is not one-dimensional, but that there are many points of view, values, goals and interests that have to be taken into account in the decision-making process. It is based primarily on conflict resolution and consensus building. It will rarely, if ever, select the 'optimal' solution, but usually selects the compromise that satisfies the largest number of different stakeholders and decision makers. The outcomes of this process will usually be limited to those aspects of the problem solution on which majority agreement can be obtained.

Typical of this approach is that the more controversial aspects of problem solution will often be deferred or ignored (placed in the 'too hard' basket). This decision-making model represents a form of power sharing. It assumes that there are many actors involved in the decision-making process with different goals and objectives. The role of planning to support this model of decision making is possibly the broadest of any of the decision-making models, because the planning support must be flexible to respond to different needs, must be able to incorporate as much information as possible in the evaluation, so that different viewpoints will be able to determine relevant impacts and consequences, and also needs to be able to respond quickly as new issues and alternatives arise.

An example of this type of decision making is to be found in the decisions on the building of the first rail transit line in Los Angeles. Los Angeles is a complex governmental area with well over a hundred different jurisdictions within its boundaries. However, the main protagonists in the decision came from the City of Los Angeles, which covers an extensive but not entirely contiguous geographic area. It includes the central business district (CBD) of Los Angeles and the surrounding urban area, but also the San Fernando Valley, which lies across the Santa Monica Mountains to the north of the CBD. From a purely technical standpoint, the obvious place to build a rail line in Los Angeles would be along the Wilshire corridor, which is a corridor that extends for about 40 kilometres from the CBD to the ocean at Santa Monica. At the time that the rail line was being planned, this corridor had the most intensive bus service of any location in the region, with articulated buses running at about 2-minute headways or less, with the buses generally being full at most times of the day. This route would have begun in the CBD, within the City of Los Angeles, but then quickly moved into several other jurisdictions on its way to Santa Monica. City council members from the Hollywood area, however, wanted to see the rail line serve Hollywood, which, despite the images from outside Los Angeles, was actually at that time a fairly depressed urban area, in need of revitalisation. Hollywood, however, was about 5 kilometres north of Wilshire Boulevard. Similarly, city council members from the San Fernando Valley were adamant that the rail line must travel to the Valley, because they assumed that, otherwise, rail would never come to the Valley.

These pluralistic decision makers were satisfied, instead, by a routeing that began in the Los Angeles CBD, travelled initially about 7 kilometres down Wilshire Boulevard, then took a north turn up to Hollywood, traversed the middle of the Hollywood area for about 2 kilometres, then turned north again, under the Santa Monica Mountains, and ended with two stations in

the San Fernando Valley, as shown schematically in Figure 3.1. Such a routing would never have arisen from a purely technical solution or probably from any of the other decision-making models. However, it came out of a political bargaining process that then provided the impetus to obtain the local, state and federal funding that made it possible to construct the line. As a footnote, it should be noted that this routing subsequently was changed because of an encounter with methane gas from old, closed oil wells in the Fairfax district of Los Angeles, close to where the line was originally intended to turn north from Wilshire Boulevard to Hollywood. This necessitated further political bargaining to agree on the final routing that was built in the 1990s.

Figure 3.1 Schematic of the Red Line Metro Rail



The mixed scanning approach

The final decision-making model considered is Etzioni's (1967) 'mixed scanning' approach, which is a commonly used way (for example) to approach strategic transport planning and policy formulation and also to undertake major transport corridor studies. The approach arose out of Etzioni's dissatisfaction with the rational and incremental models. Mixed scanning recognises bounded rationality and makes fewer demands than the rational approach. It combines '(a) high-order, fundamental policy-making processes which set basic directions and (b) incremental ones which prepare for fundamental decisions and work them out after they have been reached' (Etzioni 1967, p. 385). The approach thus combines a broad scan across issues or options at a high level, to identify those matters that seem critical to a particular policy concern, with detailed investigation of those elements. This process may involve several layers of analysis, or scanning, at increasing levels of detail.

In practice, the approach recognises that there are usually a small number of critical factors that need to be understood in detail to make good decisions, in terms of community outcomes, and that a good decision-making process needs to identify these factors, through its broad scanning process, and understand them in some detail. As noted, this approach is particularly well suited for long-term land use/transport strategic planning and is, in essence, the approach taken in the 2012–13 Melbourne Metropolitan Planning Strategy update, which includes a long-term transport plan for the city.

Conclusions on decision-making models

As noted at the outset, this is by no means the only possible set of models of decision making. Furthermore, any one decision-making process is likely to involve elements of different decision-making models, and different models will be more or less appropriate under different circumstances. This review of decision-making models also leads to the conclusion that the rational actor approach is infeasible, while all of the other approaches will have certain limitations, and only the satisficing and mixed scanning models are really capable of being proactive, while the other models will generally be reactive. Somewhat of concern is that the different models of decision making all have different requirements from the planning process. It is therefore necessary to think more about what the reality is of political decision making, and how this might relate to an effective planning procedure.

The nature of real decision making

Decision making, which leads to the formation and acceptance of policy, is quite different from the view of the planners and engineers of the 1950s. We can identify five different aspects of real decision making that are at odds with the rationalistic view of decision making. Actual decision making is:

- pluralistic, dealing with many jurisdictions, many publics and many agencies, with no single identifiable public interest;
- resource-allocative, so that policy is inevitably concerned with how best to distribute limited available resources among competing problems and issues;
- consensus-seeking, because there will be conflict among the different publics, different agencies and different jurisdictions, so that the goal is to satisfy as many stakeholders as possible;
- problem-simplifying, because urban problems with linkages to land use, economic development, environmental impacts and so on are highly

complex and it is beyond the power of most human beings to comprehend the full complexity of the issues; and

- uncertainty-avoiding, because politicians always wish to avoid uncertain outcomes and will tend to focus on the period of time until the next election. Rapid change is a cause of considerable political discomfort, because it contains the seeds of extreme uncertainty.

Hence, real decision making has to be based on compromise, negotiation and bargaining, will be influenced by powerful interest groups and will be involved in contending with many different issues. Real decision making will tend to favour short-term solutions, because of the period of election to a political position, and is therefore rather unlikely to deal effectively with problems that are ten or more years in the future. Planners and analysts who can increase the policy decision-making focus on the long term will be doing their communities a service, in our view!

Supporting real decision making

The process for supporting real decision making should still be modelled on the normative case of the rational actor approach to decision making, in that goals and objectives should still be set up at the outset to guide the formulation of policy. In turn, these goals and objectives will assist in identifying the issues or problems that need to be tackled, and will help lead to options that should be considered. Ideally, planning and policy should be proactive, but often will have to be reactive to a degree. However, somewhat along the lines of Etzioni's (1967) mixed scanning approach, it may be necessary to undertake these tasks at different levels of detail. For example, a broad look might be taken at the metropolitan area as a whole, from which decision makers may wish to zoom in on a particular area or sub-region, and then further to a particular corridor. As the geographic level of detail increases (i.e. zooming in), the level of analysis and supporting information will increase in detail. The types of solutions that will be likely to be generated will be satisficing or incremental, and will usually require modification to build consensus. Provided the mixed scanning approach has been well framed to drive outcomes towards identified social goals, this is realistic and reasonable. Recognising that programming and budgeting are key concerns of most government departments, the planning support must also examine the impacts of potential solutions on these facets of departmental function.

Necessarily, there will be an evaluation phase, in which the consequences of adopting alternative solutions are evaluated against the goals and objectives,

and also are evaluated against the different viewpoints and priorities of the decision makers. Iterative processes will enable high-level directions to be set, with more detailed solutions explored within this framework, along mixed scanning lines. Once a solution is selected for implementation, determining how and when it will be implemented and over what period of time is also a critical component of the support of real decision making. Too often in the past, planners have ignored the time it will take to implement a solution, and have tended instead to evaluate projects as though they would be implemented instantly, whereas the reality, depending on the magnitude of the project, could be that implementation will range from months to several years. The policy cycle is discussed in more detail in section 3.5.

PESTEL – a framework for strategic thinking in complex situations

As we have noted in this chapter, the actual decision-making context in transport policy is complex and subject to multiple viewpoints, multiple publics, multiple jurisdictions and multiple agencies. Therefore one may wonder how it could be possible to set about identifying goals and objectives (or, in other words, strategies) and then analysing the issues and problems that require decision making. PESTEL provides a framework in which this can be done. PESTEL is a method to organise one's thinking concerning six groups of factors that need to be considered in strategic thinking:

- Political;
- Economic;
- Social;
- Technological;
- Environmental; and
- Legal.

PESTEL was originally designed as a business strategy for managers of private companies, but works equally well for setting policy in a complex governmental arena. It must be remembered that the way in which an issue is defined will affect extensively the solutions proposed and the skills required for analysis and solution.

PESTEL is first and foremost a method of organising thinking to consider the influences and constraints that act on a specific problem area. For example, in setting transport policy, political factors might include the lack of political will for radical change (the 'don't rock the boat' mentality), the

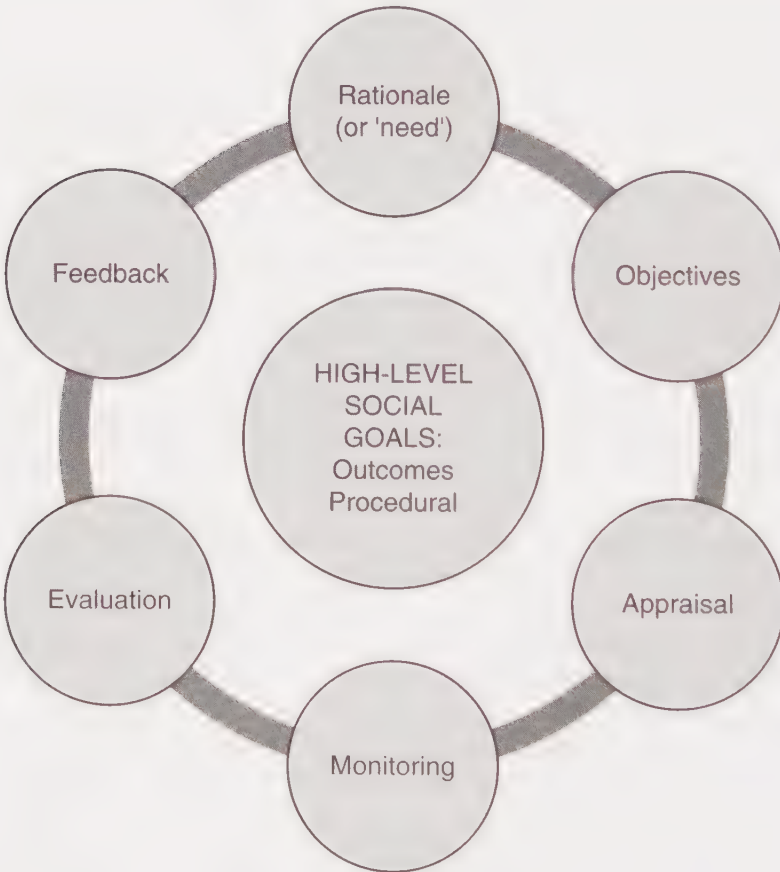
short-term and incremental political culture, the aversion to risk of the civil service, the attitudes and electoral importance of specific population segments, and the terrorism or security environment. Economic factors may include the globalisation of the world economy, the attitude of Treasury towards valuing infrastructure and rationing capital, an ideological fixation on competition, and the increasing costs of congestion, among others. Social factors might include consumer attitudes to transport, the perceived right to mobility, evolving lifestyle trends, and the status conferred by car ownership, especially in less developed countries. Influences and constraints of technology may include the fact that there is no apparent technological fix, appraisal and modelling techniques are limited and subject to manipulation, and there is a reliance on the expertise of consultants and financiers who have their own agendas. Environmental factors could include carbon emissions and climate change, the impact of climate change on transport infrastructure, the uncertainty of the impacts of carbon trading, local air quality issues, attitudes to transport-related noise, and so on. Finally, legal factors might include constitutional issues between different jurisdictions, developing law areas that pertain to transport, such as new regulations and rules, impacts of health and safety legislation, impacts of disability legislation, land use planning legislation, and so forth.

Simply listing these factors will not be a sufficient action by those supporting the decision-making process. It is necessary to consider these and to see how they may affect the decisions and policies to be made. For example, a political factor might be the need to involve multiple government agencies and also some private firms and population groups – in other words, identifying the stakeholders who need to be involved in decisions. Having identified the various stakeholders, one can then ask questions such as: Who will be affected? What will the concerns be of those who are affected? What policy solutions will they desire as a result of their concerns? Another political factor may raise questions as to what will be the consequences if nothing is done (this relates directly to the lack of political will for radical change). Also, politicians and civil servants will want to know about the risks and the unintended consequences.

In summary, PESTEL provides a useful framework or checklist for thinking about issues and problems and developing appropriate strategies in a policy context. It may be necessary to perform a PESTEL analysis at frequent intervals during the implementation of strategies and policies, because the external environment changes continually, and issues and constraints that may have been important at the outset may diminish in importance, while new issues and constraints may arise.

3.5 Generic approach to a policy cycle

The valuable UK Treasury *Green Book* (HM Treasury 2012, p. 1) argues that 'All new policies, programmes and projects, whether revenue, capital or regulatory, should be subject to comprehensive but proportionate assessment, wherever it is practicable, so as best to promote the public interest.' To assist that process, it sets out a broad policy cycle for appraisal (ex ante assessment of options) and evaluation (ex post assessment). Starting at 'Rationale', the outside circles in Figure 3.2 show that cycle, which is pursued in a clockwise direction. Figure 3.2 adds an important central focus on high-level social goals to the UK Treasury cycle, to emphasise that these should be the starting point for all policy interventions and should provide a focus for attention



Source: Based on HM Treasury (2012, p. 3)

Figure 3.2 A policy cycle

throughout the entire policy process, since it is the high-level outcomes that are ultimately to be sought through policy interventions.

High-level social goals related to societal outcomes are usually characterised in triple-bottom-line economic, social and environmental terms. It is common to see governmental transport policy statements affirm this focus. For example, the Ministerial Foreword to Norway's *National Transport Plan 2010–2019* says: 'It is about making everyday life easier; about sustainable development; making society more inclusive and universally accessible; and strengthening the competitiveness of Norwegian commerce and industry' (NMTC 2011, p. 3).

The Council of Australian Governments (COAG) has set a national objective to ensure Australian cities are 'globally competitive, productive, sustainable, liveable, socially inclusive and well placed to meet future challenges and growth' (COAG 2009).

Ensuring that a transport policy process keeps a focus on high-level social (policy) goals does not mean that detailed prescriptions of those goals, and trade-offs between them, are needed prior to the start of the process. A sufficient level of detail might simply be that the goals indicate that priority is attached to, for example:

- promoting economic productivity (or efficiency) and competitiveness;
- promoting social inclusion; and
- reducing the ecological footprint associated with transport.

Specific focus may also be given, for example, to a desire to improve health and safety and to promote regional development. A narrow set of definitions could encompass these under a triple-bottom-line focus but, because of the widespread community interest in such outcomes, specific nomination as goals is often pursued.

The Institute for Transport Studies at the University of Leeds has surveyed a number of European cities and indicated they are generally pursuing the following high-level objectives ('goals' in our language) for a sustainable transport system: economic efficiency; protection of the environment; liveable streets and neighbourhoods; safety; health; equity and social inclusion; contribution to economic growth; and intergenerational equity (KonSULT 2012). This is consistent with the preceding goal listing, but the Leeds addition of 'intergenerational equity' (which they suggest was not common) is useful, particularly in the context of climate change. 'Liveable streets and

neighbourhoods' is a useful goal or objective at city level but not for higher jurisdictions. It is arguably covered by the goals of social inclusion and safety, being one of the means to the achievement of these goals.

Listing high-level social goals towards which a transport system and services should ultimately be targeted, and against which transport and related initiatives should be judged, does not say anything particular about how achievement in one goal area might be traded against achievement on another, to advance society's well-being. Detailed information on such trade-offs between various goals can emerge as the policy process takes place and decision points are reached, because such trade-offs will generally be circumstance-dependent. For example, the amount of traffic delay that a government might be willing to contemplate in the interests of slowing traffic to improve road safety will depend *inter alia* on the starting levels of congestion and safety, prospective changes that might be achievable and how the local populace feels about such choices.

Procedural goals are included in Figure 3.2 because many people are concerned about matters such as freedom, justice, fairness and rights (Hausman and McPherson 2006). These are important matters in establishing the scope of transport policy and how one conducts the policy process. Procedural goals might relate (for example) to an expectation that people have the right to input in a policy process on matters that affect them, whether directly or indirectly. They might relate to a desire to improve the situation of the least advantaged, through structuring of suitable policy initiatives and supporting their engagement in the policy process, although such a focus could also follow from a social inclusion agenda on the outcomes side. Other procedural matters that might be important include accountability and transparency of process, to support concerns with freedom, justice, fairness and rights. As an aside, one of us once had an illuminating discussion with a former Australian cabinet minister about whether road pricing constituted an infringement of a right to free and unfettered access to the road system.

Rationale (or need)

This (first) step in the cycle is to identify the need that is to be the subject of attention. In the transport policy context, this need will usually be some particular transport or transport-related situation that is having a perceived adverse impact on achievement of the high-level social goals or a beneficial opportunity that is being forgone but might possibly be realised. The need might relate, for example, to greenhouse gas emissions from international aviation, a high national road toll, traffic congestion in a city, air pollution along a busy road,

or the unequal distribution of the benefits of transport accessibility between different groups in the community. Transport policy issues typically span this wide range of interest/jurisdictional levels and there is often a requirement for cross-jurisdictional cooperation to develop integrated policy directions. Such cross-jurisdictional work may involve *vertical integration* between different levels of government working on a shared problem (for example, national and various state or provincial governments working on a road safety strategy, where each might have different roads for which they are responsible) and/or *spatial integration* between governments at the same level that share a common concern (for example, neighbouring local municipalities may share a major congested arterial road, or state or provincial governments may decide on a shared approach to regulating truck driver hours of work in the interests of safety, recognising that truck movements cross state or provincial borders).

Transport and transport-related needs may be identified through the monitoring stage in the transport policy cycle (Figure 3.2). They might result from advocacy campaigns run by interest groups, such as political parties, public transport users, an automobile association, an environmental or social welfare peak body or the media. They also commonly result from detailed research by policy analysts, including academics, who have examined transport system and/or service performance and the consequences thereof, now and projected into the future, and identified deteriorating outcomes against goals or objectives.

Objectives

In this step, the desired transport system/service level outcomes from tackling the issue or problem of concern are set out. These need to be indicators that, if met, provide some comfort that advancement of the high-level goals will be promoted. To that extent, they need to be comprehensive, but tempered by the scale of the problems or issues at hand. Setting down a list of objectives has the useful benefit that it assists the search process for options to tackle the problem(s) or issue(s) of concern and provides a foundation for subsequent performance assessment.

Table 3.5 lists indicators from the European Union's PROPOLIS research project. The scope of these issues illustrates the range of matters likely to be important in policy, programme and/or project evaluation, recognising that the level of effort devoted to any particular evaluation process needs to be commensurate with the scale of the problems or issues under review. The PROPOLIS listing is a little dated in some sections now, such as in its treatment of social exclusion and agglomeration (see Chapters 6 and 12).

Table 3.5 PROPOLIS indicator system

Indicators	Theme	Indicator
Environmental indicators	Global climate change	Greenhouse gases from transport
	Air pollution	Acidifying gases from transport
		Volatile organic compounds from transport
		Consumption of mineral oil products by transport
	Consumption of natural resources	Land coverage
	Environmental quality	Fragmentation of open space
Social indicators	Health	Quality of open space
		Exposure to PM from transport in the living environment
		Exposure to NO ₂
		Exposure to traffic noise
		Traffic deaths
		Traffic injuries
	Equity	Justice of distribution of economic benefits
		Justice of exposure to PM
		Justice of exposure to NO ₂
		Justice of exposure to noise
		Segregation
	Opportunities	Housing standard
		Vitality of city centre
		Vitality of surrounding region
		Productivity gain from land use
	Accessibility and traffic	Time spent in traffic
		Level of service of public transport and slow modes
		Accessibility to city centre
		Accessibility to services
		Accessibility to open space
		Transport investment costs
		Transport user benefits
Economic indicators	Total net benefit from transport	Transport operator benefits
		Government benefits from transport
		Transport external accident costs
		Transport external emissions costs
		Transport external greenhouse gases costs
		Transport external noise costs

Source: KonSULT (2012)

Programme budgeting processes over the past 30-plus years have distinguished between:

- inputs = resources used in a particular application (e.g. buses and drivers added to a route service);
- outputs = indicators of actions taken in that application (e.g. change in route bus service kilometres);
- intermediate outcomes (e.g. bus mode share, change in mode share between bus and car); and
- outcomes = consequences of the actions, for high-level social goals (e.g. congestion cost savings, social inclusion benefits, improved safety, lower greenhouse gas emissions, lower air pollution levels).

Because transport demand is a derived demand, in the sense that people or goods usually do not move simply for the sake of it but to achieve other purposes, intermediate transport outcome indicators are particularly important for transport policy. In our terminology, *outcome indicators* should relate to *high-level social goals*. *Intermediate outcome indicators* relate to *transport system/service level objectives*. It is at the intermediate outcome level that much transport policy direction will concentrate. For example, Stanley and Barrett (2010) have argued that, to maximise prospects for successful high-level goal achievement, Australian land transport policy should aim to:

- reduce the demand for travel;
- achieve a mode shift to lower-impact modes;
- improve vehicle utilisation;
- reduce vehicle emissions intensity;
- increase mobility opportunities; and
- create a more sustainable freight network.

Policy performance indicators are needed at both the high level and at intermediate levels. Ratios between output and input measures provide indicators of relative technical efficiency (e.g. bus service kilometres per bus). Use of outcome indicators leads to effectiveness measures (e.g. congestion cost savings per bus kilometre).

Appraisal (or evaluation)

This step involves identifying a number of ways to tackle the identified issue or problem and assessing the expected outcomes of these various alternatives. Chapter 4 deals with this step in detail, since it is possibly

the single most important part of the policy analysis cycle. UK terminology for this step is 'appraisal', with 'evaluation' being used to describe a post-implementation assessment of performance. Elsewhere, 'evaluation' is widely used to describe both the ex ante ('appraisal' in Figure 3.2) and ex post ('evaluation' in Figure 3.2) steps. This step will typically involve several iterations to identify a potential preferred solution. Depending on the practices of the relevant jurisdiction, the process may involve considerable, little or no input from the political decision maker(s). In line with the importance we attach to the individual preferences value judgement, we see public participation as a vital ingredient in this step, from need identification and clarification, through objective setting, to option identification and evaluation or appraisal. This is an integral part of exploring trade-offs and refining options, to identify solutions that are likely to best meet the needs of multiple stakeholders.

Monitoring

One aspect of support for real decision making that is often neglected is that of monitoring. The question needs to be asked as to whether the implemented policy or project solved the problem or dealt with the issues for which it was proposed as the solution. This requires two things. It requires a budget for ongoing monitoring following implementation, and it also requires an honest assessment that can recognise when solutions have actually not resulted in the expected consequences. If more monitoring took place, our understanding of why things happen the way that they do could be greatly enhanced, and future policy making would become more effective. Unfortunately, because this post-project monitoring is rarely done, our understanding of consequences remains quite limited, and unforeseen consequences will arise all too frequently. Table 3.5 includes some examples for monitoring national land transport system performance.

Evaluation

Drawing on the monitoring process, evaluation is a post-implementation assessment of outcomes to judge delivery and to identify possible ways of improving outcomes and lessons for the future.

Feedback

Feedback is part of a process of continuous improvement, to improve the quality of outcomes.

3.6 Generating alternative solutions

In transport policy, programme and project evaluation (using this term to include appraisal), considerable focus is usually put on *how to evaluate one or more options* that have been identified to resolve some particular identified need. It is far less common to see much time spent discussing *how to select a set of options* that is likely to provide a good solution to the identified need (project generation). A useful project in this regard, for urban transport policy, has been the University of Leeds KonSULT project (KonSULT 2012). The user has the option to explore a comprehensive range of possible policy measures, including land use, information provision, infrastructure provision and management, attitudinal measures, and pricing. These various instruments are broadly rated for their likely effectiveness in contributing to the achievement of a number of transport objectives. While the particular policy instrument ratings and impacts embedded in the current software might be most appropriate for a UK setting, and the researchers appear to have a liking for road pricing solutions, the generality of policy issues and instruments that are included makes the tool a useful aide-memoire for generating possible solutions to be explored by policy analysts more broadly. The likely impacts in any particular setting will be case-specific and should be verified for that setting. The set of policy instruments included in KonSULT is set out in Table 3.6.

Any given policy instrument in this list might have a vastly different impact depending on how and where it is used, but the checklist of instruments should be very useful to transport policy analysts, to start exploring possible solutions to transport policy issues that are of concern.

It should be noted that the list of measures in Table 3.6 is essentially urban in scope. It does not include, for example, broader (sectoral) policy instruments such as:

- vehicle emission and fuel economy standards, which are an important element in lowering air pollution and greenhouse gas emissions;
- heavy vehicle driving/working hours regulations, which are part of a policy package to improve road safety; and
- road rules, which are also critical in road safety.

However, the listing is very useful for urban transport policy instruments, particularly as this relates to people movement.

It is important to recognise that most of the policy instruments set out in Table 3.6 will affect achievement of multiple high-level and intermediate-level

Table 3.6 KONSULT policy instruments

Land use measures	Attitudinal and behavioural measures
Development densities	Individual marketing to reduce car use
Encouraging public transport use through land use planning	Flexible working hours
Development pattern	Car clubs
Parking standards	Telecommunications
Developer contributions to infrastructure financing	Company travel plans
Value capture taxes	Ride sharing
Infrastructure measures	Infrastructure management measures
New road construction	Road maintenance
New off-street parking	Conventional traffic management
New and upgraded rail lines	Urban traffic control systems
New stations	Intelligent transport systems
New rail services on existing lines	Accident remedial measures
Guided bus systems	Traffic-calming measures
Park and ride	Physical restrictions
Terminals and interchanges	Regulatory restrictions
Cycle routes	Parking controls
Pedestrian routes	Public transport service levels
Pedestrian areas	Bus priorities
Lorry (truck) parks	Bus fleet management systems
Trans-shipment facilities	High-occupancy vehicle lanes
	Cycle lanes and priorities
	Cycle parking provision
	Pedestrian crossing facilities
	Lorry (truck) routes and bans
	Road freight fleet management systems
Information provision	Pricing measures
Conventional direction signing	Private parking charges
Variable message signs	Road user charging
Real-time driver information systems and route guidance	Vehicle ownership taxes
Parking guidance and information systems	Parking charges
Conventional timetable and other service information	Fuel taxes
Real-time passenger information	Fare levels
Trip planning systems	Fare structures
Operation information systems	Concessionary fares
Static direction signs	
Tactile footpaths	

Source: KonSULT (2012).

outcomes, and rarely will only one instrument be sufficient to produce the best result. Some instruments complement others. The implication is the need for *an integrated approach* to formulation of transport policy, seeking combinations of policy instruments that will maximise multiple goal achievement.

The particular actions that are most appropriate in any specific city or other location will need to be locally determined, although the types of initiatives tend to have commonality in similar geographic areas. Thus, for example, congestion is as common in Bangkok as in Paris or Shanghai, and a number of similar policy instruments will be used in each case, though with differing intensity. Evaluation approaches are vital in developing the best combination of instruments for any particular set of circumstances.

The importance of integration across and between levels of government was noted above. The list of policy instruments reinforces this, because not all instruments are necessarily under the control of the same level of government. It also underlines the importance of integration across agencies, particularly land use and transport, but also integration with agencies with responsibilities in areas such as education, health, welfare, economic development and energy. This cross-agency integration is sometimes called *horizontal integration*.

A critical area for integration in transport policy is with land use. We discuss integration at this level in Chapter 5 (on transport planning) and, in more detail, in Chapter 16 (on an integrated land transport policy).

3.7 Participation

To the cynic, public participation in transport policy (and programme- or project-level work) might be seen as being about keeping your friends close and your enemies even closer. To us, however, providing a role for public participation is no more than providing people with their right to be involved in decisions that affect them and/or about which they are concerned. The various steps in the policy cycle in Figure 3.2 provide opportunities to use public participation to (for example) help define needs, enrich the process of information provision, refine option development, evaluate alternative options, overcome barriers to implementation, and refine the process of selecting a preferred way forward. There are, therefore, considerable functional benefits, in addition to what we see as an important matter of rights to involvement.

The International Association for Public Participation has identified five main levels of participation, with the significance of the role passed to the public increasing as one moves down the list below (IAP2 2012):

- **inform:** providing the public with balanced and objective information to assist them to understand the problem, alternatives, opportunities and/or solutions;
- **consult:** obtaining public feedback on analysis, alternatives and/or decisions;
- **involve:** working directly with the public throughout the process to ensure concerns and aspirations are consistently understood and considered;
- **collaborate:** partnering with the public in each aspect of the decision, including development of alternatives and identification of the preferred solution; and
- **empower:** placing final decision-making power in the hands of the public.

The importance of the public being clear about the role it is being asked to play is emphasised by the Association. Our preference is to go as far as possible in terms of participation as often as possible, within the constraints that are set by the scale of the problem under consideration. In setting land use and transport strategies for a city, for example, the 'collaborate' level seems likely to be a suitable level. In setting design standards for a local public space, the 'empower' level may be appropriate. Information provision is a minimalist approach that is often likely to spur pressure for a higher level of engagement. Choice of the most appropriate level of engagement in any particular circumstance will be ultimately a political decision, often in the face of considerable public pressure to be involved.

3.8 Conclusions

The need for transport (and other sectoral) policy largely derives from failures in free markets always to deliver outcomes that are necessarily in the best interests of a society. Policy interventions therefore imply some notion of what is in the best interests of society, the policy process being crucial in informing relevant decisions. The many complexities in policy analysis and associated decision making outlined in this chapter argue strongly for open and transparent policy processes, with proactive engagement of a wide range of stakeholders, commensurate with the decisions at hand. While we believe that governments should generally deliver effective and efficient transport policy outcomes, transparency and openness are a useful way of helping to ensure this is the case and to keep the blowtorch on effective government.



REFERENCES

- Arrow, K. (1963), *Social Choice and Individual Values*, New York: John Wiley.
- Borck, R. (2007), 'Consumption and social life in cities: evidence from Germany', *Urban Studies*, **44** (11), 2015–2021.
- Button, K. (2010), *Transport Economics*, 3rd edn, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- COAG (Council of Australian Governments) (2009), Communiqué, 7 December.
- Coase, R.H. (1960), 'The problem of social cost', *Journal of Law and Economics*, **3** (October), 1–44.
- Etzioni, A. (1967), 'Mixed scanning: a "third" approach to decision-making', *Public Administration Review*, **27** (5), 385–92.
- Hardin, G. (1968), 'The tragedy of the commons', *Science*, New Series, **162** (3859), 1243–8.
- Hausman, D.M. and M.S. McPherson (2006), *Economic Analysis, Moral Philosophy, and Public Policy*, 2nd edn, Cambridge, UK: Cambridge University Press.
- Hitch, C.J. and R.N. McKean (1967), *The Economics of Defense in the Nuclear Age*, New York: Atheneum Press.
- HM Treasury (2012), *The Green Book: Appraisal and Evaluation in Central Government*, Annex 5: 'Distributional impacts', London: TSO, available at: http://www.hm-treasury.gov.uk/d/green_book_complete.pdf (accessed 10 April 2012).
- IAP2 (2012), *IAP2 Spectrum of Public Participation*, International Association for Public Participation, available at: http://www.iap2.org/association/4748/files/IAP2%20Spectrum_vertical.pdf (accessed 30 April 2012).
- KonSULT (2012), *The KonSULT Knowledgebase*, available at: <http://www.konsult.leeds.ac.uk> (accessed 22 April 2012).
- Lindblom, C. (1959), 'The science of muddling through', *Public Administration Review*, **19**, 79–99.
- Meyer, M.D. and E.J. Miller (2001), *Urban Transportation Planning: A Decision-Oriented Approach*, New York: McGraw-Hill.
- Nash, C., D. Pearce and J. Stanley (1975a), 'Cost-benefit analysis criteria', *Scottish Journal of Political Economy*, **22** (2), 121–34.
- Nash, C., D. Pearce and J. Stanley (1975b), 'Criteria for project evaluation techniques', *Journal of the American Institute of Planners*, **41** (2), 83–9.
- NMTC (2011), *National Transport Plan 2010–2019*, English version, Oslo: Norwegian Ministry of Transport and Communications, available at: http://www.regjeringen.no/upload/SD/Vedlegg/NTP/Binder1ntp_engNY.pdf (accessed 5 May 2012).
- Nussbaum, M. (2001), *Upheavals of Thought: The Intelligence of Emotions*, Cambridge, UK: Cambridge University Press.
- Pearce, D. and C. Nash (1981), *The Social Appraisal of Projects: A Text in Cost-Benefit Analysis*, Ringwood, Hampshire: Macmillan.
- Rawls, J. (1971), *A Theory of Justice*, Cambridge, MA: Harvard University Press.
- Sen, A. (2009), *The Idea of Justice*, London: Penguin Books.
- Simon, H. (1957), *Models of Man: Social and Rational*, New York: John Wiley.
- Stanley, J. and S. Barrett (2010), *Moving People: Solutions for a Growing Australia*, Canberra: Australasian Railways Association, Bus Industry Confederation and International Public Transport Association.
- Varian, H.R. (2006), *Intermediate Microeconomics: A Modern Approach*, 7th edn, New York: Norton.

4

Economic and sustainability foundations



CHAPTER OVERVIEW

Cost–benefit analysis (CBA) is a widely used tool to inform transport policy development (and public policy in other sectors) in many countries, and for project- or programme-level assessment. This chapter provides an overview of the CBA approach. It outlines various technical elements associated with applying the cost–benefit approach in a transport policy setting and, importantly, discusses some of the value judgements involved in application. These value judgements are often neglected in technical explanations of CBA, but they are fundamental to the meaning derived from application. The discussion encompasses issues such as benefit and cost valuation, discount rate selection, distributional implications (including issues of impacts on future generations) and risk and uncertainty. Multi-criteria analysis is also introduced, being sometimes used to complement or replace the CBA approach.

4.1 Scope

Chapter 3 outlined a broad policy cycle (at Figure 3.2) that provides a useful framework for developing and refining transport policies, programmes and projects. ‘Appraisal’ and ‘evaluation’ were two of the steps in that process, the former describing an *ex ante* (before the event) assessment of one or more initiatives and the latter an *ex post* (after the event) assessment. In the current chapter we describe both as ‘evaluation’. This chapter seeks to provide an understanding of the most common form of policy evaluation, economic evaluation in the form of cost–benefit analysis, suggesting some ways this can incorporate a sustainability perspective. The presentation is necessarily summary, but references are included to facilitate further investigation of concepts and approaches by interested readers. The chapter builds on much of the material in Chapter 3. Some reference is also made to other evaluation techniques, such as cost-effectiveness analysis and multi-criteria analysis.

The importance of the idea of *sustainability*, of seeking to meet ‘the needs of the present without compromising the ability of future generations to meet their own needs’, as the Brundtland Commission (WCED 1987, p.8) expressed it, is picked up in the triple-bottom-line (economic, social and environmental) focus to goal setting discussed in Chapter 3. The sustainability focus on ‘future generations’ draws attention to treatment of intergenerational considerations in CBA, or in evaluation more broadly. The idea of sustainability also raises issues of how distributional and other social considerations might be incorporated in CBA as it affects the current generation and issues of how environmental considerations can be taken into account.

4.2 Economic evaluation using cost–benefit analysis

Economic evaluation tools, particularly CBA, are usually well regarded by governments to help guide decision making, to the point where they are often a formal requirement of part of what is unfortunately termed a ‘business case’ for government support. This language is unfortunate, since it may be read by some as implying a concentration on implications for business, as distinct from the real intent, which is that it refers to implications for the business of government.

As pointed out in Chapter 3, the fundamental *value judgement* behind CBA is that ‘individual preferences should count’. In general terms, CBA reflects this value judgement by seeking strength of preference measures of policy or project impacts using willingness-to-pay (WTP) measures of benefit, and comparable cost indicators, to suggest whether a particular course of action (e.g. policy or project) is likely to result in net economic benefits over time. Because of the narrow scope of the concept of a *Pareto improvement*, as outlined in Chapter 3, CBA is largely founded on the Hicks–Kaldor hypothetical compensation test (Hicks, 1939; Kaldor, 1939). That test suggests that, provided the gainers from some particular course of action or policy *could* compensate the losers, then the change could be regarded as improving economic efficiency and *potentially* adding to economic welfare. Proposals that pass this test are regarded as providing a *potential Pareto improvement*, ‘potential’ referring to the lack of actual compensation to losers. The fairness implied by this criterion has been criticised by many commentators.

CBA can be seen as a form of applied utilitarianism (see Chapter 3), with money being the measuring rod used to add up impacts on utility levels of different people and willingness-to-pay measures providing the basis for monetary benefit estimation (a proxy for utility gains) and, conversely, willingness-to-accept (WTA) compensation measures for estimation of

costs (utility losses). This approach implies that it is possible to obtain some approximation of how the marginal utility of income or consumption changes as between different groups (as embedded, for example, in the α weights in the weighted sum-of-utilities welfare function from Chapter 3), such that money measures are an indication of changes in utility or welfare.

The market-based approach reflected in CBA is recognised as being helpful in choosing between alternative possible courses of action on economic efficiency grounds but as often confronting difficulties in three key areas:

1. A frequent lack of direct markets in which to value particular costs and benefits in monetary terms (the lack of such markets was a key rationale for the development of CBA).
2. The distributional problems of valuing benefits and costs to people whose circumstances are different (for example, their income levels may vary considerably, they may have different physical capacities and or they may live at different points of time). The general failure to pay compensation to losers from a course of action means that value judgements about the desirability of the relevant distributional consequences are needed to draw conclusions about the overall merit of the action(s) under consideration.
3. The validity of individual preferences as a guide to values.

We deal with the first two of these concerns here and the third in section 4.3.

Various analytical approaches have been developed to impute monetary values to costs and benefits that are not directly traded through markets, to tackle the first problem. Relevant examples in the transport field include valuation of time savings, air pollution damage, carbon emissions and accidents. Stanley et al. (2011, 2012) have developed a way of placing money values on part of the social dimension of transport – the value of reducing risk of social exclusion associated with transport disadvantage. Some examples of imputing monetary values to impacts of transport policies or projects are explored further in other chapters.

Chapter 3 identified triple-bottom-line (economic, social and environmental) goals as the ultimate outcomes from transport policy interventions, with additional goals sometimes being included (for example, health, safety and regional development, together with procedural goals). By implication, in an evaluation framework based on individual preferences, the implication is that these are matters that are highly valued by people in the community.

Economic evaluation is about those outcome consequences of policies or projects that can be measured in *money terms*. This includes the monetary evaluation of some effects that might primarily be classified as social or environmental in terms of the impact pathway. For example, air pollution caused by transport has environmental and health consequences, but some of the impacts of air pollution can be priced in money terms, giving them an economic dimension. Some social and environmental impacts, however, are not amenable to assessment in monetary terms. That does not make them irrelevant to policy evaluation. So long as some individual cares about these impacts, they are within scope in terms of the individual preferences value judgement. It means that another method of assessment is needed. In our view, it is still appropriate to apply the term 'CBA' to such wider assessment, since the purpose is still systematically to identify and evaluate all the costs and benefits of the various policy options under consideration, even if money is not the common metric for all. We illustrate how this might be approached later in this section. Environmental impact assessment techniques can also be used, and may be a legislative requirement in some jurisdictions, to deal with some environmental impacts, additional to any evaluation through CBA.

With respect to distributional implications, various benefit/cost weighting techniques have been used to allow for the likelihood that the value (marginal utility) of a dollar will differ between different groups of people (for example, whose income levels differ), mirroring the weighted sum-of-utilities approach outlined in Chapter 3. Incorporating such distributional weights is, in effect, an attempt to specify a particular social welfare function that makes explicit value judgements about the relative worth of gains or losses to different groups. The UK *Green Book* on project evaluation sets out a range of such approaches for transport evaluation applications (HM Treasury 2012). Some examples of such weighting approaches are outlined below, in the subsection 'Distributional weights'.

Intergenerational distributional considerations are also important in CBA and are fundamental to the sustainability definition. The selection of the discount rate for use in CBA and approaches to treatment of the environment, discussed in section 4.6, are important in dealing with intergenerational considerations.

The basic CBA formula

CBA seeks to quantify, in money terms, the costs and benefits of particular courses of action over time. Eigenraam et al. (2000) note that a CBA describes:

- a comprehensive assessment of different effects;
- the distribution of costs and benefits;
- project alternatives; and
- risk and uncertainties connected with the project.

The basic CBA formula is shown in equation 4.1:

$$\text{Net present value of benefits (NPV)} = \sum (B_t - C_t) / (1 + r)^t \quad (4.1)$$

where:

B_t = benefits in year t in money terms

C_t = costs in year t in money terms

r = the discount factor (to bring current and future values of benefits and costs into a present value equivalence)

t = the time period of analysis (frequently the maximum effective economic life of any initiative being assessed).

With this formula, the presumption is that a positive NPV indicates the project has passed a test of acceptance.

Evaluation results may sometimes be expressed as a benefit cost ratio (BCR), in which the denominator includes the direct costs attributed to the initiatives under assessment (capital costs and operating costs; not negative benefits, which are part of the numerator). Benefits are in the numerator. Any welfare losses expected from the initiatives are usually treated as negative benefits in the numerator.

When policies or projects of different size are being compared (different cost levels), the NPV formula is not good at ranking alternatives. Several complementary small projects undertaken together, for example, may generate a greater NPV than one large project, even though that large project has a bigger NPV than any of the smaller projects taken in isolation. A better way of ranking alternatives is by their BCRs, which show the ratio of their respective gross benefits to gross costs. Pearce and Nash (1981) show that ranking by BCR is not infallible in a constrained budget situation.

A hypothetical road project CBA in a rural area might look something like Table 4.1, in summary.

All options have benefits greater than costs, which means positive NPVs and, consequently, BCRs that exceed 1. Project selection based on choosing the option with the highest NPV would result in option 4 being chosen

Table 4.1 Hypothetical rural road improvement analysis

Option	Present value of costs (\$m)	Present value of benefits (\$m)	Net present values (NPV)	Benefit/cost ratio (BCR)
1	10	18	8	1.8
2	25	35	10	1.4
3	5	10	5	2.0
4	40	60	20	1.5

(NPV = \$20 million). This is also the most expensive option by a considerable margin. Project selection based on maximising benefits relative to costs would lead to the cheapest option, option 3, being chosen. It has a BCR of 2, but costs \$35 million less than option 4. In short, selecting option 3 frees up \$35 million that could possibly be used for some other productive purpose, perhaps in the transport policy/project area.

Distributional weights

Pearce and Nash (1981) explain that the use of this CBA formula implies two value judgements:

1. Individual preferences should count.
2. These preferences should be weighted by some 'intensity factor' that is correlated with the individual's income.

There is a long-standing debate in economics about whether utility or well-being is measurable between individuals, as is required by the utilitarian method to form a view about relative desirability of alternative options. The growing use of subjective well-being measures has rekindled interest in this question (see, for example, van Praag and Ferrer-i-Carbonell 2004). If the aim is to identify impacts on social welfare, as embedded in a measure of utility, then money is no more than a proxy, confronted with the problem that an additional dollar is likely to improve the well-being of a poor individual more than a rich one – reflecting diminishing marginal utility of income.

Distributional weighting schemes have been developed to tackle this problem, usually by estimating how the value of additional consumption falls as the total level of consumption increases (Pearce and Nash 1981). HM Treasury (2012) sets out a methodology for deriving possible weights, based on quintiles of equivalised net (or gross) income (which allow for household

composition) and empirical evidence that the utility function has the form shown in equation 4.2:

$$U = \log C \quad (4.2)$$

where:

U = utility

C = consumption.

In this case, the marginal utility of consumption is $1/C$, which implies that marginal utility is halved as income doubles. These assumptions led the UK Treasury to propose use of the distributional weights set out in Table 4.2.

Table 4.2 Illustrative UK distributional weights

Quintile	Range (equivalised net income)	Range (equivalised gross income)
Bottom	1.9 – 2.0	2.2 – 2.3
2nd	1.3 – 1.4	1.4 – 1.5
3rd	0.9 – 1.0	1.0 – 1.1
4th	0.7 – 0.8	0.7 – 0.8
Top	0.4 – 0.5	0.4 – 0.5

Source: HM Treasury (2012, p. 94).

These distributional weights are significant because they imply that a dollar of benefit to a household in the top quintile is worth only about one-fifth as much as a dollar of benefit to a household in the bottom quintile. Evans (2005) argues for a higher value for the elasticity of the marginal utility of consumption, of about 1.4, which would widen the range of distributional weights. Application of such weights results in a monetary version of what was called a weighted sum-of-utilities welfare function in Chapter 3.

Applying distributional weights requires a detailed analysis of the incidence of benefits and costs of a project among the various groups that are identified. In some cases, this requires consideration of how impacts might pass through the economic system. For example, part or all of the travel-time savings from a major road project may ultimately end up as benefits to landowners in the form of higher land prices and higher rental prices to tenants. Distributional analysis should consider such flow-through impacts. Groups are usually distinguished on an income basis, but other criteria may also be used, such as personal capacities, although many such alternatives are likely to be correlated with income. The UK Treasury is firmly behind distributional weighting,

indicating that, ‘Where appraisers decide not to adjust explicitly for distributional impacts, they must provide a justification for this decision’ (HM Treasury 2012, p.25). Other jurisdictions are less hard-line, although the importance of identifying who wins and who loses is a common requirement in CBA, which assists the application of distributional weights if desired.

A simpler approach to dealing with distributional impacts, an approach that has long standing in transport policy or project evaluation, is to value benefits and costs as if all affected people had the same income level. Thus, for example, a standard value for non-working-time savings is commonly used in project evaluations, irrespective of the income level of beneficiaries or losers. Nash et al. (1975) explain the basis for this approach.

Valuing benefits in money terms

Consumers’ surplus

The essence of CBA is the process of valuing costs and benefits. In line with the individual preferences value judgement, anything that someone values that is affected by the policy problem and potential solutions under consideration is *prima facie* within scope. National boundaries might be put on this calculation procedure in some cases (e.g. in an evaluation from the perspective of a particular country), but wider regional boundaries may be relevant to particular issues, including where these wider regional impacts feed back to the welfare of people in particular countries.

Benefits are usually measured by willingness to pay (WTP) for a gain and costs by willingness to accept compensation (WTA) for a loss. Economic theory indicates that there are a number of potential measures that might be used to value benefits and losses, and interested readers are referred to Pearce and Nash (1981) and Button (2010) for a discussion of these different measures. The *compensating variation* measure is widely recognised as the most appropriate benefit measure, since it aligns with the idea of the Hicks–Kaldor hypothetical compensation test (if the gainers could compensate the losers and remain better off in aggregate, then the change is worth undertaking). For a price fall (such as a lower price of car travel following a road improvement), the compensating variation is the maximum amount that consumers who benefit could pay for the benefit and remain as well off as they were before the price fall. In practice, however, there is little difference expected between the various possible measures of benefit (and loss), and consumers’ surplus measures are the most commonly used measures of benefit or loss. Figure 4.1 explains, while Button (2010) presents a more

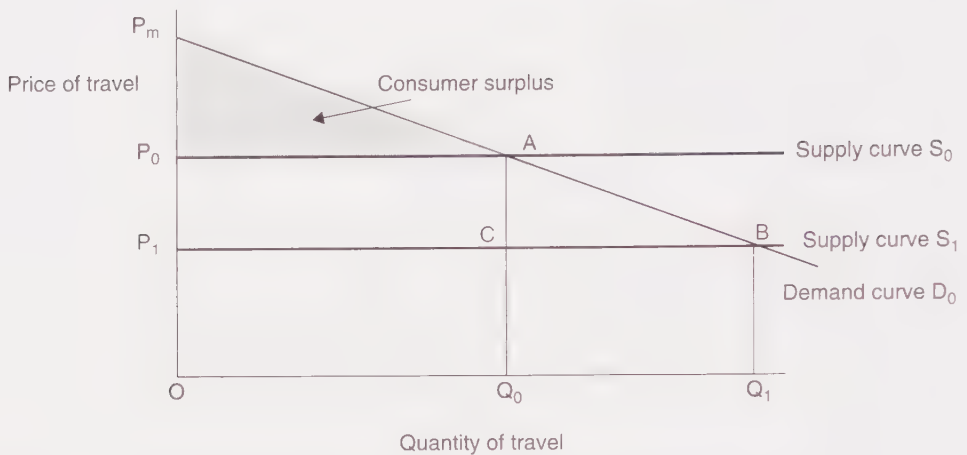


Figure 4.1 Consumers' surplus measurement

detailed exposition of broader transport demand and supply considerations that underpin such measures.

The demand curve for travel in Figure 4.1 shows the amount of travel that people demand for any given price. At any given quantity of travel, the area under the demand curve to that point is a money measure of people's aggregate willingness to pay the relevant quantity of travel. The higher the price, the less travel will usually be demanded, so the demand curve slopes down to the right.

A number of factors other than price also influence the market demand for travel. These factors include (for example) population numbers, income levels, prices of products that compete with or are substitutes for travel, consumer tastes, and suchlike. These other factors are assumed constant when showing how the quantity of travel that is demanded varies with its price. Should any of these other factors change, the demand curve shifts. For example, an increase in population is expected to move the whole demand curve for travel to the right.

The relative responsiveness of travel to a change in price is measured by the price elasticity of demand, which is equal to the percentage change in quantity demanded divided by the percentage change in price. Similar elasticity measures apply (for example) to the responsiveness of demand to changes in income (called the income elasticity of demand) and to changes in prices of competing and complementary goods (called cross-price elasticities of demand).

The supply curve for travel in Figure 4.1 shows the amount of travel that is supplied at various prices. It is shown as horizontal for simplicity. This assumes no congestion impact as travel volumes increase, an assumption we relax shortly. Congestion leads to an upward-sloping supply curve.

The initial demand and supply curves for travel in Figure 4.1 lead to an equilibrium quantity of OQ_0 trips at price OP_0 per trip. Travellers would be prepared to pay OP_mAQ_0 to undertake OQ_0 trips. They only have to pay OP_0AQ_0 , leaving them with a consumers' surplus equal to the triangle P_mAP_0 . This is a money measure of their net benefit from undertaking OQ_0 trips. It equals the difference between what they are willing to pay (WTP) and what they have to pay.

At this point we divert slightly to introduce an important pricing principle. Chapter 3 introduced the idea of an efficient allocation of resources being one where it is not possible to make someone better off without someone else being made worse off, sometimes called a *Pareto optimum*. The intersection of the demand and supply curves in Figure 4.1 represents such a position. At travel quantities less than OQ_0 , the demand curve is above the supply curve, showing that the additional benefit received by travellers from increasing trips (reflected in the demand curve) is greater than the cost of undertaking those trips (as represented by the supply curve). Beyond OQ_0 the additional cost exceeds the additional benefit, so travel should be reduced to OQ_0 to maximise net benefits. This point occurs where marginal private benefits (MPB), given by the demand curve, equal marginal social costs (MSC), given by the supply curve. In the example shown in Figure 4.1 we are assuming that no external costs (such as congestion costs) exist. If there were external costs, then these would need to be added to the costs shown in the supply curve (converting it to a marginal social cost curve) to derive the efficient output level.

Benefit maximisation (and an efficient output level) requires that equation 4.3 holds true:

$$MPB = MSC \quad (4.3)$$

where:

MPB = marginal private benefits

MSC = marginal social costs.

This is an important concept for transport policy, particularly as it relates to pricing.

Returning to Figure 4.1, assume that a transport improvement increases travel speeds and lowers costs, which causes a movement downwards in the supply curve to S_1 , with the equilibrium price of travel falling to OP_1 , generating additional travel. The number of trips increases by Q_0Q_1 . Consumers' surplus increases to P_mBP_1 , an increase of P_0ABP_1 , and this increase is the direct user benefit attributable to the relevant transport improvement, for a particular time period. This benefit needs to be calculated each year over the expected life of the improvement.

It should be noted that the increase in consumers' surplus comprises two elements:

- P_0ACP_1 = the benefit to the traffic that used the road both before and after the improvement; and
- ABC = the benefit to traffic that was *generated* by the road improvement.

On the assumption of a straight-line demand curve, generated traffic benefits per unit of travel (for example, a trip) are half those achieved by users who were on the road both before and after the improvement.

Total user benefits (TUB) can be expressed as equation 4.4:

$$\text{Total user benefits} = \frac{1}{2}(OQ_0 + OQ_1).(P_0 - P_1) \quad (4.4)$$

Pearce and Nash (1981) point out that this formula can be applied at both route and network levels and across modes, where many prices may change (for example) following a major road or other transport network improvement. Thus, for example, a major road improvement may generate additional traffic on the improved route and on feeder routes and may lead to some people switching from public transport to road. The TUB equation can be used across all routes and modes to estimate network-wide user benefits. Distributional weights may then be applied to benefits estimated for particular groups, with specification of the basis for such weighting being an important input to the evaluation, recognising the value basis for such choice.

Producers' surplus

In a competitive marketplace, economic theory suggests that the amount producers are prepared to offer for sale at any price will be determined by their marginal costs of production. Marginal cost is the change in total cost for a one-unit change in output. In the short run, this mainly means labour

and materials costs. In the long run, capital equipment costs may also be involved. In a transport setting, for example, increasing the supply of kilometres of bus route service in the short run will probably mean paying for some extra drivers and mechanics to work overtime and some additional fuel and parts costs. In the longer term, it may require additional vehicles and perhaps even larger depots. Figure 4.2 illustrates this point, introducing an upwards-sloping supply curve to reflect increasing short-run marginal costs of increasing supply.

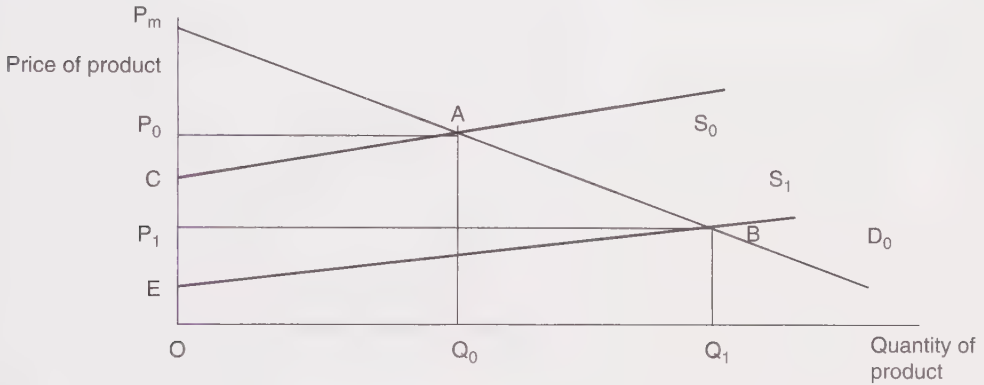


Figure 4.2 Consumers' and producers' surplus measurement

The demand curve (D_0) and initial supply curve (S_0) produce an equilibrium quantity of OQ_0 product sold at price OP_0 . An improvement causes the supply curve to fall to S_1 , with marginal costs being lower for all amounts of product offered for sale. Consumers' surplus is initially P_mAP_0 . It increases to P_mBP_1 , the value of the difference being the increase in consumers' surplus, or benefit, from the initiative that led to the supply curve falling. Producers' surplus is initially P_0AC (the difference between revenue and marginal costs), increasing to P_1BE after the improvement, the difference being a measure of the gain in producers' surplus (benefit to producers). Total surpluses are P_mAC before the change and $P_mB E$ after, the difference being divisible into the gains in consumers' and producers' surpluses.

Pearce and Nash (1981) make the point that this supply analysis does not work in a situation of monopoly, where the supply curve does not represent marginal supply costs. Nash (personal communication) adds that the area above a competitive supply curve cannot be trusted as a measure of profit and economic rent. Essentially the reason is that in perfect competition there are zero (supernormal) profits, so that changes in the area designated as producers' surplus will not necessarily measure a change in profits. Analysts need

to look behind the reasons for the shift in the supply curve to see if it will lead to a change in some other economic rent – for example, it may be due to cutting workers' wages, as happened with UK bus deregulation in the 1980s. Pearce and Nash (1981) emphasise that changes in profit and economic rent (to workers as well as owners of property) should all be measured. This is likely to be most important for very large projects, where computable general equilibrium (CGE) modelling techniques are helpful. Chapter 11 introduces CGE modelling, but the bus example indicates that it may also be an issue in other circumstances.

Road travel supply curve

In a road traffic setting, the travel supply curve has an unusual property. For a given road, it shows the generalised costs (which consist primarily of time and money costs) incurred by road users along that road at any given traffic level, reflecting the underlying speed/flow curve for the route of interest (see Chapter 5). At low traffic volumes additional vehicles do not affect the speed (and associated costs) of others. However, as traffic builds towards a route's practical capacity, congestion starts to slow traffic speeds and both time and fuel costs usually rise. The supply curve for road traffic reflects the *average private costs* (APC) of using the route, not the marginal costs. Because traffic congestion causes speeds to slow and fuel costs to rise, all users' average private costs increase. For this to happen, the costs of an additional vehicle in the total traffic stream, over and above the cost incurred by that additional traveller, must be higher than the average cost for all travellers (otherwise the average would not rise). This additional cost is the *marginal social cost* to the traffic stream.

Because individual motorists only take account of their own travel costs when choosing whether or not to use a route, traffic level is determined at the point where demand (or marginal private benefit) equals average private cost ($MPB = APC$). This is not an efficient traffic level. Because marginal social cost exceeds average private cost, we have $MSC > MPB$. The area between the marginal social cost curve and the demand curve is often called the 'dead-weight loss' from congestion. A reduction in traffic to the point at which marginal private benefit equals marginal social cost ($MPB = MSC$) is required for an efficient outcome. This is the basis for the argument in favour of congestion pricing.

Figure 4.3 illustrates this situation, with the upwards-sloping supply curve (S_0) that reflects increasing congestion as traffic volumes increase. The market-determined equilibrium is at travel volume OQ_0 , where $MSC >$

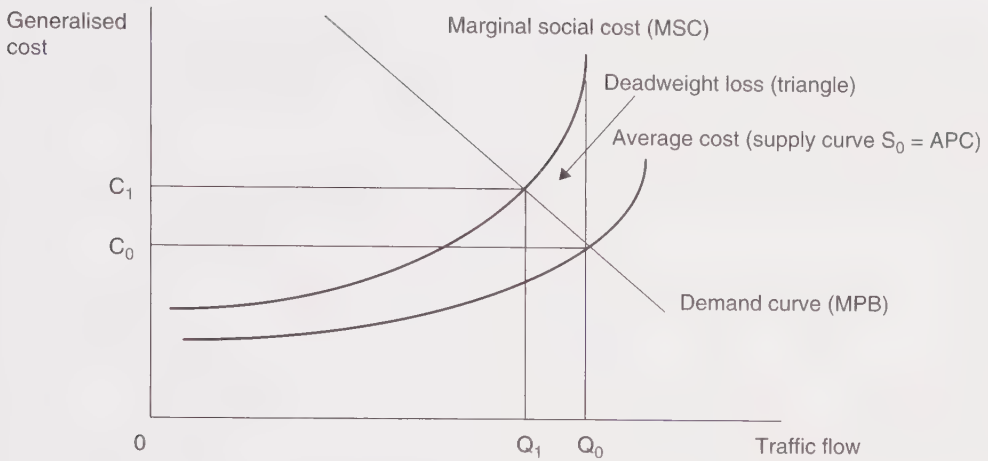


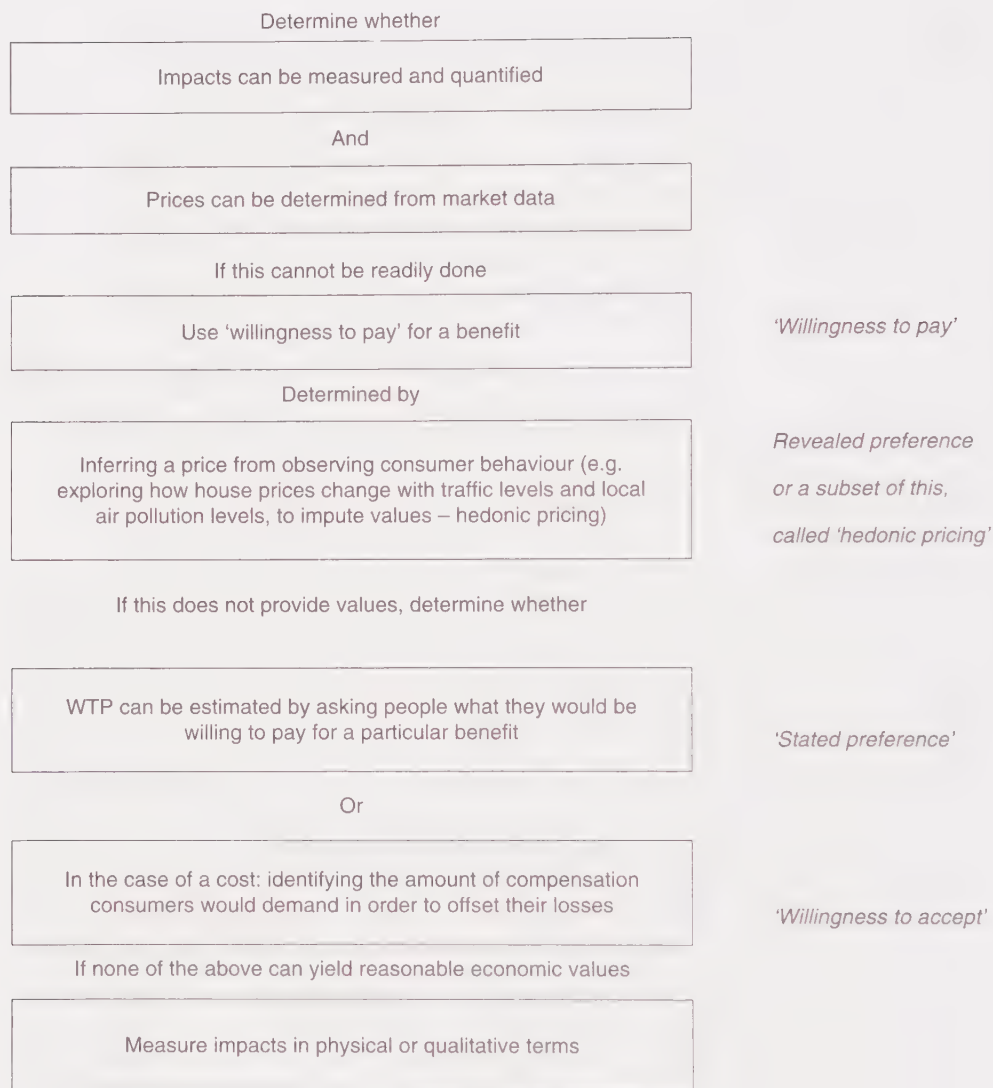
Figure 4.3 Marginal costs along a route

MPB. Only at the lower traffic volume OQ_1 does $MPB = MSC$, which is the economically optimum traffic flow. Chapter 12 discusses road pricing reform to achieve such a reduction in traffic, extending the analysis of Figure 4.3.

Approaches to monetary valuation

The preceding discussion has implied the existence of a marketplace, with observable demand and supply curves, which facilitate the process of benefit valuation. One of the rationales for cost–benefit analysis is the occurrence of market failure, frequently reflected in the *lack* of direct markets through which people can express preferences or values. In the absence of direct markets, analysts frequently seek ways to impute monetary values to particular possible consequences of transport policies, programmes and/or projects, as part of the economic evaluation process. HM Treasury (2012) sets out a neat description of the process an analyst might use, as shown in Figure 4.4, to which we have added another (final) step.

The main ways in which analysts seek to impute monetary values is through use of *revealed preferences* or *stated preferences*. Revealed preference techniques examine choice behaviour in related contexts to assess whether these imply anything about how people might value particular attributes. For example, analysis of the determinants of house prices (hedonic pricing) may shed light on how people value noise costs. Dube et al. (2011), for example, show how a bus rapid transit service in Quebec City (Canada) increased the value of single-family residential homes by \$35 million and municipal revenues by more than \$6 million, both in present value terms. They note that homes



Source: Based on HM Treasury (2012, p. 23)

Figure 4.4 Valuation techniques

closest to the route did not gain in value, reflecting local external costs such as noise, and that gains were limited to properties within walking distance of the service.

Travel mode choices are regularly used to impute various values of time. Where such revealed preference data are lacking, survey techniques may be

used to explore preferences, with stated preference techniques being increasingly applied. The inclusion of a price component among choice sets that are explored through stated preference techniques opens the possibility of putting money values on some specific variables, which may have significance for transport policy decisions. For example, stated preference techniques could be used to assess the value people place on comfort in a mode choice context.

Non-monetary valuation

If valuation in money terms is not possible or not worth the cost in the context of the decisions under consideration, then the last stage in Figure 4.4 proposes use of physical measures of impacts and a description of those impacts that are inherently qualitative, commensurate with the significance of the issue under consideration.

Physical quantification of impacts will be in units of measurement that are most applicable to the matter being considered, a step that is also often required when monetary values are to be assessed (using revealed or stated preference methods). For example, the monetary valuation of air pollution impacts of a transport policy requires an estimate of the impact of the policy on physical pollutant concentrations, such as parts per million of particulate matter, allowing for possible dispersion, and an assessment of the numbers of affected people – using techniques such as the ‘impact pathway’ (Maibach et al. 2007) – and then the assignment of monetary values to these specific impacts. Evaluation of impacts in monetary terms is thus often associated with, and accompanied by, an indication of relevant physical impacts that have been valued, particularly given uncertainties surrounding some unit benefit values. As noted previously, such quantification may also be required by environmental impact legislation in some jurisdictions.

Where neither monetary nor physical impact measurement is possible, such as for describing the impact of a major new road or railway on the aesthetics of a pleasant rural setting, a qualitative description of the before and after situation can often be undertaken and stated preference (or other) approaches may perhaps be used to imply ratings of alternatives, which may be able to be monetised if a money (price) variable is included in the relevant choice sets. At a lower cost level, people can be invited to comment on what they think about particular impacts, as part of a participation programme.

The implication is that a comprehensive transport policy evaluation will produce:

- some *monetary measures* of benefit and cost;
- some *physical measures* of benefit and cost, a subset of which may have been taken through a monetisation process (using revealed or stated preference techniques); and
- some *qualitative measures* (which may include ranking) of benefit and cost (or impact), a subset again of which may have been taken through a monetisation process (stated preference).

The range of such policy or project consequences and difficulties of placing monetary values on many policy or project impacts lead many analysts and jurisdictions to favour cost-effectiveness analysis (CEA) over CBA. CEA typically takes the policy or project purpose as given and looks for the most cost-effective way to achieve that outcome, measuring impacts in the most direct units (e.g. changes in pollutants rather than willingness-to-pay measures). Emphasis is placed on discussion and negotiation about the trade-off values implicit in choosing one alternative over another and, if the policy or project purpose is not given, in discussing that as well, relative to impacts and costs.

Techniques such as multi-criteria analysis can be used to try to rank policies or projects across all three types of assessments listed above. In line with the individual preferences value judgement, we would expect to see such assessment done by various affected or interested stakeholders, before the matter is decided by the relevant decision maker(s).

Table 4.3 provides a simple illustration of how this might be done, for an evaluation where there are two possible options (a larger number is easily accommodated). It assumes each participant in the ranking process is given 100 points to allocate across the two options (reflecting a view about fairness) and that each option has been evaluated across the three types of impact areas indicated: monetary; physical (being careful to identify any parts of the physical evaluation that have also been included within the monetary evaluation, to avoid risks of double counting); and qualitative (with the same precaution). Each person ranking the alternatives would allocate their budget of 100 points across the two alternatives in terms of how well each meets that person's assessment of each impact area's significance. Each of the three impact rows could include a number of specific itemised impacts, which would be needed to enable an informed assessment by those doing the rating. Participants might be selected to be representative of groups who are expected to be affected by the policy under examination, or a more open process could be used (even to the point of inviting online assessment, but recognising that this excludes certain types of people who do not have

access to or interest in such processes, even though they may be significantly affected by the policies in question). At the end of the day, it is about providing a rich source of information to a political decision maker, to assist an informed opinion. Other multi-criteria approaches could also be used.

Table 4.3 A way of multi-criteria ranking

Impact area	Option 1	Option 2	Points allocation to option 1	Points allocation to option 2
CBA monetary impacts	Gross PV of benefits Gross PV of costs	Gross PV of benefits Gross PV of costs		
Physical, non-monetary impacts	Physical descriptions	Physical descriptions		
Qualitative impacts	Qualitative description	Qualitative description		
Overall points allocations			Total across the two must equal 100	

Many advocates of multi-criteria evaluation techniques seek to use the decision maker's preferences to weight relative impacts, with such weightings often being determined *before* impact assessment is undertaken. Prior weighting of types of impact implies constant rates of trade-off between types of impact, irrespective of impact scales, which is a particularly restricting assumption and unacceptable in our view. Our preference is to have trade-offs determined by people through the assessment process, in the first instance as part of a comprehensive public participation process, in line with the individual preferences value judgement, and then by the decision maker(s), who is likely, of course, to be engaged through the process but will make final judgements on trade-offs at the end of the process, with the benefit of full information on impacts and how people might value such impacts.

4.3 Preferences and a wider context for money and other values

Nash et al. (1975) explore the value basis for use of the individual preferences value judgement, as it might be applied in cost–benefit analysis, and ask:

- **Whose** preferences should count? For example, how should one deal with future generations and children? Should criminals' preferences be included? Should only the preferences of those directly affected by a policy or project be included? How might the interests of non-humans be reflected?
- **Which** preferences should count? Harsanyi (1955) distinguishes between 'ethical preferences' and 'subjective preferences', the former being preferences expressed on the basis of impersonal social considerations (perhaps derived through Rawls's 'veil of ignorance') and the latter being what the individual personally prefers. We would argue that the former should be used for transport policy. Chapter 3 pointed out, for example, that provision of *merit goods* is typically based on such considerations. Should anti-social preferences count? If the preferences of oppressed people derive from their oppression, how can these preferences reflect their welfare (Hausman and McPherson 2006)?
- **When** should individual preferences count? If individuals lack the knowledge and information to evaluate particular matters, should their preferences be used to guide decision taking? Our view is that every effort should be made, commensurate with the scale of the matter under consideration, to ensure that people have the best information available on which to express preferences.
- **How** should individual preferences be aggregated (as discussed previously in this chapter and in Chapter 3)?

These questions suggest that the extent to which individual preferences should be taken as a guide to values (positive or negative) in policy evaluation requires careful thought if the purpose is to draw conclusions about whether society is in some way better off if a particular policy direction is taken. We agree with the view expressed by Hausman and McPherson that 'It is more plausible to maintain that well-being is the satisfaction of suitably "laundered" self-preferences than to maintain that it is the satisfaction of actual preferences' (Hausman and McPherson 2006, p. 128).

Pearce and Turner (1990) explore issues of values based on private *individual preferences* within an environmental context, particularly in relation to questions of *public (or social) preference values* and *functional physical ecosystem values*.

Following Harsanyi (1955), any one person can be expected to have a number of values, and such values are likely to include *social or public preferences*. These are preferences about what the individual believes should be the case, rather than necessarily what the individual desires or wants. They

may reflect social norms and may sometimes be reflected in legislation or regulatory standards. The discussion in Chapter 3 about merit goods is relevant in this context, as a reflection of public or social preferences. A 'standards' approach is often very important in the transport policy response to such matters. For example, standards exist in areas such as driver licensing, driving-hours regulation, vehicle emissions, vehicle structural requirements, and so on. It is always open to a policy analysis to identify situations where such public or social preferences may be considered relevant to an evaluation at hand and argue the case for particular options on this basis, provided this is transparent and debatable by interested parties.

The other environmental value context discussed by Pearce and Turner (1990) relates to values associated with physical processes and systems. This raises important questions. For example, in terms of the sustainability focus on the intergenerational distribution of well-being, what obligations does the current generation have, if any, to future generations? This brings in fundamental issues such as whether there are ecological limits within which economic activities should be undertaken. Turner (1991), for example, has argued that the passing on of a stock of undiminished natural capital would help fulfil intergenerational equity and sustainability objectives, an idea he argues could be seen as consistent with the Rawlsian ethic, applied in an intergenerational context. It might be seen as 'justice as opportunity', with the stock of natural capital providing that source of opportunity. Such an approach could involve policy making, and associated CBA, including environmental compensating proposals, if a particular policy initiative reduces the stock of natural capital. Some jurisdictions have implemented environmental legislation that might be seen as a way of giving currency to such a concept.

The focus on physical processes and systems also poses the important question of whether nature has *intrinsic value*, unrelated to human valuation processes or human enjoyment (i.e. the values are inherent in the assets in question rather than being values ascribed by humans) and what value should be placed (for example) on species preservation. Unless there are specific legislative obligations that prescribe how to treat such high-level issues in particular policy contexts, a way to deal with the issue where potential impacts may be significant is to describe relevant impacts as fully as possible and include the matter within the public participation process, so that the political process can make a choice about importance informed by available data and opinions (environmental impact legislation may also be relevant). The idea of seeking to conserve a stock of natural capital is consistent with the idea of intrinsic value, since the protection of habitats, water quality and so on assists the process of species diversity and survival. Relevant biodiversity

protection legislation is also supportive and provides a constraint that needs to be reflected in policy or project evaluations, where it exists.

In terms of the implications of such issues for transport policy, the importance of the policy analyst considering a suite of possible values is emphasised. The individual preferences value judgement is fundamental to policy evaluation but has many limitations and should not be exclusive of other values, such as public or social preferences and ecological values.

4.4 Total economic value

The preceding material has distinguished *user values*, expressed through individual preferences, from *intrinsic values*. Pearce and Turner (1990) point out, in an environmental setting, that people may be prepared to pay something to (for example) preserve endangered species, unrelated to any concept of use value (an example of a *non-use value*). This value is sometimes called *existence value*, which can be seen as an economic component of *intrinsic value* but not the whole, since part of the essence of intrinsic value is that it need not have anything to do with what humans value.

Existence value may reflect a person's wish to ensure a supply of natural environments to future generations (altruism) or sympathy towards other creatures. It may be about rights. Pearce and Turner (1990) note that existence values are fuzzy but are related to conditions of irreversibility, uncertainty and uniqueness. Valuation is inherently difficult, but the concept serves to promote a cautious approach to development and attention to the basis of what is regarded as in scope in policy analysis, the justification for such scoping, and consideration of how included matters are to be treated. Existence values may be relevant to transport policy or project evaluation at a big-picture scale, such as relating to biodiversity losses associated with climate change, in which transport is a significant contributor. It may also be relevant at a smaller scale, such as where a major new road or railway line may destroy a scarce habitat that is home to a threatened species.

Economic non-use values more generally may also be relevant to transport policy. For example, people may value the continued existence of a good or service, such as a transport service for disadvantaged people, even though there is no prospect of personal use of that service. This value may derive, for example, from altruistic motives. Department for Transport (2007) points out that there is a risk of double counting if non-use values are included in transport appraisals or evaluations, and that only non-use values that derive from altruistic motives do not involve such double counting.

Pearce and Turner (1990) also point to *option value* as a potential part of economic value. Supply uncertainties and risk aversion may mean that someone is willing to pay more than their expected consumer's surplus to preserve an option for possible future use. Department for Transport (2007) defines option value, in a transport context, as follows: 'An option value is the willingness-to-pay to preserve the option of using a transport service for trips not yet anticipated or currently undertaken by other modes, over and above the expected value of any such future use' (Department for Transport 2007, n.p.).

Option value is most relevant to transport evaluation in a situation where there may be a substantial change in service availability as a consequence of the issue(s) under examination.

Total economic value can thus be defined as *actual use value* + *option value* + *non-use value*.

All three elements may be relevant to the *economic assessment* of a transport policy or major project, particularly if that policy or project is comprehensive and significant in scale, such as introducing a new service or taking away an existing service or involving major quantum changes in service availability. Quantification is difficult, and there are not many examples of application. Department for Transport (2007) discusses suitable evaluation methodologies and provides some examples, indicating their limited applicability (because of a lack of suitable case studies that might be generalised).

4.5 Cost concepts

Policy evaluation using tools such as CBA requires a detailed understanding of the costs of the alternative instruments under examination and how they might impact on costs elsewhere in the system. A basic understanding of a few cost concepts is important in this regard. Relevant concepts include the following:

- **Fixed costs** = costs that do not change with respect to the level of output of a good, service or activity in a specified time period. Major infrastructure costs are usually fixed for a relatively long period, the assets in question being immobile. Costs of mobile assets (e.g. buses, trucks) tend to be fixed over relatively shorter time periods and are able to be varied in the longer term.
- **Average fixed costs** = fixed costs divided by output level. Because fixed costs are given in the short run, average fixed costs decline continually as output increases.

- **Variable costs** = costs that change with the output level of a good, service or activity in a specified time period.
- **Average variable costs** = total variable costs divided by output level.
- **Total costs** = fixed costs plus variable costs.
- **Average total costs** = total costs divided by output level.
- **Marginal costs** = the change in total costs for a unit change in output in a time period, which equals the change in total variable costs in that period, because fixed costs do not change.
- **Avoidable costs** = costs which can be avoided if a particular course of action is followed (another definition of marginal costs).
- **Attributable costs** = costs that are attributed to a particular activity or output level, that activity being a contributor to the costs in question but not the only such contributor (e.g. policing costs and their connection to truck traffic). These costs would not all disappear if the activity ceased. The term 'common costs' is sometimes used to describe these costs.
- **Joint costs** = costs of producing two or more outputs which are necessarily linked (e.g. the costs of a truck's inward and outward journeys, which can be thought of as joint products, the production of one journey leg usually requiring the production of the other).
- **External costs** = third party costs (losses) that arise as an incidental outcome of a particular activity and which are not priced (compensated) as part of that activity. The existence of external costs is an indicator of market failure.
- **Marginal external costs** = the change in total external costs associated with a unit change in output of a particular good, service or activity.

Policy, programme and project evaluation is primarily about *externalities* and *changes*. Externalities are a result of market failure, which may attract a policy intervention. *External costs* and *marginal external (or social) costs* are thus a critical concept for policy, including pricing policy, because efficient pricing generally requires prices to be set at marginal social cost. *Attributable costs* are also frequently important in pricing policy, where cost recovery objectives might lead to the inclusion of such costs within the scope of costs to be recovered, as discussed in Chapter 12.

4.6 Discounting

Within the basic CBA formula, the discount rate plays an important role in enabling benefits and costs that arise at different points in time to be brought to a single present value. This assists the process of forming an overall view of the desirability of particular possible courses of action. Because it deals with effects over time, intergenerational considerations are involved.

In line with the individual preferences value judgement, if individuals judge that benefits and costs that arise at different points in time have different values, then adherence to that value judgement suggests these preferences should be taken into account in valuing benefits and costs over time. The concept of diminishing marginal utility of consumption, which was introduced above in relation to distributional weighting, is important in this regard, because it reflects judgements about the relative worth of consumption to people at different levels. These different levels might relate to changing levels of consumption over time, as an individual's income and consumption levels increase (as they usually, but not necessarily, do).

The rate at which individuals discount the future is called their marginal time preference rate, which has nothing to do with inflation. Individuals usually prefer to receive goods and services sooner rather than later. The same is usually true of society as a whole, and the social time preference rate (STPR) is an indicator of society's preference for the present over the future. The UK Treasury (HM Treasury 2012) proposes a value of 3.5 per cent for discounting, based on its assessment of the value of the STPR. This follows use of Ramsey's (1928) formula, shown in equation 4.5:

$$r = \rho + \mu g \quad (4.5)$$

where:

r = the social rate of time preference

ρ = the utility discount rate, reflecting the rate at which individuals discount future consumption, assuming no change in per capita consumption is expected (assumed to be 1.5 per cent by the UK Treasury)

μ = the elasticity of the marginal utility of consumption (assumed to be 1.0 by the UK Treasury)

g = the expected increase in per capita consumption (assumed to be 2 per cent by the UK Treasury).

In the UK Treasury case, this gives $r = 1.5 \text{ per cent} + 1.0 \times 0.02 = 3.5 \text{ per cent}$. Pearce and Nash (1981) discuss a number of issues around choice of the components of the STPR. Evans (2005) suggests values of 1.0 per cent, 1.35 and 2.1 per cent respectively for the UK for ρ , μ and g respectively, giving an STPR of 3.8 per cent, slightly higher than the 3.5 per cent Treasury value but close enough to be of little concern.

A positive value for ρ effectively implies that the utility of future generations should be discounted, because they are in the future, a value judgement some would find unacceptable. HM Treasury (2012) points out that the value of ρ

consists of two elements. The first is an allowance for *catastrophic risk*, where that risk would effectively eliminate all returns from the initiative under examination. Stern's (2007) climate change review adopted a value of 0.1 for this component, generating considerable debate about adequacy. Weitzman (2007), for example, is critical of treating this value as close to zero, although he recognises it is a position supported by many well-known economists.

The second component is *pure time preference*, effectively an indication of individual impatience. A number of commentators suggest this should be excluded from determination of a *social time preference rate*, on the basis that it reflects irrational preferences. Pearce and Nash (1981), for example, argue in this way, adding that adherence to the individual preferences value judgement does not preclude diverting from following such preferences if they are judged as being irrational and the context is one of taking social decisions.

Garnaut, in his 2011 climate change review update for the Australian government, used what he called a normative discount rate (equivalent to an STPR) of 1.35 per cent and 2.65 per cent. This resulted from effectively ignoring the utility discount rate (ρ), a positive value for which he argued flies in the face of the principle of treating all generations alike.

Weitzman (2007) argues, in the context of long-term climate change, that one could start with a higher value for the STPR of 6 to 7 per cent, rather than 3 to 4 per cent, based on values of about 2 for each of the three variables in the STPR formula, and still conclude that a discount rate of 2 to 4 per cent is appropriate, because of the effect of uncertainty over the long term. The possibility of catastrophic climate change would lead to considerably lower rates again, as future consumption collapses.

A focus on uncertainty over the long run is picked up in the HM Treasury (2012) approach, which accepts that the discount rate should decline over very long evaluation time periods, reflecting uncertainty about the future. Thus HM Treasury (2012) proposes that an STPR discount rate of 3.5 per cent be used for analysis periods of 1 to 30 years, 3.0 per cent for years 31 to 75, 2.5 per cent for years 76 to 125, 2.0 per cent for years 126 to 200, 1.5 per cent for years 201 to 300 and 1.0 per cent for analysis periods beyond 300 years. This is particularly relevant to policy analyses involving climate change but also for long-life rail projects.

The actual rate that is selected to discount future benefits and costs can be a very significant influence on the relative desirability of alternative options. The higher the discount rate, the less value are initiatives that produce

benefits well into the future. Table 4.4 shows the present value of a dollar's benefit that arises at various years into the future, at a range of discount rates. Thus, for example, at a 3 per cent real discount rate, \$1 is worth 41 cents in 30 years' time. However, this reduces to 13 cents at 7 per cent and to just 0.6 cents at 10 per cent. Benefits beyond year 30 have little value at discount rates of around 7 per cent or higher. This is particularly significant for the evaluation of major transport infrastructure initiatives, such as new rail lines or motorways, where benefits beyond 30 years will be significant, but only if the chosen discount rate does not effectively wipe them out!

Table 4.4 Real discount factors (examples)

Year	3% discount rate	7% discount rate	10% discount rate
0	1.0000	1.0000	1.0000
10	0.7441	0.5083	0.3855
20	0.5537	0.2584	0.1486
30	0.4120	0.1314	0.0573

Many countries do not adopt an STPR for discounting future benefits and costs, adopting a social opportunity cost (SOC) of capital approach instead. This measures the estimated value forgone if the sacrificed expenditure had been used in the private sector instead, in an activity of comparable risk. The SOC is usually expected to be higher than the STPR, especially over usual project evaluation time scales, Harrison (2010), for example, estimating a rate of 8 per cent for Australia, based on a national accounts measure of the before-all-taxes real rate of return on private capital, adjusted to reflect the impact of tax distortions and foreign borrowing. He largely favours the SOC approach on the practical basis that it is easier to calculate than an STPR, a position that (1) ignores the distortions introduced to project selection by using relatively higher rather than lower rates (i.e. higher rates disadvantage projects whose benefits are more distant in the future) and (2) moves away from the fundamental individual preferences value judgement.

Important benefits of using an STPR-based rate, in our view, are that: (1) it broadly continues the general adherence to the individual preferences value judgement in evaluation through to discount rate selection, even if one concludes that pure time preference should be ignored; and (2) it introduces a consistency in distributional weighting between and across generations. From a practical transport policy perspective, an STPR also gives greater recognition to benefits and costs arising in the more distant future, which is important for many major transport policy or project decisions (for example,

where long-lived assets may be involved), especially if this is in accord with how society values the future.

In practical terms, uncertainty about the basis for discount rate selection suggests that evaluations should conduct sensitivity testing across a range of discount rates, which probably range between 3 and 7 to 8 per cent, the former being of the order of an STPR and the latter probably indicative of the SOC of capital. Our preference is for the major focus to be on an STPR rate, for reasons outlined above, with sensitivity testing of higher rates. Beyond 30 years, there are strong grounds for the discount rate declining slowly.

Risk and uncertainty

Risk and uncertainty intervene to deliver policy, programme and project outcomes that diverge from the expected values that are estimated in relevant evaluations. A risky situation is one where the analyst has some idea of the probabilities of various outcomes, whereas uncertainty refers to a situation where no such probabilities can be assigned. Concern with these matters has increased considerably over the past decade, as the scale of major projects increases and the consequences of poor decision making grow alongside.

UK Department for Transport appraisal advice identifies a range of risks that are likely to be encountered in a project and, by extension, in policy development (Department for Transport 2011). It lists the following types of project risks that might cause costs to exceed expectations (or, in fortuitous circumstances, to be less than expected):

- Policy risk:
 - legislative risk (e.g. legislative change causing costs to increase); and
 - policy risk (e.g. policy change causing costs to increase).
- Risk on delivering the asset:
 - construction risk (e.g. costs of delays or specification creep);
 - planning risk (e.g. failure to get required approvals); and
 - residual value risk (e.g. technical change may render residual asset value obsolete).
- Risk on operating the asset:
 - operational risk (e.g. operational efficiency and costs may diverge from expectations);
 - inflation risk (e.g. real cost changes may diverge from expectations); and
 - maintenance risk (e.g. sustaining asset condition may be harder than expected).

- Risks of demand and revenue:
 - demand risk (e.g. traffic forecasts may be inflated);
 - design risk (e.g. design standards may not deliver expected service quality);
 - availability risk (e.g. service availability may be less than expected, for reasons such as maintenance problems);
 - volume risk (e.g. risks that actual use may vary from forecast); and
 - technology risk (e.g. technological change may affect relative service quality and use).

Prior experience with similar policies or projects may also enable an assessment of probabilities (objective or subjective) that may affect estimation of particular benefits and costs. This can lead to expected value estimates, where single values are replaced by a number of values with associated probabilities. Table 4.5 provides a simple example.

Table 4.5 Probabilistic expected value calculation

Estimate	Travel-time savings benefits (\$m)	Probability	Expected value of benefits (\$m)
1	15	0.1	1.5
2	20	0.3	6
3	25	0.4	10
4	30	0.2	6
Expected value			23.5

Risk management procedures are well developed and should form a normal part of project evaluation, to help ensure development of a robust case. In some jurisdictions there are formal requirements in this regard, as an integral part of funding bids (such as for UK transport proposals seeking government funding support). Risk management procedures provide a structured approach to identifying, assessing and, most importantly, controlling risks that might arise during the period of a policy, programme or project, including identification and implementation of various risk mitigation approaches.

To minimise both transaction costs and overall risk premiums, it is important to allocate each risk to the party best able to manage that risk. After risk allocation and treatment (which may involve doing nothing in some cases, if the expected benefits of actions to control a risk are expected to be less than the costs involved – called tolerating risk), there are still usually a number of

residual risks, which often relate (for example) to the wider policy or project environment (for example, the economic situation).

Risk assessment involves seeking to quantify the likelihood that a particular event (risk) will occur and, if it does, quantifying the consequences. This is sometimes done by using two simple sets of ratings of 'low, medium and high', the first rating for likelihood of particular risk occurrence and the second for scale of impact if particular risks do occur. If probabilities can be reasonably assessed, this enables more refined approaches, including calculation of expected values, as illustrated in Table 4.5.

Optimism bias describes the well-documented propensity of policy or project promoters to overestimate the good (such as benefit flows) and underestimate the bad (such as cost), inflating expected proposal outcomes. It is often dealt with by adding a margin to the capital costs of a proposal and by sensitivity testing for key assumptions that underpin benefit estimates. Risk management procedures should mean that optimism bias reduces as a proposal becomes more firmly established.

Uncertainty is more difficult than risk, because there is no knowledge of probabilities. Various decision theory models can be used to assist decision taking in such circumstances, such as maximax, maximin, index of pessimism, minimax regret, and Monte Carlo simulations. The decision maker's attitude to risk taking is important in choice from among such possibilities, because some are risk-averse approaches and others are risk-neutral.

Sensitivity testing is commonly used to deal with uncertainty in transport policy evaluation. This involves testing the implications of changing various evaluation inputs for evaluation outcomes. In a transport policy setting, this will usually involve sensitivity testing of both demand-side (for example, traffic forecasts, unit benefit values) and supply-side influences (for example, capital cost). Our discussion on the value basis of CBA/evaluation, and the absence of uniquely correct value judgements, provides a further reason for sensitivity testing, in this case with respect to changes in value judgements (value sensitivity analysis). Backcasting and dynamic adaptive policy approaches may also be relevant in situations of uncertainty (Marchau et al. 2013).

In summary then, looking at the cost side of transport policies or projects, the recognition of risk, optimism bias and uncertainty will often see a policy or project base capital cost increased to incorporate a quantified risk adjustment, increased again for optimism bias and sensitivity-tested for uncertainty.

Operating costs and benefit-side considerations will also be subject to treatments for risk and uncertainty.

4.7 Conclusions

Cost–benefit analysis started as a narrow and rather limited way of evaluating the economic impacts of policies, programmes and projects, with transport being a major field of application. Improvements in technique have broadened its applicability, particularly through extending the ability to value impacts in money terms and to make allowance for distributional impacts, including between generations. Thus, some policy consequences that might be categorised as social or environmental are now included within the ambit of monetary assessment, which nominally gives them an economic dimension and facilitates inclusion within monetary CBA. Some policy impacts, however, remain beyond the scope of a monetary valuation, either because it is simply too hard or perhaps because the analyst's value position regards monetary valuation as unacceptable. Valuation of some other elements will usually be subject to wide margins of uncertainty.

The chapter has emphasised the value basis of CBA, and implicitly of all evaluation, and argued that it is important for value judgements that are embedded in any evaluation to be clear and transparent, for others to consider and test the implications of alternative value positions, should they so desire. It has argued that value sensitivity analysis should be undertaken as a normal part of the CBA policy evaluation process, along with sensitivity testing to help manage uncertainty on more technical inputs (e.g. costs).

Sustainability perspectives emphasise the importance of the environment, of considering intergenerational issues and of distributional considerations within generations. Discount rate selection is particularly important for intergenerational questions. The question of whether some aspects of the natural environment should perhaps be treated as constraints within which economic considerations should be expected to operate is also important, with intergenerational relevance. This again raises important values questions. The idea of conserving the stock of natural capital, part of which may be sometimes given legislative mandate, can assist in contributing to both environmental and intergenerational sustainability. It leads to the concept of *environmental compensation* as being within the scope of policy.

If a means of adding up both those expected policy outcomes that are quantifiable in money terms and those that are not (but which may perhaps be measurable in other physical terms or in terms of subjective ratings or

rankings) is desired, then the public participation process can include a way of accomplishing this, using a form of multi-criteria analysis.

The individual preferences value judgement is fundamental to the approach to evaluation set out in this chapter. It underpins CBA and supports the wide use of public participation processes throughout the policy cycle. There are grounds for moving from the individual preferences approach for some matters, such as where public or social preferences might be different from individual private preferences or where private preferences might be based on a lack of information. Where such positions are taken in an evaluation, this should be made clear, to support discussion about relevant value positions.



REFERENCES

- Button, K. (2010), *Transport Economics*, 3rd edn, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Department for Transport (2007), *The Options Values Sub-objective: TAG Unit 3.6.1*, January, available at: www.dft.gov.uk/webtag/documents/expert/unit3.6.1.php (accessed 5 May 2012).
- Department for Transport (2011), *The Estimation and Treatment of Scheme Costs: TAG Unit 3.5.9*, April, available at: www.dft.gov.uk/webtag/documents/expert/unit3.5.9.php (accessed 9 May 2012).
- Dube, J., F. Des Rosiers, M. Theriault and P. Dib (2011), 'Economic impact of a supply change in mass transit in urban areas: a Canadian example', *Transportation Research Part A*, **45** (1), 46–82.
- Eigenraam, C.J.J., C.C. Koopmans, P.J.G. Tang and A.C.P. Verster (2000), 'Evaluation of infrastructural projects: guide for cost–benefit analysis', Section 1: 'Main report', report prepared for Netherlands Ministry of Transport, Public Works and Water Management and Ministry of Economic Affairs.
- Evans, D. (2005), 'The elasticity of marginal utility of consumption: estimates for 20 OECD countries', *Fiscal Studies*, **26** (2), 197–224.
- Garnaut, R. (2011), 'Update paper 1: weighing the costs and benefits of climate change action', Garnaut Climate Change Review, Canberra, 3 February.
- Harrison, M. (2010), 'Valuing the future: the social discount rate in cost–benefit analysis', Visiting Researcher Paper, Productivity Commission, Canberra, April.
- Harsanyi, J. (1955), 'Cardinal welfare, individualistic ethics and interpersonal comparisons of utility', *Journal of Political Economy*, **63** (4), 309–21.
- Hausman, D.M. and M.S. McPherson (2006), *Economic Analysis, Moral Philosophy, and Public Policy*, Cambridge, UK: Cambridge University Press.
- Hicks, J.R. (1939), 'The foundations of welfare economics', *Economic Journal*, **49** (196), 696–712.
- HM Treasury (2012), *The Green Book: Appraisal and Evaluation in Central Government*, Annex S: 'Distributional impacts', London: TSO, available at: http://www.hm-treasury.gov.uk/d/green_book_complete.pdf (accessed 10 April 2012).
- Kaldor, N. (1939), 'Welfare comparisons of economics and interpersonal comparisons of utility', *Economic Journal*, **49** (195), 549–55.
- Maibach, M., C. Schreyer, D. Sutter, H.P. van Essen, B.H. Boon, R. Smokers, A. Schroten, C. Doll, B. Pawlowska and M. Bak (2007), 'Handbook on estimation of external cost in the transport sector', produced within the study Internalisation Measures and Policies for All External Cost of Transport, CE Delft, 19 December.

- Marchau, V., J. Annema, W. Walker and J. van der Waard (2013), 'Transport futures research', in B. van Wee, J. Annema and D. Banister (eds), *The Transport System and Transport Policy: An Introduction*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Nash, C., D.W. Pearce and J.K. Stanley (1975), 'An evaluation of cost-benefit analysis criteria', *Scottish Journal of Political Economy*, **XXII** (2), 121–34.
- Pearce, D.W. and C.A. Nash (1981), *The Social Appraisal of Projects: A Text in Cost-Benefit Analysis*, London: Macmillan.
- Pearce, D.W. and R.K. Turner (1990), *Economics of Natural Resources and the Environment*, Hemel Hempstead, Hertfordshire: Harvester Wheatsheaf.
- Praag, B. van and A. Ferrer-i-Carbonell (2004), *Happiness Quantified: A Satisfaction Calculus Approach*, Oxford, UK: Oxford University Press.
- Ramsey, F. (1928), 'A mathematical theory of saving', *Economic Journal*, **38** (152), 543–59.
- Stanley, J., D. Hensher, J. Stanley, G. Currie, W.H. Greene and D. Vella-Brodrick, D. (2011), 'Social exclusion and the value of mobility', *Journal of Transport Economics and Policy*, **45** (2), 197–222.
- Stanley, J.K., J.R. Stanley and D.A. Hensher (2012), 'Mobility, social capital and sense of community: what value?', *Urban Studies*, **49** (16), 3595–609.
- Stern, N. (2007), *The Economics of Climate Change: The Stern Review*, Cambridge, UK: Cambridge University Press.
- Turner, V.K. (1991), 'Environment, economics and ethics', in D. Pearce (ed.), *Blueprint 2: Greening the World Economy*, London: Earthscan Publications.
- WCED (World Commission on Environment and Development) (1987), *Our Common Future*, Australian edn, Melbourne: Oxford University Press.
- Weitzman, M. (2007), 'A review of the Stern Review on the economics of climate change', *Journal of Economic Literature*, **XLV** (September), 703–24.

5

Traffic theory and transport planning foundations



CHAPTER OVERVIEW

A predominant theme of transport policy in the twenty-first century is dealing with congestion. In this chapter, the fundamentals of road traffic flow are explored, as a means to create understanding of how congestion arises and, as a consequence, how one might set about formulating policies that can respond to a desire to control or even reduce congestion. Following a treatment of basic traffic flow relationships, the chapter also outlines the concept of performance levels for highways and then explores briefly the role of travel-demand forecasting in supporting policy formulation. The chapter then turns to the estimation of congestion costs, based on the preceding traffic flow relationships, and finally the policy implications of the material in the chapter are discussed.

5.1 Rationale for traffic theory

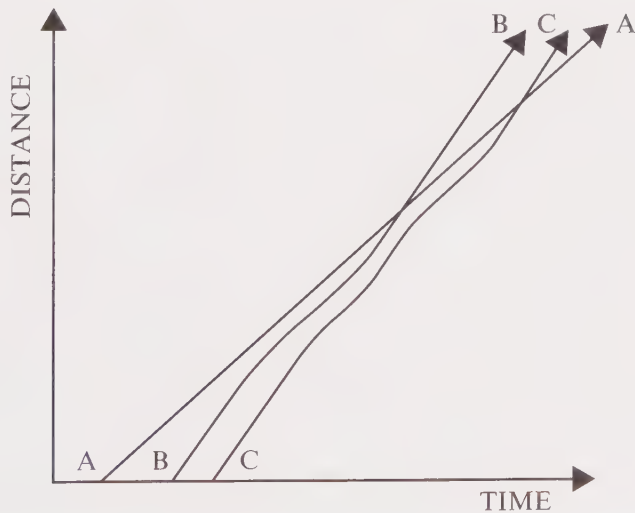
Much of the concern with transport policy focuses on the issue of congestion and how to limit or even reverse congestion. Governments around the world attach a high policy priority to dealing with congestion, which is not surprising, given that the costs of congestion are typically estimated at around 1 per cent or more of GDP. The ways in which congestion arises must be understood so as to be able to formulate sensible policy relating to congestion. Further, transport policy should be aiming at improving the performance of the transport system, or maximising the performance of the investments already made. Therefore, it is necessary to understand how to measure performance of the transport system and, more particularly, to understand why and how congestion arises, and how the system behaves once congestion occurs. The traffic theory explained in this chapter is concerned with these two aspects of the transport system – how to assess the performance of the system from the viewpoint of the system provider (usually government) and how to understand and manage the phenomenon of congestion.

Most traffic theory has been developed in the context of roads and highways. Notwithstanding this, the concepts and ideas are actually more broadly applicable across most forms of transport. Therefore, we make no further apology about developing basic ideas of traffic theory from the highway perspective, but urge the reader to see how these theories actually continue to work in more or less the same way in other areas, such as bus transport, rail and even intercity air. All are subject to potential congestion, and all can be assessed in terms of performance levels that relate to the extent of interference in the performance of one vehicle compared to other vehicles in the system.

5.2 Space and time

A good starting point for understanding traffic movements through a transport system is the space–time diagram. This is a diagram on which the horizontal axis is time and the vertical axis is distance, both time and distance being measured from some arbitrary reference point. On these axes, one may then represent the trajectory of individual vehicles, moving through a portion of a transport system. Figure 5.1 shows a simple space–time diagram.

Figure 5.1 A simple space–time diagram



In Figure 5.1, three vehicle trajectories are shown, for vehicles A, B and C. Vehicle A passed the spatial reference point a short time after the beginning of the observation period. It travelled at a constant speed, as is shown by the straight-line trajectory, which shows that the rate of progress of this vehicle in space and time is constant. Vehicle B passed the spatial reference point a short time later than vehicle A, but travelled faster, with some speed variation

a short while after starting, but then returning to a steady speed. Vehicle B travelled faster than vehicle A and therefore passed it at a certain point in time and space. Vehicle C passed the reference point even later than vehicle B, initially travelled at about the same speed as vehicle B, then had more variation in speed, shown by the wavy line, and eventually also passed vehicle A, and appeared to be travelling at about the same speed as Vehicle B. Indeed, one of the interesting things we can immediately note from Figure 5.1 is that the speed of the vehicle is given by the slope of the line at any point that describes the trajectory of that vehicle. Thus, we can conclude that vehicle A is travelling more slowly than either of the other two vehicles, and hence it is logical that Vehicle A is eventually passed by each of the following vehicles.

Potentially, there is yet more information that can be obtained from this diagram. This is illustrated in Figure 5.2, which shows the trajectories of six vehicles.

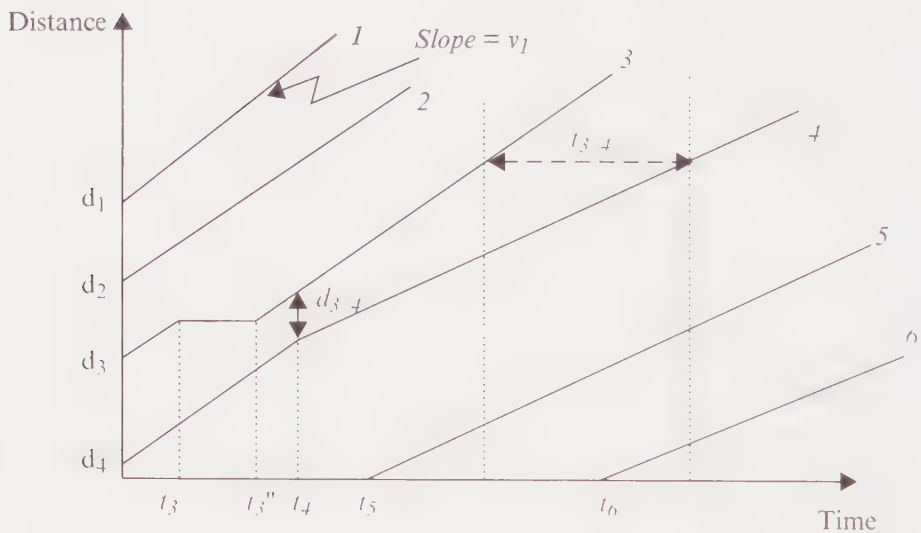


Figure 5.2 Space-time diagram showing six vehicles

Figure 5.2 shows four vehicles that had already passed the spatial reference point when the time started, and were travelling in the same direction. These vehicles were respectively at distances d_1 , d_2 , d_3 and d_4 from the reference point when the time of observation began. Because the slopes represent the speed, indicated as v_1 for the first vehicle, we can see that vehicles 1 and 2 were travelling at a constant speed, with vehicle 1 being slightly faster than vehicle 2. Vehicle 3 had to stop between time t_3 and time t_3'' . Somewhat

unrealistically, this vehicle is shown as having a constant speed up to the stop and a constant speed immediately after the stop, where in fact these lines would be expected to curve, showing deceleration and acceleration.

Two other concepts are shown in this diagram. These are the time headway and the distance headway between vehicles. The time headway is the time from when a vehicle passes a point until the following vehicle passes the same point. In Figure 5.2, it is shown for vehicles 3 and 4 at a certain distance from the reference point as t_{3-4} . Time headways are particularly important for public transport, where most passengers will want to know the time headway between buses or trains serving a particular route. Time headways are also important in safety, where they may be used to guide drivers on how far apart their vehicle should be from the one in front. Distance headways are also important from a safety viewpoint. Vehicle spacing for safe operation should be greater than the distance required for driver reaction time plus the braking time required to come to a complete stop. Distance headway is shown between vehicles 3 and 4 at a certain point in time in Figure 5.2 as d_{3-4} . The letter h is usually used to indicate the time headway, while the letter s is used to indicate the space headway.

The time–space diagram is particularly useful for studying the performance of traffic signals, as well as understanding something about the way in which individual vehicles are performing in a transport network. For example, the time–space diagram can readily allow the planner or engineer to determine the amount of time that vehicles incur as delay at a traffic signal, by finding the stops and adding up the amount of stop time by vehicles from the trajectories. By plotting time–space diagrams for the different directions of traffic at a signalised intersection, one can also determine the amount of vehicle delay on different approaches and the volumes of traffic on each approach. This information can be used to assist in deciding on the appropriate timing for the signals. With the line indicating the trajectory of the vehicle through time and space, the slope of the line at any point indicating the speed of the vehicle, and the distance vertically between two vehicles showing the space headway and horizontally the time headway, a considerable amount of information is available from such a diagram.

5.3 Flow, density and speed

At its most basic, a transport system is devised to move people, goods or vehicles from one point to another. One would normally expect that the desire would be to move people, goods or vehicles from point A to point B efficiently and safely. From this, it follows fairly logically that the fundamental

concept of traffic that is of interest in the context of policy is that of flow, where flow is measured as the quantity of traffic that moves from one point to another. Of course, as we shall develop shortly, flow alone is not a sufficient criterion to use to describe efficient and safe movement of traffic, but it is fundamental. The engineer defines flow specifically as the number of people, goods or vehicles passing a point in a set period of time. To make the discussion a little easier to comprehend from this point forward, we will concentrate on the movement of vehicles, and assume that this implies movement of people and goods as well, and that each of these is substitutable with the others. However, the growing focus on moving people and goods, rather than concentrating just on the vehicles, is the right way to go in formulating policy.

Hence, flow or volume is the first basic concept that we need to consider as a means of understanding performance of the system and also congestion. Flow or volume is represented mathematically by the letter q , derived from the fact that flow measures quantity. Flow is usually measured in vehicles per unit of time, for example vehicles per second, vehicles per minute or even vehicles per hour.

The second basic concept that is important in this respect is that of the density of traffic, where density is measured as the concentration of vehicles in a length of roadway and is usually measured as vehicles per lane kilometre at a point in time. A very low density of traffic would imply that there is too much available space in the transport system for the use being made of it, and one would conclude that such a transport system is inefficient. Very low densities are also associated with very low volumes, such as one might experience when travelling on a roadway at, say, 2 a.m. In all likelihood, there are few, if any other, vehicles in sight, and vehicles pass a given point very infrequently. The density is very low, where density is measured in vehicles per kilometre of roadway, at a given point in time. If one were to imagine the view of a segment of roadway from a satellite and taking a picture of that roadway, the count of vehicles along the roadway would be small, and the density, therefore, very low.

On the other hand, when density becomes excessively high, this would represent a traffic jam situation, where vehicles are very close to one another and hardly moving. In fact, if a segment of the transport system becomes totally jammed, so that no vehicles are moving and the vehicles are as close to one another as is safe, then the density is a maximum, but the flow is zero. Flow has dropped to nothing because of the fact that density has increased to its maximum value.

Density is represented mathematically by the letter k . Density is measured in vehicles per unit of distance, for example vehicles per kilometre or vehicles per metre.

The third basic concept for understanding traffic movements is speed. However, speed can be measured in at least three ways. First, there is the instantaneous speed of a vehicle, which is the slope at a given point of the vehicle trajectory in the space–time diagram. This is the speed at any moment for a specific vehicle. However, we are often more interested in the average or mean speed. In this case, there are two ways to average speed. One may average the speed of all vehicles over a period of time at a particular point on the transport network. This is the time mean speed, usually written as $\bar{v}_t$. The other way is to average the speed of all vehicles in a length of the network at a particular moment in time. This is the space mean speed and is usually written as $\bar{v}_s$. Under the conditions where the density is very low and flow is very low, the space mean speed and time mean speed will normally be at their maximum value. This is called the free-flow speed. At the extreme, when there is but a single vehicle on a length of roadway for a specified period of time, both the space mean speed and the time mean speed will be the same and will be equal to the free-flow speed. As the flow begins to increase, and the density also increases, the drivers of individual vehicles will need to begin to adjust their speeds to the speed of other vehicles, or decrease their speed to maintain safe distance headways. As density continues to increase, speeds will continue to drop until, at jam density, the speed is zero. Likewise, as flow increases, the speed will decrease, mainly slowly over a significant range of flows. However, when flow begins to be reduced by increasing density, the speed will drop more dramatically, but will also become unstable, with the gradual development of stop-and-go conditions, where the speed will vary considerably.

As noted above, speed is represented by the letter v , which represents the speed of a single vehicle. Time mean speed and space mean speed are written with an overbar on the letter v , with a subscript t to represent time mean speed, and subscript s to represent space mean speed.

There are, in fact, both mathematical and graphical relationships between flow, density and speed. Assuming that the traffic flow is observed for T seconds, during which time n vehicles are observed passing the observation point, then the flow rate in vehicles per hour would be given by equation 5.1:

$$q = \frac{n \times 3600}{T} \quad (5.1)$$

Similarly, if n vehicles are observed to occupy a length of the roadway of L metres at a point in time, then the density in vehicles per kilometre will be given by equation 5.2:

$$k = \frac{n \times 1000}{L} \quad (5.2)$$

Flow and density are related to the time and space headways. Time headway is the inverse of the flow, as shown in equation 5.3:

$$h = 1/q \quad (5.3)$$

Similarly, the space headway is the inverse of the density, as shown in equation 5.4:

$$s = 1/k \quad (5.4)$$

The time mean speed is obtained by averaging the speeds of the vehicles observed at a specific place over a defined period of time. If the speed of the i th vehicle is denoted v_i , then the time mean speed is given by equation 5.5:

$$\bar{v}_t = \frac{1}{n} \sum_{i=1}^n v_i \quad (5.5)$$

The space mean speed is determined from the harmonic mean of the individual vehicle speeds observed at a point in time over a length of roadway, or by multiplying the length of the road segment by the number of vehicles and dividing this by the sum of the lengths of time taken by each vehicle to traverse the segment, as shown in equation 5.6:

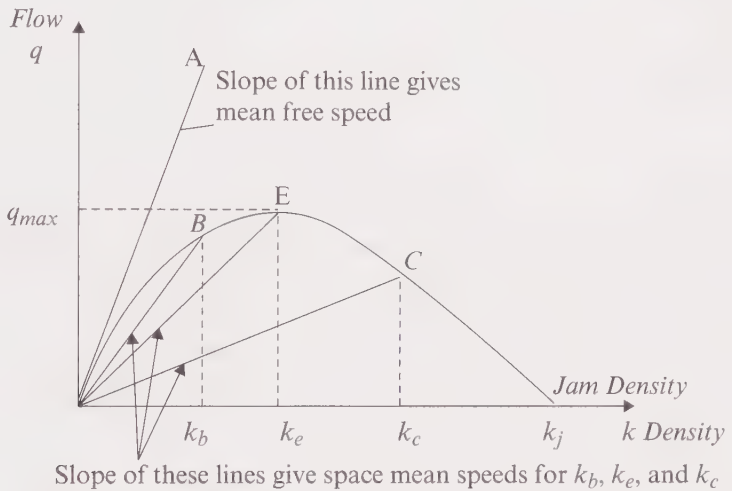
$$\bar{v}_s = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{v_i}} = \frac{Ln}{\sum_{i=1}^n t_i} \quad (5.6)$$

These six equations represent the relationships among the basic concepts of flow, density, speed and headway.

The relationships between flow, density and speed can also be shown graphically. These graphs are readily deduced from the descriptions of the inter-relationships provided a few paragraphs earlier. Figure 5.3 shows what is known as the fundamental diagram of traffic flow, which shows the relationship between flow and density. As described earlier, at zero flow, the density is also zero. At jam density, the flow is again zero. As flow increases initially, so does the density. However, when point E is reached, while the density

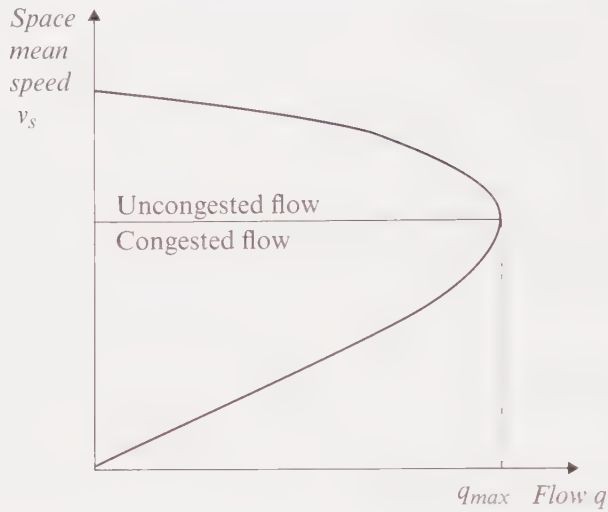
continues to increase, the flow now decreases. Thus, point E corresponds to the density that produces the maximum flow. Also, as noted earlier, the speed gradually declines as density increases. The space mean speed of the vehicles is the slope of the line that connects the origin of the graph to any point on the curve. Thus, the line from the origin to point B has a slope of v_b , which is the space mean speed at a density of k_b , at which speed and density the volume or flow would be q_b . The same would be true at point C. The line to A, which is the slope of the curve at the origin, represents the free-flow speed or mean free speed.

Figure 5.3 The fundamental diagram of traffic flow



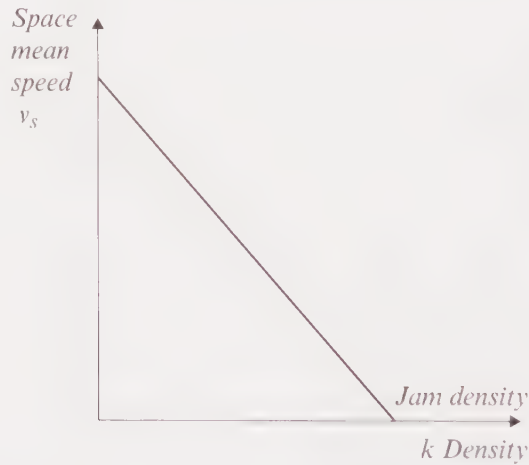
From this diagram, we can then derive the next important graph, which shows the relationship between flow and space mean speed, and is shown in Figure 5.4. As can be deduced from Figure 5.3, as the flow increases from zero, the speed initially decreases very little, but gradually decreases more as the flow continues to increase. When maximum flow is reached, the speed starts to drop quite dramatically and the output volume also drops, as traffic moves into the realm of what traffic planners and engineers call congestion. To transport economists, congestion sets in as soon as vehicles start to impede each other and speeds fall below free speed – this distinction in the meaning of congestion, as between different transport professionals, is seldom recognised. Although not shown in Figure 5.4, experimental observation provides points that lie on or very close to the line from zero flow and maximum (mean free) speed up to maximum flow. However, when the curve returns on itself, the observed points become widely scattered, owing to the instability of the flow of traffic, under congested conditions that include stop-and-go traffic.

Figure 5.4 The speed–flow diagram



The final relationship of interest graphically is that between density and speed. In this case, the relationship is linear, as shown in Figure 5.5. At zero density, the space mean speed is a maximum and equal to the mean free speed, or free-flow speed. As density increases, the speed steadily drops, reaching a value of zero at jam density. Point E from the fundamental diagram of traffic flow (Figure 5.3) is at the midpoint of density and speed on Figure 5.5.

Figure 5.5 Relationship between space mean speed and density



There are several conclusions that can be drawn from these relationships, shown both graphically and mathematically. Perhaps the most important conclusion is that, in terms of technical flow efficiency, it is desirable for transport systems to operate at or near maximum flow. However, because instability and congestion arise rather quickly once maximum flow is achieved, it is probably

desirable to maintain a design flow that is a little bit below the maximum flow, at speeds a little higher than those that occur at maximum flow and at densities that are below the density that corresponds to maximum flow. Second, maximum flow is also quite clearly what is meant by the capacity of any part of the system. Capacity cannot be exceeded, and attempts to put more into the system than there is capacity to allow to flow through the system will result in congestion and a decrease in the actual output volumes, as shown by Figure 5.4. Speeds will also drop rapidly in this realm and, as a result, the time and cost of travel will both increase dramatically. Third, the idea of traffic flow maximisation is not the same as the economic idea of maximising the net social benefits that are associated with traffic flow. We have pointed out in Chapter 4, for example at Figure 4.3, that this point occurs where the demand curve intersects the marginal social cost (MSC) curve, which is derived from the speed–flow curve.

In Chapter 4, the idea was introduced of applying the microeconomic concepts of supply and demand to transport. By looking at these graphs in Figures 5.3 to 5.5, and particularly 5.4, we can also deduce the shape of the supply curve for transport. Speed relates closely to the cost of travel. The higher the speed, the less time it takes to travel from one point to another. Hence, the less expensive is travel. At the same time, most vehicles have decreasing operating costs with increasing speed, up to a point, after which the costs will start to increase again. However, in the range that is normally consistent with safely designed transport systems, costs of operation of vehicles are generally higher when the vehicle is moving slowly, and lower when it is moving quickly. Therefore, Figure 5.4 would lead to arguments for a supply curve for transport that would appear much like Figure 5.6. More correctly, it should be stated that, in the case of transport networks, this is not strictly a supply curve, but is rather a price–volume curve, that is, a curve that shows the relationship between the price that will be incurred and the volume of use being made of the system.

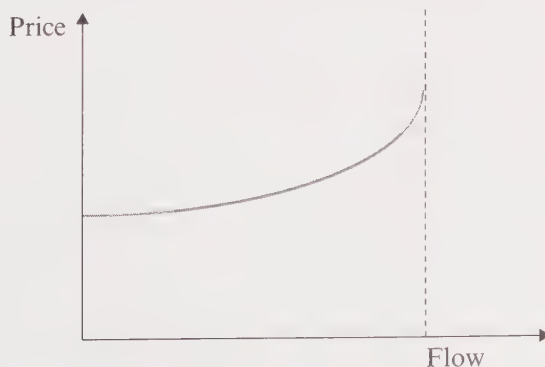


Figure 5.6 Supply curve derived from the speed–volume curve

It behaves very much like the supply curve. In this context, price of travel includes time costs, money costs, inconvenience, unreliability, discomfort and so forth and is usually referred to as the generalised cost of travel.

5.4 System performance

Having now developed these basic concepts and relationships for traffic flow, we can turn our attention to a second important aspect of traffic flow – performance. The performance of the system is important to understanding the system and to developing policy about the system. Traffic engineers use a system of performance measures that have been developed in the USA, and which have largely been adopted around the world (TRB 2010).

Capacity

As has been shown in section 5.3, capacity is the maximum output flow rate that can be achieved within a given time period. It is usually measured by observing traffic flow for a 15-minute period. Capacity is affected by a number of different characteristics of the road, its users and its environment. Elements of design that affect capacity include:

- number of lanes;
- lane width;
- shoulder width;
- horizontal and vertical alignments;
- design speed;
- queuing space at intersections; and
- lateral clearance.

The familiarity of drivers with the roadway, as well as other aspects of driver competence and confidence, also affect capacity, as does the mix of vehicles using the roadway. The maximum capacity of a roadway is achieved in a long segment of roadway with no intersections, no significant gradient, and a straight horizontal alignment, normally found on freeways or motorways. With traffic composed mainly of passenger cars and with drivers who habitually use the roadway and are very familiar with it, a maximum capacity will be achieved. The presence of intersections, traffic signals and so on, as well as a mix of traffic that includes trucks and buses, will reduce the capacity of a segment of roadway. Capacity of urban roads with frequent intersections is normally determined by the capacity of the intersection itself, not the capacity of the lanes between intersections. For example, consider a major urban arterial road, with traffic signals at intersections, where the green time for

the urban arterial is 70 per cent, or 42 seconds out of every minute: such a condition will reduce the capacity of the roadway to, at most, 70 per cent of the capacity of an equivalent roadway with no intersections. If there are also vehicles entering and exiting from driveways along the roadway (such as entrances to shopping and commercial facilities), the capacity will be reduced yet further. In addition, if vehicles are changing lanes, in order to position themselves for right and left turns out of the roadway, this will again decrease capacity. Another cause of reduction in capacity is that of drivers looking for parking spaces, especially when these are along the street. The manoeuvre of getting into or out of a parking space will also reduce capacity.

As a result of these impacts on capacity, all of which also affect level of service of the roadway, the ideals for level of service are couched in terms of a long length of freeway, with no near intersections or interchanges, minimal horizontal and vertical curves, minimal requirement to change lanes, and a predominantly passenger car vehicle population with drivers who use the roadway habitually. Under such conditions, depending on the design speed of the roadway, capacity may run as high as 2400 passenger car equivalents per lane per hour. A passenger car equivalent (pce) is a concept that allows any type of vehicle to be expressed in terms of equivalent passenger cars, based on its effect on traffic. For example, a large, heavy truck negotiating a steep upward gradient may be equivalent to as many as six or more passenger cars. A bus on a freeway, without stops, may be equivalent to no more than about 1.5 passenger cars. Because of the mix of vehicles in the traffic stream, capacity and levels of service are usually expressed in pces.

Levels of service

The level of service provided by a facility is a qualitative measure that is intended to reflect how the users of the system perceive the operating conditions and provides at least one possible indicator of a need for some sort of improvement to the system or policy change in how the system is operated. In a highway context, the level of service relates to the perception by drivers of the highway operating conditions. The *Highway Capacity Manual* (TRB 2010) defines six levels of service. These can be defined in terms of density of traffic, speed, average space headway and maximum achievable capacity. The levels of service range from A, which represents free-flow conditions, to F, which represents congestion and breakdown of flow. These levels of service can also be related to the speed–flow curve of Figure 5.4.

However, as noted previously, a number of factors affect the level of service of a roadway. Principal among these are lane width, lateral clearance, traffic

composition, grade, speed and the driver population. Lane widths of 3.7 metres or greater have no effect on level of service. Lane widths less than 3.7 metres will have the effect of slowing traffic, as drivers become more cautious about potential collision with vehicles in adjacent lanes. Obstacles at the side of the roadway have a similar effect on driver behaviour when they are within 2 metres of the edge of the roadway. Beyond 2 metres, they have no noticeable effect on level of service. Obstacles may include road signs, parked vehicles, barriers, trees and shrubbery, bicycle lanes and so on. As already noted, the presence of trucks and buses in the traffic stream affects levels of service, because the presence of large vehicles (buses, trucks, etc.) reduces the maximum flow because of their size, operating characteristics, and interaction with other vehicles. Gradients of less than 3 per cent do not affect levels of service, unless they are maintained for more than 1 kilometre. Steeper gradients than 3 per cent, and any gradient that is maintained for longer than 1 kilometre will affect the level of service. The driver population also affects the level of service, in that weekday commuters who know the route and conditions well represent ideal conditions. Large proportions of recreational drivers and other categories of drivers will reduce the level of service.

- **Level of service A.** Level of service A is free-flow operation with no interference from other vehicles. The density is less than or equal to 7.5 passenger cars per kilometre per lane, so the space headway is not less than 134 metres between successive vehicles in any lane. At a speed of 110 km/h, the safe distance between vehicles is about 90 metres, so the average headway is about 50 per cent greater than the safe distance. Hence, vehicles do not interfere with each other. Under these conditions, an incident, such as a disabled vehicle or an accident, may cause a reduction in the level of service, but recovery is rapid once the location of the incident is passed.
- **Level of service B.** Level of service B still represents free-flow operations, but with a higher density and some interference from other vehicles. The density will be between 7.5 and 12.5 passenger cars per kilometre per lane, or a space headway of 134 to 80 metres. Given that, at 110 km/h, the safe distance between vehicles is 90 metres, it is clear that, at the upper end of densities, cars will now have to adjust their speeds to maintain a safe spacing. At a speed of 100 km/h, the safe spacing drops to 83 metres, so that this becomes closer to the maintainable speed under this level of service. Nevertheless, if an incident occurs, recovery to the original level of service will still be rapid, and the reduction in level of service will not normally be large as a result of an incident.
- **Level of service C.** Level of service C has speeds at or near free-flow speed, but there is reduced freedom to manoeuvre because of the

presence of other vehicles. The density ranges from 12.5 to 18.8 passenger cars per kilometre per lane, or a space headway of from 80 to 53 metres. At a speed of 90 km/h, the safe distance between vehicles is about 75 metres, so that sustained speeds over 90 km/h are not safe at this level of service. Greater vigilance is needed by drivers under these conditions; an incident will cause a significant decrease in level of service, and it will take longer for recovery to the original level of service.

- **Level of service D.** Level of service D has speeds that are now declining, and freedom to manoeuvre is much more restricted. The density ranges from 18.8 to 26.3 passenger cars per kilometre per lane, or a space headway from 53 to 38 metres. At these headways, a safe speed¹ would be about 80 km/h. Minor incidents can create queuing at this level of service, which causes a decline to level of service F. Recovery from such incidents will be much slower than at higher levels of service.
- **Level of service E.** Level of service E represents conditions near to maximum flow rate with few if any acceptable gaps. The density ranges from 26.3 to close to half of jam density – up to about 41.9 passenger cars per kilometre per lane, or a space headway of 38 to 24 metres. Safe speeds have now dropped dramatically, with a safe speed of less than 60 km/h if a safe distance is maintained between vehicles. Minor incidents result in immediate and extensive queues, with degradation to level of service F, which will take a long time to recover.
- **Level of service F.** Level of service F represents breakdown conditions, and uniform flow is not maintained. Densities will be between 41.9 passenger cars per kilometre per lane and jam density (up to 80 passenger cars per kilometre per lane). The space headway will range from 24 metres to 12.5 metres, or bumper-to-bumper conditions. Upstream queues occur because the rate of vehicle arrivals exceeds throughput. This is a condition of congestion, as defined by traffic engineers and planners. To a transport economist it would represent extreme congestion.

The various regimes of level of service are shown approximately in Figure 5.7 on the speed volume curve. Table 5.1 shows the variation in conditions for the five levels of service above F for different free-flow speeds on a freeway or motorway. The different blocks of the table refer to different possible speed limits or design speeds. Considering Table 5.1, Figure 5.7 and the discussion on levels of service, it is clear that somewhere around level of service C to D probably represents the ideal operating conditions in terms of safety, service and utilisation of road space. Operation at level of service E is problematic, because of the rapidity with which operation can slip into congested conditions, with all of the negatives that the condition implies. It should be noted that a similar table to Table 5.1 for arterial streets and for lower speed

limits is not available, because, under these conditions, intersection design is usually the limiting factor. In addition, maximum speeds will depend on intersection spacing and other conditions, rendering the derivation of a table of this type much too complex to be useful.

5.5 Travel-demand forecasting

Travel-demand forecasting is an essential part of the planning that is required to develop transport policy and to inform policy makers about the implications of alternative policies. Policy without planning is much like a cart without a horse, but this is also true of planning without policy. The two are necessary elements of the policy-setting process, as was set out in Chapter 3. So, one may ask what the purpose of travel forecasting is in this context. Specifically, travel forecasting has two main purposes: first, to provide a means to understand and measure the current situation in the transport network (because its size and complexity are generally too great to allow data collection and analysis alone to paint a clear picture of what is happening); and, second, to provide forecasts of the likely future situation and of the expected response of travel demand to changes in the generalised cost of travel that would be caused by pursuit of any specific policies.

Travel-demand forecasting is based on the economic concept of ‘utility’, which can be understood as the usefulness or value that is derived from a good, service or experience that is consumed. Chapter 4 and the discussions earlier in this chapter have already proposed the idea that travel can be considered to behave somewhat like other economic goods. However, there is a difference with travel. Whereas an economic good, such as a car or a membership in a gym, usually confers some value or usefulness directly to the consumer, travel is somewhat different. With the exception of such travel as walking the dog, jogging or a possible recreational activity of taking a drive, travel is usually not desired for and of itself, but rather as a means to engage in some activity that either cannot be engaged in at home or can be engaged in with greater value elsewhere than at home. As a result, travel must be regarded more generally as a ‘derived’ demand, that is, a demand that can only be derived by knowing also the demand for activities outside the home. Similarly, the price that a person is willing to pay for his or her travel depends on how much utility is either obtained from the non-home activity or added to the activity when it is engaged in outside the home.

Unfortunately, it is generally very difficult, if not impossible, to ascertain the utility of activities, or the added utility conferred upon some good or service

by consuming it away from home. This immediately makes travel-demand forecasting a difficult proposition. The way in which the derived demand nature of travel has been handled in contemporary modelling and analysis is to use the purpose of the travel as a surrogate for utility. At the simplest level, trip purpose may be categorised to work and non-work. However, the broad category of non-work can be subdivided into such other purposes as education, shopping, recreation, social and other. In some instances, travel forecasters have used as many as 25 or more categories of trip purpose, while others may have used only four or five. Clearly, the more that trip purpose is disaggregated into narrower and narrower categories of purpose, the more accurately these categories will reflect a useful utility grouping. On the other hand, the more that purposes are disaggregated, the greater are the demands on data collection and also on an understanding of how to categorise the purposes.

The second major challenge for travel-demand forecasting is how to model the process mathematically. For some people in some situations, the decisions on why, how and when to travel may be made more or less simultaneously, such as a decision that 'We need to go shopping for groceries this afternoon by car and we will shop at such-and-such a supermarket.' Other travel decisions may be habitual, such as the decisions on travel to and from work, which may be engaged in so routinely that no decision is made on a daily basis about when to leave for work, what means of travel to use, where to go and how long to stay there. Yet other decisions might be made sequentially, for example a decision to go to the city centre to visit a museum, where, after it is decided to go and to visit the museum, a decision may next be made on what means of travel to use, and then a decision made on the route to travel. Somewhere among those decisions, a decision will also be made as to the time of day at which to go. With this mix of decision sequences and habit, the design of a modelling procedure is clearly challenging. More than 50 years ago, when travel-demand forecasting first commenced, a decision was made to model these complex decisions and habit by a sequential four-step modelling process. While many professionals have decried this over the past half-century, it still remains the principal paradigm that is used. The four steps consist of:

- trip generation – the determination of how many trips will be made;
- trip distribution/destination choice – the determination of where the travel will be made to (and, effectively, where the travel is from);
- mode choice – what means of travel to use from point A to point B; and
- route choice or network assignment – the choice among alternative paths through the appropriate network.

While there are increasingly modifications of this paradigm that may be used, especially a move towards considering activities or tours (travel from home and eventually back to home again) as the basic units for the modelling, the broad structure of the four-step models continues to be used. Trip generation may, for example, be replaced by tour or activity generation. The same will occur for the other steps in the process, but the use of a sequential modelling paradigm remains.

In addition, there should be a land use model, which would permit forecasts to be made of the location and intensity of land uses into the future. Because trip generation is usually a function of attributes of the population, and the location of population and employment, these latter two would be the outputs of the ideal land use model. However, despite numerous attempts over many years to develop integrated land use/travel-demand models, such models are rare and rarely used. Many metropolitan regions do not use any land use model, although they may forecast the location and intensity of population and employment using non-modelling procedures. There are a few cases where land use models are used, but they are usually independent of the travel-demand models, and use little or no information from them. This may be a reflection of the silo mentality of many governments with responsibility for transport and development, with each of these two functions – transport and land use development – sitting in different departments, with little or no communication between them.

It is beyond the scope of this book to describe the way in which these models are operationalised, and there are many other books that do this (e.g. Meyer and Miller 2001; Ortúzar and Willumsen, 2001), but it is important for this present context to describe a few aspects of these models. First, the models are developed from local data about socio-demographic characteristics of the population, data about land use patterns (especially the location of population and employment) and data about the supply of transport, that is, the transport networks and their levels of performance.

Second, the models were originally designed to provide forecasts of travel at the regional level and for a forecast horizon of about 20 years or so. The models were not intended to be used for local and small-area forecasts, and were also not designed to provide accurate forecasts for the short term, for example periods of three to five years into the future. There is no explicit accounting for time in the models, which are also based on the assumption that the transport system is in supply–demand equilibrium. Whether or not such an equilibrium state can or does ever exist has been the subject of considerable debate in recent years.

Third, the models, although purporting to be demand models as would be defined by a microeconomist, fail to include a fundamental property of a demand model. In fact, only the trip generation model (or its equivalent in activity and tour models) is supposedly a demand model, because the other models are actually only allocation models: trip generation estimates the total amount of travel; the other models allocate the demand to origin–destination pairs, to means of travel and to routes through the networks. This fundamental failing of the trip generation model is that it does not include any aspect of price or cost. The model is usually constructed solely in terms of characteristics of population and employment and does not include network characteristics. There is a reason for this. Because, at this point in the modelling paradigm, the origin–destination linkage is not known, nor is the means of travel or the route, only something very generic in the way of network characteristics could be included. Past attempts to incorporate, for example, a general measure of accessibility have proved to be of no value, because, in modern metropolitan areas, there is little difference from one locality to another in general accessibility. As a result, there is too little variability in accessibility for a statistical relationship to be detectable.

Perhaps the most telling aspect of these models relates to the inputs to the models themselves. First, the models rely on survey data to provide a means to understand and model the relationships on which forecasts of travel are based. However, the data pose numerous challenges for collection and are subject to enormous potential error (Stopher 2012). Traditionally, the largest of these errors arises from the reliance of most surveys of the past 60 years on self-reporting of travel by the survey respondents, in addition to errors introduced by low response rates, poor sampling, and other issues that call into question the representativeness of the data. In addition to this, however, the forecasts of travel rely on forecasts of population, employment, land use, and values of the various input variables, such as fuel prices, public transport levels of service and fares, parking costs, and so on. On the one hand, forecasts of these input variables are often as hard to make as the forecasts of travel itself. The forecasts are subject to a great deal of error. Partly, these errors stem from the methods used to generate the forecasts and partly and more importantly from the fact that the forecasts are treated as being exogenous to the transport system, where they are, in fact, endogenous. On the other hand, the forecasts of these input variables can also be used to skew the results from the travel forecasting models, so as to justify some political agenda, or preconceived idea of the desired outcome of the forecasting.

Travel-demand models are quite often misused. They are used to provide short-term forecasts, for which they were never intended. They are also used

to produce forecasts at the sub-regional level – again a use for which they were not designed. However, without good alternatives for such forecasts, it is inevitable that they will be used this way.

In summary, the models provide reasonable results for long-term, region-wide forecasting of travel, particularly under circumstances where there are no major changes in the transport system or in the socio-demographics of the population. The models are, however, likely to overestimate response to changes in the transport system and other contextual variables, because the models do not capture habit and resistance to change that are endemic in the population. The models are probably of greatest use in comparing among alternative policy directions, where they are likely to provide reasonable estimates of the relative responses among a set of alternatives, but the absolute numbers produced by the models are generally not accurate. For short-term forecasts and forecasts for sub-regional areas and corridors, the model results should be treated with considerable caution, because these are not the uses for which the models were designed. The models are based on substantial simplifications of reality, often omit important causal variables, and are based on faulty data. However, as Box and Draper (1987) stated, relating to statistical models or response surface methodology, ‘All models are wrong, but some are useful’, and this is certainly an appropriate statement about travel-demand models. They are wrong, but, if carefully designed and developed, they are still useful and can be extremely helpful in policy formulation. However, they must be used knowledgeably and carefully, with recognition of the errors inherent in any modelling.

5.6 From speed–flow to congestion costs

The traffic density and speed–flow data presented above leads directly to estimates of congestion costs. The primary impact of congestion is to increase travel times and associated time costs, but there is also an adverse impact on fuel use and on some other external costs (e.g. air pollution, greenhouse gas emissions). This linkage from speed–flow to congestion costs is illustrated below for an urban arterial road operating at a volume/capacity ratio (VCR) of up to 1. Only travel-time costs are assumed to be affected by congestion, for simplicity. BTRE (2007), in a comprehensive assessment of Australian congestion costs, estimates that \$7.1 billion of a total \$9.4 billion Australian road congestion costs in 2005 are additional travel-time costs (split almost equally between private and business time costs), so this simplified presentation is indicative of broader cost consequences.

Two speed–flow curves are used, to show how results can differ substantially as these are changed. The first speed–flow curve is the standard Bureau of Public Roads (BPR) (1965) equation (equation 5.7):

$$S = S_0/[1 + \alpha(x)^\beta] \quad (5.7)$$

where:

S = average link speed (miles per hour in this presentation)

S_0 = free-flow link speed (assumed to be 40 mph)

x = link volume/capacity ratio, where maximum flow is assumed to be 1800 vehicles/lane hour

$\alpha = 0.15$

$\beta = 4$

The second speed–flow curve, derived from Dowling and Skabardonis (2006), uses the same functional form as equation 5.7, but changes the parameter values of α to 2.248 and β to 1.584. Dowling and Skabardonis find that the latter values are a far better representation of a set of US sample urban data that they analyse than the standard BPR formula, for a number of roads with $VCR < 1$.

Figure 5.8 presents the two speed–flow curves, showing estimated speed against VCR. The standard BPR curve shows little impact of traffic volume on speed until the VCR reaches about 0.5, but, even at a VCR of about 1, the predicted decline in average speed is only from 40 to about 35 mph. The revised (or ‘fitted’) BPR curve shows a much stronger effect of increased lane volume on estimated average speed, speeds falling throughout the VCR range to a projected 12.3 mph at $VCR = 1$. This is a much stronger congestion effect.

Figure 5.9 then converts these speed–flow effects to marginal congestion (time) costs, showing the effect of an additional vehicle mile at various lane traffic volumes. A weighted average time value of \$33/hour has been used in this translation from speed–flow to flow–cost. This value was derived by taking Austroads (2008) time values for 2007 for private time (\$11.49/person hour), business time (\$36.76/person hour) and truck driver time (\$24/hour), applying Austroads’ vehicle occupancy rates (1.6 for urban private car, 1.4 for urban business car, 1.0 for truck) and then updating time values from June 2007 to May 2012 by using the increase in Australian average weekly ordinary time earnings for persons over the period (ABS 2012). A traffic mix of 90/10 for car/truck was assumed, with 75 per cent of cars being assumed to be on private use and 25 per cent business, which is broadly

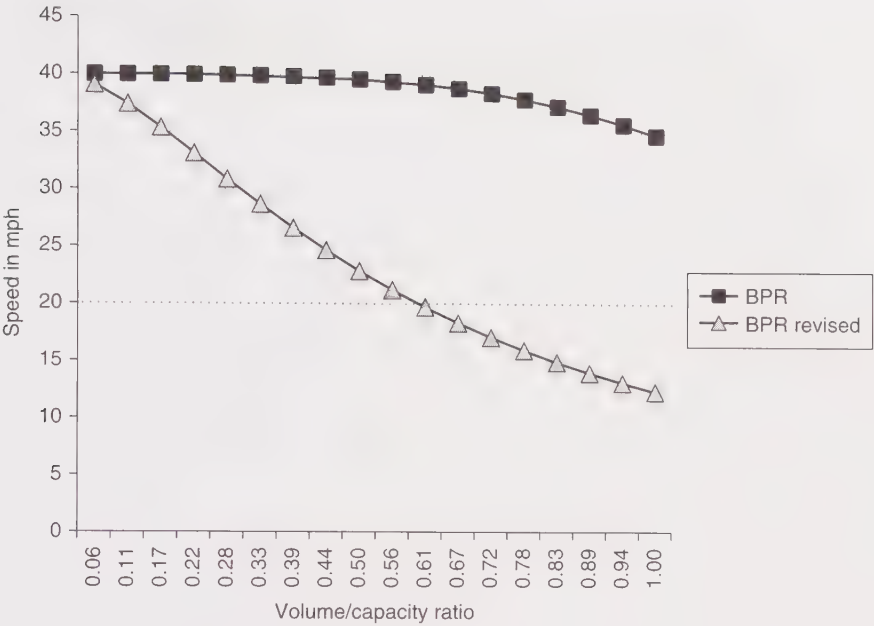


Figure 5.8 Illustrative speed versus volume/capacity ratios (for $VCR < 1$)

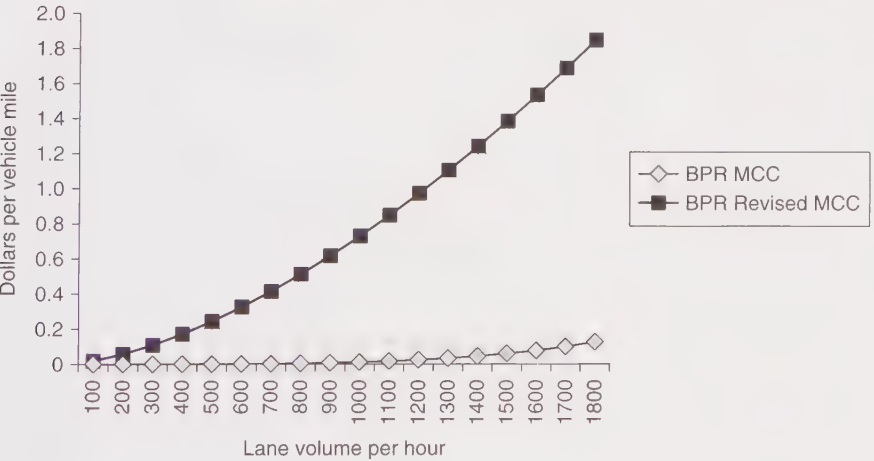


Figure 5.9 Marginal congestion (time) costs associated with speed-flow curves

representative of Australian urban operation. Figure 5.9 shows marginal congestion costs (MCC) reaching 12.7 cents per mile at $VCR=1$ for the standard BPR speed-flow curve, but being much higher for the revised ('fitted') BPR curve. In the latter case, marginal congestion costs (time only) reach

\$1.86/mile at $VCR = 1$. These calculations underline the critical importance of careful modelling of the speed–flow relationship if reliable estimates of congestion costs are to be derived.

5.7 Implications for policy

Understanding how transport network (or component) performance varies under traffic conditions is a crucial ingredient in assessing network condition and in formulating possible ideas for improvements. Performance ratings of the kind presented in this chapter are a common way to flag network or component segments that might need improvement. When potential improvements are identified it is then important to project how this improvement will affect network or component performance in traffic engineering terms and then to translate that through to economic terms, encompassing particularly user costs but also relevant capital and maintenance costs. Various important elements in this process have been identified.

Using speed–flow curves that reflect the road operating conditions in the actual networks under consideration is vital, because of the large variation in marginal congestion costs that may be associated with different speed–flow curves. The implication is the scope for large variations in potential benefit estimates. The slope of the speed–flow curve and rate of increase in congestion costs with traffic volume suggest that only small traffic reductions may be needed in congested conditions to produce significant user benefits.

The chapter has highlighted a language problem to do with the concept of ‘congestion’. The traffic engineering and transport planning definition is not the same as the economists’ definition. The former work with the concept of congestion as being at or beyond maximum flow, whereas the latter regard congestion as any reduction in speed below free or regulated speed that is attributed to traffic interruption. Policy analysts need to be aware of this distinction. However, in a policy process that includes efforts to monetise potential impacts of traffic volumes on speed (whether this is called congestion or not), the actual definition of congestion loses policy significance, and the benefits and costs of alternative possible courses of action become the focus.

NOTE

- 1 A safe speed is usually defined as the speed at which drivers can maintain a safe distance between cars, such that reaction time and braking time will permit the vehicle to be stopped without a collision with another vehicle.



REFERENCES

- ABS (2012), 'Average weekly ordinary time earnings', Australian Bureau of Statistics, Cat. 6302.0, August, available at: <http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6302.0May%202012?OpenDocument> (accessed 25 October 2012).
- Austrroads (2008), 'Update of RUC unit values to June 2007', Report IR-156/08, Austrroads, Sydney.
- Box, G.E.P. and N.R. Draper (1987), *Empirical Model-Building and Response Surfaces*, New York: John Wiley & Sons.
- BTRE (2007), 'Estimating urban traffic and congestion cost trends for Australian cities', Working Paper No. 79, Bureau of Transport and Regional Economics, Canberra.
- Bureau of Public Roads (1965), *Traffic Assignment Manual*, Washington, DC: Department of Commerce, Bureau of Public Roads.
- Dowling, R. and A. Skabardonis (2006), 'Urban arterial speed/flow curves for travel demand models (draft)', paper for presentation and publication at the Innovations in Travel Modeling Conference, Transportation Research Board, Austin, TX, April.
- Meyer, M.D. and E.J. Miller (2001), *Urban Transportation Planning*, 2nd edn, New York: McGraw-Hill.
- Ortúzar, J. de D. and L.G. Willumsen (2001), *Modelling Transport*, 3rd edn, New York: John Wiley & Sons.
- Stopher, P.R. (2012), *Collecting, Managing and Assessing Data Using Sample Surveys*, Cambridge, UK: Cambridge University Press.
- TRB (2010), *Highway Capacity Manual 2010 (HCM2010)*, Washington, DC: Transportation Research Board.

6

Social exclusion



CHAPTER OVERVIEW

The social dimension of transport policy is probably the least understood. This chapter briefly discusses social policy goals that might be relevant to transport policy and introduces some of the important concepts that need to be understood when exploring links between transport policy and social policy. These concepts include transport disadvantage, mobility, social inclusion and well-being. Some findings from recent research are summarised, to show the high value of improving mobility opportunities for people at risk of social exclusion deriving from poor mobility. The implications of the findings are explored in terms of minimum public transport service levels to support inclusion. The chapter also introduces some new research on the possible role of social enterprise business models in delivering improved mobility opportunities in regional or low-density areas.

6.1 Some context

Triple-bottom-line (economic, social and environmental) approaches to goal setting for transport are increasingly common. Economic and some environmental goals have long standing in this regard. The economic goal has received a stimulus in recent years from work on agglomeration economies. Environmental goals in transport have received elevated status because of the current focus on climate change. The social dimension is the least developed. This applies with respect both to impact assessment (ex ante and ex post) of specific transport policies, programmes and/or projects and to the generation of such policies, programmes and/or projects specifically to target social goal achievement. The lack of any clear definition of just what is/are the social goal/goals of transport is a major reason for this situation.

Markovich and Lucas (2011) highlight part of the problem about the social dimension of transport, pointing out that there is ambiguity about what constitutes a social and/or equity impact and distributional effect. By extension, there must be similar ambiguity about the social goal of transport. Markovich and Lucas review available literature on what they call social and

distributional impacts of transport, including many impacts that might initially be classed as environmental (such as emissions and noise). They follow the Guers et al. (2009) comprehensive definition of social impacts: ‘changes in transport sources that (might) positively or negatively influence the preferences, well-being, behaviour or perception of individuals, groups, social categories and society in general (in the future)’ (Guers et al. 2009, p. 71).

This approach mirrors the approach taken by economists when seeking to measure ‘social costs’, whereby monetary values might be imputed to (for example) emissions and noise. In our view this definition is too broad to be helpful in terms of providing clarity about the social goal(s) of transport, although it is helpful in terms of defining the scope for what might need to be considered in a policy analysis or evaluation, in line with the individual preferences value judgement (see Chapter 3).

When they are considered, social drivers of transport initiatives tend to cluster around a few particular issues, such as public transport fare concessions to assist low-income earners, physical access to public transport for those with a disability, the fairness or equity of various road pricing arrangements such as congestion taxes and, more recently, transport measures to assist access for people at risk of social exclusion. The growing interest in transport and social exclusion, in particular, carries a perhaps implicit suggestion that promoting or fostering social inclusion is an appropriate social policy goal for transport. For example, the Norwegian *National Transport Plan 2010–2019* states that it is about (*inter alia*) ‘making society more inclusive and universally accessible’ (Norwegian Ministry of Transport and Communications 2009, p. 3). The ways to realise inclusion are not spelt out, except perhaps via improved infrastructure for pedestrians and cyclists and universal design features. Similar language and policy intentions are reflected in many transport plans at a national level and/or at subsidiary levels (e.g. provinces/states and/or local authorities), depending on where jurisdictional transport responsibilities lie. Such topics are worthy issues to pursue but cover only a part of the place of transport in achieving broad social policy goals.

Current interest in connections between transport and social exclusion largely derives from the important pioneering work undertaken in the UK by the then Social Exclusion Unit. In its report on transport, it identified five groups of transport-related barriers to social inclusion (SEU 2003):

1. the availability and physical accessibility of transport;
2. the cost of transport;
3. services located in inaccessible places;

4. safety and security, or fear of crime; and
5. travel horizons – people on low incomes were found to be less willing to travel to access work than those on higher incomes.

The SEU suggested a number of ways in which these barriers could be tackled. These identified barriers are important issues of themselves, but remain as 'topics' rather than a cohesive structure or framework leading to social goal achievement through transport.

Social exclusion and its link to well-being is a more useful lens through which to view the social goal in transport. A number of authors are now beginning to research this area (see, for example, Mollenkopf et al. 2005; Spinney et al. 2009). A major recent Australian study has explored links between mobility, transport disadvantage, social exclusion and well-being (see, for example, Currie 2011; Stanley et al. 2011). Several well-accepted measures of well-being were used in that research, as set out in Table 6.1. That table reports average scores reported for a sample of Victorian residents who exhibited none of a possible five indicators of risk of social exclusion, compared to scores for those respondents who had three or more of the five possible risk factors (these five risk factors are elaborated later in this chapter).

Table 6.1 Well-being measures for groups with three varying levels of social exclusion

Well-being measure	Three or more SE risk factors (N=139)	Total sample (N=1019)	No SE risk factors (N=355)
Personal Wellbeing Scale (range 1–10)	5.5	7.1	7.7
Satisfaction with Life Scale (range 1–7)	3.8	4.9	5.4
Positive affect (range 1–5)	3.3	3.5	3.7
Negative affect (range 1–5)	4.8	1.8	1.7

Table 6.1 shows that the well-being of those with three or more risk factors of social exclusion was considerably lower than for the sub-group with no social exclusion risk factors, on both the Personal Wellbeing Scale and the Satisfaction with Life Scale. Z-tests indicated that the differences are significant at the 1 per cent level. For the Satisfaction with Life Scale, the world average sits at 70 per cent $\pm$ 5 per cent (Cummins 2011), which was the average for the full sample in this study. The percentage score for the high-risk group in this survey is well below this level, at 54 per cent, suggesting those with three or more social exclusion risk factors have decidedly lower well-being.

For positive affect (PA) and negative affect (NA), the measurement scales ranged from 1 to 5 (Table 6.1). Those with higher levels of risk of social exclusion had slightly lower positive affect and a much greater level of negative affect than those exhibiting no risk factors. Z-tests again indicated that the differences are significant at the 1 per cent level, again suggesting that people with a high risk of social exclusion have a lower quality of life.

In the Australian research, those with three or more risk factors also reported more frequent difficulties accessing activities as a result of transport problems, suggesting a link between transport and risks of social exclusion. More detailed econometric analysis confirming these connections is reported in Stanley et al. (2011) and summarised in section 6.5. In short, the mobility, social exclusion and well-being linkage is a useful way to explore the social leg in a triple-bottom-line approach to transport, with well-being being the ultimate social goal and the transport/social inclusion connection a key pathway.

Lucas (2012) notes that the transport/social exclusion research agenda has grown considerably over the past decade. This growth now includes a stronger focus on well-being (for example, Mollenkopf et al. 2005; Spinney et al. 2009; Ettema et al. 2010) and on links between transport, social exclusion and well-being (Stanley et al. 2011).

This chapter gives a brief overview of social goals and summarises some recent work that tries to better understand the place of transport in achieving social policy goals, through links to social inclusion and well-being. The inevitable conclusion, however, is that the transport field has yet to fully understand and define the full set of social policy goals where transport plays a role.

6.2 Some definitions

A number of concepts are critical to understanding this subject area. In shorthand terms, our understanding of the major concepts discussed in this chapter is as follows:

- **social exclusion** = the existence of barriers which make it difficult or impossible for people to participate fully in society;
- **accessibility** = the ease with which a person, from a particular place, can get to particular services, locations and/or other people;
- **mobility** = the ease with which a person moves around;
- **social capital** = the benefit a person derives from social networks, trust and reciprocity within a community;

- **transport disadvantage** = a situation where people experience a shortage of transport options and/or have restricted abilities to use available options, which restricts their mobility and hence their access to goods, services and relationships;
- **sense of community** = the strength of a person's sense of attachment to where he or she lives; and
- **well-being** = a person's rating of his or her quality of life.

Social capital and sense of community, together with mobility (and some other factors, such as household income), are significant contributors to a person's risk of being socially included or excluded, and this, in turn, has a significant impact on well-being (Stanley et al. 2011).

6.3 Social policy goals

Well-being is commonly seen as the target goal for social policy and is achieved when human needs are met. Human needs were defined by Maslow (1954) and, while there are variations on these, they still stand strong. Maslow defined these needs as: physiological – the ability to breathe, have food and water, and sleep; safety – health and freedom from violence; love and belonging – friendship, family, and sexual intimacy; esteem – self-esteem, achievement, confidence and respect; and self-actualisation – morality, creativity, acceptance and lack of prejudice.

To achieve these needs and maximise well-being, certain conditions are required. These conditions enable people to obtain needs directly, such as forming friendships, as well as facilitating the means to achieving needs. For example, a minimum level of income is required to purchase food and shelter, and therefore people need the capacity to generate income, which may require a job and the means to access the job. This is described in the idea of a person having capabilities to achieve goals (Sen 1992). Capabilities denote a person's opportunity and ability to generate well-being, taking into account relevant personal characteristics and skills, external factors and societal resources, such as the provision of infrastructure and services.

Transport can play a role in both the direct achievement of good outcomes for people (e.g. having the ability to travel to see a friend) and indirectly, through assisting the achievement of intermediate goals, such as enabling the procurement of income to meet basic physiological needs. However, the ability or capability to be mobile is just one component of the mix that facilitates well-being. The relative importance of transport is yet to be fully defined

and understood. This is not an easy task, as those with low well-being or high levels of social exclusion commonly experience multiple disadvantages along with transport disadvantage, with many adversities having compounding or cumulative effects.

6.4 Transport disadvantage

We have defined transport disadvantage as a situation where people experience a shortage of transport options and/or have restricted abilities to use available options, which (in turn) restricts their mobility and hence their access to goods, services and relationships. Transport disadvantage may arise from a single reason but is more likely to be from a number of linked issues and to occur with people who are at risk of social exclusion. The causes of transport disadvantage can be classified into three broad categories (Stanley 2011):

1. **Institutional barriers or facilitators.** This would include factors like the first and third barriers identified by the SEU – the availability/ accessibility of transport and/or the absence of particular services, such as medical or educational services. These issues are commonly spatially based.
2. **Individual barriers or facilitators.** These include issues such as personal characteristics (e.g. personality attributes, health, language and culture).
3. **External impacts.** This includes broad trends external to the personal and structural conditions, such as climate change and environmental conditions, population growth, the political environment, government ideology and policy, and international economic trends, such as recession.

It is possible (for example) for people to be able to move around freely (have a high level of mobility) but not have good access to shops and personal business outlets such as banks, because of the lack of such services near where they live or the lack of transport options to reach the services where they are available. Equally, a person may live near a bank but not be able to access this facility because of a personal disability and uneven footpaths that prevent wheelchair travel. Transport disadvantage can reduce a person's mobility and/or access to goods, services and relationships, and thereby inhibit the achievement of social inclusion and well-being. These examples underline the important role played by accessibility in supporting social inclusion and draw attention to the roles played by both transport and land use in providing access to opportunities.

Spatially based research on transport disadvantage frequently uses distance-based, time-based or generalised cost-based measures of relative accessibility to particular services (such as jobs or health services) to suggest relative degrees of transport disadvantage. Separate measures are sometimes estimated for cars and public transport, typically showing the relative disadvantage experienced by those without car access.

Research on transport disadvantage usually focuses on particular groups thought most likely to be in this situation, such as older persons, youth, unemployed people, and people with a disability (see, for example, Mollenkopf et al. 2005; Stanley and Stanley 2007; Spinney et al. 2009; Currie et al. 2010). People experiencing risk of social exclusion may be transport disadvantaged, but not all people who are transport disadvantaged are necessarily at risk of social exclusion. For example, some high-income people may choose to live in attractive rural environments and be prepared to put up with poor, or no, public transport services. This is likely to mean they are frequently required to act as chauffeurs for others (such as their basketball-playing or dancing adolescent children), and there will be times that their children will not be able to pursue some of their interests that require travel. The children in this case are likely to be relatively transport disadvantaged. Care is thus needed not to confuse the two concepts: mobility-related social exclusion is more relevant for social transport policy. Most people in this situation are also likely to be transport disadvantaged, but not all transport-disadvantaged people are at risk of mobility-related social exclusion.

6.5 Mobility, social inclusion and well-being

Stanley et al. (2011) have attempted to establish connections between mobility (measured as trip making and aligning with engagement in activities), social exclusion (measured as the number of social exclusion risk thresholds, out of the possible five listed below, exhibited by people) and well-being (measured by the Personal Wellbeing Index, or PWI) in an integrated model. Their indicators for a person's risk of social exclusion are:

- **household income** – less than a threshold of \$A500 gross per week;
- **employment status** – not employed, in education or training, not looking after family nor undertaking voluntary activities;
- **political activity** – did not contribute to or participate in a government political party, campaign or action group to improve social or environmental conditions, or to a local community committee or group in the past 12 months;

- **social support** – not able to get help if needed from close or extended family, friends or neighbours; and
- **participation** – did not attend a library, sport or exercise event (participant or spectator), arts or cultural activity, hobby, leisure and/or interest group in the past month.

Separate models were estimated for the Melbourne metropolitan area (a city of 4 million people) and for a regional area in Victoria (population about 100 000, spread between four main urban centres and rural areas). The first part of the model sought to explain risk of social exclusion as a function of a person's household income, trip making, social capital, and age; the second part, in turn, sought to explain people's well-being as a function of their risk of social exclusion and a range of measures of psychological well-being, together with their sense of community.

The Melbourne analysis suggested that a lower relative risk of social exclusion is associated with people having:

- contact with members of their close family more frequently than once a year but less than once a month;
- contact with members of their extended family;
- trust in people in general;
- relatively higher income;
- relatively higher trip rates; and
- higher levels of extroversion.

In the second part of the model, people's personal well-being (as measured by the PWI) was found to be likely to be greater:

- the lower their risk of being socially excluded;
- the greater their sense of attachment to the community;
- the greater their sense of mastery over their environment;
- the greater their positive relationship with others;
- the more positive their self-acceptance; and
- the greater their age.

Social capital variables were not significant in the regional model, but sense of community remained a significant contributor to personal well-being, with the coefficient suggesting a stronger link than in the Melbourne results.

The research suggests that, the lower a person's level of realised mobility (and hence the fewer activities in which the person is likely to engage),

the higher the likelihood that the person is at risk of social exclusion, particularly if that person is also socially disadvantaged. Research supports a conclusion that undertaking travel may improve a person's likelihood of social inclusion and his or her well-being. This might occur directly and/or through a mediating influence of building social capital and connection to the community. While personal characteristics (for example, *locus of control* and *affect*) are related to the uptake of these activities, it would seem that, without the ability to be mobile, many opportunities simply cannot be taken up.

This research confirms the important mediating factors of networks and connections to the community for social inclusion and self-rated well-being, which in turn confirms the importance of the ability to have mobility. In Sen's (2009) terms, this suggests a role for mobility (as a means of achieving accessibility to people and activities) as an important *capability* that should be pursued through transport (and social) policy.

Household income and trip making were significant influences on risk of social exclusion in both the Melbourne metropolitan and regional Victorian models. The relative coefficient estimates can therefore be used to impute the value of an additional trip. The implied value of an additional trip was A\$24.40 (2008 prices) in the Melbourne study and a very close A\$19.40 for an additional regional trip, in both cases being valued at the relevant sample mean household income (Stanley et al. 2011). Modelling included household income squared as the best way to express that variable, which means the value of an additional trip is higher for those with lower household incomes (halving household income doubles the value of an additional trip). This aligns closely with relative equity weights that were discussed in Chapter 3 for a utility function of the form $U = \log C$ (where U = utility and C = consumption).

These are high monetary values, which can be used to value the benefits of increased trip making as it contributes to reducing the risk of social exclusion. The values essentially mean that, in terms of seeking to reduce social exclusion, facilitating another trip by an 'at-risk' person is effectively equivalent to giving him or her about A\$20. Application of the Melbourne value (A\$24.40) in a case study of the benefits of that city's route bus system suggested that social inclusion benefits are the largest single benefit from those services, being greater than the total cost of service provision (Stanley and Hensher 2011). Congestion-cost savings were the second-largest benefit. The high unit trip value was assessed as being relevant to about one in three trips on the Melbourne bus network.

A very important implication of this research is that transport policy initiatives that seek to reduce the external costs of car use, such as greenhouse gas emissions, should focus on trip lengths rather than trips, particularly for people likely to be at risk of social exclusion. Trips have high value because of the activities they facilitate. The primary aim should be to enable this value to be realised with shorter travel distances. This underlines the importance of the land use/accessibility dimension.

The unit trip value derived in the Australian research was not mode-specific. An implication is that, in regional and rural areas, for example, where public transport service availability is usually thin and car dependence high, additional travel (by whatever means) by people at risk of social exclusion has high value. This may be travel that relies on a lift giver in a private car. The high value has strong intuitive sense in that setting. Equally, transport improvements that enable cycling where it was not possible before may also activate high values for new trips by 'at-risk' people.

6.6 Minimum service levels

One important policy implication of this research for urban areas is that an important way to reduce risks of social exclusion that have their origins in mobility is to ensure that there is a reasonable base level of public transport service available. A reasonable service level is one that enables most people to access most of the things they need for a fulfilled life at most times.

Enabling all people to access all things at all times might be an aspiration but will never be better than that! The appropriate service level in any particular circumstances will depend, *inter alia*, on the land use/transport/demographic circumstances of particular locations, but should be embedded in integrated land use/transport plans for cities and regions.

In low-density cities, minimum public transport service levels on the fringe are particularly important to support social inclusion, especially when the interaction of housing and transport markets means low-income households are 'forced' to the fringe in pursuit of affordable housing. This is the reality in Australia's major cities in 2012. Affordable housing does not align with affordable living, because of the increased transport costs (and reduced opportunities) on the fringe. Minimum public transport service levels can provide a bridge to reduced risks of social exclusion and to improved well-being.

Analysis of public transport patronage levels on services of different standards can help identify possible minimum acceptable service levels. Loader

and Stanley (2009) analysed patronage growth rates for Melbourne bus routes by area of operation/service standard, after a substantial increase in fuel prices. Their conclusion was that only routes with reasonable service levels (operating seven days a week with at least an hourly frequency for 15 hours a day on the fringe of Melbourne, but preferably a half-hour minimum frequency) will be successful at attracting new patronage, even when general conditions are likely to favour public transport growth.

The idea of minimum service levels can be given an economic interpretation. Figure 6.1 shows conventional demand and supply curves for public transport travel (measured in terms of seats/time period). Equilibrium is initially at OQ_1 seats at price OP_1 . If a 'needs analysis' suggests that the minimum service level should be fixed at supply curve $Supply_2$ (with OQ_2 seats/time period provided), this is effectively arguing that the social value of public transport seats is reflected in a social demand curve $Demand_2$, which reflects a measure of socially assessed value, rather than the privately expressed demand curve $Demand_1$.

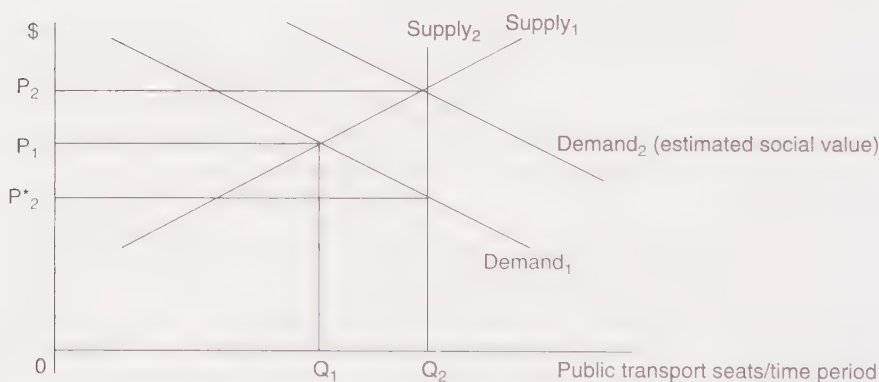


Figure 6.1 Service allocation by need of minimum service level

6.7 Regional mobility policy: a social enterprise model

In regional or rural areas in developed countries like Australia, Canada and the US and in Europe, public transport service levels are typically lower, travel distances longer and car reliance higher than in metropolitan areas. For those without ready availability of a private vehicle, mobility opportunities can be severely restricted. The Australian research summarised above, for example, indicates:

- the importance of mobility for social inclusion and well-being in regional areas;
- the greater difficulties of travelling in regional areas;
- the lesser availability of public transport in regional areas; and
- the high value of additional trip making by those at risk of social exclusion in regional areas.

Regional groups more likely to be both transport disadvantaged and at risk of social exclusion from a mobility origin include young people (especially those living out of town and reliant on lift giving), older folk, people with a disability, the unemployed, and those on a low income. The activities they are less likely to be able to undertake are most commonly associated with visiting family and friends, engaging in recreation, and so on. These are activities that build social capital, social inclusion and personal well-being and thereby reduce future costs associated with exclusion, such as welfare and mental health costs. Public policy initiatives that deliver cost-effective improvements in regional personal mobility are likely to be valuable both to individuals and to the community generally. Where the beneficiaries are likely to be at risk of social exclusion, this value is likely to be particularly high, provided mobility solutions are provided cost-effectively.

In regional areas, there are often a number of publicly supported transport opportunities, from regular public transport services and community transport services (CT) to school buses and subsidised taxis. However, regional mobility services are frequently restricted in availability by:

- **regulation:** for example, which may define the catchment area for student use of school buses;
- **institutional restrictions:** for example, which may limit carriage of other passengers on school buses, or restrict availability of various community transport options to particular client groups, such as people with certain physical conditions or of certain ages;
- **provider attitudes:** for example, where some community groups with their own (often publicly provided) vehicles may be unwilling to make them available for use by others; and
- **restrictions** in use as a result of funding arrangements.

A particular concern with government funding of regional mobility solutions is that these solutions operate almost entirely in silos. Silos reflect traditional functional administrative frameworks and encourage behaviour that protects territory and self-interest, whereas accessibility or mobility problems arise on a place basis, which cuts across functions. Administrative systems have

often adapted inadequately. Place-based approaches are likely to be more effective, cutting across functional silos.

The political response to transport funding constraints is usually to exhort 'doing more with less'. In many environments this is defensible, and the constraints outlined above suggest there may be some opportunities in regional or rural areas in this regard. Different ways of providing regional or rural mobility services should be explored, to maximise the public value that is achieved from those services, with a particular focus on efficiently meeting the needs of transport-disadvantaged people who are at risk of social exclusion.

Matching (1) those with mobility needs with (2) potential mobility providers is likely to be most successfully achieved if both sides have a sense of ownership of the problem. Given the significant role played by attachment to community in regional well-being (Stanley et al. 2011), social enterprise-type business models for tackling mobility-related social exclusion at local levels, involving service providers, people needing transport, client agencies (e.g. hospitals and welfare agencies) and volunteers, are likely to be an effective way forward for reducing regional transport disadvantage and improving social inclusion and well-being. This requires local empowerment, ideally with strong positive support from higher levels of government, including some freeing up of any existing funding flows to enable local decision taking on priorities.

A social enterprise is a business with mainly social objectives, whose surpluses are primarily reinvested in pursuit of those objectives, in the business or community. It draws on the entrepreneurial spirit of the business sector, using this for community purposes. This is likely to work well in a region with a strong sense of community and committed community leadership. A social enterprise model in such a region should be able to:

- integrate better the range of existing regional mobility opportunities and leverage community development more broadly in the process, to improve social capital and sense of community, reduce social isolation and improve well-being;
- make better use of existing community mobility resources (e.g. vehicles, drivers, volunteers), capturing synergies across agencies and increasing specialisation and coordination in service planning and delivery, resulting in more efficient and effective client service; and
- deliver increased transport opportunities to a wider range of people, particularly those at risk of social exclusion from mobility origins.

In a regional setting, the objectives for a mobility-oriented social enterprise should include:

- improved transport or mobility opportunities for those whose needs are not presently met by mainstream public transport;
- better understanding of mobility needs and current options, leading to more cost-effective transport arrangements; and
- provision of employment or training opportunities for some people and volunteering opportunities for others (e.g. driving, office, website, etc.).

Key components of such a social enterprise model should be (1) removal of administrative and governance barriers between transport modes and (2) making the needs of the travelling public the central issue. The social enterprise model may not be able to offer a number of travel alternatives for all regional or rural transport-disadvantaged people, including those at risk of social exclusion, but it should have a greater chance of doing this than current agency-based transport arrangements. A case study of a social enterprise model is currently being trialled in the southwest Victorian (Australia) town of Warrnambool (population 35 000).

6.8 Some other 'social' matters

Markovich and Lucas (2011) discuss a range of 'social' issues that are associated with transport. This chapter has taken a narrower approach to 'social' questions, believing that this gives more clarity to the elements of the triple bottom line. However, it is important to acknowledge that there are a number of important transport impacts of a social nature, discussed by Markovich and Lucas (2011), which are very important for transport policy. In this regard we highlight the following:

- The adverse impacts of transport initiatives (such as traffic noise and local air pollution) tend to have a disproportionate impact on people who are already relatively less well off, many of whom are at greater risk of social exclusion. This is something of a double whammy and is an important equity policy question.
- Transport facilities can sever communities, reducing social capital and sense of community in the process. A likely consequence is reduced well-being.
- Transport safety issues might be considered as part of the social dimension, although the approach to defining 'social' pursued in this chapter leaves safety as a fourth leg of the goal-setting process (economic, environmental, social, health and safety), not part of the social goal.

6.9 Policy directions

Transport opportunities affect the extent to which people are able to participate in their society (i.e. social inclusion or exclusion), which affects well-being. The connection between mobility, risk of social exclusion, and well-being underlines, in our view, the primary nature of the social goal that should be pursued through transport policy and why that social goal is important. However, it is still very early days in understanding of the social role played by transport, and much remains to be done to understand the linkages between transport and well-being.

Implementation of minimum public transport service levels, which will vary by location, can play an important role in tackling problems of social exclusion, supported by a wider range of measures to ensure transport opportunities are available, accessible (including dimensions of personal safety and security, able to meet the needs of persons with a disability, etc.) and affordable. In regional and rural areas, new governance models on service delivery are worth exploring. This chapter has suggested implementation of a social enterprise approach to reducing mobility-related social exclusion in rural and regional areas. Successful implementation will require a funding framework that encourages the social enterprise approach and a light touch to regulation, rather than excessive bureaucratic control. This is about empowering local communities to solve their own problems in a supportive environment.

In both urban and regional settings, land use/transport policy integration should recognise the multiple dimensions of accessibility, such that social inclusion and well-being can be promoted by improving transport opportunities, improving the local availability of activities, or cost-effectively improving elements of each. This ties social inclusion and well-being firmly back to land use/transport integration in the longer term.



REFERENCES

- Cummins, R. (2011), 'What makes us happy? Ten years of the Australian Unity Wellbeing Index: Survey 25.0', Deakin University, Melbourne.
- Currie, G. (ed.) (2011), *New Perspectives and Methods in Transport and Social Exclusion Research*, Bingley, UK: Emerald Group Publishing.
- Currie, G., A. Richardson, P. Smyth, D. Vella-Brodrick, J. Hine, K. Lucas, J. Stanley, R. Kinnear and J. Stanley (2010), 'Investigating links between transport disadvantage, social exclusion and wellbeing in Melbourne – updated results', *Research in Transportation Economics*, **29** (1), 287–95.
- Ettema, D., T. Garling, L. Olsson and M. Friman (2010), 'Out-of-home activities, daily travel, and subjective well-being', *Transportation Research Part A*, **44** (9), 723–32.
- Guers, K.T., W. Boon and B. van Wee (2009), 'Social impacts of transport: literature review and

- state-of-the-practice appraisal in the Netherlands and the United Kingdom', *Transport Reviews*, **29** (1), 69–90.
- Loader, C. and J. Stanley (2009), 'Growing bus patronage and addressing transport disadvantage – the Melbourne experience', *Transport Policy*, **16** (3), 106–14.
- Lucas, K. (2012), 'Transport and social exclusion: where are we now?', *Transport Policy*, **20** (1), 105–13.
- Markovich, J. and K. Lucas (2011), 'The social and distributional impacts of transport: a literature review', Working Paper No. 1055, Transport Studies Unit, School of Geography and the Environment, Oxford University.
- Maslow, A. (1954), *Motivation and Personality*, New York: Harper.
- Mollenkopf, H., F. Marcellini, I. Ruoppila, Z. Szeman and M. Tacken (eds) (2005), *Enhancing Mobility in Later Life: Personal Coping, Environmental Resources, and Technical Support. The Out-of-Home Mobility of Older Adults in Urban and Rural Regions of Five European Countries*, Amsterdam: IOS Press.
- Norwegian Ministry of Transport and Communications (2009), *National Transport Plan 2010–2019*, available at: http://www.regjeringen.no/upload/SD/Vedlegg/NTP/Binder1ntp_engNY.pdf (accessed 10 April 2012).
- Sen, A. (1992), *Inequality Re-examined*, London: Oxford University Press.
- Sen, A. (2009), *The Idea of Justice*, London: Penguin Books.
- SEU (Social Exclusion Unit) (2003), *Making the Connections: Final Report on Transport and Social Exclusion*, available at: <http://www.socialexclusionunit.gov.uk/publications/reports/html/transportfinal/summary> (accessed 23 May 2012).
- Spinney, J.E.L., D.M. Scott and K.B. Newbold (2009), 'Transport mobility benefits and quality of life: a time perspective of elderly Canadians', *Transport Policy*, **16** (1), 1–11.
- Stanley, J. (2011), 'Social exclusion', in G. Currie (ed.), *New Perspectives and Methods in Transport and Social Exclusion Research*, Bingley, UK: Emerald Group Publishing, pp. 27–44.
- Stanley, J. and D. Hensher (2011), 'Economic modelling', in G. Currie (ed.), *New Perspectives and Methods in Transport and Social Exclusion Research*, Bingley, UK: Emerald Group Publishing, pp. 201–22.
- Stanley, J. and J. Stanley (2007), 'Public transport and social exclusion: an operator's perspective', in G. Currie, J. Stanley and J. Stanley (eds), *No Way to Go: Transport and Social Disadvantage in Australian Communities*, Clayton, VIC: Monash University Press, 13.1–13.17.
- Stanley, J.K., D. Hensher, J.R. Stanley and D. Vella-Brodrick (2011), 'Mobility, social exclusion and well-being: exploring the links', *Transportation Research A*, **45** (8), 789–801.

7

Tackling the externalities – environment



CHAPTER OVERVIEW

The formulation of public policy is largely a matter of dealing with the externalities of the provision of goods or services, such as transport. This is the first of several chapters that discuss specific externalities, as a way to understand the issues that face the formulation of policy. In this chapter, environmental externalities are the focus, with three examples being used, following a more general discussion of governmental response to environmental externalities. The first example is that of air pollution from transport, which examines the major pollutants and how they are produced. The second example is greenhouse gas (GHG) emissions, and examines the role of greenhouse gases, their production and policies that may help to reduce them. The third example is noise. Again, the discussion looks at the properties of transport-related noise and at mitigation strategies. The chapter concludes with a section on valuing environmental externalities for use in a cost–benefit analysis.

7.1 Historical background

It is probably true to say that environmental externalities are those externalities that have been recognised the longest (apart, possibly, from congestion) and that have been the subject of legislation longer than any other similar externality, and also that environmental laws for transport externalities are some of the most ancient environmental laws. Possibly the oldest environmental law in existence may be one from China, relating to the use of natural resources, and passed in the Qin dynasty in about 200 BC (Sanft 2010). A more recent one, specifically relating to transport, was a law passed in Rome in about 50 BC which forbade the driving of chariots through certain residential areas after 9 p.m. The Emperor Justinian (c.482–565) had environmental laws passed that related to protection of the seashore. Some later non-transport-related environmental laws were a part of Magna Carta of 1215. The next reference to environmental law appears in the early 1800s,

when, under Napoleon, laws were passed restricting sewage disposal from Paris to certain areas outside the Paris environs. Since that time, there have been many other environmental laws. However, one of the noticeable things about environmental laws is that they tend to be a feature of well-developed economies and are less common in developing economies. In other words, environment often takes a back seat whilst economic development is the leading issue, but tends to move to the fore as nations become wealthier and their economies become well developed.

In modern times, the Western world experienced an environmental awakening that seems to have begun in the late 1950s or early 1960s. This was coupled with increasing personal freedoms in many Western countries and an awakening to the idea that government was not always right and did not always make decisions that were in the best interests of the people. Added to this, the technological revolution of the 1960s onwards started to make it increasingly possible to measure the effects of environmental degradation.

In transport, environmental awareness seems to have surfaced most clearly with the case of the Embarcadero Freeway in San Francisco. This freeway was originally planned in 1955 and would have been an elevated double-decker freeway running along the Oakland Bay shoreline from south of San Francisco, eventually linking to the Golden Gate Bridge to the north of San Francisco. While there were already major concerns raised about this freeway in the late 1950s, and the San Francisco Board of Supervisors voted against this and six other freeways in the San Francisco region in 1959, a revised plan was put forward by the California Department of Highways in 1964. This revised plan led to a public protest by more than 200 000 people, which was successful in stopping the construction of the Embarcadero Freeway, part of which had already been built (at both the north and south ends of the planned freeway). This freeway revolt was perhaps the most visible and most dramatic of what was emerging as a series of revolts in various states of the USA, as well as revolts in other countries, most notably the UK. In the US, as a partial response to the increasing activism against freeway building, the Federal Highway Administration (FHWA) promulgated Policy and Procedure Memorandum 20-8 (PPM 20-8), which set out a two-hearing process required for any highway involving federal funds and also listed a large number of environmental effects that were to be taken into account in the process of planning and designing a federally funded highway. The first hearing was to take place at a corridor level of planning, to select among potential alternative corridors for the highway, and the second hearing was to take place to select an alignment within the preferred corridor.

In effect, PPM 20-8 mandated public involvement in the process leading to the design and construction of a new highway and also mandated that a lengthy list of environmental impacts, including those on air and water quality, noise, aesthetics, neighbourhood cohesion, flora and fauna, and so on, must be considered in the planning and design process. Later in the same year – 1969 – the US Congress passed the National Environmental Policy Act (NEPA), which was followed in 1970 by the Clean Air Act Amendments (CAAA), which also affected transport quite extensively.

These federal acts in the US affected only projects that were partially funded by the federal government. As a result, for projects funded by US states, each state, over the next decade or so, passed its own environmental legislation, which dictated similar requirements to those of NEPA. In some cases, state law was more stringent than the federal law, while in other cases it was less stringent. However, each state's legislation was largely patterned after the federal legislation. Similar legislation was passed in Canada and New Zealand in the early 1970s, and in Australia in the late 1970s. The UK and most Western European countries followed in the 1980s. Similar legislation has now been passed in many other countries around the world. Interestingly, most countries have modelled their environmental legislation on the US NEPA, although various local considerations have involved changes to the legislation that has been passed by each country.

Australian environmental legislation

It is interesting and useful to compare and contrast the US and Australian legislation on environmental impacts. This is particularly interesting because both countries have a federal government and individual state and territory governments, so that it is easy to compare and contrast the way in which the legislation has been enacted. Under NEPA, the first step in the process is to issue a notice of intent (NOI), which outlines the project that it is intended to plan and for which environmental assessment will be undertaken. The NOI may result in a categorical exclusion if it is determined that the proposed project will not affect the environment or result in consumption of scarce resources that might have been used for other actions or projects. This means that no further environmental studies are required. In Australia, federal legislation requires first a determination as to whether the impacts of a project will be indiscernible or not. If they are, no further action is required on environmental grounds. This is determined by setting a threshold of environmental significance, which is the responsibility of the minister in charge of overseeing the environmental legislation. All projects that are deemed to fall below this threshold require no reporting on environmental issues.

However, if there are expected to be impacts of environmental significance, then an NOI is the next step that must be completed. This documentation is similar to that under the US legislation.

The next step under NEPA is to perform an environmental assessment (EA). The purpose of the EA is to establish whether the environmental impacts will be significant or not. In the latter case, a finding of no significant impact (FONSI) is issued. If, however, the EA determines that there will be significant impacts, then an environmental impact statement (EIS) must be prepared. This document must report on all potential environmental impacts of at least two alternatives – a no-build or do-nothing alternative and the ‘locally preferred alternative’ (LPA) – and additional alternatives that could meet the purpose and need should also be documented in the EIS. It is a requirement under NEPA that the documentation not only assesses the environmental impacts, but also documents the steps that will be taken to mitigate these impacts, or else documents that the impacts cannot be mitigated.

In Australia, the situation is a little different. There is a second threshold, this time between acceptable and unacceptable impacts. If a project is deemed to fall below the threshold of unacceptability, then a public environment report (PER) is required. This will document the assessment of environmental impacts and show that those impacts that will occur are below the threshold of unacceptability. However, in the event that there will be serious impacts or dangerous impacts, these being the categories of impacts that fall above the threshold of unacceptability, then an EIS must be prepared.

In the USA, if a project involves both federal and state money, the environmental documents that are prepared are required to meet the legislative requirements of both the state and federal jurisdictions. In the event that all state requirements are equal to or less stringent than the federal, then the one document will usually suffice. In the event that state requirements are more stringent than federal, the document will need to deal with each of these sets of requirements. In contrast, in Australia, where a project might come under both federal and state or territory jurisdiction, or where a project involves more than one state or territory, a memorandum of understanding (MOU) is prepared in which it is documented as to which jurisdiction’s laws will be followed, and only one set of laws is then required to be followed.

A further contrast between the US and Australian legislation has to do with the documentation of comments and responses received through the public hearing processes. In the US, it is required that all comments received on the environmental documentation, as the planning process proceeds, must be

documented and become a part of the environmental documentation, along with the responses to each and every comment received. In Australia, there is no such requirement. The environmental document will be modified in response to comments received, as and where necessary, but the comments and responses themselves do not become a part of the overall documentation of the PER or EIS.

7.2 Environmental impacts

Legislation such as NEPA lists a large number of potential environmental impacts of transport projects. It is important to note that impacts occur not only from the constructed project, but also during construction itself, and these latter impacts may be more severe than those arising from the operation and maintenance of a project. This entire book could be devoted to discussion of these various impacts. However, this is not the purpose of this book. Instead, we use three areas of land transport environmental impact as case studies, partly because these three areas tend to be almost universally an issue for transport, and partly because these three areas are among those of great current concern. The three that are discussed in the balance of this chapter are air quality/air pollution, greenhouse gas emissions, and noise.

Air quality/air pollution

Motorised vehicles that are used in transport emit a number of potential air pollutants. Not all vehicles emit the same pollutants, and certainly not to the same extent. However, there are five primary pollutants of the air that are emitted by a majority of vehicles, which are: carbon monoxide (CO), volatile organic compounds (VOCs) also referred to as hydrocarbons (HCs), oxides of nitrogen (NO_x), oxides of sulphur (SO_x), and particulate matter (PM). Among the latter, particles of 10 microns or less (referred to as PM_{10}) and particles of 2.5 microns or less ($\text{PM}_{2.5}$) are the main elements of concern.

Carbon monoxide is a highly toxic, odourless, colourless gas that is heavier than air and is emitted from internal combustion engines as a product of combustion. In modern-day vehicles, most of the CO is removed from the vehicle exhaust by means of a catalytic converter. However, to be effective, the catalytic converter has to be hot, so that, in the initial stages of operation of a cold engine, CO is not removed effectively. CO is also produced in large quantities at low speeds, and reduces as speeds increase.

Volatile organic compounds are also produced by internal combustion engines, and largely consist of unburnt fuel that has been vaporised and is

removed from the engine through the exhaust. Similarly to CO, VOCs tend to be very high at low speeds, and to diminish quite rapidly with increasing speed, up to about 70 km/h, after which they begin to increase again. Some VOCs are carcinogenic. However, the major concerns with VOCs relate to their effect in combination with oxides of nitrogen and also with $PM_{2.5}$.

Oxides of nitrogen (NO and NO_2 primarily) are formed by any process that involves combustion in the atmosphere, especially at high temperatures. They are also formed by lightning and, to a lesser extent, by electric arcing from third rail and overhead wire electric trains and trams. NO_x tends to be fairly low at low speeds, and increases with higher speeds, particularly where higher speeds involve a higher temperature of combustion. NO is not itself harmful, but it is slowly oxidised to NO_2 in the atmosphere. NO_2 is a toxic reddish-brown gas, which can often be seen in the atmosphere on a clear, sunny day. It can cause respiratory problems. However, it is most serious in combination with VOCs in the atmosphere in the presence of sunlight, where the reactions between NO_2 and the VOCs produce ozone (O_3), which is also a toxic gas and the main constituent of ozone smog.

Oxides of sulphur, mainly SO_2 , are a by-product of burning fossil fuels that have some sulphur content. To a large extent, most developed countries have reduced the sulphur content of petroleum products, so that oxides of sulphur are no longer a significant component of transport emissions in those countries. However, in countries where there is still a high sulphur content in the fuels, SO_2 is likely to be a continuing problem. SO_2 manifests itself in two forms. First, in the presence of sunlight and moisture, it produces sulphuric acid, which then falls as acid rain and acid snow. Second, oxides of sulphur can also react with the VOCs and the oxygen in the air to produce hydrogen sulphide, which becomes a component of smog, giving the smog the typical bad-egg smell and greenish-yellow colour.

Particulate matter is the fifth major pollutant that is emitted from transport vehicles. PM_{10} is generally visible as black soot, particularly from diesel engines. $PM_{2.5}$ is generally not visible and is also produced most significantly by diesel engines. However, both sizes of PM are also produced from wear and tear of moving parts in vehicles and from friction between the wheels of the vehicle and the travelled way (roads or rails). Particulate matter, by itself, is not particularly problematic, although inhalation of large quantities of particulate matter can lead to reduced lung capacity and related respiratory problems. Of a more serious nature, however, is the potential for other pollutants in the air to attach to the particles, thereby providing a means for the introduction of other pollutants into the lungs. PM_{10} is generally less of

a problem, in that particulate matter larger than about 2 microns in size will generally be removed from the lungs by coughing or swallowing. However, the finer particulate matter – $PM_{2.5}$ – will travel down into the smallest part of the lungs – the alveoli – where any adhering pollutants will penetrate and damage the lungs and also may move into the bloodstream. Because $PM_{2.5}$ particles readily attract VOCs, some of which are carcinogenic, $PM_{2.5}$ can become a vehicle for introducing cancer through the lungs.

There are many factors that affect the production of emissions from vehicles. Speed and operating mode are probably the most significant of these factors. Emissions of VOCs and carbon monoxide tend to be higher to much higher at low speeds, while emissions of NO_x tend to rise with increasing speed. PM, VOCs and, to a lesser extent, NO_x increase with acceleration, while VOCs and PM increase with braking and deceleration. There are five operating modes for petrol engines that affect their emissions quite markedly: cold start, hot start, warm stabilised, hot soak and cold soak. A cold start occurs when a vehicle is started after two hours (if equipped with a catalytic converter) or five hours (if there is no catalytic converter). A cold start lasts for 550 seconds. It is characterised by high emissions of VOCs and CO (because the catalytic converter is not yet hot enough to reduce CO to CO_2 and water vapour). A hot start occurs if the engine is restarted after two (five) hours. It has much lower emissions of both VOCs and CO than a cold start. NO_x emissions are generally low during either start condition. Warm stabilised is the mode of operation of the vehicle after the end of the cold start or warm start, when the vehicle continues to be used. Under this operating mode, speed becomes more important in determining emissions from the engine. Low speeds and stop-go driving will result in increased VOCs, PM and CO, while high speeds will result in increased NO_x and PM, with lower VOCs up to a certain speed, followed by an increase at highest speeds. The warm soak operating mode is characterised mainly by evaporative emissions of unburnt fuel and is therefore almost solely VOCs. The cold soak occurs after two (five) hours and has very low emissions, still mainly evaporative in nature.

Other factors that affect emissions from the internal combustion engine include ambient air temperature, fuel, vehicle mix, and road conditions (such as steep gradients). Probably the lowest emissions would be produced from a vehicle on a level, straight road, under level of service A or B conditions (see Chapter 5), with no intersections or interchanges, operating on a low-sulphur fuel, at an ambient temperature around 25° C, and running at a steady speed of around 70 km/h. Because such ideal conditions will occur very rarely, emissions will almost always be higher than this minimum condition.

To determine the impact of different transport policies (and other related policies) on vehicular emissions, it is necessary first to estimate the emissions production under both no-build and project conditions. This is usually done by using computer models that take into account all of the variables that affect emissions production. For much of the past several decades, the US EPA has supported a series of computer models with the generic name of MOBILE. Recently, this has been replaced with a new computer software called MOVES (Motor Vehicle Emissions Simulator). MOVES outputs the quantities of CO, CO₂, VOCs, NO_x, PM₁₀ and PM_{2.5} produced under the conditions input to the software (EPA 2009). Because few other countries have developed standard software for this purpose, MOVES is probably the best available software for such estimation. Outputs can be in total mass (tons, pounds, kilograms or grams) or in grams per mile.

The estimation of the total emissions output from a transport project is, however, still not the final answer required. There is a further modelling step required called airshed modelling. This step determines what happens in the atmosphere to the emissions. Among other things, the airshed model will estimate the formation of atmospheric pollutants, their concentration and their dispersion under various weather and wind conditions. Thus, to evaluate the air pollution consequences of alternative actions and policies, it is necessary to estimate the traffic conditions that would arise under each such alternative action or policy, and then to use the models to estimate the actual effects on air quality of each.

Similar approaches to this are used to set vehicle emission standards for a city, state or possibly even country. Initially air quality standards may be proposed, and then the contributory factors on an airshed basis are examined across a region or country. From this examination, the proposed standards may be adopted as proposed, or modifications may be proposed and examined further, to determine if they are now appropriate and acceptable.

It should be clear from this discussion that emissions are likely to increase significantly under congested traffic conditions. They will also be much higher when vehicles are used only for very short trips and then remain parked for more than two hours. As a general rule, any actions that will reduce congestion levels, or will reduce use of the car, will result in reduced emissions. Thus, while additions to capacity of highways may seem likely to increase emissions as a result of increased travel, to the extent that vehicles perform under more optimal conditions, the emissions impacts may represent a reduction from the do-nothing or no-build alternative. Each case must be estimated, however, to determine what the eventual impacts will be on air

quality. These impacts are likely to change over time, as congestion levels build again, which typically will be the case.

One of the difficulties that often confronts transport policy is determining the emissions impacts from public transport improvements. While electric vehicles, such as electric trains, produce almost no emissions as they move along the transport system, the electricity generation at a fixed source or sources removed from the transport system may produce significant emissions. Therefore, a complete assessment of the air quality impacts of policies and investments that would increase use of public transport need also to take into account remote, fixed-source emissions from power generation.

Policies to reduce air pollution

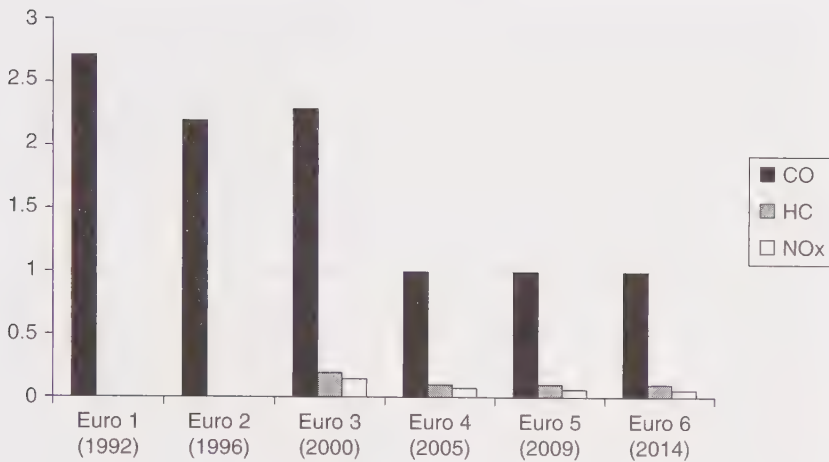
Looking at what policies might be considered if a goal is to reduce air pollution from vehicles, the most important transport policy to achieve this is setting vehicle emission standards that vehicle manufacturers must meet, such as the set of Euro standards and their equivalents in the USA. Over the past several decades, vehicle emission standards have been steadily tightened in Europe and North America and, because of vehicle imports to other areas of the world, throughout much of the world. Examples of the changes over time to these standards for Europe are shown in Figure 7.1 for diesel passenger cars, Figure 7.2 for petrol passenger cars and Figure 7.3 for heavy-duty diesel engines.



Note: In grams per kilometre

Source: http://en.wikipedia.org/wiki/European_emission_standards

Figure 7.1 EU emission standards for diesel passenger cars



Note: In grams per kilometre.

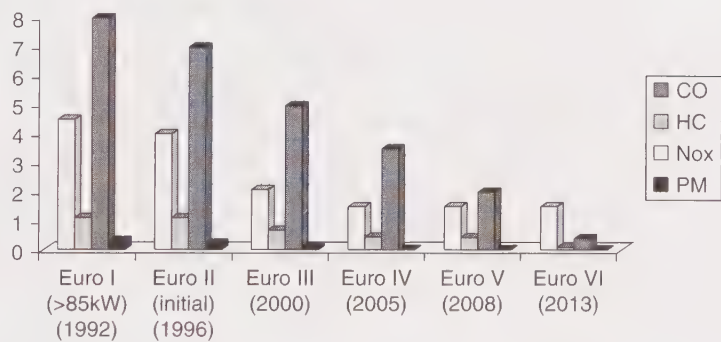
Source: http://en.wikipedia.org/wiki/European_emission_standards

Figure 7.2 EU emission standards for petrol passenger cars

Note: In grams per kilometre.

Source: <http://www.dieselnet.com/standards/eu/hd.php>

Figure 7.3 EU emission standards for heavy-duty diesel engines



Clearly any policy that will reduce the need to travel or the extent of travel is likely to reduce vehicular air pollution. Similarly, policies that encourage use of public transport, walking and bicycling and discourage use of personal internal combustion engine vehicles will also be likely to reduce air pollution. Both of these may be brought about through successful implementation of voluntary travel behaviour change (VTBC) programmes. Speeding up congested vehicular traffic is also fairly likely to reduce air pollution, but this must be balanced against the potential of increased capacity to generate additional travel. To a lesser degree, travel demand management (TDM) actions and transport system management (TSM) actions that serve either to reduce or reschedule demand and to increase supply without significant investments will have an effect on reducing air pollution and improving air quality. Policies that encourage or require the use of alternative-fuelled vehicles are also likely

to be effective in reducing air pollution. Hybrid vehicles using a combination of diesel and electric or petrol and electric propulsion systems will have a moderate effect on air quality, while fuels such as hydrogen and solar are non-polluting, the former producing only water vapour and the latter no exhaust at all. In addition, technological solutions that reduce vehicular emissions from conventional fuels would also be likely to be effective.

There also may be policies that can be implemented that will lead to improvements in transport-related air quality that may have nothing to do with transport per se. An interesting one to consider is that of the change of clock time. Ozone is readily formed from vehicular exhausts because the oxides of nitrogen and the VOCs are already mixed and warm as they emerge from the exhaust pipe. These gases must then remain in the atmosphere in the presence of sunlight to produce ozone, a process that takes a few hours. In urban areas, a substantial number of vehicles are started in the early hours of the day (say between 6 a.m. and 8 a.m.) in cold start mode, are then driven to work, where they remain in warm soak and then cold soak conditions for about eight hours, and are then started again in cold start mode and driven home. If the cars in the morning are started before it is light, the exhaust gases produced are quite likely to be dispersed before the sun has become strong enough to begin the process of catalysing the VOCs and nitrogen oxides into ozone. In the later afternoon, when these same vehicles are used to travel home, darkness will fall before there is sufficient time for the sun to convert the exhaust emissions into ozone. In most locations around the world, the emissions are likely to disperse long before the next sunrise. Thus, introducing or extending daylight savings time may result in reduced ozone production in the atmosphere, without any direct intervention on vehicular travel.

One of the things to look for in policy solutions is win-win situations, or, in other words, policies that not only provide a gain in an area such as air pollution, but also provide gains in other respects. For example, almost all of the policies listed above will also increase physical activity, which is likely to lead to improved health. Most of these policies will also reduce the consumption of non-renewable resources and therefore contribute to sustainability.

Greenhouse gas emissions

There are five principal greenhouse gases present in the earth's atmosphere. In order of quantity, they are water vapour, carbon dioxide, methane, nitrous oxide and ozone. There are a number of other greenhouse gases that are present in much lower concentrations. Without the greenhouse gases, the earth would be uninhabitable by human beings and most of life as we know

it upon the earth. Figure 7.4 illustrates the effects of greenhouse gases in the atmosphere. The sun emits short-wave radiation that passes through the earth's atmosphere and warms the earth's surface. This short-wave radiation has relatively little effect in warming the atmosphere.

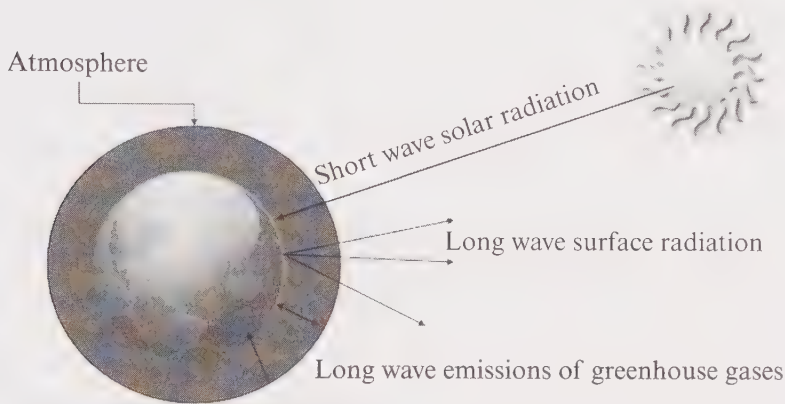


Figure 7.4 Schematic of the effect of greenhouse gases in the earth's atmosphere

The warmed surface of the earth then emits long-wave radiation, which passes back through the atmosphere and out into space. The long-wave radiation is more effective at warming the earth's atmosphere, but this effect is substantially reduced if greenhouse gases are not present. The greenhouse gases have the effect of trapping these long-wave emissions and bouncing them back to the earth's surface, allowing them to further warm the atmosphere and the earth's surface – hence the name greenhouse gases, as they operate like the glass of a greenhouse to trap warmth inside the atmosphere. Without any greenhouse gases in the atmosphere, it is estimated that the earth's surface would have an average temperature of about -19°C , rather than the actual average of $+15^{\circ}\text{C}$. The contributions of each of these greenhouse gases to the warming of the atmosphere are shown in Table 7.1.

Table 7.1 Contributions of greenhouse gases to atmospheric warming

Greenhouse gas	Contribution to atmospheric warming
Water vapour (not including clouds)	36–70%
Carbon dioxide	9–26%
Methane	4–9%
Ozone	3–7%
Other greenhouse gases	Very small effects

Source: Kiehl and Trenberth (1997).

The effect of greenhouse gases is readily apparent to those living in temperate latitudes. When the humidity is very low, as is the case over deserts, but also over many other areas of land at certain times of the year, there is rapid cooling once the sun sets, and very low temperatures occur overnight, because there is insufficient water vapour (in the form of humidity) to trap the earth's radiation once night falls. The large variation in the contribution of water vapour shown in Table 7.1 is a direct result of the variation in relative humidity. This also then results in changes in the contributions of other greenhouse gases to atmospheric warming, because they are subject to much less variation in the atmosphere.

There are a number of anthropogenic sources of greenhouse gases. Burning fossil fuels produces both CO₂ and ozone. The ozone that is produced, however, remains in the troposphere, whereas the ozone that acts as a greenhouse gas and also as a protection against excessive ultraviolet radiation from the sun is in the stratosphere. There is no way for ozone produced in the troposphere to find its way into the stratosphere, so this ozone is largely a pollutant and contributor to ozone smog, as discussed earlier. Livestock and some farming and waste disposal systems generate methane, and much is tied up in permafrost. Nitrous oxide is produced from agriculture, while the other major anthropogenic greenhouse gases are CFCs, HFCs and PFCs, used in refrigeration and air conditioning.

Based on ice-core measurements, it is estimated that, prior to the Industrial Revolution, carbon dioxide concentrations in the atmosphere were around 260–280 parts per million (ppm) (World Bank 2012). As of August 2012, according to the NOAA (2012), the atmospheric concentration is about 392 ppm, and emissions are continuing to increase. Global mean temperature has continued to increase and is now about 0.8 °C above pre-industrial levels.

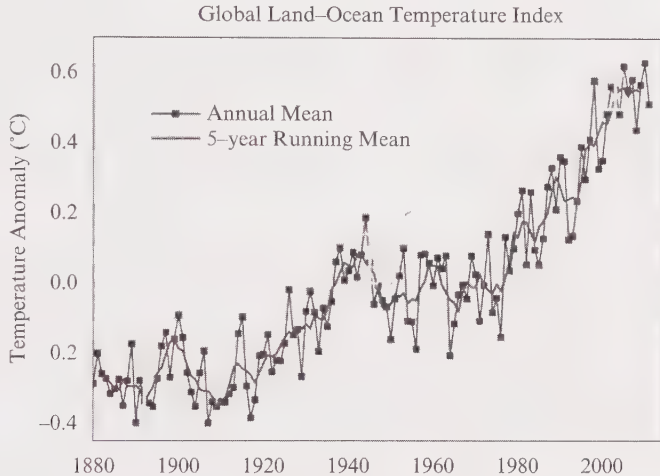
There appears to be rather little doubt that the increases in CO₂ in the atmosphere are anthropogenic. It is estimated (OECD 2010) that about 59 per cent of greenhouse gas emissions (primarily CO₂) in 2005 were from the burning of fossil fuels. Of this total, about 23 per cent globally of the emissions from burning fossil fuels came from transport, with road transport being by far the dominant source. Perhaps of even more significance is that transport emissions of GHGs have been growing rapidly, showing growth from 1990 to 2007 of about 45 per cent (OECD 2010).

Looking at recent data, there are clear signs that temperatures have been rising over the geological short term of the past 100 to 200 years,

although decadal data suggest that global temperatures may have dropped or remained level in the decade from 1998 to 2008, as can be seen in Figure 7.5. (The grey bars in the figure indicate the reliability of the estimates.) However, Foster and Rahmsdorf (2011) have shown that, after removal of known factors that affect short-term temperature variations (solar variability, volcanic aerosol effects, the El Niño/Southern Oscillation events), the warming trend is continuing, the last three decades providing a strong warming signal.

Source: NASA (2008).

Figure 7.5 Variation in global temperatures for the past 130 years



The president of the conservative World Bank, Dr Jim Yong Kim, has recently stated that the science is unequivocal that humans are the cause of global warming (World Bank 2012, p.ix). While the global community has committed itself to holding warming below 2 °C, to prevent dangerous climate change, a business-as-usual approach which includes current reduction commitments has been estimated as having the world on a trajectory for warming of well over 3 °C, with a 20 per cent probability of being associated with a 4 °C warmer world by 2100. The prospective extreme consequences of such a scenario, which will mean monthly summer temperature rises of more than 6 °C in the subtropical region, have been pointed out by the World Bank (2012).

Nevertheless, the authors of this book are not in complete agreement about whether reducing CO₂ concentrations will mitigate climate change, or whether it is possible for humans to limit temperature increases to below 3 °C. However, we are in agreement that a risk management approach is an appropriate policy response, particularly where such an approach also produces other obvious and needed benefits.

A compelling case for such an approach has been presented by Weitzman (2012), who highlights uncertainties around temperatures, climate, comprehensive damages, and the associated welfare links. His analysis emphasises the roles played by the shape of the damages function and the disastrous consequences of a fat-tailed probability density function (PDF) for warming (which increases the probability of extreme warming, for any given median warming estimate). Weitzman's analysis leads him to conclude that:

When relatively fat-tailed PDFs are combined with a reactive damages function, then seemingly modest changes in target levels of GHGs can sometimes have very big welfare consequences. In such conditions, the primary purpose of keeping down GHGs is to prevent large damages from extreme warmings in the 'bad' tail, which is a much more powerful incentive to target low GHG levels than trying to keep down the relatively modest damages from median temperatures. (Weitzman 2012, p. 241)

As noted earlier in this chapter, transport plays a substantial role in the emission of greenhouse gases, principally CO₂. Indeed, the catalytic converter that removes CO from the exhaust increases the amount of CO₂ that is emitted. As discussed for air pollution, we advocate seeking out win-win policies. Indeed, almost all of the potential policy directions discussed for cleaner air will have similar benefits in reducing greenhouse gas emissions, increasing sustainability of transport and improving human health. A risk management approach strongly supports pursuit of such policy measures. Reducing dependence on the car, improving alternatives to the car, seeking alternative fuels that do not have other undesirable side effects or unintended and negative consequences, and seeking technological changes that provide more sustainable and efficient transport are all likely to be of long-term benefit under any circumstances.

To illustrate what this might mean in practice, Stanley et al. (2011) have developed some scenarios to indicate the broad scale of changes that might be needed to cut Australian road transport GHG emissions by 2020, to a level that is 20 per cent below 2000 levels, and then 80 per cent below the 2000 levels by 2050. The range of measures included in their analysis includes both behaviour change and technological solutions. Table 7.2 shows three alternative scenarios that deliver the 2020 reduction target, with reductions in emissions intensity being critical but a range of behaviour change variables being equally important in combination. However, if the emissions intensity of the vehicle fleet improves relatively slowly, then changes required in other policy variables increases substantially. The policy conclusion is the importance of taking an integrated approach to cutting transport GHG emissions.

Table 7.2 Alternative options to deliver a 20 per cent GHG reduction on 2020 levels for Australian road transport

Measures	Units	2007	2020 (A)	2020 (B)	2020 (C)
1. Fewer/shorter urban car trips	Fewer car kilometres	n.a.	10%	20%	25%
2. Shift urban car to walking/cycling	Active transport mode share (urban)	16%	26%	34%	39%
3. Increase urban public transport mode share	PT share of trips (all urban trips)	8%	15%	20%	21%
4. Increase urban car occupancy	Passengers/car	1.4	1.6	1.8	1.9
5. Freight efficiency	Less fuel than forecast	n.a.	30%	30%	30%
6. Car emissions intensity	Less than 2007	n.a.	30%	18%	13%
7. Truck emissions intensity	Less than 2007	n.a.	18%	13%	8%

Note: The 2007 numbers refer to relevant numbers/shares in Australia for that year.

Source: Drawn from Stanley et al. (2011).

The commonality between the policy measures to cut transport GHG emissions and those required to improve air quality and reduce congestion also draws attention to the importance of an integrated approach to land transport policy more generally. Our experience suggests that, if such an integrated approach is taken, then specific GHG benefits will typically only account for a small part of the total quantified benefit stream. Evaluating transport GHG policy in isolation of the complementary benefits will, therefore, severely undervalue the relevant initiatives.

Noise

A third environmental impact that is considered in this chapter for illustrative purposes is that of noise. As with each of the previous two impacts, we start by considering what we know about noise itself and in relation to transport, then discuss how it is assessed and compared among alternative projects, and finally consider what mitigating policies and directions can be followed.

Noise can be defined as unwanted sound. It is always intrusive, but can also be physically harmful if it is sufficiently loud or persistent. Humans perceive noise through their ears, where the actual mechanism of hearing involves vibrations that are received by hair cells in the inner ear. Sudden, very loud noises, and prolonged noises above a certain level can damage these hair cells, which the body is unable to replace. Therefore, hearing damage is permanent.

Noise is measured as relative sound pressure in decibels, which are on a logarithmic scale. Thus, an increase of 10 decibels represents a doubling of the noise level. However, a simple measurement of noise in decibels does not completely define noise as a nuisance. Noise is affected by pitch and frequency and also by the capability of the human ear to perceive it. The human hearing response to noise is measured using what is known as the A-scale of decibels, usually abbreviated as dBA. Noise has properties that are quite similar to those of light. Noise is attenuated (reduced) by distance, and is transmitted through the atmosphere as waves. These waves travel in a straight line. Noise, like light, may be reflected, refracted and absorbed. Barriers can also be used to attenuate noise through reflection and absorption.

The human ear can distinguish noises from about 20 dBA upwards, but differences of less than 3 dBA are not normally detectable. Noises are also not additive. There are some simple rules that can be applied when there are two different noise sources. If the two noises are equal, the resulting noise level will be 3 dBA higher than either noise on its own. Therefore, the increase in noise by adding a second equal noise is barely detectable by the human ear. If two noises are more than 10 dBA different, the resulting noise is at the level of the louder of the two sources. If two noises are within 1 dBA of each other, then the resulting combined noise will be 3 dBA higher than the louder of the two sources. If the two sources of noise are between 2 and 4 dBA different, then the resulting noise is 2 dBA higher than the louder noise. Some reference noise levels are provided in Table 7.3.

Table 7.3 A selection of reference noise levels

Source	Noise level
Quiet back street with no traffic	60 dBA
Man's voice at 1 metre	65 dBA
Telephone ringing at 3 metres	73 dBA
Busy street	85 dBA
Electric trains on steel bridge at 6 metres	100 dBA
Unsilenced motorcycle at 0.7 metres	108 dBA

Noise from land transport and from air transport are different and are measured on different scales. Aircraft noise is measured in terms of the effective perceived noise level, or EPNL. This is similar to the decibel scale, but is weighted differently. Land transport noise depends on such things as the speed and volume of traffic, the type of road or rail surface, and the traffic mix. For road noise, there is a fairly simple relationship that captures the

mean noise level from traffic, as shown in equation 7.1, although this does not take into account the traffic mix:

$$\bar{L} = 10 \log_{10} q - 10 \log_{10} d + 20 \log_{10} u + 20 \quad (7.1)$$

where:

q = traffic volume in vehicles per hour

d = distance from the sound source in metres

u = speed of traffic in km/h.

Given the properties of noise levels, the human perception of a busy road at 85 dBA, with the addition of a truck also emitting 85 dBA, would result in the perception of a noise of 88 dBA. If the truck were emitting 86 dBA, the resulting perceived noise level would be 89 dBA, while, if it were 87 to 89 dBA, the perceived noise level would be 89 to 91 dBA. If the truck were emitting 95 dBA, then 95 dBA would be the perceived noise level. Building walls generally reduce noise levels by about 10 dBA if the doors and windows are open. If they are closed, the noise reduction is about 20 dBA. Inside a car, the reduction of noise is about 10 dBA, while vegetation alone can reduce noise levels by around 3 dBA.

In a similar approach to that for air quality/air pollution, the first step is to determine current ambient noise levels by taking measurements at different times of the day and night. It is important to take measurements at different times, because typical noise levels in an urban environment can vary from as low as 35 dBA at night to as high as 70 dBA or more in the peak traffic periods of the daytime. There are some noise standards that can be used as a rough guide to what is acceptable and unacceptable, which are shown in Table 7.4.

For broad policy setting, a similar approach of setting vehicle noise standards can be taken as to emissions standards from vehicles in the case of air

Table 7.4 Example noise standards

Rating	General external exposure
Unacceptable	>80 dBA 60 minutes in 24 hours >75 dBA 8 hours in 24 hours
Normally unacceptable	>65 dBA 8 hours in 24 hours Loud repetitive sounds
Normally acceptable	<65 dBA more than 8 hours in 24 hours
Acceptable	>45 dBA less than 30 minutes in 24 hours

pollution. For project-specific contexts, for each alternative, including the no-build or do-nothing alternative, estimates are then made of the expected noise levels at similar times of the day. An equation such as equation 7.1 can be used to estimate the expected noise levels at various distances from the proposed projects or facilities, as well as for the no-build situation with traffic forecasts for some point in the future. For rail projects, actual noise measurements would need to be made from similar vehicle–rail configurations, because there are no simple equations that will provide estimates. The addition of noise would use the rules for adding sound sources, to arrive at new predicted noise levels.

Aircraft noise presents significant issues that cannot be dealt with in detail in this book. Runway expansion of existing airports and construction of new airports require the development of noise exposure contours around the expected flight paths for take-off and landing aircraft. Even operational changes at an airport may affect aircraft noise exposure: increased service frequency or changes to landing and take-off patterns around an airport will have an impact on local noise levels. However, newer aircraft are generally showing lower noise levels than older ones, so that changes to the aircraft fleet mix using any particular airport can also result in significant changes to noise exposure at airports.

Mitigating unacceptable noise levels

For road transport, the most effective ways to mitigate unacceptable noise levels are to put the roadway in tunnels or in deep cuttings. Because of the transmission of sound, these are effective ways to reduce the received noise levels for buildings and persons on the surface. Adding barrier walls along the top of cuttings may be very effective when the depth of the cutting is not enough to reduce the noise sufficiently, as a result, perhaps, of the width of the right of way. When such topographic noise abatement is not possible, the next method is to construct sound walls that will both absorb and reflect sound, as well as prevent direct transmission of the noise from the roadway to sensitive receptors alongside the roadway. Road surface also has an effect on noise, with jointed concrete roads tending to generate the highest noise levels and asphalt surfaces the lowest noise levels. Good surface maintenance of roadways can also reduce noise levels that are exacerbated by the jolting of vehicles as they pass over rough road sections.

For rail, similar strategies can be employed: namely, the use of tunnels and cuttings, and the construction of sound walls. A particular concern for rail is the use of rather tight horizontal curves, which may cause excessive noise

from the wheel flanges contacting the rail. Larger-radius curves should be used in design and construction to avoid this issue. For rail, as for road, speed may also affect the production of noise, so that lower speeds may be necessary as a mitigation measure.

While the pursuit, in the case of noise, of win–win strategies is less imperative than if there were questions about whether or not noise from transport was a negative impact, the strategies suggested here can also be viewed as win–win strategies. For example, putting road or rail facilities in tunnels or cuttings may have aesthetic merit, as well as reducing the tendency of such facilities to cause neighbourhood disruption and contribute to social exclusion.

7.3 Valuing environmental externalities

Chapter 4 discussed the general subject of valuing environmental impacts, particularly with respect to their inclusion within a cost–benefit analysis framework. A number of studies have produced estimates of relevant unit costs for particular externalities, to assist the process of application in policy, programme or project evaluation (see, for example, Australian Transport Council 2007; Maibach et al. 2007; HM Treasury 2012). Maibach et al. (2007) provide a very comprehensive source of emission cost estimates for use in transport evaluations. They follow the impact pathway approach and present *inter alia* marginal external cost estimates for air pollution, greenhouse gas emissions and noise, the three externalities considered in this chapter, for European countries.

Air pollution costs typically include health costs (the largest single cost component), building and material damage, crop losses and costs for ecosystem damage (biodiversity, soil and groundwater). By way of example, Maibach et al. (2007) provide air pollution costs on a per-tonne-of-pollutant basis for a range of pollutants, then expressed per vehicle kilometre and per train kilometre. The costs per vehicle kilometre take account of vehicle type (passenger car or truck, by engine size for cars and vehicle mass for trucks), fuel type (petrol or diesel), the level of emissions control equipment on the vehicle (relevant Euro standard) and the operating environment. Separate cost estimates are thus presented for metropolitan roads, urban roads, inter-urban roads and motorways, and an average figure. Similar detail is available for rail transport. For motor vehicles in Germany, the highest air pollution cost estimates are for large trucks (over 32 tonnes) with Euro 0 emission controls in metropolitan operation (€0.38 per kilometre in 2000 prices). This air pollution cost estimate is only €0.052 per kilometre for a Euro 5 truck (over 32 tonnes).

Estimation of marginal **GHG emission costs** is difficult. Estimating GHG emissions as a function of travel activity is not difficult, but selecting an appropriate unit cost value to apply to these emissions is very complex. Global damage costs are the most appropriate way to approach unit costing of the externality, but, as Weitzman (2012) points out, estimation of such damage costs is subject to layer upon layer of uncertainty. As a result, there is a wide range of relevant estimates. Avoidance costs are frequently used, with values in the range of €50–€100/tonne of CO₂. Maibach et al. (2007) proposed central values of €25/tonne for 2007–10, €40/tonne for 2020, €55/tonne for 2030, €70/tonne for 2040 and €85/tonne for 2050. The UK government requires use of £56/tonne CO₂ for GHG emissions in the non-traded sector (which includes petrol and diesel), in evaluations of public proposals in that country for 2012 (HM Treasury and DECC 2011).

Noise costs include health costs and annoyance costs. Key cost drivers are time of day, the receptor density close to the source of the noise, and existing noise levels. Maibach et al. (2007) again present marginal noise cost estimates for motor vehicles and trains (passenger and freight). For cars, urban noise costs (2004 prices) have a central value of €0.0076 per vkm during the day and €0.0139 per vkm at night. On suburban roads the comparable figures are €0.0012 per vkm (day) and €0.0022 per vkm (night). Rural noise costs are negligible at €0.0001 per vkm (day) to €0.0003 per vkm (night). For heavy goods vehicles, the highest noise costs are estimated at €0.12–8 per vkm for night-time operation on urban roads (€0.0701 per vkm in daytime). There is a considerable band around these central cost estimates, related to traffic density.

7.4 Conclusions

Transport policies and investments will create a number of potential environmental impacts. In this chapter, we have looked at three such impacts – air pollution, greenhouse gas emissions, and noise. In each case, the process that is required is to understand the nature of the impacts, determine whether impacts will be significant from specific policies and investments, and then assess the magnitude of the impacts. An extremely important part of this is always to include the impacts of doing nothing, because such a policy may often have more severe environmental impacts than doing something about the transport situation.

A second lesson that is important, and continues as a theme through the subsequent chapters on other externalities, is to seek for win–win strategies and policies. This is particularly important in instances where there may be

some question about the magnitude or extent of the impact or externality. We define a win-win policy as one that creates a number of benefits besides the one of mitigating the particular impact under study. These policies may be considered to be robust policies, because the overall benefits from them will meet other objectives that may be more obvious and uncontroversial.



REFERENCES

- Australian Transport Council (2007), *National Guidelines for Transport System Management in Australia*, Vol. 3: *Appraisal of Initiatives*, Canberra: Commonwealth of Australia.
- EPA (2009), 'EPA releases MOVES2010 mobile source emissions model: questions and answers', US Environmental Protection Agency, Office of Transportation and Air Quality, December, available at: <http://www.epa.gov/otaq/models/moves/420f09073.pdf> (accessed 18 September 2012).
- Foster, G. and S. Rahmsdorf (2011), 'Global temperature evolution 1979–2010', *Environmental Research Letters*, **6** (4), 044022.
- HM Treasury (2012), *The Green Book: Appraisal and Evaluation in Central Government*, London: TSO, available at: http://www.hm-treasury.gov.uk/d/green_book_complete.pdf (accessed 10 April 2012).
- HM Treasury and DECC (2011), *Valuation of Energy Use and Greenhouse Gas Emissions for Appraisal and Evaluation*, October, London: HM Treasury and Department of Energy and Climate Change.
- Kiehl, J.T. and K.E. Trenberth (1997), 'Earth's annual global mean energy budget', *Bulletin of the American Meteorological Society*, **78** (2), 197–208.
- Maibach, M., C. Schreyer, D. Sutter, H.P. van Essen, B.H. Boon, R. Smokers, A. Schroten, C. Doll, B. Pawlowska and M. Bak (2007), 'Handbook on estimation of external cost in the transport sector', produced within the study Internalisation Measures and Policies for All External Cost of Transport, CE Delft, 19 December.
- NASA (2008), 'Goddard Institute for Space Studies Surface Temperature Analysis (GISTEMP)', 11 January, available at: <http://data.giss.nasa.gov/gistemp/> (accessed 2 February 2013).
- NOAA (2012), 'Trends in atmospheric carbon dioxide', US Department of Commerce, National Oceanic and Atmospheric Administration, August, available at: <http://www.esrl.noaa.gov/gmd/ccgg/trends/#mlo> (accessed 19 September 2012).
- OECD (2010), *Reducing Transport Greenhouse Gas Emissions: Trends and Data 2010*, OECD/TIF, available at: <http://www.internationaltransportforum.org/Pub/pdf/10GHGTrends.pdf> (accessed 19 September 2012).
- Sanft, C. (2010), 'Environment and law in early imperial China (third century BCE – first century CE): Qin and Han statutes concerning natural resources', *Environmental History*, **15** (4), 701–21.
- Stanley, J.K., D.A. Hensher and C. Loader (2011), 'Road transport and climate change: stepping off the greenhouse gas', *Transportation Research Part A*, **45** (10), 1020–30.
- Weitzman, M.L. (2012), 'GHG targets as insurance against catastrophic climate damages', *Journal of Public Economic Theory*, **14** (2), 221–44.
- World Bank (2012), *Turn Down the Heat: Why a 4°C Warmer World Must Be Avoided*, report for the World Bank by the Potsdam Institute for Climate Impact Research and Climate Analytics, Washington, DC: World Bank.

8

Tackling the externalities – health and safety



CHAPTER OVERVIEW

Continuing the theme of examining the externalities of provision of transport facilities, this chapter looks at health and safety. The chapter begins with a consideration of traffic accidents from the specific viewpoint of their causes, and actions and policies that may be adopted to reduce the severity and quantity of traffic accidents. Next, the chapter provides a discussion of issues of personal safety and also looks at policy directions that may be considered for improving both the reality and the perception of personal safety while travelling. The balance of the chapter is concerned with the effects of transport on health, looking first at concerns of the relationship between transport and obesity, second at exposure to pollutants, and third at the externality known as road rage. In each case, potential policy directions for mitigation are discussed. The chapter concludes with a discussion of valuing health and safety improvements.

8.1 Health and safety externalities

There are both health and safety externalities for transport. Safety has two components – safety from accidents and personal safety, such as from crimes against the person while using transport facilities. Health externalities relate to three components – obesity, exposure to pollutants, and road rage. In this chapter, each of these components is discussed, together with potential policy directions that may reduce the severity of the externalities. The chapter is focused principally on road-related matters, and Chapter 15 on freight deals with some specific freight-related safety issues. Safety for rail and air is not discussed here.

8.2 Transport accidents

Travelling by any means of transport opens up the risk to property damage, injury or death. While all forms of transport are subject to accidents, by

far the most accidents occur in relation to road traffic. Indeed, the World Health Organization ranked road traffic accidents in 2004 as the ninth most common cause of death in developed countries (WHO 2006). Worldwide, there are about 1.2 million deaths annually from road traffic and between 20 million and 50 million non-fatal injuries (WHO 2009). More than half of those killed in road accidents around the world are aged between 15 and 44 years. In fact, for 5- to 14-year-olds it is the second most common cause of death, and is the leading cause of death for those aged between 15 and 29, and the third most common cause of death for those aged 30 to 44 years (WHO 2009).

Road traffic accidents are normally reported as rates, with the two most common such rates being per 100 000 population and per 10 000 vehicles. This makes the rates more directly comparable among different regions and nations and also provides a better measure of exposure or risk. However, the absolute number of accidents, by country and by region, provides a clear idea of the incidence of road traffic accidents throughout the world. Another useful measure, which relates directly to accident exposure, is accidents per million vehicle kilometres. However, not many agencies report statistics in this latter form.

According to a report to the UN General Assembly in 2003 (UN 2003), traffic accidents were responsible for 25 per cent of all deaths due to accidents. The same report indicated that the worldwide cost of road accidents was estimated to be around \$518 billion per year, with over \$100 billion of this occurring in developing countries, for which it represented more than twice the amount of development aid received. These costs represent between 1 and 3 per cent of GDP for most countries around the world (WHO 2009). Various sources estimate that road traffic accidents will become the fifth leading cause of death by 2030, if there is no change in recent trends. The incidence of traffic accidents is much higher in low- and middle-income countries than in high-income countries, with 2002 statistics shown in Table 8.1.

Table 8.1 Estimated road traffic injury-related deaths, 2002

Locality	Number	Rate per 100 000 population	Proportion of total (%)
Low- and middle-income countries	1 065 988	20.2	90
High-income countries	117 504	12.6	10
Total	1 183 492	19.0	100

Source: WHO (2009).

Exacerbating these differences is the fact that road traffic accidents are tending to fall in many high-income countries, while they are growing in middle- and low-income countries, largely as a result of increasing motorisation in those countries. In 2009, WHO estimated that 91 per cent of traffic accidents occurred in low- and middle-income countries, while those same countries accounted for only 48 per cent of the world's registered motor vehicles. Many countries in Africa, South America and South-East Asia show rising fatality rates from traffic accidents, while many countries in Europe, North America and Australasia show falling rates of fatalities and injuries from traffic accidents. For example, between 2002 and 2011, annual deaths due to road accidents in Australia declined by 24.7 per cent, even though this decade saw a growth both in population and in registered vehicles (BITRE 2012). In terms of exposure, the decrease has been even more significant, with a decline from 8.73 deaths per 100 000 population in 2002 to 5.71 in 2011, a decrease of 34.6 per cent over the decade. Another measure of exposure is the rate per 100 million vehicle kilometres of travel, and this also has declined from 0.84 in 2002 to 0.58 in 2011, which is just under a 31 per cent decrease.

Apart from deaths, road accidents also cause injuries and property damage. Injuries from road accidents are often more severe than other types of accident injuries and often leave the victims with permanent disabilities, sometimes ones that prevent them from being gainfully employed for the rest of their lives. The costs of these injuries and costs resulting from property damage are not included in the estimates given previously, partly because of the problems of reporting of these accidents and resulting injury and damage.

To further demonstrate the seriousness of the problem, the number of annual deaths, over a decade, would equal approximately half of the population of Australia, while the number of injuries may be as high as half of the population of India. Clearly, road safety is a serious issue that needs transport policy to address it. While accidents in air traffic, on railways and at sea also occur, the extent and exposure of the population to these safety issues are very much smaller than for road traffic. For example, for air there were, in 2011, 117 accidents worldwide, which took 828 lives. The accident rate for aircraft is about one-hundredth that of the car, while the total number of fatalities is about 0.07 per cent of the rate of road traffic fatalities. Rail is also very low in developed countries, although figures for less-developed countries are apparently difficult to find. The only figure that seems to be available suggests about 0.4 deaths per billion passenger kilometres of travel (PKT) and about 15 injuries per billion PKT. These figures are also far below those for road traffic.

Causes of road accidents

There are a number of factors that lead to road accidents. By far the largest contributory factor is driver behaviour, with speed, alcohol and drugs, and aggressive driver behaviour being the major driver behaviour factors. There are also other causes, which include equipment failure, roadway design and poor roadway maintenance. However, these are far less significant factors than those relating to driver behaviour. In most middle- and high-income countries, equipment failure, road design and road maintenance contributions to road traffic accidents have largely been minimised. In low-income countries and, to a lesser extent, middle-income countries, there remains work to be done in this respect.

As a framework for discussing the causes of road accidents, one can divide the basic causes into three: the driver, the vehicle and the road environment. We use this framework here and again in Chapter 15 on freight safety issues.

Driver

In all countries, driver behaviour remains a major factor in road traffic accidents. Therefore, policies are needed that concentrate on modifying driver behaviour in various ways.

Speed is clearly one of the major factors, with speeding being implicated in a substantial proportion of road traffic accidents. Statistics from around the world are not readily available to indicate the causes of accidents, but anecdotal data indicates speed as one of the major driver behaviour aspects of road traffic accidents. However, speed is a complex issue. In part, the issue of speed relates to the posted speed limit, but, in part, it also relates to driver willingness to obey posted speed limits. In most high-income countries and many middle-income countries, appropriate posted speed limits in both urban and rural areas have been established. However, there are at least two factors that can lead to inappropriate driver behaviour even when appropriate speed limits are posted. First, if drivers perceive a major discrepancy between the posted speed limit and the apparently safe speed to operate a vehicle, with the latter being higher than the posted limit, there will be a great tendency for drivers to ignore the posted speed limit, especially if enforcement is lax. For example, a roadway that is designed for safe operation at 110 km/h being posted at 70 km/h is likely to lead to a majority of drivers driving well above the posted speed limit, unless enforcement is very strict. A second problem with posted speed limits relates to frequency of changes. When speed limits change frequently and by small increments

(e.g. 10 km/h) drivers will often ignore the changes and drive at a higher speed in many speed zones.

A report prepared for the FHWA (M.R. Parker and Associates 1992) makes several important points with respect to speed. First, it points out that setting arbitrary, unrealistic or non-uniform speed limits has created a social acceptability to driving at speeds above the speed limit. Researchers have reported that it is safest to drive at the same speed as the majority of other drivers. Therefore, when the majority of drivers are driving above the speed limit, the driver who attempts to stick to the speed limit is actually exposed to a much greater accident risk than those drivers who are driving at the prevailing speed of the traffic. Traffic engineers generally advocate setting speed limits at the 85th percentile of the speeds at which the majority of safe and reasonable drivers would drive. Traffic engineering design manuals usually specify safe speeds for roadways, having consideration for horizontal and vertical curves, lane widths, presence of obstructions, and environmental conditions (such as the presence of pedestrians, cyclists, schools, etc.). These speeds are usually based on the 85th percentile of the speed distribution that experience shows would be driven by safe and reasonable motorists.

Setting arbitrary and non-uniform speed limits tends to arise from two misconceptions. First, when accidents occur on a stretch of road, there are often demands to lower the speed limit, from the belief that higher speeds lead to more accidents. While it is true that higher speeds will result in more severe accidents, with the risk of fatalities being significantly higher, research has shown, as mentioned previously, that the biggest risk of accidents occurs for motorists who travel at speeds much lower or much higher than those of the majority of motorists (M.R. Parker and Associates 1992). This also suggests that a severe mismatch between design speed of a roadway and posted speed limit will lead to a greater potential for accidents than simply the speed itself. A review of research on accidents and speeds indicates that there remains considerable controversy on the exact nature of the relationship. However, it seems clear that frequent changes in speed limits and setting speed limits that are well below the driver perception of safe speed will both be likely to result in speeding behaviour by motorists and, as a result, to increases in both the severity and the number of accidents.

Using alcohol and recreational drugs, as well as some prescription drugs, interferes with driving skills and leads to increased chances of accidents. Studies have shown clearly that reducing the permitted level of alcohol in the blood and enforcing the lower limits leads to a significant reduction in road traffic accidents. The problem with alcohol and many drugs is that they slow

reaction times and also often give users a false sense of their own skills and abilities. The slowing of reaction times means that there is a greater chance that a driver misjudges the need for evasive action, such as braking, waiting too late and as a result causing a collision between the vehicle that the driver is operating and either other vehicles or stationary objects on or beside the road. Adding to this is the change in perception of the driver's ability to handle a situation, which further impedes reactions to external situations.

There is no doubt that reducing permitted levels of alcohol and banning the use of recreational and illicit drugs by drivers are instrumental in reducing accidents. According to the World Health Organization (WHO), a blood-alcohol level above 0.04 g/dl results in impairment of driving (WHO 2012a). The WHO also recommends random breath testing as a cost-effective method to reduce the occurrence of drink-driving violations, especially where the maximum permitted level of blood-alcohol is 0.05 g/dl or lower.

There are a number of other driver behaviour factors that have an impact on road accidents. These include aggressive driver behaviour, wearing seat belts and using child restraints, wearing helmets for motorcyclists, and, increasingly, driver distractions, such as use of mobile phones while driving. All of these factors influence road traffic safety. In part, these can be addressed through improved driver training programmes. Drivers need to be trained in good, defensive driving, and discouraged from aggressive driving, such as driving too close to the vehicle in front, frequent lane changing, speeding (as discussed previously) and so on. Recent statistics show that using mobile phones, especially for texting, is a major distraction that is implicated in increasing numbers of road traffic accidents. It is also suspected that using hands-free mobile telephones is not much safer than using a hand-held mobile telephone.

Vehicle

There is no doubt that vehicles also play an important role in traffic safety. Mechanical integrity and correct response by the vehicle to the demands of the driving situation are clearly significant influences on safety. In 2013, most vehicles are designed to operate safely in normal driving conditions, and many are equipped with the means to deal with unusual and safety-threatening situations. Modern cars, for example, are fitted with many additional safety devices, such as sensors (both rear and front in some vehicles) warning of the proximity of objects, anti-locking braking systems (ABS), a variety of airbags and three-point seat belts, head restraints, and many more. Many of these are capable of protecting occupants during an accident or

pre-empting the driver in some way to help avoid an accident. As technology continues to improve, more and more such devices are likely to be added to vehicles. Many vehicles are also constructed in such a way as to provide added protection to occupants in the event of a collision, by designing engine compartments that will not intrude into the passenger space in the event of a crash, and so on. Also, because pedestrian fatalities are a very significant component of on-road fatalities, some recent vehicles now include an exterior airbag on the front of the vehicle that will provide some degree of protection to a pedestrian hit by the vehicle.

However, a number of these safety systems that are built into the vehicle must be correctly maintained for full effectiveness. This is also true of other safety aspects of a vehicle. Poor maintenance, or neglect of certain aspects of the vehicle may lead to the manufactured safety devices and systems no longer affording protection. Hence, correct maintenance and upkeep of the vehicle are necessary to maximise the effectiveness of these various safety devices.

Road environment

The road environment is the third aspect that affects safety. This includes the quality of the road construction, including the road surface. It also includes such things as the horizontal and vertical curves, the width of lanes, the existence and width of shoulders, the control of intersections, and the existence of objects alongside the road, such as trees, signs and light poles. Other aspects of the road environment that affect safety include such things as the presence of pedestrians, the density of driveway entrances along the roadway, the traffic mix, the posted speed limit, and so on. All of these can potentially affect safety either through good design that minimises the potential of the road environment to create hazards or through neglect that leaves the road environment as a significant contributor to safety problems.

The road environment is also likely to have quite different impacts on safety between night and day. A design that may be quite safe in daylight may become very unsafe at night, because of poor visibility, glare that disorients the vehicle operator, and so forth. For example, lane markings that may be clearly visible in normal daylight may become invisible at night, especially when it is raining or the road is wet. Pedestrians and bicyclists may become very vulnerable at night because they are hard to see. In some countries, pedestrians wear small reflectors that are easily picked out by vehicle headlights and make the pedestrians much more visible. However, such reflectors are by no means universal, as also is the case with proper lighting for a bicycle at night.

Additional policy directions for road traffic safety

The chapter has already indicated a number of potential policy directions. There is also discussion on some aspects specific to freight movements in Chapter 15. A very important aspect of policies aimed at changing driver behaviour, whether of freight or passenger vehicles, is that of enforcement. If laws are passed and rules made that are aimed at changing driver behaviour, these will have relatively little success unless they are followed through with effective enforcement. For example, in Australia, there is extensive use of fixed speed cameras. However, it has been estimated that 90 per cent of Australian drivers drive at speeds of at least 10 km/h over the speed limit (personal observation), except in the immediate vicinity of speed cameras. The locations of fixed speed cameras are known and clearly signed in Australia. Drivers routinely slow down in the vicinity of these cameras (often to speeds that are at least 10 km/h below the posted speed limit). However, as soon as the camera is passed, the speeding resumes. Effective enforcement of speed limits is essential for speeding to be reduced. This is true for other changes to laws and rules of the road that are aimed at reducing accidents. For example, laws that prohibit the use of mobile phones whilst driving, that require the wearing of seat belts by all occupants of the vehicle, that prohibit tailgating behaviour, and so forth are all effective only to the extent that they are enforced and carry with them significant penalties.

As has already been mentioned, there are also some obvious bad policies in road traffic safety. These include varying speed limits within short segments of roadway, enacting unenforceable rules and laws on driving, and lowering speed limits far below engineering safe speeds. It may be necessary, in some instances, to effectively reduce the safe driving speed on some roads, in concert with lowered speed limits, to overcome a mismatch between apparent safe speed and the speed limit. This can be done by introducing roundabouts, traffic islands and speed bumps, permitting parallel parking along the side of the road, introducing bicycle lanes and bus lanes to reduce the capacity of some roads, and similar strategies that decrease the perceived safe speed. Referring back to Chapter 5, it can be seen that any of the strategies mentioned there that affect capacity and speed could be used effectively in conjunction with lowering speed limits. Planting trees along the kerbside, for example, is another strategy that can help reduce the perceived safe speed of a road.

Personal safety

Another aspect of safety that is sometimes overlooked is that of personal safety while travelling. In a number of countries, there are concerns with

potential mugging, robbery, assault, and other aspects of personal safety while travelling on public transport vehicles, especially at night. In some countries, there are also concerns with such crimes as carjacking. Personal safety concerns while travelling may be a major factor in pushing people away from some alternative means of transport, as documented by Currie et al. (2010) from a number of cities around the world. For example, there have been a number of studies in the US that have indicated that concerns about personal safety have a major effect on people's willingness to use public transport in some metropolitan areas (e.g. Loukaitou-Sideris 1999).

Research reported in the literature indicates that it is the perception of personal safety on public transport that has most effect on people's attitudes to riding on public transport, while at the same time indicating that these perceptions are often quite different from the facts. A person who has either been the victim of an attack on public transport or has witnessed an attack on another person is the most likely to consider public transport as being unsafe and actively to avoid using public transport. However, media reports of safety issues on public transport seem likely to exaggerate the risks, so that the risks that people perceive are not in line with the reality of those risks.

Literature relating to risks and actual personal safety issues for the car are much harder to find in the literature. Indeed, to a large degree, references to personal safety in private transport on the internet are limited predominantly to advice on how to maintain personal safety when using a car, but do not point to any research that has been done to associate perceptions of risks to personal safety that affect car use. The primary issues appear to relate to the location in which a car is parked and securing a car when it is parked. In a small number of countries around the world, there are serious concerns with carjacking, and advice is provided on ways to avoid high-risk areas where such crimes might occur. However, there is little else to be found on this topic.

Policy directions for personal safety

There should be at least three potential directions for policy relating to personal safety while travelling, especially for public transport. The first of these is to provide appropriate safeguards for persons using different forms of transport. For public transport, this is often the provision of CCTV or other surveillance equipment to deter criminal activity and to make it easier to identify the perpetrators, the presence on board vehicles of an alarm of some type that can be easily activated in the event of a situation arising that risks personal safety, clear display of instructions on what to do in case of emergency,

and ensuring that spaces through which people pass (e.g. on the way to a rail station or bus stop) and where they must wait are well lit, patrolled if necessary, and readily accessible to help. Parking places, especially those that are used after dark, should be well lit and should not afford easy places to hide for people who may have inappropriate intentions towards vehicle users.

The second policy direction is to put in place the means to discourage persons who might attack patrons using means of transport. As with issues on accident safety, enforcement is a strong requirement of policy in this regard. This means that there needs to be an adequate system in place that can prevent attacks from taking place and that can increase the public's awareness that they are not at risk when using public or private transport. This also leads directly into the third policy direction, which is to acknowledge that much of the concern about risks to personal safety may be perceptual and based on psychological issues. Hence, further research should be encouraged on how these perceptions are developed and what steps can be taken to change perceptions to fit the facts better. Part of this effort may also entail ensuring that the public is properly educated about the real risks and the steps taken to ensure personal safety while travelling.

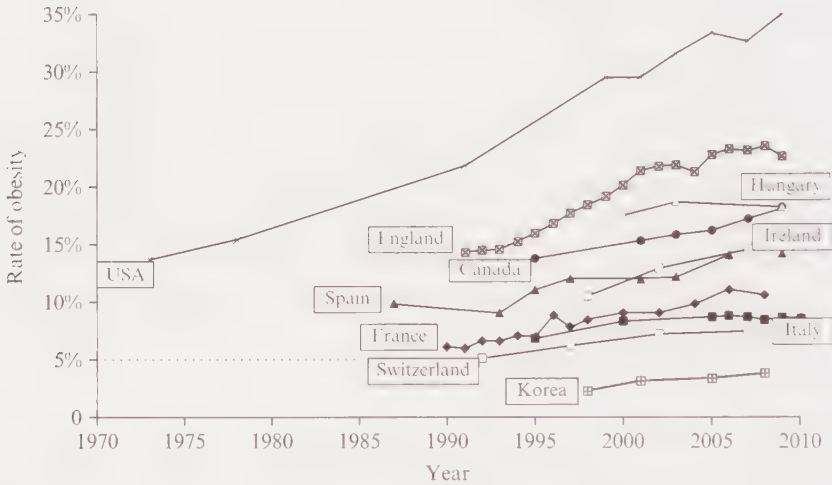
In addition to the above approaches, personal safety in the context of using public transport needs to be placed in a broader urban land use setting. Land use policies that increase the presence of 'eyes on the street' will tend to increase personal safety, and perceptions thereof, from the proximity of other people. Measures such as transit-oriented development that increase urban densities should contribute positively to personal safety outcomes on and around public transport.

8.3 Transport and health

Obesity

Obesity is defined as having a body mass index (BMI) greater than or equal to 30, where the BMI is calculated as the person's weight in kilograms divided by the square of their height in metres. There is no question that, around the world, there is a growing proportion of the population that is obese. The WHO defines a BMI greater than 25 as overweight and a BMI over 30 as obese. According to statistics from WHO (2012b), in 2008 about 0.9 billion adults (defined as those 20 years of age and older) were overweight, while a further 0.5 billion adults were obese, with more women than men being in this category (300 million women to 200 million men). The WHO also points out that overweight and obesity, which used to arise almost entirely in

high-income countries, have now spread to low- and middle-income countries. Overweight and obesity are implicated in more deaths than underweight, and are becoming increasingly an issue in health, with the incidence of diabetes, heart disease and cancer being linked to overweight and obesity to some degree. There is no question, then, that obesity and overweight are an increasing problem throughout the world. Figure 8.1 shows the growth in obesity for a number of countries for the past 20–50 years.



Source: OECD (2012)

Figure 8.1 Trends in obesity for the past 20–50 years in selected countries

At the same time that overweight and obesity have been growing in the population, motorisation of the population has also been increasing. In these same countries, car ownership has increased at annual rates of anywhere from 9.9 per cent per annum in Spain to 2.8 per cent in the USA over the 42-year period from 1960 to 2002 (Dargay et al. 2007). A number of researchers and commentators have linked growing obesity to growing car dependence around the world (e.g. Ewing et al. 2003; Frank et al. 2004; Pucher et al. 2010; Bassett et al. 2011). Before proceeding further, however, it is important to recognise some important facts about weight, exercise and statistics.

First, the idea that weight and exercise are related to one another is an idea that became popular only in the 1960s. Prior to that time, there was no significant credit given to the idea that exercise and weight are linked. Interestingly, it appears that this link was first put forward at the point in time where overweight was beginning to be seen as an issue in high-income countries. In previous generations, adults were not fixated on exercise as a part of a weight loss

programme. Indeed, such programmes were solely focused on what a person consumed in food, specifically intake of kilojoules or calories. Second and as a direct corollary to this, there is a statistical correlation between weight and exercise, but correlation does not measure cause and effect. To look at this more closely, a reduction of 1 kilogram of weight requires a person to burn 35 000 kilojoules of energy. Normal food intake for a day is about 11 200 kilojoules for a normal adult. Running 1 kilometre burns up 4 kilojoules per kilogram of body weight. Thus, a person weighing 80 kilograms will burn about 320 kilojoules by running 1 kilometre – to lose 1 kilogram of weight will require this person to run over 100 kilometres.

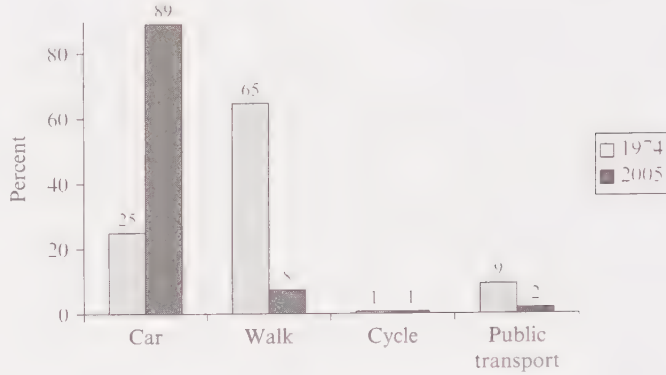
Third, the link between weight and exercise was based on limited evidence from experiments on rats that showed that exercise could lead to lower weight. However, several studies in the 1980s and 1990s demonstrated that weight and exercise are not related: some people lose weight through exercise, some gain weight, and some will have their weight remain unchanged. Indeed, there is also a suspicion that the relationship may go the opposite way: people who are overweight are less likely to exercise, while those who are not overweight may be inclined to get more exercise. Therefore, weight may be more a cause of how much exercise a person is willing to undertake, rather than exercise causing a reduction in weight.

The hypothesis put forward by proponents of the idea that car dependence and obesity are linked states that, by driving cars, people limit their exercise, compared to walking, bicycling, or using the bus or train. In turn, by limiting their physical exercise, those who use their cars will experience increasing weight gain as a result of their dependence on the car. Hence, one way to reverse the trend to increasing obesity in the population is to discourage the use of, and particularly dependence on, the car. Of even more importance, it is pointed out that children are being driven to school in greater proportions, especially in those countries where there is evidence of increasing obesity among children. An example of this trend is shown by data from the Melbourne, Australia suburb of Essendon, shown in Figure 8.2. As this figure shows, children being driven increased from 25 per cent to 89 per cent over the 31-year period, while walking to school dropped from 65 per cent to 8 per cent, and riding on public transport dropped from 9 per cent to 2 per cent. This is a massive shift from what are called active travel modes to passive (i.e. the car).

The record is quite clear that obesity is increasing worldwide. Car dependence and car ownership are also increasing worldwide, while walking, bicycling, and using public transport are generally in decline worldwide, although

Source: Peddie and Somerville (2005)

Figure 8.2 Change in modes of transport to school in Essendon, 1974 and 2005



recent statistics in some developed countries, such as Australia, show increases in public transport use and even public transport share in some cities. The important policy question is whether or not these facts are linked by cause and effect. There are other implications of overweight and obesity, however, that need to be considered beyond the possible link between car dependence and weight gain. Lavallière et al. (2012) suggested that obese people are more likely to be involved in road traffic accidents and that their injuries would be more severe. This is particularly the case because most cars are designed for a person weighing about 74 kilograms. When people are significantly heavier than this, they may not be adequately protected by the safety systems designed into the car.

There is a clear cause-and-effect link between food intake and weight, and especially the number of kilojoules and their source. It appears to be much more likely that most of the obesity epidemic that is sweeping the world is aided by both income growth and the car, in that these two factors lead to more time commitments, a willingness to eat fast food in place of healthier meals, ease of accessing fast-food outlets, and the income to support eating out and eating fast food more often. China provides a good example of this. There has been a rapid growth in obesity over the past 10 to 20 years in China, with just less than 5 per cent of the population now being obese. This represents a growth of about 97 per cent over the past decade. Most obesity is found in metropolitan areas, where also fast-food outlets have sprung up rapidly and the diet of urban Chinese people has changed from one that relied heavily on rice as a staple to a diet that includes an increasingly high proportion of energy-dense, nutrient-poor foods that are high in saturated fats and sugars. This diet change has probably by far the most to do with the increasing obesity in the Chinese urban population. While bicycles are still more prevalent in many Chinese cities than in European and North American cities, they are no longer the dominant form of transport, but have

largely been replaced by cars, which represent as much a status symbol or a symbol of wealth as they do a means of transport. It is arguable as to whether the rapid proliferation of fast-food outlets and other elements of Western diet would have occurred, however, without the increasing ownership of cars and the abandonment of the bicycle as a means of transport. To this extent, the car is clearly implicated in increasing obesity.

Policy directions

As with some of the other possible externalities of transport, it may not be that important if the link between obesity and car dependence is not a cause-and-effect process. If car dependence is a cause of obesity and policies are needed to reverse the trend of increasing obesity in the population, then it clearly follows that the policies that should be considered are ones that would shift people from car dependence into more active modes of transport. Coupled with this would be policies aimed at increasing the amount of exercise that people get as part of their normal daily life. What should be clear here is that, even though the link between car dependence and obesity may be indirect or non-existent, the policies that should be adopted if there were such a direct link will offer other benefits.

While we may question the extent of the link between exercise and weight, it is clear that moderate exercise is likely to improve overall health and longevity. Hence, transport policies that move people towards transport that offers more opportunity for physical activity, such as walking, cycling and public transport, are likely to improve the overall health of the population. Likewise, policies that encourage the use of walking, bicycling and public transport will have benefits in reduced use of carbon-based fuels, reduction of air pollution, reduced dependence on imported oil products, possibly reduced congestion levels, and even a reduction in the amount of investment required in the road system. All of these are beneficial results of such policies. As we have noted before, these are win-win policies – policies that confer benefits whether or not a controversial link exists.

Pollutant exposure

As is discussed in Chapter 7, most motorised vehicles generate various air pollutants. As a result, travel along vehicular rights of way, and possibly travel in vehicles along those rights of way, can be expected to expose travellers and those who live along the routes in question to pollutants. There are several factors in pollutant exposure, with the primary pollutants of concern being carbon monoxide (CO), nitrogen oxides (NO_x), particulate matter

(PM₁₀ and PM_{2.5}), and volatile organic compounds, especially those that are carcinogenic. The main factors that will affect the exposure to pollutants are:

- the volume and mix of traffic;
- the population in close proximity to the traffic;
- the local environment that affects dispersion or concentration of pollutants, including such things as urban canyons, weather conditions, and altitude;
- the time spent in the polluted environment; and
- the level of exertion of the subject in travelling through the environment.

Travelling along a roadway in a car that has its vent closed and with all windows closed will produce different exposure characteristics per trip than would travelling on a bicycle along the same street at the same time. In the first place, many pollutants would be effectively excluded from inside the car, compared to being on a bicycle. In the second place, the cyclist is expending much more effort to travel than the car driver, and is therefore inhaling larger amounts of air whilst travelling, and also moving more slowly through the polluted air, both of which aspects will tend to result in increased exposure levels. The concentrations of pollutants also are affected by location. For example, the heaviest concentrations of CO are likely to be found at the exits from car parks and parking garages, especially in the late afternoon, when cars that have been parked all day are exiting and are operating in cold start mode. A pedestrian waiting to cross a driveway at such an exit is likely to experience very heavy concentrations of CO. In addition, because CO is heavier than air, the concentrations do not disperse as readily as other pollutants, so the nature of the pollutant itself can have a significant effect on exposure.

A complicating factor, however, is that it is probably impossible to separate out the effects of vehicular pollution from other air pollution, except in the case of pollutants emitted only by vehicles. The heaviest concentrations of vehicular pollutants will often occur in those parts of the urban area where other air pollution sources are also concentrated, so that the pollution exposure issue is not simply a vehicular one.

In 1950s London, 'pea-soup' fogs were fairly common in the winter. Of course, in the London area, there was a heavy concentration of vehicles and vehicular traffic, so that part of the pollution trapped in these fogs (which today would be called smogs) came from vehicles. However, a major component of these smogs was actually a result of burning coal fires to heat homes, producing NO_x, carbon compounds, soot (PM₁₀ and PM_{2.5}) and SO_x. The pea-soup fogs were largely eliminated by the introduction of clean-burning

coke in place of coal, and changes to fireplaces to burn coke efficiently. The lesson from this is that non-vehicular pollutants in this case were probably the major constituents of exposure to pollution for Londoners, even though vehicular traffic was certainly a contributory factor. Nevertheless, removal of sulphur from vehicular fuels, as well as from heating fuel, and reductions in the overall level of vehicular pollutants were certainly important changes and resulted in a reduction of pollutant exposure for the London region over the ensuing years.

Much can be discussed about the relative pollution levels experienced when travelling by different means of travel. There is again some controversial research in this area, where it has previously been thought and generally accepted that travel by car creates the greatest exposure to pollution. However, a number of recent studies have found high pollutant concentrations in underground railways (and other railways that spend long stretches in tunnels), for bus passengers on buses that do not have closing doors on the passenger entrances and exits, and for cyclists and pedestrians in busy central city locations. In addition, exposure to vehicular pollutants is not limited to the travelling public. Those who live or work along major traffic arteries may experience prolonged exposure to vehicular pollutants. There is some UK research suggesting that lower-income groups are more exposed, because that is where they are more likely to live (Markovich and Lucas 2011).

Policy directions for mitigating pollutant exposure

In this book, it is not deemed useful to undertake a lengthy discussion of what environments are the most polluted, nor to look at the different merits of different means of travel. It probably is important to note that, while car travel may well involve extensive exposure to pollutants, resulting both from open vents and windows in the vehicle while travelling, and the concentrations of vehicular pollutants along busy roadways, especially where these roadways are congested, trains, buses, bicycling and walking, all of which are often seen as desirable substitutes for the car, are also themselves subject to varying levels of pollution exposure. Changing mode from car to rail is likely, in most cases, to result in reduction of the levels of most pollutants to which the traveller is exposed. However, if the rail is underground, much higher levels of PM_{10} and $PM_{2.5}$ may be experienced than on any form of surface transport.

On the other hand, it is clear that congested roadways will result in increased levels of pollutant exposure as a result of two factors: first, that vehicles that are moving in stop-and-go conditions will generally emit much higher levels of almost all pollutants; and, second, that congested conditions will result in

high concentrations of vehicles in the road space, thereby further increasing the exposure to high pollutant levels.

Mitigating pollutant exposure from traffic suggests five possible avenues. First, any policy that will result in lower concentrations of vehicles and less congestion will necessarily reduce exposure to vehicular pollution. Whether this is achieved by diverting travel from cars to other forms of transport, by other methods of reducing car travel or by increasing capacity for car travel is immaterial to this debate (assuming of course that capacity increases could be achieved in such a way as to lead to a permanent reduction in congestion). Second, reducing the levels of pollutants emitted by vehicles will also reduce the levels of pollutant exposure both for those who are engaged in travelling and for those who live or work in close proximity to heavily travelled roadways. This can be achieved by reducing emissions from all vehicles and/or changing the vehicle mix towards lower-emitting vehicles, such as by giving tax incentives for cleaner vehicles. Because such a policy is also effective in reducing overall levels of pollution, providing for cleaner air and reducing the incidence of smog and acid rain, such efforts are again capable of achieving multiple policy goals. Third, for those who are travelling and are inevitably exposed to vehicular pollutants, finding means to reduce the amount of pollutants that are likely to be inhaled will also mitigate this externality. For example, fitting cars with filters for air that is drawn into the vehicle, where these filters remove high percentages of pollutants from the air, and fitting buses with similar filters and ensuring that buses have doors that are opened and closed only at stops, will also reduce pollutant exposure for the travelling public. Fourth, because tunnels have been found to be major sources of concentrated pollutants, whether these are roadway or railway tunnels, adequate tunnel ventilation systems and pollutant removal systems are appropriate for situations where vehicular tunnels are used. Finally, careful attention to the types of developments permitted alongside heavily travelled roadways, use of vegetation to screen and absorb pollutants, and other similar strategies can also reduce pollutant exposure for those who live and work close to busy transport corridors.

Road rage

Because road rage can result in injury or death and is clearly an externality of vehicular travel, it is included here as a health-related externality. Road rage is defined as including a number of aspects of behaviour:

- sudden acceleration, sudden braking and close tailgating;
- cutting off other drivers, or preventing them from merging;

- excessive use of the horn and flashing headlights or headlight high beams;
- rude gestures; and
- shouting verbal abuse or threats.

These behaviours may also include intentionally causing a vehicle collision, getting out of the car to start a confrontation or damage another vehicle, threatening to use or actually using a firearm, and throwing objects from a moving vehicle with the intent of causing damage to other vehicles. The term ‘road rage’ appears to have come into use around the late 1980s in the US, and specifically in relation to some incidents of freeway shooting that occurred at that time in the Los Angeles metropolitan region. However, road rage probably existed many decades prior to that, if not hundreds of years earlier. There are certainly accounts from centuries past of one rider or driver of a horse whipping or otherwise trying to startle a horse or horses of another rider or driver, possibly leading to a horse bolting. However, recent years have seen an escalation of road rage in the various forms described above, especially in the US, but also increasingly in other countries around the world. It is currently estimated that there are at least 1200 reported road rage incidents per year in the US, although there are likely to be far more that are never reported. About 300 of these incidents in the US end in serious injuries or even fatalities (Rathbone and Huckabee 1999).

However, a major problem with the concept of road rage is that it has not been well defined, and statistics on its occurrence are not collected in consistent ways so that the extent of the problem can be clearly analysed. There are anecdotal stories that appear to indicate an increasing level of road rage. There are, however, unquestionably more drivers on the roads, so that, even if road rage were not increasing in terms of a rate per driver or vehicle, there would be an increase in the number of incidents.

A congressional hearing in the US concluded that congestion and time pressures are the primary causes of road rage (US Congress 1997). Others have speculated that overcrowded roads, inadequate roads, and the time pressures of modern life are all potential causes. However, all of this is speculative, because clear statistics and cause-and-effect links are not apparent. There are also anecdotal reports that using mobile phones, applying make-up, and other activities that are often done while driving may be contributory factors.

Policy directions for road rage

It is certain that, for this and other safety reasons, one of the primary strategies should be a strong emphasis on defensive driving behaviours. It is also

fairly probable that such an emphasis will not remove the problem, although it could possibly reduce it. Similarly, laws that prohibit use of mobile telephones while driving and also penalise other activities that are frequently seen to be done while driving may help to reduce accidents and incidents of road rage. There also is a need for increased policing of behaviours such as persistent tailgating, frequent lane changing, and inappropriate use of headlights and horns. Increased severity of penalties on conviction for damage or injury as a result of road rage may also help as a deterrent, but only if enforcement of the appropriate driving rules and laws is adequate. Finally, because road rage seems to arise with greater frequency under congested conditions, reduction of congestion will also be a potential policy direction that could reduce the incidence of road rage.

8.4 Valuing safety policies and programmes

Transport safety is an area where there has been extensive use of cost–benefit analysis to evaluate the merits of policy and programme initiatives, including major transport safety programmes implemented via regulatory changes (often evaluated in regulatory impact statements). The general approach is in line with that summarised briefly in Chapter 7 with respect to environmental impacts. This involves tracing out expected impact pathways for relevant policies and/or programmes on various stakeholder groups and placing unit money values on such impacts. For example, in seeking to evaluate the safety benefits and costs of strengthening bus and coach seats, a policy analyst would need to estimate (*inter alia*) the probability of bus accidents and how such seat strengthening is likely to affect the number of fatalities and serious injuries in the event of collision. Unit values for lives lost and for serious bus injury accidents would then be applied to these expected changes in fatalities and serious injuries, to estimate relevant benefits. Seat manufacturers would be able to provide relevant cost estimates for seat strengthening, with bus body builders also needing to be involved in costing, because of possible impacts on vehicle body requirements.

Perhaps the most contentious issue in economic evaluation of transport safety improvements, which is also relevant in evaluation of some health impacts of transport environmental effects, concerns the valuation of lives saved or lost. Some jurisdictions value life in terms of lost production, estimating the loss of economic output if (for example) someone dies in a traffic accident. This approach has, in our view, some morally reprehensible implications, since it implies that the death of a retired person might have positive benefits!

Our preferred approach to valuing human life in transport safety evaluations is to apply the willingness-to-pay approach outlined in Chapter 3. In a safety setting, this approach looks at risk situations and estimates people's willingness to pay to reduce risk. Hensher et al. (2009) derived a value of A\$6 million per life lost in an Australian transport setting, using this approach.

8.5 Conclusions

In the areas of health and safety, there are externalities that are indisputable and others that are questionable. Use of any form of transport can result in property damage, injury and death, both to users and to bystanders. Similarly, use of transport may, in certain circumstances, involve risks to personal safety. Whether use of the car contributes to obesity or not is still a somewhat controversial issue, with some professionals convinced of the link and others who are dubious or who even outright deny the existence of such a link. Exposure to pollutants is clear, but the extent of it and how to separate effects from other sources of air pollution is much more difficult to determine. Road rage may or may not be increasing in incidence, and its root causes remain somewhat obscure.

Nevertheless, there appear to be a number of policy directions that may be effective in reducing several of these potential externalities. Reductions in congestion, improved driver behaviour training, and shifts from driving to alternative modes, such as bus, train, cycling and walking, all seem likely to have a positive benefit on most if not all of the health and safety externalities. These same policies have been cited as potentially beneficial for other externalities of transport as well. The policy maker who seeks to find policies that will have benefits in a variety of areas and in mitigating various externalities is unlikely to go wrong in pursuing such policy directions as these, irrespective of whether or not they can be proven to mitigate specific externalities of transport or shown to be appropriate in a cost–benefit analysis, because they achieve other laudable goals for transport development.



REFERENCES

- Bassett, D.R., J. Pucher, R. Buehler, D.L. Thompson and S.E. Crouter (2011), 'Active transportation and obesity in Europe, North America, and Australia', *ITE Journal*, **81** (8), 24–8.
- BITRE (2012), *Road Deaths Australia 2011 Statistical Summary*, Canberra: Department of Infrastructure and Transport, Bureau of Infrastructure, Transport and Regional Economics.
- Currie, G., A. Delbosc and S. Mahmoud (2010), 'Perceptions and realities of personal safety on public transport for young people in Melbourne', conference paper delivered at the Australasian Transport Research Forum, Canberra, Australia, 29 September – 1 October.
- Dargay, J., D. Gately and M. Sommer (2007), *Vehicle Ownership and Income Growth, Worldwide:*

- 1960–2030, available at: http://www.xesc.cat/pashmina/attachments/Imp_Vehicles_per_capita_2030.pdf (accessed 26 September 2012).
- Ewing, R., T.L. Schmidt, R. Killingsworth, A. Zlot and S. Raudenbush (2003), 'Relationship between urban sprawl, and physical activity, obesity, and morbidity', *American Journal of Health Promotion*, **18** (1), 47–57.
- Frank, L., M.A. Andresen and T.L. Schmidt (2004), 'Obesity relationships with community design, physical activity, and time spent in cars', *American Journal of Preventive Medicine*, **27** (2), 87–97.
- Hensher, D.A., J.M. Rose, J. de Dios Ortúzar and L.I. Rizzi (2009), 'Estimating the willingness to pay and value of risk reduction for car occupants in the road environment', *Transportation Research Part A*, **43** (7), 692–707.
- Lavallière, M., G.A. Handrigan, N. Teasdale and P. Corbeil (2012), 'Obesity, where is it driving us?', *Journal of Transportation Safety and Security*, **4** (2), 83–93.
- Loukaitou-Sideris, A. (1999), 'Inner city commercial strips: evolution, decay or retrofit?', *Town Planning Review*, **68** (1), 1–29.
- Markovich, J. and K. Lucas (2011), 'The social and distributional impacts of transport: a literature review', Working Paper No. 1055, Transport Studies Unit, School of Geography and the Environment, University of Oxford, March.
- M.R. Parker and Associates, Inc. (1992), *Effects of Raising and Lowering Speed Limits*, Report FHWA-RD-92-089, October, Washington, DC: US Department of Transportation, Federal Highway Administration.
- OECD (2012), *Obesity Update 2012*, Paris: Organisation for Economic Co-operation and Development, available at: <http://www.oecd.org/els/healthpoliciesanddata/49716427.pdf> (accessed 26 September 2012).
- Peddie, B. and C. Somerville (2005), 'The ghost of TOD's past: schools reconnecting', paper presented at the Transit Oriented Development Conference, Fremantle, WA, Australia, July.
- Pucher, J., R. Buehler, D.R. Bassett and A.L. Dannenberg (2010), 'Walking and cycling to health: a comparative analysis of city, state, and international data', *American Journal of Public Health*, **100** (10), 1986–92.
- Rathbone, D.B. and J.C. Huckabee (1999), 'Controlling road rage: a literature review and pilot study', prepared for the AAA Foundation for Traffic Safety, 9 June.
- UN (2003), *Global Road Safety Crisis: Report of the Secretary-General*, United Nations General Assembly, 7 August, available at: http://www.searo.who.int/LinkFiles/whd04_Documents_un_secretary_general_report_on_road_traffic_injuries.pdf (accessed 21 September 2012).
- US Congress (1997), *Road Rage: Causes and Dangers of Aggressive Driving*, hearing before the Subcommittee on Surface Transportation of the Committee on Transportation and Infrastructure, House of Representatives, One Hundred Fifth Congress, first session, 17 July.
- WHO (2006), *The Global Burden of Disease: 2004 Update*, World Health Organization, available at: http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html (accessed 20 September 2012).
- WHO (2009), *Global Status Report on Road Safety: Time for Action*, World Health Organization, available at: http://whqlibdoc.who.int/publications/2009/9789241563840_eng.pdf (accessed 20 September 2012).
- WHO (2012a), 'Road traffic injuries', Fact Sheet No. 358, World Health Organization, Media Centre, September, available at: <http://www.who.int/mediacentre/factsheets/fs358/en/index.html> (accessed 25 September 2012).
- WHO (2012b), 'Obesity and overweight', Fact Sheet No. 311, World Health Organization, Media Centre, September, available at: <http://www.who.int/mediacentre/factsheets/fs311/en/index.html> (accessed 26 September 2012).

Tackling the externalities – congestion



CHAPTER OVERVIEW

The third chapter on externalities is concerned with congestion. The chapter starts out by defining what is meant by congestion, noting that there are differences between the way in which traffic engineers and economists define congestion. Two types of congestion – incident-related and recurring – are discussed. The discussion continues by defining mobility and accessibility and noting the effects on each of different policies relating to mitigating congestion. The balance of the chapter examines the implications of microeconomic theory for congestion, discusses briefly the effects of market forces on congestion and concludes with a discussion of potential policy directions to deal with congestion.

9.1 Defining congestion

The topic of congestion was introduced in Chapter 5 on traffic theory foundations. That chapter presented definitions of congestion and pointed out that economists and traffic engineers have different definitions. These different definitions can influence assessment of the significance of congestion as a transport problem, because (for example) they impact on the measurement of the scale of congestion costs. However, it is common for congestion costs to be of the order of 1 per cent of gross domestic product, which represents a huge economic waste. For example, Australian road congestion costs in 2005 were estimated at A\$10 billion (based on a deadweight loss measure of congestion costs). This was close to 1 per cent of national GDP at the time (BITRE 2007).

For policy clarity, it is important to be clear about exactly what is being defined and measured as congestion. The discussion in Chapter 5 is important to such clarification. Whichever definition is ultimately used, however, congestion is a clear externality of the transport system, in that the costs paid by users do not reflect their contributions to the costs experienced by

others that are due to congestion. This externality characteristic and the scale of congestion costs, however measured, make congestion a prime focus for transport policy.

In this chapter on congestion, we start by defining two different types of congestion – incident-related congestion and recurrent congestion. Before defining these two types of congestion, a general definition of congestion would be helpful. Congestion can be defined as extreme use or overuse of a facility.

In traffic terms, traffic congestion arises when the input volume exceeds the output capacity. This could be illustrated rather readily by considering the action of pouring a liquid into a bottle. Assuming that the bottle has a neck that is narrower than the bottle itself, then 'congestion' arises when one tries to pour liquid into the bottle at a rate faster than the capacity of the neck allows. In this case, liquid then spills over the sides of the bottle. Hence comes the term in traffic congestion (and elsewhere) of a 'bottleneck', which is usually the cause of traffic congestion. In the case of traffic, unlike the bottle example, if the input volume exceeds the capacity of the neck, then traffic backs up waiting to get through. The back-up of traffic can be long and severe if the input volume is much higher than the output capacity, and if the high input volume lasts for some time. Outside of this, there is no such thing as 'link' congestion, because congestion can only arise through the input volume exceeding the output capacity. This occurs either because of a narrowing of a facility or because of the intersection of another facility, such that the downstream capacity is lower than the combined capacities of the two upstream facilities.

Two types of congestion

Incident-related congestion

Non-recurring congestion is also described as incident-related congestion. It is usually caused by such things as traffic accidents, roadworks, special events, temporary road closures and so on. This type of congestion is not usually related to infrastructure problems. It often occurs at unpredictable times and locations. It is best managed by quick response methods, but is not generally susceptible to long-term solutions relating to infrastructure, unless it arises from traffic accidents that are caused by design faults in the infrastructure.

There are a number of potential directions that can be used to reduce incident-related congestion. One of the best methods that is being employed

increasingly around the world is the use of intelligent transport systems (ITS) capabilities to provide variable message signs and in-vehicle messages, and broadcast information to drivers, such as through radio stations. However, these methods require that there is a high level of vigilance of the road network and that the messages and signs are kept meticulously up to date. Variable message signs and in-vehicle messages that warn of incidents that have been cleared already are of little value and tend to instil in motorists an attitude of ignoring the information provided as being unreliable. It is imperative that, as soon as the incident is dealt with, the messages concerning the incident are immediately removed. A lag of minutes in so doing can already start to persuade motorists of the unreliability of the system.

A second strategy to deal with incident-related congestion is to improve response times by police and tow trucks, so that accidents and breakdowns are cleared out of traffic lanes as rapidly as possible. Recent efforts in some countries to provide additional capacity to major roadways by utilising the shoulders and medians for traffic lanes unfortunately militate against rapid clearance of traffic accidents and breakdowns, by removing the space that could be utilised to deal with inoperable vehicles.

In the Sydney region in Australia, there is a good example of a strategy to reduce incident-related congestion through partially an infrastructure solution. To the north of Sydney, there is a freeway known as the F3 freeway. The freeway traverses areas of hills and bushland, with few access roads and extremely limited parallel routes, making access to incidents difficult and alternative routings largely impossible. This freeway seems to experience more than its fair share of traffic accidents and incidents, possibly because of the numerous gradients along the route. Because of the overall narrowness of the right of way, an accident or breakdown often has the capability of completely shutting down one direction of travel, sometimes for long periods of time. Until a few years ago, there were almost no locations along the freeway where crossover lanes were provided that could allow traffic in the blocked direction to be rerouted into one or more lanes of the opposing direction. However, the decision was finally made to build a number of crossovers to allow contra-flow operation to be put into place in the event of a serious blockage in one direction. While building such crossovers was a multimillion-dollar investment, the ability to provide contra-flow operation to get around a serious incident has saved enormous amounts of congestion delays in the relatively short time that these crossovers have been available to use.

Other strategies that can be used to mitigate incident-related congestion include changes in laws to permit damaged vehicles to be moved out of traffic

streams quickly, use of road sensors and CCTV cameras to allow controllers to spot incidents sooner, improved management of roadworks, so that they do not impinge on heaviest traffic flow periods, and improved management of special-event traffic. Again, in many parts of Australia, roadworks on major roadways are undertaken only at night, with traffic restored to using the full facility during daylight hours. This means that, in many instances, roadworks-related congestion is much more limited than would be the case if the roadworks were performed in daylight hours.

Thus, policy directions for incident-related congestion have primarily to do with introducing more rapid detection, better advance warning and more rapid response, and investing in the means to maintain traffic flow in both directions in the face of major incidents. In addition, scheduling of necessary road maintenance and repair activities to low-volume hours, and improved management of incidents will also contribute to ameliorating the effects of incidents on traffic congestion. However, it is clear that incident-related congestion cannot be eliminated, given the human factor in vehicle operation, the necessity for maintenance and repair of rights of way, and the occurrence of special events that impose unusual traffic flows on the system.

Recurring congestion

Recurring congestion is the topic that the remainder of this chapter covers. Most often, recurring congestion occurs wherever there is a bottleneck in the system, such as:

- a permanent reduction in the number of lanes;
- a reduction in lane widths, which results in a reduction in capacity, even though the number of lanes may remain unchanged;
- the merging of two or more roadways into one roadway, where the latter has fewer lanes than the former;
- traffic lights that are in failing cycles (a failing cycle is one where the queue of traffic at the red light is not completely cleared during the following green phase); and
- any other permanent reduction in road capacity.

These situations alone are not sufficient to result in congestion, but will do so when the input volumes are persistently higher than the output capacity. For example, two two-lane roadways may merge into another two-lane roadway. If the combined input volume on each of the two roadways is less than the capacity of a two-lane roadway, congestion will not normally occur. Another

example of where one of these conditions is not likely to result in congestion occurs on a radial freeway in a metropolitan area, where the number of lanes decreases incrementally as the distance from the CBD increases. Such a freeway might have eight lanes (four in each direction) near the downtown area of the city, decrease to six lanes at some significant distance from the downtown, and then decrease again to four lanes in the outlying suburbs. Such decreases as one leaves the centre of the city are not likely to cause congestion if well planned, because the traffic volumes are also decreasing as the distance from the CBD increases.

It is interesting to note that, in most instances outside transport, congestion would be considered to be a sign of success, for example if it occurred at a sports venue, a theatre or a shopping centre. In such cases, congestion would be indicative of levels of use that exceeded expectations and would most probably be accompanied by increased profitability. Some writers have recently questioned whether or not congestion is necessarily as negative as it has often been painted (Taylor 2002), and researchers at the University of California at Davis found a significant proportion of commuters who valued congestion positively (Redmond and Mokhtarian 2001). Among the benefits cited for congestion or lengthy commutes was the fact that the time spent in the car driving between home and work often represented the only time for these commuters that they were alone and free to listen to their type of music, smoke, think things through without interruption, have decompression time from work to home, or planning time from home to work, and talk on the phone to whomever they chose.

In recent years, there has been continuing urbanisation of the population throughout much of the world, so that populations that are growing in most countries anyway are to be found more and more within urban areas. This increasing urbanisation, coupled with growth in population and, in many countries, growth in car ownership, leads to an increasing level of traffic congestion in many urban areas. Yet, despite the increases in congestion, people do not choose, in most cases, to change to public transport, walking or bicycling, but rather continue to seek car ownership and to use their cars. In many countries around the world, from the USA to China, growth in car traffic has outpaced growth in the road network for at least the past two to three decades. In addition to urbanisation and car ownership growth, there has also been a trend for cities to decentralise to a fairly substantial degree, so that many of the movements within cities are becoming suburb-to-suburb, rather than suburb-to-CBD. This change makes it harder for established public transport, especially rail, to serve urban area movements.

Traffic congestion is also neither a developing nation nor a developed nation phenomenon, but rather occurs in major cities of almost every country. Los Angeles, New York, Miami and Baton Rouge all experience traffic congestion of varying severity. So also do Mexico City, Beijing, Shanghai, Bangkok, Manila, Santiago, London, Paris, Berlin and many other cities, too numerous to mention. Thus, congestion appears to be widespread and to affect most conurbations around the world.

Effects of recurring congestion

There are a number of effects of recurring congestion that represent some of the externalities of traffic congestion. Most obviously, traffic congestion causes traffic delays and introduces increasing unreliability in journey times. One of the authors of this book spent two weeks working in Bangkok, and found that congestion there was so severe that it could take an hour or more to get from one side of the CBD to the other, but the variability in travel time was such that it could sometimes take as little as 30 minutes and at other times as much as 90 minutes. If one had to attend an important meeting, considerable time could be wasted if traffic was lighter than usual and one arrived an hour early for the meeting, because it was necessary to allow for the worst case.

Traffic congestion also causes increases in vehicular emissions, because stop-and-go driving results in very inefficient operation of most vehicular engines, with heavy production of unburnt VOCs, NO_x and particulate matter, especially from wear on the brakes and gears. There are also impacts on the living and working environment for those living or working in close proximity to a congested road. These include noise, vibration, air pollution, changes to the character of the neighbourhood, difficulty in crossing the road, causing possible neighbourhood disruption and social exclusion, and so forth. On the positive side, however, traffic congestion also indicates a high level of use of the congested facilities, which is, to at least some extent, an indicator that the roadways are placed where they are needed.

Assuming that many people find congestion to be stressful and also that it wastes time and creates unreliability, one would assume that people would attempt to avoid congestion, by driving at different times, by changing one or both ends of their trips or by changing their mode of travel, so that congestion would not be encountered. However, not only does this not seem to happen, but instead congestion continues to increase. On the commercial side, over the past several decades, there has been a pronounced move of freight traffic from rail to road and also a shift to just-in-time freight systems,

which add to the congestion issues on existing roads. Despite mounting congestion in many urban areas around the world, causing increasing unreliability for road-based freight movements, especially within urban areas, this trend seems to continue, with more traffic moving away from rail and on to the roads.

9.2 Historical background to recurring congestion

Recurrent traffic congestion unquestionably occurs throughout the world today. As noted previously, examples of traffic congestion can be drawn from almost every country of the world and from most major and many minor cities. Traffic congestion outside urban areas tends to be somewhat less common, but still occurs, as witness the 2010 multi-day traffic jam in China (Chang 2010), which spanned a distance of approximately 100 kilometres in Inner Mongolia and Hebei province on the Beijing–Zhangjiakou highway. However, it is reasonable to ask the question as to whether traffic congestion is a modern phenomenon or not. Figure 9.1, from 1807 in London, shows that, more than 200 years ago, traffic congestion was already a serious issue. There is evidence to suggest that ancient Rome experienced traffic congestion, with Julius Caesar banning wheeled traffic within the walls of Rome between dawn and dusk. Historically, congestion seems to be a consistent feature of humans organising living into conurbations.



Figure 9.1 Traffic congestion in London in 1807

Some more recent examples of traffic congestion are provided by some interesting quotations. In 1837, Asa Greene wrote:

But broad as Broadway is, and exceedingly broad as it was doubtless thought by our fathers and grandfathers, it is now quite too narrow for the immense travel, business, and locomotion of various kinds, of which it is the constant scene . . . [T]he attempt at crossing is almost as much as your life is worth. To perform the feat with any degree of safety, you must button your coat tight about you, see that your shoes are secure at the heels, settle your hat firmly on your head, look up street and down street, at the self-same moment, to see what carts and carriages are upon you, and then run for your life.

Continuing with Manhattan, Hood (1995) noted that the major complaints regarding congestion in the 1860s in Manhattan had to do with vehicular congestion on the streets and overcrowding in public transport vehicles. He notes that one commentator described peak-period streets in Manhattan as 'a solid mass of braying, animal-powered vehicles'. He also quotes Walt Whitman as describing 1860s Manhattan as '[similar to] military battles, where regiments and platoons clashed in violent disarray. Wagons, lorries, carriages, and omnibuses moved at different speeds, manoeuvred in and out of traffic, and dodged from one side of the street to another' (Hood 1995, p. 40).

Searching the internet will provide numerous pictures from the late nineteenth and early twentieth century showing congestion in London, New York and many other cities. It is clear from these quotations and pictures that congestion is not a modern phenomenon, but has been an attribute of major cities for many years in the past. It seems appropriate, then, to pose the question as to why the solution to congestion has not been found. Very probably, the reason that a solution has not been found is that property rights have not been allocated, so that congestion becomes a clear illustration of Hardin's (1968) tragedy of the commons.

9.3 Responses to traffic congestion

Many politicians bemoan the congested state of roads in various urban areas. Candidates for public office are frequently heard to claim that, if elected, they will reduce or even eliminate congestion. Many transport plans over the years have also promised relief of congestion. Most often, the route proposed to reduce or eliminate congestion is that of adding capacity to the road system. However, for reasons that become clear from the following discussion, this is unlikely to reduce or eliminate congestion.

Accessibility and mobility

To understand the reason that adding capacity will be unlikely to mitigate congestion, it is first appropriate to discuss the twin concepts of accessibility and mobility. Often these two terms are used incorrectly as meaning the same thing. However, they actually describe two different concepts, and understanding these differences is extremely important. Accessibility is related to the ease of reaching specific destinations. Thus, accessibility is a function of both the performance of the transport system and the locations of available destinations. To help explain this, consider two situations. The first is a small town in England in the 1950s. The high street has numerous small shops and stores along it, each one selling different categories of items, such as a green-grocer, a grocer, a butcher, a baker, a hardware store and so on. There is also a post office, bank branches, hairdressers and stylists, dentists' and doctors' surgeries, and many other services, scattered in among the various shops and stores. Residences are located on streets running off the high street, and a majority of residents can reach the high street on foot within five to eight minutes. The shops and stores are also located in a fairly compact area, so that the pedestrian shopper is able to go from one shop to another within a rather short space of time. Residents of this small town largely do not own cars, although a number may have bicycles. However, these same residents have a high level of accessibility for shopping for consumables. Neither distances nor times of travel are substantial, and the lack of a car does not impede the residents from accessing the goods and services required. These residents would be described as having a high level of accessibility.

Now, we move forward in time to the 2010s. Most of the individual shops and stores exist no longer. Instead, these have been replaced by a large regional shopping centre, located 15 kilometres away. Similarly, many of the health and medical services have moved and are located in medical centre complexes, others are in the regional shopping centre, and others are located only in the next larger town some 20 kilometres away. People residing in the same homes in the small town now need to travel by car or public transport to access these goods and services. Much longer distances must be travelled, and probably the time taken is longer, especially if residents need to visit two or three of these more remote 'centralised' business locations. The residents no longer enjoy such a high level of accessibility, even if they have subsequently acquired cars, because the distances and times required to travel have been increased. Of course, there may now be a greater variety of goods and services that can be accessed, so the argument could be made that the quality of the goods and services on offer, as well as the quantity, has increased. Nevertheless, this has come at the expense of accessibility.

The concept of mobility, however, has to do simply with the ease of travelling. A person who has no other means of travel than his or her feet is limited in mobility, because of the speed of movement and the potential range that can be covered. Gaining a wheeled vehicle, provided that there are appropriate rights of way on which it can be used, will usually add mobility. Once a wheeled vehicle is owned or available, then mobility becomes a function of the extent of the network available over which to operate the vehicle.

Let us consider again the resident of the 1950s small town in England. If this resident had no bicycle, motorcycle or car, and there was a restricted or no bus service available, then he or she would have very little mobility. Living close to the high street, this person would have good accessibility but poor mobility. Obtaining a bicycle or a car would have added significantly to mobility, as would provision of bus services to link to other places. However, such additional mobility might not have improved accessibility if the places to which the person could travel did not offer possible consumable shopping opportunities that were greater than those already available on the high street. In the 2010s, if the resident now owned a car or cars, then mobility would be greatly increased, especially if there was an extensive road network, and possibly also freeways or motorways available for use. Hence, the person might now have less accessibility, because of the agglomeration effects on consumable shopping outlets and services, but greater mobility, because of the extensive road system, including motorways and other high-performance roadways.

In short, mobility can be ascertained simply by determining access to vehicles (both freight and passenger) and the level of performance of the network on which the available vehicles can be operated. Accessibility, on the other hand, is a function of both the level of performance of the appropriate network and also the land use development pattern that determines what destinations are available. Mobility can be measured in terms of *cost per kilometre of travel*, whereas accessibility is measured in terms of *cost per destination*.

Given this understanding of mobility and accessibility, the effects of congestion and capacity increases can be understood more clearly. As congestion increases, the costs of travel per kilometre must necessarily increase, especially because the costs considered here are the generalised costs of travel, that is, those costs that include not only the monetary outlay, but also such things as time, convenience and reliability. Similarly, if there is no change in the land use pattern, the costs per destination will also increase. Therefore, congestion decreases both mobility and accessibility. An increase in capacity, in the short run, will lead to a reduction in the costs per kilometre and per

destination, for the same reasons that congestion increased both. Thus, in the short run, capacity increases will result in improvements to both mobility and accessibility. The long term will, however, see a different result. To understand the longer-term effects of capacity increases, it is necessary to return to the underlying theory of travel.

Microeconomic theory and capacity increases

As discussed in Chapters 4 and 5, it is possible to consider the behaviour of person travel in the microeconomic paradigm of supply and demand. As derived in Chapter 5, the supply or, more correctly, price–volume curve for a road is an upward-sloping curve, identical to the supply curve for most goods and services. The curve implies that, at low volumes of use, the price of travel remains at or near the minimum for the road. As volumes increase, however, the price of travel will start to rise, which is a sign of congestion starting to have an impact on flow and on user costs. As one approaches the capacity of the road, the price starts to rise rapidly and capacity may be reached. The demand curve is a downward-sloping curve, indicating that, all other things being equal, as the price decreases, the amount of travel demanded will increase. Key factors that are taken as fixed in any representation of a demand curve for road travel include, for example, population, incomes, tastes, and the price and quality of complementary and competing goods and services (public transport service levels and fares are included in the latter). These standard curves are shown in Figure 9.2.

The initial demand (D_1) and price–volume (supply) (S_0) curves are the ones that intersect in Figure 9.2 at V_0 and P_0 . As time goes by, and wealth

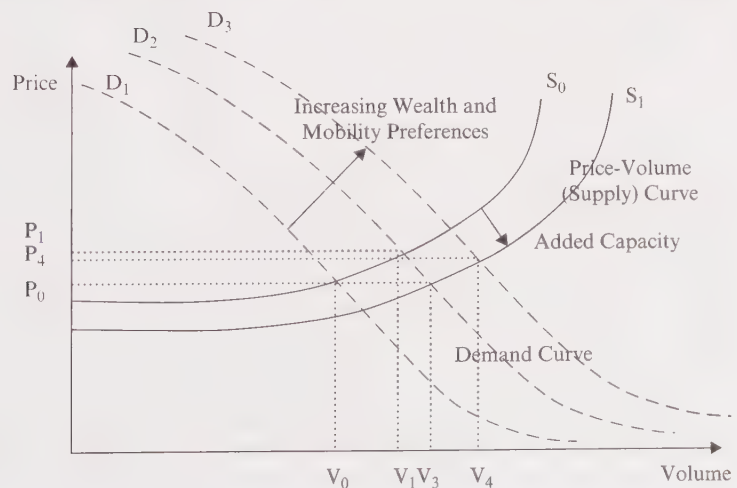


Figure 9.2
Microeconomic theory
applied to road traffic

increases, along with changing preferences for mobility, the demand curve shifts upwards and to the right in Figure 9.2 to D_2 . At some point in time, the new intersection of demand and price–volume shifts upwards to P_1 and V_1 . This represents increasing congestion – the price of travel is increasing and affecting both mobility and accessibility. If capacity is now added to the roadway, this will have the effect of lowering the price–volume (supply) curve, as shown by the lower one of the two price–volume curves (S_1), and a new intersection between demand and price–volume will occur at P_0 and V_2 . In short, for at least some time, the initiative that lowered the price–volume curve has enabled a greater traffic flow without increasing average trip costs.

However, with the lowered price of travel, there will now be an increase in demand, which is shown by the initial change from V_1 to V_2 . As wealth and mobility preferences continue to change, the demand curve can usually be expected to continue moving to the right, so that, in time, the new intersection of demand and price–volume will occur at P_1 and V_3 . As these further increases take place in wealth and mobility preferences, the actual price of travel will eventually exceed that at which the capacity was originally added, and the volume will now be much greater. Thus, the addition of capacity has made possible a substantially larger volume of travel, but the continued escalation of price has not been stopped and is likely to continue, with further increases in demand over time.

Interpreting this process in the light of the earlier discussion on mobility and accessibility, it is clear that the initial effect of a capacity increase is an increase in both mobility and accessibility, resulting from the reduction in price of travel. However, as time goes by, mobility continues to increase (more people can travel on the enlarged facility), but accessibility starts to decline. Eventually, a situation is reached in which accessibility is reduced to the same or a lower level than prior to the capacity increase, although mobility has been increased, as evidenced by the larger volume of travel. One might ask if this phenomenon is just a theoretical development from microeconomics, or if there is reality in it. A good example is provided by the situation in San Francisco around the time of the opening of the Bay Area Rapid Transit (BART) line across the San Francisco Bay from San Francisco to Oakland. Prior to the opening of the BART tunnel from San Francisco to Oakland, the Oakland Bay Bridge was heavily congested in both peak periods. When BART was opened across the bay (a line running parallel to the bridge), the cross-bridge volumes of travel dropped significantly and immediately. At the same time, a significant volume of travel was being carried by the new BART connection across the Bay, representing travel diverted from the bridge to BART. It was reported, however, that within six

months the volumes of travel across the Oakland Bay Bridge had returned to their pre-BART levels, although, by then, about 55 000 rides per day were being carried by BART across the Bay. Clearly, this increase in capacity had resulted in an increase in mobility (55 000 additional trips per day across the Bay), but the price of travel on the Bridge had returned to previous levels, so there was no improvement overall in accessibility, only in mobility.

The demand curve moving to the right is partly generated by real increases in wealth, but is also a function of mobility preferences. In particular, it can be postulated that these mobility preferences represent a combination of a desire to be able to travel and an increasing tolerance of the levels of congestion. The fact that increasing congestion has rarely been seen to lead to decreases in the amount of travel, but rather that congestion continues to increase, suggests that this tolerance for congestion is also increasing around the world. However, a decline in per capita car use has been observed in some countries in recent years (e.g. Australia, the UK and the US), with increasing congestion arguably one of the factors contributing to this change in use.

If tolerance for congestion continues to grow, and acceptance of increasingly serious congestion also grows, then there is not likely to be a possibility of ever reducing congestion levels permanently into the future, other than through strong congestion pricing measures. Growing acceptance and tolerance of congestion will mean that efforts to reduce congestion by building more capacity will be doomed to fail, because more travel will be undertaken to return the expanded facilities to similar and eventually worsening levels of performance to those that existed prior to the capacity addition.

In short, it is not possible to build one's way out of congestion. Capacity increases will lead to increases in mobility, but continued deterioration in accessibility.

A closing comment on this subsection relates to the subject of Chapter 12 in this book. We have stressed here that capacity increases alone are not likely to provide a long-term solution to congestion reduction or elimination. However, it is important to note that capacity increases that are provided with a toll or other road user charge associated may have the potential to reduce traffic congestion. Looking at Figure 9.2, it should be clear that, if new capacity is provided with an associated charge, then the price of travel at any given volume will be higher as a result of the road user charge. If the road user charge is equivalent to the price reduction created by the added capacity, then volumes of travel will be largely unchanged with the capacity addition, but travel will be faster and less congested. If the charge that is levied

is more than the cost difference, then the amount of travel can be decreased by the capacity addition, a phenomenon that has probably occurred in three instances of late in Australia. The Sydney Cross-Town Tunnel, the Lane Cove Tunnel in Sydney, and the Airport Link Tunnel in Brisbane have all failed to raise the expected revenue from the tolls imposed, most probably because the amount of the tolls exceeded the price reduction arising from the capacity increase. Traffic through each of these tunnels failed to reach the projected levels of travel that would have permitted the tolls to repay the construction costs. While this outcome is looked on as an economic failure, in the sense that the tolls imposed are not raising the necessary revenue to repay the costs of building the tunnels, this outcome is quite clearly a success in terms of reducing congestion, because traffic using these tunnels largely enjoys much less congested travel as a result of the deterrent effect of the tolls. Thus, it can be concluded that adding capacity and imposing road user charges of some amount can lead to a reduction in congestion if the charges are sufficiently high.

More generally, setting road user charges such that users are confronted by their marginal congestion costs will produce a situation where the resulting level of congestion is economically efficient. The relevant price will not lead to travel taking place at free speed, but congestion levels will be less than in a situation where pricing solutions are not used. Pricing congestion is probably the only effective way to achieve lasting reductions in congestion levels, a matter to which we return in Chapter 12.

Congestion and the marketplace

Given an understanding of what occurs with congestion and also capacity additions, it is appropriate to consider next how the marketplace tends to react to increasing congestion. Although the preceding discussion suggested that congestion is likely to continue to grow, there is a range of different responses of the marketplace to continuing congestion. The most immediate and often the least disruptive change that people make is to change their times of departure so as to avoid the worst of the congestion. Because usually only some people have the flexibility and also the desire to change time of departure, this results in the phenomenon of peak spreading, where the length of time of the peak becomes longer, although the peak may be slightly less peaked. A second strategy to cope with congestion is to change one's route of travel to avoid the most severe congestion. This, however, leads to congestion on other routes, so results in congestion spreading. A third strategy is to choose public transport over private transport, especially when rail or bus on its own right of way is present as an alternative mode to the car.

If buses are subject to the same congestion as the car, then switching modes to avoid congestion will largely not be possible and will not take place. When less congested public transport is available, such as rail on its own right of way and buses on their own rights of way, then switching mode from private to public may be seen as a strategy to avoid the worst of congestion.

All of the above strategies are short-term strategies. In the longer term, there are strategies adopted by both residents and businesses. Residents in an urban area may decide to change home locations or workplaces. This is partially a strategy that can contribute to urban sprawl, with people moving further out of the congested city and finding workplaces around the periphery of the city, so as to avoid the congestion within the major parts of the city. In like manner, businesses may find it increasingly difficult to attract the calibre of worker desired if they are located within the more congested areas of the city, so businesses also seek to relocate to less congested areas, usually on the outskirts of the city. This again contributes to urban sprawl. In the long run, if not carefully managed, these relocation decisions will actually culminate in more congestion, not less, as peripheral areas of the city become increasing populous and the transport infrastructure is found to be inadequate for the new demands being placed upon it. In addition, the moves to the outer areas may also generate longer travel distances and therefore increase the overall density of traffic, to the point that congestion arises once more. New urban economic modelling approaches are starting to provide insights into how residents and businesses might change their locational and travel behaviours in response to congestion and to policies intended to mitigate congestion, such as congestion pricing (see, for example, Anas and Hiramatsu 2013).

9.4 Policies for mitigating congestion

As has been noted in preceding discussions, congestion is very frequently an issue that is raised in a policy context. Politicians running for elected office often make promises around reducing congestion, and politicians in office often discuss strategies for reducing or eliminating congestion. Many of the consequences of congestion are also clearly undesirable: increased fuel consumption, increasing air pollution, noise and vibration, degradation of the urban environment, peak spreading, urban sprawl, decline of city centres, and so forth. Governments also quite frequently publish estimates of the costs of congestion. For example, in Australia, the Bureau of Infrastructure, Transport and Regional Economics (BITRE) has estimated the annual cost of congestion in 2005 in the capital cities alone (Adelaide, Brisbane, Canberra, Darwin, Hobart, Melbourne, Perth and Sydney) to have been \$9.4 billion (BITRE 2007). This cost is the 'avoidable' cost of congestion,

meaning that it includes only those costs where the benefits of travelling in congested conditions are less than the costs imposed on other road users and the community. They also take into account only the economic costs and do not include such costs as those relating to neighbourhood character, amenity, comfort and so on. These costs were, at the time of the study, estimated to rise to at least \$20 billion by 2020. While recent growth in personal travel has slowed compared to the middle of the decade when this report was prepared, continuing population growth in Australia has tended to maintain the overall pace of growth in congestion and its associated costs. Thus, despite the fact that the existence of congestion might be seen as a sign of success from some viewpoints, overall, increasing congestion is seen as something undesirable and it is considered that policies are needed to mitigate congestion.

Congestion mitigation policies should draw on knowledge of the underlying determinants of road travel demand. With the demand for road use being inversely related to its price, increasing the price of road use by congestion pricing or by supply-side measures that slow road traffic (for example) is an obvious way to seek to reduce congestion costs, by discouragement. Chapter 12 explores road pricing in more detail. The demand for car travel is also influenced to a modest extent by the generalised cost of using public transport. Lowering the relative generalised cost of public transport will reduce car use (because of the positive, albeit usually low, cross-price elasticity of demand for car travel with respect to the generalised cost of public transport). Policies intended to reduce congestion thus usually include some measures to improve the attractiveness of public transport (e.g. adding bus priority lanes to speed up bus travel or improving public transport service frequencies), to get people out of their cars. Reducing the number of car parking spaces is another way that car traffic, and associated congestion costs, might be reduced. The University of Leeds KonSULT website, which was discussed in Chapter 3, lists a range of policy measures that may help to reduce congestion or at least slow its growth. In general, a package of measures is needed, with road pricing the most likely to be effective (albeit politically difficult, as explained in Chapter 12).

Given the position suggested by this chapter that congestion is unlikely to be eliminated, and may also be very difficult to reduce, one of the complementary policy directions that could be followed is to reduce some of the negative aspects of congestion, that is, assume that congestion will continue but focus on how to reduce some of its impacts. Two of the impacts that could be considered worth focusing on are fuel and air pollution. These two impacts can be tackled together by the introduction of low-emission and

zero-emission vehicles. Low-emission vehicles include those with more efficient and less polluting petrol and diesel engines, and hybrid fuel vehicles, while zero-emission vehicles are principally electric and fuel cell vehicles. At the time of writing this book, these vehicles have still not penetrated the market that well, especially in countries such as the USA, Canada the UK, Australia and much of Europe. Introducing such vehicles initially is relatively costly, because the economies of scale of large production runs are not immediately available. At the same time, governments appear generally hesitant to introduce subsidies or other financial incentives to people to purchase such vehicles or to manufacturers to produce such vehicles. To a large degree, the technology for such vehicles is now available, and there is increasing availability of such vehicles – at a price.

Of course, governments are also placed in something of a dilemma in supporting such a transition. Many governments around the world currently derive substantial revenues from fuel taxes. Reductions in fuel consumption are already reducing the governmental income stream from fuels, in some cases very significantly. This places government in a catch-22 position with respect to low- and zero-emission vehicles. If such vehicles are introduced in large numbers, fuel tax revenue will suffer substantially. For about 20 years, following the fuel crises (oil embargoes) of the 1970s, several countries legislated fleet fuel efficiency increases. To a large degree, fuel-based revenue did not decline, but continued to increase, mainly as a result of substantial increases in vehicle kilometres of travel (VKT) over the same period. However, in the late 1990s and the early years of the twenty-first century, the pressure on vehicle manufacturers to increase fleet fuel efficiency evaporated, and, more recently, the trends in growth of VKT declined. Recently, there has been the beginning of a resurgence in the need to legislate for improved fuel efficiency, as a consequence of the rapid increase in fuel consumption worldwide and concerns with the environmental effects of the use of fossil fuels. In countries such as the US, the UK and Australia, this is contributing to declining fuel tax revenues (the lack of indexation of fuel taxes has also been a significant contributor to this revenue decline in countries such as the US and Australia).

In the meantime, it is still frequently the case that governments propose increases in road capacity as a means to tackle congestion issues. In the event that there are genuine bottlenecks in the system that can be removed, capacity increases may be appropriate. However, as has been discussed at length in this chapter, capacity increases will largely not bring about reductions in congestion in the long term. If the goal of the policies is to increase personal mobility, then capacity increases may be an appropriate direction

to pursue. However, if increases in accessibility are desired, and long-term reductions in congestion are also desired, then capacity increases, by and of themselves, will not achieve these goals. In the recently issued plan for New South Wales (Australia), there is still stated the goal to 'Deliver road infrastructure to relieve congestion' (NSW 2011), while the *NSW Long Term Transport Master Plan* also has a substantial component in it of adding capacity to the road network (Transport for NSW 2012). These are illustrative of the fact that governments are still not convinced that congestion cannot be tackled by adding more road capacity. Instead, perhaps the focus should be more on land use and understanding of the linkage between land use and demand for travel. This could potentially be far more effective in reducing future congestion than trying, ineffectively, to build one's way out of congestion.

A number of other potential policy directions for reducing congestion are put forward in Chapters 12 to 14, which deal with strategies for solving the externality problem of transport, especially those that result in congestion. They are not discussed further in this chapter.

9.5 Some concluding comments on congestion

In conclusion, it seems appropriate to state quite strongly here that congestion is not likely to be eradicated, and is also not likely to be reduced significantly in the future unless there is a willingness to reduce mobility (and also possibly accessibility). More importantly, neither is it efficient to seek to remove congestion entirely, because optimising the efficient use of congested facilities is consistent with some remaining congestion.

The tolerance of both people and businesses to increasing congestion seems likely to increase over time, so that the willingness to continue to operate in increasingly congested conditions seems likely to be the hallmark of the future of transport in urban areas around the world. One can clearly point to many examples of cities that have achieved really high levels of congestion and contrast these to many other congested cities that are far below such congestion levels. Tolerance for congestion seems not to have reached its limits even in the most congested areas of the world.

It is also potentially useful to consider the other side of the congestion coin. Reducing congestion may not be the best way to go in all cases. It is interesting to consider whether speeding up traffic flows may be part of the negative cause that leads to deterioration of city centres. An interesting experiment took place in Winter Park, Florida some years ago. At the time, it was noted

by city officials that the downtown of Winter Park was declining. It was suggested that one of the reasons for this was that the accumulated policies of prior years had been to speed up traffic through the downtown area, resulting in a negative impact on businesses and the environment of the downtown area. As a consequence, it was decided to introduce policies to slow traffic in the downtown area, through a number of mechanisms. Lower speed limits were implemented. Traffic lanes were removed and converted to angled parking, making it relatively easy for drivers to park at the kerbside along the main streets of Winter Park. Sidewalks were also widened, and trees were planted along the kerbs as a buffer between the commercial uses and the roadway, with occasional tree plantings between angled parking spaces. Merchants were encouraged to use the widened sidewalks for such things as additional marketing and display space, sidewalk cafés, and other improvements. This enhancement of the environment of the main streets created a welcoming area that encouraged through traffic to slow down, stop, and take advantage of the opportunities on offer in the downtown area. The result was a reversal of the decline in the downtown area and success in creating an environment in the city that was conducive to impulse visits. As stated in the 'Transportation element' of the Winter Park plan, it is intended that Winter Park 'is and will continue to be a walkable, pedestrian-friendly, sustainable, treed, relaxed, beautiful, safe, urban village that promotes neighborliness and courtesy among its citizens and visitors' (City of Winter Park 2008). Speeding up traffic by increasing capacity and reducing congestion in the short term may not be the ideal way to go.



REFERENCES

- Anas, A. and T. Hiramatsu (2013), 'The economics of cordon tolling: general equilibrium and welfare analysis', *Economics of Transportation*, **2** (1), 18–37.
- BITRE (2007), *Estimating Urban Traffic and Congestion Cost Trends for Australian Cities*, Working Paper 71, June, Canberra: Bureau of Infrastructure, Transport and Regional Economics.
- Chang, A. (2010), 'China's 60-mile traffic jam to last weeks', *China on MSNBC News*, 24 August, available at: http://www.msnbc.msn.com/id/38828837/ns/world_news-asia_pacific/t/chinas-mile-traffic-jam-last-weeks/#.UKQxs6XEO-I (accessed 15 November 2012).
- City of Winter Park (2008), *Comprehensive Plan*: Chapter 2: 'Transportation element', City of Winter Park, Florida, Planning Department, available at: http://cityofwinterpark.org/Docs/Departments/comp_plans/CompPlan_TransportDIA.pdf (accessed 16 November 2012).
- Greene, A. (1837), *Glance at New York: Embracing the City Government, Theatres, Hotels . . .*, available at: http://books.google.com.au/books/about/A_glance_at_New_York.html?id=3ToZAAAAYAAJ&redir_esc=y (accessed 15 November 2012).
- Hardin, G. (1968), 'The tragedy of the commons', *Science*, **162** (3859), 1243–8.
- Hood, C. (1995), *722 Miles: The Building of the Subways and How They Transformed New York*, Baltimore, MD: Johns Hopkins University Press.
- NSW (2011), *NSW 2021: A Plan to Make NSW Number One*, September, Sydney: Government of New South Wales, Department of Premier and Cabinet.

- Redmond, L.S. and P.L. Mokhtarian (2001), 'The positive utility of the commute: modeling ideal commute time and relative desired commute amount', *Transportation*, **28** (2), 179–205.
- Taylor, B.D. (2002), 'Rethinking traffic congestion', *Access*, **21** (Fall), 8–16.
- Transport for NSW (2012), *Draft NSW Long Term Transport Master Plan*, September, Chippendale, NSW: NSW Government, Transport for NSW.

10

Tackling the externalities – fuels and technology



CHAPTER OVERVIEW

The fourth and final chapter on tackling negative externalities of transport focuses on fuels and technology, particularly those used for road-based traffic. The chapter deals first with the rather difficult issue of the estimation of available oil supply, including a discussion of oil reserves and the concept of peak oil. Policy implications of these aspects of fuel are discussed, and the topic of alternative fuels is introduced. The chapter discusses briefly a number of alternative fuels and the pros and cons of introducing them. The chapter concludes with a discussion of policy directions relating to fuels and vehicle technology.

10.1 Introduction

The primary externalities associated with fuels are sustainability (the continuing use of fossil fuels), energy security for countries that are net importers of fuels, pollutants from burning fossil fuels, and greenhouse gas emissions from burning fossil fuels. The topic of fuels is combined with technology, because of the potential issue of technology finding a potential replacement for fossil fuel engines. At present, transport is almost solely fuelled by derivatives of petroleum oil:

- cars – petrol and diesel;
- trucks and buses – mostly diesel, but with some compressed natural gas (CNG), liquid petroleum gas (LPG) and other fuels for buses;
- trains – some diesel and some electric, but electric trains may run on electricity that is generated by burning fossil fuels;
- planes – kerosene; and
- ships – diesel and related engines.

There are also, of course, taxis and cars that operate on LPG or CNG, and also electric/petrol hybrids. In some countries, electric trains receive their

electricity from fossil fuel generators (coal or oil), so that even these means of transport use fossil fuels. Some countries have processes that derive petrol from coal, but this remains a fossil fuel derivative, although coal deposits are substantially more plentiful than known oil deposits.

Known oil deposits are concentrated in a relatively small number of countries:

- the Middle East, for example Saudi Arabia, Kuwait, Iran and Iraq;
- North America (especially Canada, but also Alaska and Texas in the USA, plus oil shale deposits in much of the western USA);
- Libya;
- Nigeria;
- the United Kingdom and Norway (North Sea);
- Russia;
- China; and
- South America, for example Venezuela and Brazil.

A substantial proportion of the known reserves of oil are located in politically unstable areas. There has also been a history of supply disruptions among some of those nations, such as the petroleum crises of the 1970s and more recent short-lived disruptions during the Iraq wars. Supply is politically volatile. Because so many countries depend on oil imports, often from these politically volatile areas, the goal of achieving some level of independence of fuel for transport is a common goal among those countries that do not have large reserves of their own.

10.2 Oil supply

One of the biggest problems in looking at the supply of oil is the complexity and the political nature of estimating 'proven oil reserves'. The amount of oil actually in the ground does not equate to oil reserves, because only a fraction of the oil discovered can actually be recovered. The fraction that can be recovered varies with different oil fields, and also varies with the extraction technologies used. It is this amount that can be extracted that represents the proven oil reserves. Only a few years ago, extraction of oil from oil shale was considered to be uneconomical and also environmentally unsustainable. This appears to have changed now, and oil from oil shale is being counted as a major source of proven oil reserves by at least some commentators. Oil reserves are also politically problematic. It is in the interest of some countries to exaggerate the extent of their proven oil reserves, and there is no independent auditing of the claims of oil reserves. All this means that the amount of oil that the world has is really unknown, and so the state of supply is largely guesswork.

There are also huge reserves in tar sands and oil shale, which, in the past, it has not been feasible to extract. However, with oil prices having been sustained at a much higher level for the past decade, it has become increasingly economical to extract both oil and gas from shale, creating a revolution in these fuels, especially in North America, where the USA is now predicted to be energy independent by 2035 or even sooner (*Economist* 2012). There are, of course, both pluses and minuses to this. On the minus side, extraction of gas and oil from shale has impacts on the environment, as does the burning of both of these fossil fuels. However, the gas from shale is cleaner than burning coal, so that, to the extent that gas replaces coal or oil for generating electricity, there are fewer pollutants and greenhouse gases emitted. Also, the potential energy independence that shale oil and gas offer to countries like the US provides the prospect of energy security to those countries. The potential extent of oil and gas shales around the globe is not yet known by any means, and this same revolution may extend to other countries over the coming years.

Oil reserves

Oil reserves are classified into three categories: proven, probable and possible. Proven reserves are those reserves for which there is about a 90 to 95 per cent certainty that the amount specified can be extracted. Probable reserves are those for which there is about a 50 per cent certainty that the amount specified could be extracted, while possible reserves are those with about a 5 to 10 per cent probability of being able to be extracted. Published oil reserves from country to country should only include proven reserves. However, it is more than likely that probable reserves may be included in the estimates of some countries.

Claiming larger reserves than are true may be politically motivated, because it may seem to confer more power on the countries that have greatest proven oil reserves. On the other hand, oil companies would profit by understating oil reserves, thereby making oil seem more scarce and driving the price up.

Peak oil

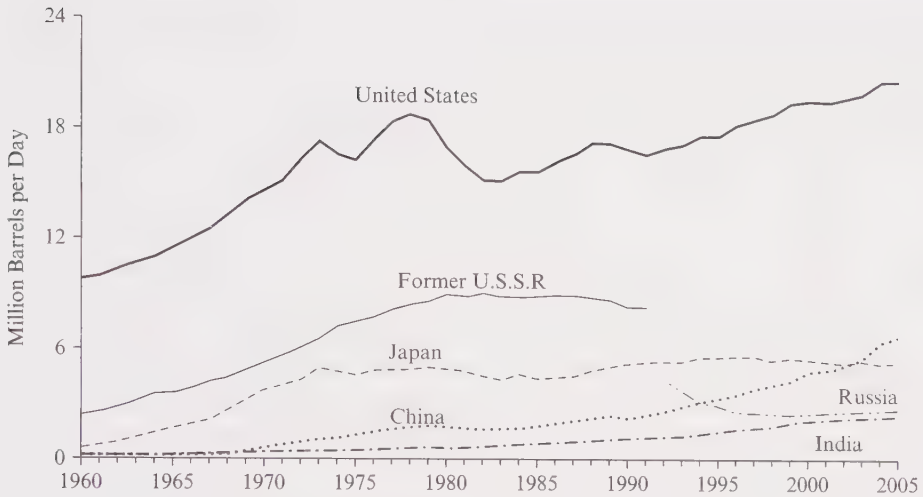
The concept of peak oil is that this is the point at which the maximum rate of global production of oil is reached. Following this point, oil production is assumed to fall, with the consequence that demand for oil will exceed the supply of oil on an ongoing and increasing basis. This would probably create an ongoing energy crisis, because of the uncertainty of supply and rapidly increasing prices, together with the economic and social disruptions that these would cause. Assuming that such a point exists, then the major

uncertainty becomes the question of when that point will be reached, or if it has already been reached. Some theorists believe that peak oil was reached in the 1950s, but there are many claims for later years, as well as claims that the peak has not yet been reached. It is also possible to find tables that indicate that various countries individually have already reached their peak of production and have since declined, but there are also other claims that these same countries, as well as other new ones, have yet to reach their peak. As a good example of the issues relating to peak oil, a few years ago the province of Alberta in Canada was believed to have proven oil reserves in the millions of barrels of oil. Most recently, it has been estimated that these reserves are now in the trillions of barrels. The Australian *Energy White Paper 2012* (DRET 2012) indicates that peak oil is not now expected to occur until 2035 at the earliest. It also acknowledges that this may yet be subject to further change.

Another complicating factor in all of this relates to the economics of oil extraction. When prices of crude oil are relatively low, the economics of recovery of oil from known oil fields may result in large amounts of oil being considered uneconomical to extract. However, when the price of crude oil rises, the economic balance can change quite rapidly. As a consequence of sharp price increases in the early 2000s for crude oil, it became much more economically feasible to extract oil from oil wells that had previously been abandoned. As a result, many oil wells in various oil fields around the world were reopened and extraction resumed, using more expensive methods. Such shifts in prices can affect the existence of peak oil or the point at which peak oil is reached.

Oil is essential to modern agriculture, medicine, the plastics industry, and transport, to name a few of the areas that oil affects. In agriculture, for example, oil powers machinery that is used in crop production, provides many fertilisers and pesticides, and powers the production of many of the inputs to agriculture. Figure 10.1 shows the growth in consumption of oil by the top consuming countries in the world. The potential for China and India to show substantial increases over the coming years is, of course, enormous, as car ownership increases rapidly in both nations. It is estimated that 75 per cent of the growth in demand for oil can be attributed to cars and trucks. Approximately 70 per cent of oil consumed in the US and 55 per cent of the oil consumed worldwide is for transport.

Clearly, if the supply of oil is limited and declines, then this will have an impact on the consumption of oil around the world. Transport and agriculture alone would suffer enormously if supplies of oil are seriously limited. The potential for worldwide conflict over the supplies of oil is also a serious risk, if peak oil is real and either has occurred already or will occur in the near



Source: http://www.eia.doe.gov/emeu/aer/pdf/pages/sec11_20.pdf.

Figure 10.1 Top oil-consuming countries, 1960–2005

future. Peak oil would threaten not only transport, but also the production of food, medicine and plastics.

There are three theories regarding peak oil. The pessimistic theory, which is largely discredited at this time, holds that peak oil production has already been reached or that it will be reached within five to ten years. Under this theory, proactive mitigation is no longer possible. As a result, global depression is likely to occur and global industrial civilisation may collapse. If, on the other hand, peak oil will be reached within 20 to 30 years, then this leads us to the second theory, or plateau theory. The plateau theory holds that oil production will reach a sustained plateau for one or more decades, before declining slowly. Under this theory, mitigation is possible, and global depression and other more serious outcomes could be prevented, by reducing current growth in demand for oil and finding alternative energy supplies. It is also theorised that rising prices will trigger a shift to alternative fuels well before a crisis is reached. In addition, new technologies for oil extraction and synthetic oil production would be expected to come online, further pushing the occurrence of peak oil into the longer-term future. The third theory is that there is no such thing as peak oil. Under this theory, it is proposed that current theories of peak oil are flawed and that the world has more than adequate oil reserves for at least the next century, if not longer. Under this theory, political and economic effects are much more important. This theory also assumes that further discoveries are likely to occur, continuing to increase the proven oil reserves. Also, similarly to the plateau theory,

it assumes that new technologies will be brought in that allow for a larger extraction percentage, and also that alternative fuels will be developed as economics and politics drive this.

It is also possible to detail when various oil-producing countries reached their peaks in production. For example, one of the earliest to peak was Japan, which was said to have peaked in 1932. The USA was said to have peaked in 1970, along with Venezuela and Libya. However, oil shale in the US has now led to new predictions that the US will not peak for another 10 to 20 years from now, or somewhere in the region of 2025 to 2035. Clearly, identifying when a country's oil production peaked is fraught with difficulties, especially because new oil deposits have been found, new technologies are already emerging for extraction of oil, and known oil deposits that are in difficult or expensive locations, such as much deeper in the earth's crust or under deep ocean waters, may become economical to extract, as oil prices change and political and economic pressures increase, as witness the revolution in shale oil and gas in North America.

10.3 Policy implications

If the pessimistic theory about peak oil is true, then there are several immediate and short-term policy implications. First, to preserve agriculture and the world's food supply, oil would need to be diverted from transport to agriculture. Suburban and ex-urban living would become unsustainable, and high-density development would be required to replace existing low-density development. This would have the effect of reducing the size of urban areas, therefore requiring less travel distance, and also would make public transport more viable, thereby further diverting transport use from cars to public transport. To maintain liveability, alternatives to oil would need to be developed at least in a decade or two prior to peak oil, so that the use of oil derivatives for transport could be phased out over two or three decades. Vehicle technologies would need to be changed and new fuel sources used. This opens up the area of alternative fuels, which is discussed in the next subsection.

If the plateau theory about peak oil is true, then the implications are somewhat similar to those of the pessimistic view, but the time frame available for the changes is longer. There is the potential that new technologies for vehicle propulsion and alternative fuels will make it possible for present conurbations to survive in their existing form.

As with other externalities discussed in this book, the authors would propose a risk mitigation approach. Such an approach would offer gains by shifting

fuel technologies from limited resources to sustainable resources, also offering reductions in the production of pollutants and greenhouse gas emissions, and increasing energy security for those countries that lack fossil fuel sources of their own. The following subsection discusses elements of such an approach.

Alternative fuels

Alternative fuels are of interest for several reasons:

- they have the potential to remove oil dependence on the limited number of oil-rich nations by those nations that have no oil, or lessen the demand in those nations;
- they are likely to reduce emissions, both those that are smog-related and those that are greenhouse gas-related;
- they are a response to the possibility that either the pessimistic or the plateau theory of peak oil is true;
- they are a way to make current forms of transport more sustainable in the long-term future.

Bio-fuels

These are primarily ethanol, butanol and bio-diesel. They require modification of current petrol engines if more than 10 per cent of the content of petrol is ethanol. Engines can be constructed to burn 85 per cent bio-fuel. However, the main problem with bio-fuels is that they are largely derived from agricultural products and therefore require diversion of land from food production to fuel production. This is already a concern (e.g. IFPRI 2008; Babcock 2011), even with the currently relatively limited production of bio-fuels. Given the increasing need for food production, it is likely that the future of bio-fuels as a substitute for oil-based fuels is fairly limited. The exception to this is bio-fuels that can be produced from waste products, especially agricultural wastes. These are termed second-generation bio-fuels and are generally produced from non-food crops and inedible waste products, such as sawdust, citrus peel, algae and so on (Scott et al. 2010).

Dual-fuel vehicles

There are already a number of dual-fuel vehicles in production. These vehicles normally have an engine that is capable of operating on two different fuels (also called bi-fuel vehicles). These vehicles have mainly been produced

in Europe and are most commonly vehicles that use petrol and either LPG or CNG.

A second type of vehicle, more often correctly referred to as a hybrid engine, has a petrol engine (or equivalent) that both powers the vehicle at certain times and is used to generate electricity to charge on-board batteries that are used under certain driving circumstances as the only propulsion source. The vehicles currently available on the market generally achieve fuel efficiencies of around 2.5 to 4 litres per 100 kilometres (94 to 59 miles per gallon (mpg)), and it is possible that this level of efficiency could be improved further in the future. Advances in battery designs and also in the petrol engines themselves could result in further improvements. The only downside of such engines is that they still require some amount of oil-based fuel.

Fuel cells

In the longer term, fuel cells may be the most promising alternative fuel system. These are based on the use of hydrogen as the primary fuel component. Burning hydrogen produces only water vapour as the emission; no other emission is produced. With current technology, the fuel cell technology requires very bulky storage. Also, production of hydrogen is still expensive and raises problems of storage and refuelling, both of which are more hazardous than for current oil-based fuels. The production of fuel cells is currently very expensive, although some cars are beginning to appear that use this technology – mainly at the upper end of the car market. Production of hydrogen from the atmosphere or water (abundant sources of hydrogen) requires a significant amount of input energy. However, the energy to do this can be derived from a number of possible sources, including nuclear, solar and hydro. If the energy source for extracting hydrogen is, itself, sustainable and not oil-based, then this technology has great promise. Several vehicle manufacturers have announced plans to begin production of fuel cell vehicles, with Mercedes-Benz having moved up production to 2014 for a fuel cell passenger vehicle (Lienert 2011). There are also about 100 fuel cell buses operating in various places around the world. In addition, there is a potential to combine hydrogen with nitrogen to produce ammonia, which is another potential fuel for the fuel cell.

Solar

Solar-powered vehicles already exist, but are not produced commercially. The issue is that the efficiency of the solar panels needs to be increased substantially. In a solar-powered vehicle, the solar radiation is used to charge

batteries, which then run an electric engine in the car. While sunshine is not necessary for the generation of solar power, dark, cloudy days substantially reduce production of energy from solar panels, so the technology is currently applied best in parts of the world where there is extensive sunshine. Solar-powered vehicles generally have low top speeds at this time and restricted operating ranges. Most vehicles that have been developed to date need a large solar panel to produce sufficient electricity, and multi-passenger vehicles have not yet been developed. There is potential, however, as solar panel technology improves for this option to become increasingly realistic.

Other potential fuels

There are at least three other types of fuel that have been considered as possibilities for vehicle propulsion systems. Nuclear fission or fusion is one option, although this is not considered currently to be safe or feasible. Compressed air is a potential mechanism, but requires energy to compress the air, and efficient compressors running on fuels other than fossil fuels are not yet available. The third type of fuel that has been considered is that of liquid nitrogen. Here, the barrier is the means to produce the liquid nitrogen. These and other possible fuels could emerge in the coming years, depending on the impetus for research into alternative fuels and technological breakthroughs.

10.4 Policy directions

As before, in areas where the externalities are controversial, or not completely known, the ideal is to seek for policies that represent win-win situations, that is, where the policies produce positive impacts, whether or not the externality under consideration is real. In the case of peak oil, where there are considerable doubts as to whether it is real or, if it is, when it will occur, the best policies to follow are again those policies that produce other benefits. Among the transport-related policies that have emerged from this discussion are reduction of dependence on oil for transport, control of urban sprawl and densification of urban areas, and reduction of vehicle kilometres of travel.

Tougher fuel economy standards are one possible policy instrument that would result in reduced consumption of fossil fuels and would also be likely to result in benefits to air quality, GHG emissions and fuel security. In 1975, the US Congress first introduced Corporate Average Fuel Economy (CAFE) regulations. From 1991 to 2011, the CAFE goal was 27.5 miles per gallon or 8.6 litres/100 kilometres. New standards have now been agreed in the US requiring increasing fuel economy from 2012 to 2025 (NHTSA 2011). Both

Japan and Australia had goals for 2010 of 6.7 litres/100 kilometres, although the standard in Australia was a voluntary one. The European Union had a goal of 5 litres/100 kilometres for 2012.

Investment in technologies that will permit vehicles to operate on non-fossil fuels have benefits far beyond those of reducing dependence on oil, especially for some countries dependent on foreign oil. Most alternative fuels have positive impacts on air quality, by reducing vehicular emissions or changing vehicular emissions into harmless substances, such as water vapour. The adoption of alternative fuels will also generally mean a reduction in the emission of greenhouse gases. These changes will also generally make transport more sustainable, although the use of bio-fuels has some potentially serious consequences on food production, which could render these fuels unsustainable.

Policies that encourage control of urban sprawl and densification of urban areas also have benefits in making public transport more viable, decreasing dependence on cars, and reducing the shift of land from agricultural and other uses to residential and commercial space. Again, this means that such policies can be expected to have positive impacts on agriculture, car dependence and other urban issues. Such policies are also likely to lead to a reduction in congestion caused by urban sprawl, and therefore have further positive impacts on fuel consumption and air quality.

Finally, policies that are directed specifically at reducing vehicle kilometres of travel will also produce benefits for air quality, greenhouse gas emissions, fuel use, and reduction of congestion. Such policies would generally be those that encourage the use of public transport, walking, bicycling and car sharing. These policies may also provide health benefits from more active transport. Reducing vehicle kilometres of travel is also likely to have an impact on traffic accidents, leading to some reduction in accidents, especially where travel is diverted from cars to safer modes such as bus and train.



REFERENCES

- Babcock, B.A. (2011), 'The impact of US biofuel policies on agricultural price levels and volatility', Issue Paper No. 35, June, Center for Agricultural and Rural Development, Iowa State University, for ICTSD.
- DRET (2012), *Energy White Paper 2012*, Canberra: Government of Australia, Department of Resources, Energy and Tourism.
- Economist* (2012), 'America's oil bonanza', *Economist*, 17 November.
- IFPRI (2008), *Biofuels and Food Security: Balancing Needs for Food, Feed, and Fuel*, Washington, DC: International Food Policy Research Institute, available at: <http://www.ifpri.org/publication/biofuels-and-food-security> (accessed 6 March 2013).

- Lienert, A. (2011), 'Mercedes-Benz fuel cell car ready for market in 2014', *Edmunds Inside Line*, 21 June.
- NHTSA (2011), *2017–2025 Model Year Light-Duty Vehicle GHG Emissions and CAFE Standards: Supplemental*, Washington, DC: US Department of Transportation, National Highway Traffic Safety Administration.
- Scott, S.A., M.P. Davey, J.S. Dennis, I. Horst, C.J. Howe, D.J. Lea-Smith and A.G. Smith (2010), 'Biodiesel from algae: challenges and prospects', *Current Opinion in Biotechnology*, **21** (3), 277–86.

11

Agglomeration and other wider economic benefits



CHAPTER OVERVIEW

Most discussion about externalities in transport policy focuses on external costs, such as congestion and accidents. Recent research suggests, however, that there also are potential external benefits from transport initiatives that support increased concentration, or clustering, of high-valued economic activities, particularly knowledge-based activities. The existence of such benefits indicates, for example, a need to look beyond direct user benefits when exploring the benefits of major transport policy or project interventions. The chapter summarises some recent research on agglomeration benefits and presents some findings about the scale of potential benefits, through reference to three recent case studies. Other potential external benefits included under the generic heading of 'wider economic benefits' are also introduced.

11.1 The relevance of agglomeration economies to transport policy

One of the newest areas of transport policy focus concerns the link between transport, urban structure, and the economic performance of a city. If such a link exists, with economic outcomes typically being a key policy goal for governments, then it is vital that transport policy understands the significance of the relevant connections and is shaped to support enhanced city economic performance. The current chapter deals with this issue, through the lens of agglomeration economies. Because this is a new area for transport policy, the chapter presents some evaluation case studies that demonstrate the potential scale of benefits that might be available. This scale demonstrates why agglomeration economies need to be understood by transport policy analysts.

Noted urban scholar Cervero (2001) has examined efficient urbanisation, considering economic performance and the shape of large cities. He finds that employment densities, urban primacy (importance of the centre) and

well-functioning infrastructure are positively associated with economic performance. There is a trade-off in city size between agglomeration benefits and the external costs of (for example) traffic congestion, pollution and noise (liveability and sustainability concerns). Agglomeration benefits are essentially those additional benefits that arise because of the clustering or scale of economic activities in space (i.e. in effect, $1+1>2$). Focusing on economic productivity, Graham and Melo (2012) describe this in terms of firms or workers benefiting from being near to other firms or workers.

Large cities that are compact and enjoy good accessibility, matched by efficient transport infrastructure, are among the most efficient urban settlements. These cities do not arise by chance but require decades of careful management and guidance. Cervero (2001) suggests that, beyond about 5–10 million population, the increasing social costs of size exceed the additional benefits. Other chapters in this book have discussed the external cost side in some detail. This chapter explores the external benefit potential of agglomeration economies and brings in the idea of what have become known as ‘wider economic benefits’ (WEB) to help show why this is a vital area of transport policy concern.

For a considerable time during the application of cost–benefit analysis to land transport projects, external benefits of any kind were essentially out of scope, regarded as non-existent. For major road projects, for example, a focus on user benefits, primarily in terms of savings in travel time, fuel use and accident costs and including benefits to generated traffic (whatever the source of this generation), was seen as sufficient for measurement of direct benefits. Because transport is an intermediate input in the production of other goods and services, the demand for transport is a derived demand, and measuring benefits at the level of product demand and supply as well as at the travel demand level was recognised as double counting. This is no longer the case, particularly (but not only) because of the potential external benefits arising from agglomeration economies.

Half of the world’s population now lives in cities. The concentration of people in cities increases productivity and liveability, through agglomeration benefits in production (essentially productivity improvements as a result of absolute size or concentration of employment and population) and consumption (e.g. a wider range of services are available to people). International research suggests that agglomeration effects in production typically range between 3 per cent and 8 per cent, with a central value of about 6 per cent, such that (for example) doubling city size can be expected to lead to output increasing by about 106 per cent (Rosenthal and Strange 2004; Melo et al. 2009). Relative

output increases in knowledge-intensive industries, many of which tend to concentrate in central business districts and other urban hubs (e.g. universities), are typically higher, but are lower in manufacturing. There is limited evidence about whether some minimum city or cluster size is needed to achieve agglomeration effects, and whether elasticity values remain constant with respect to city or cluster size.

The origins of agglomeration economies have been well understood for some time (e.g. Marshall 1890). They include such sources as improved access to inter-industry information flows (information spillovers), improved labour matching, and better access to specialised services (e.g. legal services, design and testing, financial services) and to public and private (tolled roads) infrastructure, together with economies of scale that may accrue to individual firms. Porter's (1990) work on clusters reflects some of the potential benefits of agglomeration, with clustering both within an industry sector and across sectors being potential sources of agglomeration benefits. The accessibility dimension to these benefits suggests that transport policy should take a close interest in these matters.

Agglomeration effects in consumption, an important element of liveability, are a relatively new area of quantitative research. However, recent German analysis (for example) indicates clear evidence of agglomeration externalities in consumption, with bigger cities (in population terms) showing benefits for residents from a larger range of service choices, across areas like restaurants and bars, concerts, dancing, theatres and museums (Borck 2007). While recognising the existence of agglomeration benefits in consumption, we concentrate in what follows on the production-side focus of agglomeration, where thinking on the role of transport has been focused.

The US Transportation Research Board (1998) linked agglomeration benefits to transport, defining agglomeration benefits as the increase in productivity, creativity and synergy among firms because of a higher concentration or density of employment, made possible by more compact transit-served development. Agglomeration economies are usually classified (Simmie 2004; Graham 2005) as:

- internal scale economies - whereby a firm reduces its production costs by increasing its scale of production (the related concept of economies of scope is also relevant here, where a firm's costs are reduced by broadening the range of its activities);
- localisation economies – these result when a firm's unit costs are lower in an urban area because of the presence of many firms in the same

sector (benefits of proximity for firms undertaking the same or a similar activity; and

- urbanisation economies – these are benefits that arise from city size or scale, not concentration in particular sectors.

Separating the three sources has proven to be complex, but studies that include both localisation economies (concentration) and urbanisation economies (size) tend to find that localisation effects are stronger. Knowledge-intensive industries (e.g. finance, insurance and real estate) and light industries tend to show urbanisation economies as relatively important. Agglomeration effects tend to be stronger for service industries than for manufacturing, because of the higher relative dependence on customers (Melo et al. 2009).

The standard approach to estimating the impact of agglomeration externalities on economic output uses a production function framework. A production function estimates the relationship between output (or productivity) attainable and the quantities of labour, capital and other input factors used to produce this output. To estimate agglomeration economies, it usually takes the general form shown in equation 11.1:

$$Y_{ij} = g(z_j) f(L_{ij}, K_{ij}, G_{ij}) \quad (11.1)$$

where y_{ij} is the output for the i th firm in the j th location, L_{ij} and K_{ij} are the labour and capital inputs to firm i and G_{ij} represents other inputs, which might include energy, materials, public infrastructure and knowledge. The shift factor, $g(z_j)$, represents the agglomeration economies and is usually measured by either metropolitan employment or population, to represent urbanisation economies, or industry employment to represent localisation economies. These agglomeration economies are positive externalities that derive from the spatial concentration of activity. Calculation is considered in section 11.2.

DfT (2005), Eddington (2006) and Vickerman (2007) have identified a number of ways in which transport can influence the clustering of firms and have a positive impact on a city's economy, while Venables (2007) developed a simple economic model to show how transport investment might affect productivity, through effects on city size. He saw linkages to improved productivity through:

- improving links between firms within a city, increasing the effective employment density of the cluster (defined in section 11.2); and
- relaxing constraints on access to the city, increasing the effective employment catchment.

The value of any productivity enhancement resulting from agglomeration economies in production, flowing from a transport improvement, is now quite widely recognised as a wider economic benefit that can be added to user benefits of major land transport projects. DOT (2012) provides a very readable general overview of this subject and associated discussion of agglomeration assessments.

11.2 How do we measure agglomeration benefits in land transport?

Graham (2007) has used the idea of effective employment density to describe how transport improvements can drive agglomeration economies. He defines the effective employment density of a location as the employment in the surrounding area, weighted by proximity to the location. The measure thus brings in both employment size and proximity components, with the latter having a deterrence effect, in the sense that more accessible locations (typically measured by distance, travel time or generalised trip cost) have a stronger influence than those that are less accessible (e.g. further away or more costly to reach).

Graham and Melo (2012) illustrate the concept of agglomeration by expressing it as an aggregation of workers (employment) in the geographical neighbourhood of each firm i , the relevant agglomeration index having the general form of equation 11.2:

$$A_{it} = \sum a(c_{ijt})z_{jt} \quad (11.2)$$

where A_{it} = an index of the relative effective employment density experienced by firm i at time t , c_{ijt} is a measure of the deterrence effect of time, cost or distance between place i and j at time t on the effective density of place i , and z_{jt} is employment, typically postcode-level employment.

Then, using a conventional gravity-type model structure for the deterrence factor c_{ijt} in equation 11.2, Graham and Melo (2012) express effective employment density as equation 11.3:

$$A_{it} = \sum d_{ijt}^{-\alpha} z_{jt} \quad (11.3)$$

where d_{ijt} is the straight-line distance between i and j at time t . The effective employment density of the firm at site i is thus a function of employment at other locations at time t (z_{jt}) and the distance between i and j .

In equation 11.3, the value of α determines the effect of distance on the strength of density externalities. If available, separate values can be used for different industry sectors. When separate values are not available for different sectors, then a common value is used. The value of this parameter should be estimated empirically but is often assumed to be unity. Some empirical work suggests a higher value may be appropriate and that the relevant value varies for different sectors, service sectors (for example) having higher values than manufacturing (Graham et al. 2010). The higher the value of α , the faster the rate of decline of agglomeration economies, an implication being that agglomeration effects for services decline more rapidly with distance (generalised travel costs) than for manufacturing.

The measures of effective employment density that result from application of this approach (or the zonal equivalent formula) are relative measures that are useful for rating different locations with respect to agglomeration opportunities. They have no absolute meaning. The UK Department for Transport's WebTAG guidance material sets out relevant formulae for calculating effective employment density in a zonal analysis system, which takes account of the employment in, and size of, zone i as well as its distance from all other zones and the employment levels in those zones (DfT 2012).

Equation 11.4 shows how the benefits of agglomeration economies from a major transport improvement (Δy) are calculated, requiring estimation of:

- how that major transport improvement affects relative effective employment density (i.e. effective density after the improvement, Aa , is divided by effective density before, Ab);
- the application of relevant agglomeration elasticities, ϵ (which show the proportional change in labour productivity, relative to a proportional change in effective employment density, as estimated back in equation 11.1, or by use of an externally provided elasticity value, such as about 1.06); and then
- multiplication by y_b , the level of economic output before the transport improvement, where economic output is typically measured as gross regional product.

$$\Delta y = [(Aa/Ab)^\epsilon - 1] y_b \quad (11.4)$$

DfT (2012) details how to calculate the relevant agglomeration benefits in a multi-sector setting, in line with the preceding logic.

In estimating values of the elasticity of productivity with respect to density (ϵ in equation 11.4), as estimated in equation 11.1, the UK WebTAG approach advocates distance-based elasticities (in contrast to, say, time-based elasticities). Graham and Melo (2012) describe the process of calculating the effect of transport improvements on generalised travel costs (or time costs) and then translating this relative change into a comparable change in distance, to suit application of the distance-based elasticities. This approach is because of a concern that agglomeration elasticities based on generalised cost are biased upwards.

The UK approach assumes fixed employment levels when calculating effective density and related agglomeration effects, with land use changes being dealt with in sensitivity tests and as part of the calculation of the wider economic benefit of 'change to more productive jobs' (as explained in section 11.3). Readers interested in the detail on estimation should read DfT (2012).

11.3 Other wider economic benefits

Agglomeration benefits are one of four benefits that the UK Department for Transport calls 'wider economic benefits', which are not captured by conventional economic analysis of land transport projects. The other three are (DfT 2012):

1. **Output change in imperfectly competitive markets:** a benefit that recognises that most markets are not competitive and that, in consequence, users may value marginal benefits at higher than the relevant marginal costs. This benefit has been incorporated into UK transport evaluations by increasing the value of business user benefits (e.g. time savings, cost savings or reliability benefits) by 10 per cent.
2. **Labour supply impacts:** this benefit recognises that transport improvements that save travel time might encourage some people to increase their hours of work, with a consequential flow-on effect on GDP. Estimating relevant benefits requires (*inter alia*) knowledge of travel cost changes, application of labour supply elasticities (to predict the change in labour participation), assumptions about the relative productivity of the additional labour supply (to predict the output response), and relevant tax rates, to net out the relevant benefit component and avoid double counting. The tax component of the predicted increase in GDP is seen as the benefit that is not picked up in conventional cost-benefit analyses of transport projects.
3. **Move to more or less productive jobs:** agglomeration benefits are

estimated on an assumption of fixed employment locations. Transport improvements may encourage firms or workers to move to locations where their productivity is different, an area in which there is a lack of research. If people or jobs move to locations where productivity is higher, there is a flow-on effect on GDP. The tax gain from these relocations is seen as an additional benefit.

The significance of such matters for transport policy depends on the scale of potential benefits involved. Evaluation case studies can shed light on this issue. Worsley (2011), in reviewing London's Crossrail evaluation, outlines the complexities in incorporating these various items in an economic evaluation. While the UK Crossrail evaluation (discussed in section 11.4) suggests that the move to more productive jobs should certainly be an important focus of close attention, given the estimated scale of this impact in that evaluation, Worsley points out that the analysis methods to assess this productivity impact are not well developed in the UK, preferably involving advanced land use transport interaction models, linked to computable general equilibrium (CGE) models (see, for example, Truong and Hensher 2012). Agglomeration economies seem, at this point in time, to be the best developed of the wider economic benefits with a potentially significant scale of impact. However, the case studies below indicate that gains from 'moves to more productive jobs' also seem potentially very important, and analytical approaches are being developed to incorporate these more systematically into transport evaluations.

11.4 Significance of agglomeration benefits and other wider economic benefits

Agglomeration benefits and other wider economic benefits are important for transport policy because they are potentially a significant additional economic benefit from major land transport initiatives whose purpose is to support agglomeration of jobs, especially high-productivity jobs. Many major land transport initiatives are intended to do just that, particularly public transport improvements serving central cities. For example, employment in Melbourne's central business district has doubled in the past 20 years and has the potential to double again in the next 20, with large areas of land suitable for precinct-scale urban renewal available. There are significant potential agglomeration/wider economic benefits available, but constraints in the city rail system's capacity to deliver people to the central area in the peak will hamper realisation of those agglomeration economies. Construction of a major new rail tunnel, with associated improvements (e.g. new stations), adds sufficient capacity to open the opportunity for realisation

of a new growth horizon. The allied agglomeration/wider economic benefits are crucially dependent on public transport capacity expansion and on the transport policy decisions that prioritise such initiatives. Two detailed case studies are outlined below, to demonstrate the potential transport significance of agglomeration/wider economic benefits. While these examples are rail-based, Abrantes et al. (2013) make a strong case that comprehensive bus networks serving central business districts of major cities such as Manchester and Leeds will support similar benefits.

London Crossrail

Crossrail, Europe's biggest infrastructure project, is described as follows in its business case summary report:

Crossrail will connect Maidenhead and Heathrow west of London with Shenfield and Abbey Wood to the east, running through a new 13 mile (21 km) twin-bore tunnel under central and east London. As well as linking commuter and suburban services, it will provide a high-frequency, high-capacity and accessible link between Heathrow Airport, the West End, the City of London and Canary Wharf. (MoL et al. 2010, p. 1)

Travel times from Heathrow to Liverpool Street, for example, are expected to be reduced from 55 minutes to 36 minutes. It is projected that there will be 200 000 morning peak passengers by 2026, and the project is intended to help the continued development of London's finance and business services sectors, and support regeneration. The economic development gains are related to the wider economic benefits that are expected. The regeneration benefits are essentially distributional in purpose, in the sense of improving job access for many socio-economic groups.

A conventional cost–benefit analysis (without wider economic benefits) undertaken by Crossrail in 2005 produced an estimated benefit–net cost ratio (BCR) of 1.80, with user benefits of £16.1 billion, estimated gross scheme costs of £13.9 billion (of which £10.6 billion was capital) and net costs of £9.0 billion (Crossrail 2005).

Crossrail provided the first major application of the wider economic benefits approach in the UK. The analysis of wider economic benefits resulted in the additional base case economic benefits shown in Table 11.1. The elasticity of productivity value used in estimating agglomeration benefits (of £3.1 billion GDP in Table 11.1) was 0.059, which is about mid-range in the 3–8 per cent range cited in section 11.1. A sensitivity test of a value of 0.075 increased the

value of agglomeration benefits to £4.5 billion, which shows the sensitivity of evaluation outcomes to choices of key parameter values. The estimated benefits of £10.8 billion associated with 'Move to more productive jobs' were larger than estimated 'Agglomeration benefits' but, as noted in section 11.3, the agglomeration benefits are regarded as more reliable.

Table 11.1 GDP and welfare benefits of Crossrail from wider economic benefits

Wider economic benefits	GDP £ million	Additional welfare benefits £ million
Move to more productive jobs	10 772	3232
Agglomeration benefits	3094	3094
Labour force participation	872	349
Imperfect competition	486	486
Total	15 224	7161

Source: Crossrail (2005, Table 5).

The inclusion of wider economic benefits thus added significantly to the total estimated project value. Worsley (2011) argues that the inclusion of these additional benefits was one important factor in getting government go-ahead for the project and in the subsequent decision to raise a business rate supplement as one funding source for the project.

London travel-time savings values are considerably higher than UK national values, reflecting income differentials. A later evaluation of Crossrail, as part of the business case preparation, showed a BCR of 2.55 if London values of time are used, compared to 1.87 with national time values (MoL et al. 2010). This highlights the importance of the distribution of gains or losses and how they are valued.

A further later evaluation, in July 2011, increased the user BCR to 1.97 using national time values, or 2.7 with London time values (Crossrail 2011). The increasing BCR on user benefits, compared to the 2005 result, is due to both lower capital cost estimates and higher revenue estimates. This later evaluation also revisited wider economic benefits and produced a project BCR of 3.97, including central estimated values of wider economic benefits. In short, including the wider economic benefits has increased the project's BCR by about 40 per cent, which is very significant and shows the potential policy relevance of these external benefits to major rail projects like Crossrail. With potential benefits of this order, transport policy makers need to be alert to the significance of wider economic benefits for shaping land use/transport policy priorities.

Sydney North West Rail

The scope of this major rail project in Sydney is still in the process of being finalised. However, the 2011 NSW government announcement was for a 23-kilometre line between Epping and Rouse Hill, including six new stations. The line will serve a population of 360 000, expected to grow to 485 000 by 2021.

Hensher et al. (2012) and Truong and Hensher (2012) have taken their integrated land use/transport model – Transport, Environment and Social Impact Simulator (TRESIS) – and linked it to a computable general equilibrium model of the Sydney economy (Sydney general equilibrium model, SGEM, with 14 zones), to evaluate the wider economic benefits of the North West Rail project. TRESIS includes key household decisions such as where to locate, dwelling type, where to work, and various travel choices. SGEM aligns its zones with those in TRESIS, each zone being a 'mini' economy trading with other zones. A major transport improvement can change the distribution of relative housing and employment opportunities among zones and lead to changes in zonal flows. Agglomeration elasticities estimated for Sydney by Hensher et al. (2012) are generally consistent with the very small number of others that have been estimated for Australia and New Zealand (DOT 2012), providing some comfort for use of the resulting values.

The relative changes in travel-time opportunities confronting those in different zones led TRESIS to predict an increase in employment in some zones and a decrease in others. With total employment numbers fixed in the analysis of agglomeration benefits at a point in time (but allowing for growth over time), rail network effects led to the somewhat unexpected prediction that employment levels at the outer end of the line may actually decline, because of the greater range of employment choices opened up elsewhere in Sydney by the new line. Zones outside those in which the new line is located may thus gain from the improvement. This finding shows the value of integrated modelling for policy purposes: for example, an awareness of this potential provides an opportunity to design a policy package that is suitably responsive if possible employment loss in the particular outer zone is of concern. The modelling outcomes should also lead the policy analyst to encourage use of integrated land use/transport/CGE modelling frameworks. These enable a more nuanced approach to understanding agglomeration economies, an approach that recognises the potential for aggregate benefits but with no assurance that all areas will share in those benefits (distributional impacts).

The change in employment distribution between various Sydney zones that flows in the short term (within a year or so) from the rail improvement project produces two main kinds of benefits in the analysis by Hensher et al. (2012):

1. the general equilibrium benefit which is measured as the gain in total consumers' surplus that results from rearrangement of work opportunities, which is effectively a measure of the traditional transport user benefits; and
2. associated agglomeration benefits, estimated with total employment fixed.

Hensher et al. (2012) also identify short-term changes in rental incomes, which are absorbed into the consumer's surplus gains, and identify some minor gains in total employment, which have a minimal impact on benefit estimation.

The agglomeration benefits are estimated to add about 17 per cent to total direct user benefits estimated through the TRESIS/SGEM model. These benefits relate only to economic productivity benefits flowing from the employment redistributive impacts of the North West Rail project, because of the assumption that total employment levels are fixed. The project should also be expected to generate some longer gains in total employment, with associated benefits, but the analysis was not intended to explore this issue.

The approach of Hensher et al. (2012) shows the value for policy purposes of using an integrated land use/transport model, linked to a CGE model, for identifying potential general equilibrium benefits of a major transport initiative. It is unlikely that the potential scale of these benefits in the Sydney example would have been apparent without the analysis, which is conservative because of its assumptions about aggregate Sydney employment levels being fixed for any given year between the do-nothing and improvement cases. The addition of a more dynamic element, to endogenise total future employment in any year in the improvement case, would seem worthwhile, probably increasing benefits from moves to more productive jobs.

11.5 High-speed rail in Australia

The final example of wider economic benefits considered in this chapter is a regional Australian case study of high-speed rail (HSR) between Sydney and Melbourne via Canberra (c.850 kilometres), with the analysis again by Hensher and colleagues from the University of Sydney's Institute of Transport and Logistics Studies (Hensher et al. 2013). This case study is of interest because it not only includes estimates of agglomeration benefits in production but also seeks to value what the authors call 'social agglomeration

benefits'. The analysis suggests that agglomeration benefits in production from HSR are small, mirroring UK findings (see, for example, Graham and Melo 2012), but that social agglomeration benefits are potentially substantial.

The usual approach to valuing benefits to non-work travel is the user benefits approach set out in Chapter 4 (see Figure 4.1), based on change in consumers' surplus for base and generated traffic. Taking an analogous approach to that used to estimate agglomeration benefits in production, Hensher et al. (2013) use zonal populations (instead of jobs), household incomes (instead of GDP), and a measure of distance based on generalised cost to estimate what they call 'household non-work activity agglomeration economies' or, in more shorthand terms, social agglomeration economies from implementation of the HSR project. These are intended to measure the benefits of changes in the ease with which household members can access services and people not engaged in economic production per se, following a transport improvement. Instead of an output (production) elasticity with respect to effective economic density, the social agglomeration approach uses a generalised cost (or accessibility) elasticity with respect to effective social density.

The Hensher et al. (2013) HSR analysis suggests that the social agglomeration benefits that will result from the HSR are worth about \$A4 billion in 2036, an implied value of about \$23 per additional trip. This seems plausible, or perhaps a bit low, in terms of typically Australian costs and benefit values associated with long-distance transport projects.

An important policy question is whether or not this value is additional to benefits to generated traffic. The authors do not suggest this is the case and recognise that their approach may actually be a different way of valuing benefits to generated traffic, based on changes in accessibility. This is a fair assessment. As this is a new area of investigation, policy analysts need to keep a close watch on developments, to understand better the most appropriate ways of measuring benefits from transport improvements.

11.6 Conclusions on agglomeration and wider economic benefits

The UK-led research on agglomeration benefits and other wider economic benefits is shedding new light on the potential value of transport projects, particularly major urban public transport projects. It adds an interesting counterpoint to the focus on external costs that has dominated much transport policy research for the past 40 to 50 years, since it is substantially about external benefits. While it is still early days in the application of such analy-

ses, there is a clear indication that the scale of benefits is significant and should be incorporated in assessments of major urban transport investments. More importantly, the scale of benefits suggests that transport policy makers should think about how transport policy can be used to promote agglomeration economies and other wider economic benefits. This will usually mean concentrating on public transport accessibility to central city areas, but the opportunity to promote such benefits in other urban clusters or nodes in a polycentric city, through well-planned transport investment, should also occupy the minds of transport policy analysts, working with their land use colleagues. Evidence is less clear for such additional benefits from high-speed rail, but transport policy should still be exploring the opportunities to extend use of this technology and identifying relevant benefits expected to ensue.

The linking of regional land use/transport models to regional computable general equilibrium models is an important step in the further development of this research area, to support land use/transport policy analysis and development. The work of Hensher et al. (2012) and Truong and Hensher (2012) has been noted above, and similar research by Anas and Hiramatsu (2013) and Safirova et al. (2007) is cited in Chapter 12. Such models will be strategic, including a relatively small number of zones across a city, but should enable new insights into the links between land use and transport and to the associated tasks of wider economic benefit assessment and transport policy formulation to maximise such benefits.



REFERENCES

- Abrantes, P., R. Fuller and J. Bray (2013), *The Case for the Urban Bus: The Economic and Social Value of Bus Networks in the Metropolitan Areas*, PTEG, February, available at: <http://www.pteg.net/NR/rdonlyres/5F26BBD3-C4A4-4052-A453-D5BFE5E0F0B8/0/ptegCaseforbusreportFINAL.pdf> (accessed 7 March 2013).
- Anas, A. and T. Hiramatsu (2013), 'The economics of cordon tolling: general equilibrium and welfare analysis', *Economics of Transportation*, **2** (1), 18–37.
- Borck, R. (2007), 'Consumption and social life in cities: evidence from Germany', *Urban Studies*, **44** (11), 2015–2121.
- Cervero, R. (2001), 'Efficient urbanization: economic performance and the shape of the metropolis', *Urban Studies*, **38** (10), 1651–71.
- Crossrail (2005), *Economic Evaluation of Crossrail*, available at: www.crossrail.co.uk/assets/downloads/742 (accessed 8 May 2012).
- Crossrail (2011), *Crossrail Business Case Update: Summary Report*, July, available at: <http://www.dft.gov.uk/publications/crossrail-business-case-update-summary-report/> (accessed 15 November 2012).
- DfT (2005), *Transport, Wider Economic Benefits and Impact on GDP*, Discussion Paper, July, London: Department for Transport.
- DfT (2012), *The Wider Impacts Sub-objective*, TAG Unit 3.5.14, August, London: Department for Transport, available at: http://www.dft.gov.uk/webtag/documents/expert/pdf/U3_5_14-wider-impacts120723.pdf (accessed 19 November 2012).

- DOT (2012), *Job Density, Productivity and the Role of Transport*, June, Melbourne: Government of Victoria, Department of Transport.
- Eddington, R. (2006), *The Eddington Transport Study: The Case for Action*, Rod Eddington's advice to Government, London: HMSO.
- Graham, D. (2005), 'Wider economic benefits of transport improvements: link between agglomeration and productivity. Stage 1 report', report prepared for Department for Transport, Centre for Transport Studies, Imperial College, London.
- Graham, D. (2007), 'Agglomeration, productivity and transport investment', *Journal of Transport Economics and Policy*, **41** (3), 317–43.
- Graham, D. and P. Melo (2012), 'Assessment of wider economic impacts of high speed rail for Great Britain', *Transportation Research Record*, **2261**, 15–24.
- Graham, D., S. Gibbons and R. Martin (2010), 'The spatial decay of agglomeration economies: estimates for use in transport investment', final report to the UK Department for Transport, Imperial College, London, October.
- Hensher, D.A., T.P. Truong, C. Mulley and R. Ellison (2012), 'Assessing the wider economy impacts of transport infrastructure investment with an illustrative application to the North-West Rail Link project in Sydney, Australia', *Journal of Transport Geography*, **24**, 292–305.
- Hensher, D.A., R. Ellison and C. Mulley (2013), 'Assessing the employment agglomeration and social accessibility impacts of high speed rail in Eastern Australia', *Transportation*, DOI 10.1007/s11116-013-9480-7, available at: <http://link.springer.com.ezproxy2.library.usyd.edu.au/article/10.1007/s11116-013-9480-7>.
- Marshall, A. (1890), *Principles of Economics*, 8th edn, London: Macmillan.
- Melo, P., D.J. Graham and R.N. Noland (2009), 'A meta-analysis of estimates of urban agglomeration economies', *Regional Science and Urban Economics*, **39** (3), 332–42.
- MoL, Department for Transport, Crossrail and Transport for London (2010), *Crossrail Business Case: Summary Report*, July, London: Mayor of London, Department for Transport, Crossrail and Transport for London.
- Porter, M. (1990), *The Competitive Advantage of Nations*, New York: Free Press.
- Rosenthal, S.S. and W.C.V. Strange (2004), 'Evidence on the nature and sources of agglomeration economies', in J.V. Henderson and J.F. Thisse (eds), *Handbook of Urban and Regional Economics*, Vol. 1, Amsterdam: Elsevier, pp. 2119–72.
- Safirova, E., S. Houde and W. Harrington (2007), *Marginal Social Cost Pricing on a Transportation Network: A Comparison of Second Best Policies*, Discussion Paper RFF DP 07-52, Washington, DC: Resources for the Future.
- Simmie, J. (2004), 'Innovation and clustering in the globalised international economy', *Urban Studies*, **41** (5/6), 1095–1112.
- Transportation Research Board (1998), *Measuring and Valuing Transit Benefits and Disbenefits*, TCRP Report 35, Transit Cooperative Research Program, Washington, DC: National Research Council, Transportation Research Board.
- Truong, T.P. and D.A. Hensher (2012), 'Linking discrete choice to continuous demand within a framework of computable general equilibrium models', *Transportation Research Part B*, **46** (9), 1177–1201.
- Venables, A. (2007), 'Evaluating urban transport improvements: cost benefit analysis in the presence of agglomeration and income taxation', *Journal of Transport Economics and Policy*, **41** (2), 173–88.
- Vickerman, R. (2007), 'Research roundtable: recent evolution of research into the wider economic benefits of transport infrastructure investments', European Conference of Ministers of Transport, Joint OECD–ECMT Transport Research Centre.
- Worsley, T. (2011), *The Evolution of London's Crossrail Scheme*, Discussion Paper No. 2011-27, December, Paris: International Transport Forum.

12

Road user charges



CHAPTER OVERVIEW

Road user charging to make users accountable for the costs of their travel decisions is a topic of considerable interest in many jurisdictions. This is both because of its potential to change traveller behaviour, away from motor vehicle use towards modes with smaller footprints, and because of its revenue-raising potential. Politically, however, it is very difficult territory, as evidenced by the small number of examples where it has been applied. The chapter summarises some of the research on road user charging, identifying estimates of relevant external costs, and then indicates the broad magnitude of charges that might result. A detailed Australian cost estimate is developed, using a 'second-best' fuel charge for application. The chapter also considers alternative ways of funding transport initiatives, including a range of beneficiary pays approaches, which seek to access a wider range of contributors than transport system or service users.

12.1 Context and some principles

Chapters 7 to 10 discussed some of the key external costs associated with land transport, which are primarily (though not solely) associated with road use. Chapter 11 considered agglomeration economies, which are a potential external benefit, particularly from urban public transport (primarily rail). Regional economic development benefits, which have similarities with agglomeration benefits, are also a potential external benefit in some cases. This chapter considers pricing of road use to take account, *inter alia*, of external costs, such as those considered in Chapters 7 to 10, recognising that external benefits also need to be considered in price setting. As was noted in Chapter 9, road user charges are probably the most effective way of allocating property rights for road space.

Transport pricing systems are usually the outcome of trying to balance four objectives:

- economically efficient resource use;
- financial cost recovery;
- social (distributional) objectives; and
- administrative feasibility or convenience.

It has long been recognised that transport costs and prices (or charges) should be connected, to give users price signals that will help ensure *economically efficient resource allocation*. The basic economics of efficient resource allocation, based on marginal social cost pricing, were discussed in Chapter 4. There is an extensive literature that modifies application of the marginal cost pricing rule, to allow for issues such as the absence of similar pricing approaches in markets for competing products or services (see, for example, Button 2010). However, the actual implementation of marginal social cost pricing models has been slow and piecemeal, such that policy in most jurisdictions is still primarily seeking to come to grips with first-order considerations. This involves resolving basic questions such as (for example):

- how to ensure that cost recovery targets are met, if a jurisdiction believes this is important, while applying a marginal social cost pricing system (where the usual answer is to price at short-run marginal social costs and raise any additional revenue required to meet cost recovery targets by charging higher prices to users who are least deterred by higher prices);
- how to calculate relevant marginal social costs, when there are frequently many joint costs involved in provision of transport services and the analytics of costing is still emerging (the European Commission has supported substantial valuable research to improve relevant marginal social cost estimates, with Maibach et al. 2007 a comprehensive source); and
- how to design a pricing scheme that will be acceptable to voters (also the subject of a wide literature, with many commentators proposing linking and dedicating revenues raised to specific transport and/or closely related applications (such dedication of revenue streams is called hypothecation).

The European Commission notes that about 90 per cent of transport external costs are associated with road transport and that about 25 per cent are due to freight movement (CEC 2008). As a result, road has been the major focus of attention in terms of applying marginal social cost pricing in transport. This is an increasing policy concern in many jurisdictions, as the apparent scale of the external costs of transport (particularly traffic congestion costs but also environmental damage) continues to increase. However, it is fair to say that ambition has generally outreached achievement to date, in terms of the application of improved road pricing systems.

In so far as road pricing systems have sought to recover marginal social costs from particular categories of user, it is the European approach to charging heavy vehicles that has perhaps progressed furthest. A small number of congestion charging schemes are also notable (e.g. Singapore, London and Stockholm).

In terms of the real-world application of road pricing systems, it has frequently been cost recovery that has been the primary driving force. This function of pricing is receiving increasing attention today, as governments around the world struggle to find ways of paying for much-needed transport infrastructure improvements, in the aftermath of the global financial crisis and governmental responses to that crisis (which have seen ballooning debt levels in many countries).

Johnson et al. (2012) have recently noted the declining fuel tax revenue flow to the UK national government, as a result of (for example) improving fuel efficiency and slow road traffic growth, and how significant this will become in total national government revenues in coming years (a one-third fall in the annual revenue flow by 2029, at current fuel taxation and vehicle excise rates). To deal with this declining revenue base and improve the efficiency of the pricing regime, Johnson et al. (2012) recommended a pay-as-you-go pricing system, with road pricing (including congestion charging) leading to reductions in fuel taxes and vehicle excise duties.

The US federal gas tax provides revenue to the US Highway Trust Fund. The federal tax rate on gasoline has been unchanged at 18.4 cents per gallon since 1993. Revenue flows into the Fund are, not surprisingly, also declining as per capita car use drops and fuel economy rates improve, resulting in less money for spending on roads and public transport (both of which receive money from the Fund). Congress has had to provide top-up funding since 2008.

US state and local gasoline taxes add an average of just over 30 cents per gallon, such that total US fuel taxes on gasoline are close to \$US0.50 per gallon, although this varies from state to state. Taxes on diesel are a little higher. By contrast, many European countries levy fuel taxes at five or more times this rate, with Germany, the Netherlands and the UK all exceeding \$US3 per gallon.

The US National Surface Transportation Infrastructure Financing Commission (NSTIFC) has proposed that the US should shift from the current US road funding system, based largely on indirect user fees in the form of federal motor taxes, toward more direct user charges, in the

form of a vehicle miles travelled (VMT) charging system (NSTIFC 2009). The Commission proposed that the US federal government commit to deploying such a system by 2020, this timeline recognising the difficulties in implementation. Shortfalls in the US Highway Trust Fund provide a sharp edge to the consideration of this matter in the US, as they do in the UK.

Evidence of the connection between transport and land prices, through the role of accessibility, has been provided by many studies (e.g. National Economics 2010). The huge cost of providing new transport infrastructure, particularly in developed urban areas, coupled with the increasing focus on integrating land use and transport planning, has encouraged a search for new ways of paying for infrastructure. Broadening of the user pays approach towards a beneficiary pays approach, with attention to the role that various land value capture techniques might play in helping to fund transport infrastructure, is now a major policy interest.

More broadly, the beneficiary pays approach recognises that users are not necessarily the only ones who might gain from transport infrastructure improvements. For example, users of a heavily congested bridge might benefit from the construction of a new tolled facility in the adjacent area, which attracts traffic away from the congested facility. It is arguable that these beneficiaries, who are not users of the new facility, should contribute to the costs of its realisation.

Distributional considerations are a frequent reason why road pricing reform is postponed, delayed or simply rejected. For example, if low-income people lack transport choices to avoid or minimise the cost increases that are likely to follow application of marginal social cost pricing, this is usually regarded as unfair and an argument against such pricing. In the absence of attention to distributional impacts, marginal social cost pricing will typically create the conditions for what Chapter 4 called a potential Pareto improvement but not an actual Pareto improvement. Attention to compensation mechanisms is then a necessary ingredient to bridge the gap and helps explain why, for example, London's congestion charging scheme was accompanied by a substantial increase in bus services.

Administrative feasibility has been a stumbling block to comprehensive reform of road pricing systems. Administrative simplicity is the main reason why fuel taxes or excise duties are the major means of raising revenue from road users. Political courage, or lack thereof, has meant failure to index tax rates in some countries, with revenue consequences as outlined above, but

it is a simple administrative matter to achieve indexation when the political courage is present. While tolling technologies have enabled closer alignment of charges and road use, the limited network coverage of such systems constrains what can be achieved in terms of better overall pricing systems. Development of GPS technologies is likely to make significantly improved pricing systems possible in a reasonable time frame, although the Dutch failure to implement such a scheme in recent years shows the political difficulties involved.

12.2 Road cost recovery including external costs

As argued in Chapter 4, economic theory recognises that, in a competitive market economy, the existence of external costs and benefits creates a situation where the market decisions of individual consumers and producers no longer add up to an efficient outcome for society. There is now an extensive literature on the costs of road use and optimal road user charges, dating back many decades. This literature has grown in recent years as congestion has worsened in many cities and the (small) number of live congestion charging schemes has increased.

For road pricing purposes, most discussion of external costs has focused on the following external costs of road use:

- congestion;
- greenhouse gas emissions;
- local air pollution;
- noise pollution;
- the external cost of accidents; and
- road damage.

For example, European Commission heavy vehicle tolling provisions as at 2012 allow for charges for noise, air pollution, congestion, and road infrastructure costs (capital, maintenance and a return), but not for accidents (CEC 2008). Efficient resource use should see efforts to internalise all the listed costs within relevant road user charges.

It is arguable that high community dependence on motor vehicles increases risks of social exclusion for many people, which suggests that there is also a social exclusion external cost of road use (Chapter 6 discusses this subject). Energy insecurity is also increasingly being considered as a negative externality of fossil fuel dependence associated with motor vehicle use (Parry and Small 2005).

Johnson et al. (2012) set out three sets of estimates of marginal social costs of road transport for the UK, as shown in Table 12.1 (not including energy security or social exclusion). If congestion costs are excluded (on the assumption that these are typically only relevant in urban areas and should be explicitly charged in such locations), two of the three estimates are broadly similar in total (at around 3p per kilometre) but not in composition. The low Sansom et al. (2001) estimate (from the earliest period shown in the table) is also of a similar magnitude when congestion costs are excluded, but their high estimate is about twice this size, exclusive of congestion costs.

Table 12.1 Estimates of the marginal costs of road transport

Type of cost	Sansom et al. (2001) (1998 prices)		DfT (2010) (2002 prices) (p/km)	Bayliss (2011) (2009 prices) (p/km)
	Low (p/km)	High (p/km)		
Congestion	9.71	11.16	13.1	4.60
Infrastructure	0.42	0.54	0.1	0.57
Accident	0.82	1.40	1.5	0.88
Local air quality	0.34	1.70	0.4	0.57
Noise	0.02	0.78	0.1	0.50
Greenhouse gases	0.15	0.62	0.3	0.64
Total	11.46	16.20	15.5	7.76

Source: Johnson et al. (2012)

Cost estimates for the US by Parry and Small (2005), drawing from a number of sources, sum to about 5 cents per mile (c.3 cents per kilometre), excluding congestion and energy security, in 2000 prices, which is a little lower than the UK estimates (on an exchange rate adjusted basis) but of a similar scale. Adding congestion costs doubles the costs per mile. The broad similarities between such sources, excluding congestion costs, provide some encouragement that cost estimates may be reasonably robust, but the substantially different magnitudes of the components of the cost estimates are less encouraging.

Theory says that a road user charging system should seek to recover the marginal social costs of road use, which ideally requires examination of how the costs indicated above will change with road use (e.g. traffic volume/mix). Maibach et al. (2007) is a widely used European source, which provides detailed external cost estimates for European countries, seeking to reflect

many of the sources of variation in costs. Button (2010) also includes a number of such estimates.

12.3 Fuel taxes as charges for road use by cars

If efficient road user charging requires a focus on the marginal social costs of use, then the question is how progress toward efficient pricing systems might be achieved (recognising that there are also other objectives for road pricing systems, as outlined above). Most countries are starting with some system of fuel excise and perhaps charges on vehicles and/or drivers, even if these charges are not specifically levied as charges for road use.

Fuel taxation or excise is generally recognised as an imperfect way of charging for road use, because many of the external costs of road use are not well correlated with fuel use. Carbon emissions (and energy security) are clearly related to fuel use, and air pollution has a connection, but noise, accident externalities, road damage and congestion costs are more closely related to distance travelled. However, given the political will, fuel taxation can be adjusted much more quickly than the time taken to implement a new and broader charging system (such as the vehicle mile tax proposed for the US by NSTIFC 2009). Recognising this benefit of administrative feasibility objective, we suggest how fuel taxes might be improved as an initial means of improved charging for road use.

Parry and Small (2005) have developed an optimising model that uses fuel taxes to recover the marginal social costs of road use. It includes three components (Johnson et al. 2012):

1. A range of external costs of road use (local pollution, greenhouse gas emissions, energy security, accidents, congestion), to enable calculation of what is sometimes called a Pigovian (externality-reducing) tax, which prices the marginal costs of the relevant externalities.
2. An adjustment to allow for the efficiency trade-off between commodity taxation and income taxation, which allows for the excess burden of different taxes within the welfare-optimising framework (often called a Ramsey (1927) tax). This component recognises that welfare-maximising revenue arising from commodity taxation should impose higher taxes on commodities with lower price elasticities of demand (such as petrol and diesel used for motoring).
3. A congestion feedback component, which relates to a positive impact on labour supply, and social welfare, of reduced congestion. The effect of this element is very small within the total.

The relevant external costs are estimated as marginal social costs, but a lack of detailed knowledge of the shape of the relevant damage functions for a number of externalities, and the requirement to strike a single fuel tax rate, inevitably means that marginal and average social costs per kilometre are effectively assumed to be equal. External costs that correlate more closely with distance than fuel use are converted to a fuel charge equivalent. The effect of higher fuel prices is allowed to feed back to vehicle choice and to the fuel economy rate, which means that the optimal fuel tax is less than might be implied by simply estimating external costs for the current vehicle fleet.

Parry and Small (2005) use their model to evaluate fuel tax rates in the UK and the US. They conclude that the optimal US rate was more than double its rate at the time of their analysis and the optimal UK rate was less than half the rate in place in that country at the time. Parry and Small emphasise, however, that the optimal fuel tax ultimately depends on how the revenue that is raised from the charge is used, the interaction between fuel taxes and the broader fiscal system impacting on ultimate efficiency outcomes. Their optimal charges assume efficient use of revenues, such as removal of distortionary income taxes.

The distributional impact of increasing fuel taxes is also an important policy consideration that needs to be taken into account. Fuel taxes are generally recognised as regressive. Using increased fuel tax revenues to benefit lower-income groups is one way to deal with this concern. This might include provision of additional public transport services, as was done in London when that city's congestion charging scheme was introduced.

A case study of optimal fuel taxation for cars in Australia

Stanley and Hensher (2011) applied the Parry–Small model to road use by cars in Australia, adopting the values and assumptions set out in Table 12.2. The table is included to show the range of inputs required in the assessment. The analysis is summarised in some detail below, to illustrate the kinds of judgements that policy analysts need to consider when forming policy proposals. Road damage costs are not included in the analysis, on the basis that marginal road damage costs relate primarily to heavy vehicle use and should be recovered only from heavy vehicles.

The carbon price assumed in Table 12.2 by Stanley and Hensher is \$A25 per tonne, broadly in line with early expected application pricing under Australia's carbon pricing scheme. The UK (non-traded) price, estimated

Table 12.2 Parameter assumptions used for base application of Parry–Small model to Australia

Parameter (units)	Base value used	Comments
Initial car fuel efficiency (miles/gallon)	21.5	Authors' estimate.
Pollution damage – distance-related (c/ml)	2.4	Clarke and Prentice (2009).
Pollution damage – fuel-related (c/gal)	32	Carbon \$25/t; energy security 10c/gal.
External congestion costs (c/ml)	10.9	Clarke and Prentice (2009).
External accident costs (c/ml)	3.5	Parry (2009).
Fuel price elasticity	–0.21	Parry and Small (2005).
VMT portion of fuel price elasticity	0.4	Clarke and Prentice (2009).
VMT expenditure elasticity	0.6	Parry and Small (2005).
Uncompensated labour supply elasticity	0.2	Parry and Small (2005).
Compensated labour supply elasticity	0.35	Parry and Small (2005).
Government spending/GDP	0.35	Authors' estimate.
Fuel production share	0.0156	Clarke and Prentice (2009).
Producer price of fuel (c/gal)	227	Clarke and Prentice (2009) on gallon basis.
Initial tax rate on fuel (c/gal)	144.4	Clarke and Prentice (2009) on gallon basis.

Notes: c/ml – cents per mile; c/gal – cents per gallon; /t – per tonne.

Source: Stanley and Hensher (2011, Table 6).

with respect to the UK meeting its emissions reductions targets, is far higher, at £56 per tonne, or about \$A80 per tonne at purchasing power parity. It is thus arguable that the greenhouse gas externalities embedded in the Stanley and Hensher estimates are too low from a long-term cost perspective. Distance-related externalities in Table 12.2 are local air pollution and noise.

The base optimal fuel (petrol) tax for Australia estimated by Stanley and Hensher (2011) is \$A0.94 per litre (Table 12.3). The external cost component of this is 44 cents per litre, suggesting that the Australian fuel excise rate of c.38 cents per litre is not sufficient to cover the external costs of road use. The Ramsey tax component in the base estimate is 50 cents per litre. Revenues to government from the optimal fuel tax are more than double existing revenues, but they increase relatively less than the increase in the fuel excise (or tax) rate, because higher fuel prices drive fuel economies.

Table 12.3 Stanley and Hensher fuel charge estimates for Australia

Basis of calculation	Externality component (per litre)	Ramsey component (per litre)	Optimal tax (per litre)	Ratio of revenue from resulting charge to current revenue
Base estimate	A\$0.44	A\$0.50	A\$0.94	2.27
Accident costs increased (higher value of life)	A\$0.57	A\$0.55	A\$1.13	2.64
No congestion costs	A\$0.39	A\$0.48	A\$0.87	2.11

Source: Stanley and Hensher (2011, Table 7)

Some sensitivity tests

Stanley and Hensher (2011) undertook sensitivity tests on accident costs and congestion costs. They argue that the result from combining these two sensitivity tests (incorporating a higher price for accident costs, to more accurately reflect a willingness-to-pay value, but excluding congestion costs, on the presumption that these should be priced on a location-specific basis, perhaps via a cordon toll – although it can be done through a varied VKT) gives the most appropriate result for policy. The resulting external cost combination implies an externality charge of about \$A0.52 per litre (ignoring the Ramsey component). This is about 14 cents per litre above the current Australian excise rate. This charge would be higher still if a higher carbon price was used, as in the UK. In terms of an optimal fuel tax that includes the Ramsey component, the result is a tax of \$1.05 per litre. This latter scale of change in the fuel excise rate seems far too large to be politically palatable (although this is a decision for the politicians, not for us!), but a 14 cents per litre increase spread over a few years might well be implementable, particularly if hypothecated.

It was argued earlier that distance-based charging is a better way to reflect the external costs of road use than fuel tax (although greenhouse gas emissions are best charged on a fuel base). Parry and Small (2005) note that the welfare gains from an optimal vehicle mile tax are higher than those from an optimal fuel tax. In the Stanley and Hensher (2011) base run, the optimal tax expressed on a per vehicle kilometre basis is 11.4 cents per kilometre. With accident costs included this increases to 15 cents per kilometre. In the 'no congestion costs' case it falls to 4.6 cents per kilometre, showing the significance of congestion costs within the total set of external costs. For Stanley and Hensher's preferred option (increased accident costs but no congestion costs), the charge would be 8.2 cents per kilometre (with an additional

premium for congestion imposed in the most heavily congested locations). Optimal fuel taxes can also be estimated for diesel, as illustrated by Parry (2009).

12.4 Heavy vehicle charging

In discussing optimal road user charges in section 12.3, road damage costs were not included, the argument being that these are primarily attributable to heavy vehicles. Space precludes a detailed discussion of heavy vehicle charging for road use, but some key points about the state of play and policy directions are important.

There has generally been a much closer policy focus on heavy vehicle road charges than on charges for light vehicles, particularly in the context of cross-border travel. This cross-border focus has been on achieving harmonised charging regimes both between states within a single country, as in Australia, and between countries in a broader union, as in the European Union. The purpose when the focus has been between states within a single country has been primarily to increase the ease of doing business and cut administrative costs. In the European Union context, it has also been more about ensuring a level competitive playing field between member states. The Eurovignette scheme and earlier Australian heavy vehicle charging schemes, for example, help ensure fair recovery of road costs from out-of-jurisdiction vehicles. The Australian scheme also covers heavy vehicle road costs within jurisdictions.¹

Setting heavy vehicle charges for infrastructure damage is typically based on fully recovering costs that can be reasonably attributed to heavy vehicle road use.² This means costs that are directly linked to road use, particularly pavement damage (exponentially related to increasing axle loading), and usually also includes a share of joint costs that are attributed to heavy vehicles (as a whole, or to classes thereof differentially, if relevant) through a cost attribution model that is agreed by the parties to the charging scheme.

In calculating the road charges to be levied on heavy vehicles, Australia (for example) goes through the following process:

- Estimate total road expenditure.
- Within this total, estimate the amount that is due to road use or policing.
- Identify cost causation for the resulting amount.
- Allocate attributable costs to users (which is done according to axle class).
- Distribute non-attributable costs.

- Sum the heavy vehicle attributable and share of the non-attributable costs to obtain the total costs to be recovered from heavy vehicles.
- Set a two-part charge, consisting of (1) a common fuel charge (per litre) plus (2) a variable registration charge, as a function of axle class, to recover costs.

Road user charges that result from this process represent about 11.7 per cent of the average operating costs of a typical six-axle articulated truck (with trailer), 8.8 per cent of the average costs of a three-axle rigid truck of over 18 tonnes (with no trailer) and 6.9 per cent for a two-axle rigid truck of 4.5 to 7 tonnes gross vehicle mass (NTC 2007).

Perhaps the most encouraging development in road use charging of heavy vehicles in recent years has been the European provision to incorporate charges for environmental externalities and congestion within the heavy vehicle charging framework, with agreed procedures for their inclusion (CEC 2008). Even though the congestion allowance is required to be revenue neutral, it will provide better price signals than charges that ignore congestion. The opportunity for varying heavy vehicle tolls, for example as a function of a vehicle's emission rating and by location and time, is a significant move to improving the economic and environmental performance of the heavy vehicle fleet and an important step along the pathway to extension to light vehicles, which is currently on the European agenda.

Graham and Glaister (2004) suggest that the price elasticity of demand for road freight movement is about -1.07 . This suggests that a 10 per cent increase in price would lead to about an 11 per cent fall in demand. The implication is that marginal social cost pricing of road freight movement, with all relevant externalities included, is likely to have a noticeable impact on road freight movement volumes (differing, of course, by location, depending on the scale of relevant costs and availability of alternatives and, most importantly, on the proportion of marginal social costs within total freight costs, which will differ significantly between different movement task sectors).

12.5 Congestion pricing schemes

Congestion costs are frequently the largest single external cost of road use, leading to proposals to charge road users for the congestion costs attributable to their road use. Cities such as Singapore, London, Stockholm and several Norwegian cities have now implemented congestion pricing schemes (area- or cordon-based charging), places such as Oregon have had trials, and the Dutch have gone close to implementing a comprehensive GPS-based pricing

scheme. These schemes, plus failed efforts at implementation in some other cities (e.g. Edinburgh, Manchester, New York), have added evidence about costs, benefits and difficulties of implementation. There are a number of detailed reviews of congestion pricing schemes (see, for example, May 2010; Anas and Lindsey 2011; Parry 2012), so only a short overview is provided here.

Congestion pricing schemes are area-based (e.g. London), cordon-based (e.g. Stockholm), facility- or link-based (e.g. charging a congestion premium for faster travel on a lane of a toll road, such as on SR 91 in southern California) or distance-based (e.g. based on comprehensively applied GPS technology, as in the proposed Dutch scheme). The Singapore scheme has a cordon and link basis.

Anas and Hiramatsu (2013) and Safirova et al. (2007) have shown that cordon tolling schemes can realise substantial benefits but less than from more comprehensive charging schemes. A mass/distance/location variant of the distance-based approach may potentially be closest to the economic ideal, in terms of seeking to reflect relevant marginal social costs, but accurate estimation of such marginal social costs and related optimal tolls at a fine level of detail is currently a barrier to the practical implementation of such an (aspirational) approach (de Palma and Lindsey 2011). Congestion pricing schemes thus inevitably involve many simplifications, to ensure feasibility (e.g. using a very small number of daily time periods for charging, to reflect different congestion levels, rather than seeking continuously varying charges by time of day and congestion level).³

May (2010) has reviewed congestion pricing schemes, Table 12.4 summarising his main findings. The evidence is compelling that there is no other initiative

Table 12.4 Impacts of congestion pricing schemes

Characteristic	Impact
Traffic reduction	14–23% from schemes intended to reduce congestion (e.g. London, Stockholm), often growing over time. Less from schemes with lower charges, where revenue raising was the main purpose.
Travel speeds	Relatively bigger than traffic reduction.
Local business in charge area	Minimal.
Emissions	Favourable, but usually relatively less than congestion impact.
Public acceptance	Usually weak before implementation, then majority support.

Source: Based on May (2010).

that can reduce road traffic levels so much, so quickly and in such a sustained manner.⁴ A pricing solution helps to ensure that traffic reductions do not attract additional traffic levels back to fill the void, such that congestion cost savings (benefits) can actually be realised from congestion pricing solutions.

Anas and Lindsey (2011) underline the importance of a logical cordon boundary (e.g. Stockholm's island setting and London's Inner Ring Road) and good-quality public transport services (to give people a choice to avoid paying the charge) to successful implementation of congestion charging. Stockholm and London also had a strong research base on which to draw, and this is seen as an important requirement for successful implementation. Anas and Lindsey (2011) note that decentralisation of jobs in US cities militates against successful implementation of congestion charging in that country, because of the way this decentralisation has helped to reduce potential congestion levels and costs.

Research by Safirova et al. (2007) is useful in contextualising congestion pricing. They use a spatially disaggregated general equilibrium model of the Washington, DC economy, integrated with a strategic transport model, to examine six road pricing schemes for the city: a downtown morning peak cordon; a wider area cordon aimed at the morning peak; the two cordons in tandem; a road toll on freeways and Potomac bridges, where charges are set by distance and time of day; a comprehensive road toll, charging users on all links; and a VMT tax. They examine benefits from congestion pricing and from broader social cost pricing, which includes a charge for air pollution, accidents, greenhouse gas emissions, oil dependency and noise, as well as congestion. They find that the largest welfare benefits are from full marginal social cost pricing, rather than from just pricing congestion. Importantly, they conclude that a VMT tax is almost as efficient as full road pricing when all social costs are included but does require a substantial charge per unit distance (similar in magnitude to the VMT charge calculated by Stanley and Hensher (2011) for Australian cities). They also show that comprehensive charging (by either a VMT tax or a comprehensive toll) has a greater impact than cordon tolling or narrower tolling in terms of reducing external costs.

While there are additional welfare benefits achievable from network-wide marginal social cost pricing, most of those benefits in the Safirova et al. (2007) case study were achievable with a simpler VMT tax (distance-based charging). They emphasise that their conclusions are case study specific. However, taking their findings in association with those of Anas and Lindsey (2011) and the European findings perhaps suggests that, in lower-density cities, a VMT may be a more effective end point than a comprehensive road

tolling system (like a GPS-based charging system). In more congested cities the more complex GPS-based approach may be the preferred end point in terms of potential benefits, but the complexity of calculating the optimal toll remains a difficulty (de Palma and Lindsey 2011).

The difficulty of implementing pay-as-you-go charging in any jurisdiction is likely to necessitate a wide-ranging community consultation, about the need for reform and the best way to implement such reform. A suitable community consultation might need a two- to three-year period to cover such issues as:

- why road pricing needs to change;
- the options for change;
- how these options will impact on various stakeholders (where scenarios would be useful in describing expected outcomes);
- what will happen to revenue raised from the charges;
- what measures might be implemented to mitigate particular adverse impacts; and
- how privacy will be protected if comprehensive mass/distance/location charging is adopted.

If road prices more closely reflect the relevant marginal social costs of the travel in question, the case for funding support to public transport, through a low-cost recovery rate, reduces. However, there will still remain strong social safety net arguments for some governmental funding support of public transport, even in a regime of marginal social costing of road use.

12.6 Land value capture for land transport funding

It has long been understood that accessibility influences land prices and is itself influenced by the quality of the land transport system. Places that have better accessibility typically have higher land prices. Transport improvements frequently improve accessibility and increase land values, benefiting landowners and developers. Reflecting the beneficiary pays principle, land value capture mechanisms can be used to convert part of this created benefit into revenue, which can be used to help fund the relevant transport initiative that generated the value gain, or other public purposes.

Land value capture is particularly relevant for major urban rail projects that support urban agglomeration economies, particularly when the external costs of road use are not properly priced. Increasing user charges (fares) to help pay for such improvements will divert passengers away from rail and increase the external costs of road use. Land value capture, on the other

hand, will extract part of the landowner or developer benefit from land price increase and reduce the need to seek additional revenue from system users. This will avoid, or at least reduce, any adverse feedback effects of fares on road externalities. Some user contribution is still likely to be required and is supportable in policy terms, being greater if external costs of road use are priced.

A number of possible approaches to land value capture are outlined below, drawing on CTS (2009). Most of these measures relate to revenue raising from specific initiatives, but some are more general.

Tax increment financing

Tax increment financing (TIF) is widely used in the US and can now be used by local government in the UK to help drive local investment and economic growth. In essence, TIF allows a (usually) local government to borrow against predicted growth in locally sourced revenues in a defined area, to help fund activities that will drive that growth. Bonds are usually issued to provide the necessary upfront finance for infrastructure or urban renewal initiatives, additional annual local tax (rate) revenues being used to fund the interest and principal repayments. TIF is particularly suited to an urban renewal context.

A key issue in relation to TIF as a possible funding source is the extent to which the infrastructure programmes being financed lead to a net increase in development-related revenues to the sponsoring government, as distinct from simply diverting revenue from one area to another (even within the same municipality). US evidence on this account is mixed, Dye and Merriman (2008), for example, finding little evidence that TIF actually led to net new development in a Chicago area case study.

Special assessments

These impose special charges on property close to a new facility, with the charges only being raised for those properties that receive a special (identifiable) benefit from the public improvement, such as a new transport facility. Committee for Melbourne (2012) uses the generic description of 'Benefitted Area Levy' for this type of funding source. For example, Melbourne's Regional Rail Link and Sydney's North West Rail project will benefit properties located adjacent to proposed stations. Some value capture in relation to such properties may also be pursued through means such as air rights development or joint development projects, as discussed below, but all properties that will clearly gain could be subject to a special assessment, to value-capture

part of the relevant uplift. This approach is widely used across the United States, typically for local infrastructure improvement projects. It has much in common with TIF. It was also used specifically for a public transport project in the case of the Red Line Metro Rail in Los Angeles in the 1980s (Stopher 1993). In this case, a special assessment was applied to commercial buildings in an area of the CBD of Los Angeles, with the intent of raising \$130 million of the costs of constructing the Red Line. The special assessment was based on the fact that the construction of the Red Line would increase property values in the vicinity of stations, up to a distance of about 1 kilometre from the CBD stations.

Metropolitan improvement levy

This is a broad-based charge related to all properties in a large area, set at a low rate and used to fund specific government services. It might be levied on a flat rate per property basis or on a proportion of property value basis. The advantage of the latter is that it implies an element of value capture and is not as regressive as a flat levy. Discussing this approach in a transport context, Committee for Melbourne (2012) calls this a 'Broad-Based Transport Improvement Levy'. It is particularly suited to help fund services that are widespread throughout the charging area, such as public transport services. Thus, for example, a metropolitan improvement levy could be used to help fund PT service costs in growing suburbs, on the argument that there are direct user benefits (including social inclusion benefits, as discussed in Chapter 6), 'option benefits' (essentially insurance benefits, to those who might possibly need to, or wish to, use the service at some future time, i.e. a form of beneficiary pays) and reduced external costs of road use from the availability of such services. In an Australian urban context, the redistribution of revenue raised from inner or middle to outer areas implied in this arrangement may have equity benefits, since most public transport services or benefits currently accrue to better-off inner or middle urban residents.

Development impact fees (developer contributions)

Development impact fees, also known as developer contributions, are one-time charges levied on new development. They are commonly used in Australia (e.g. for greenfields development and major precinctual urban projects such as the Melbourne Docklands) and are mainly levied on new development, to help recover costs of public infrastructure or services, growth-related public service costs, such as new rail level crossings (if a development creates a need for such a facility), parks or open space and perhaps local public transport. These charges have some similarities to negotiated

exactions but differ in so far as development impact fees are usually determined by a formula, related to expected public service costs attributable to a level of new development, rather than through the less formal negotiation processes typically used with negotiated exactions. Levying such charges on a consistent basis across all new urban development is appropriate, particularly with urban infill being expected to play a bigger role in most cities.

Negotiated exactions

Negotiated exactions might cover similar types of costs to a development impact fee but are subject to negotiation, rather than being the outcome of a formulaic process. They may be in-kind contributions (e.g. of open space), instead of money.

Joint development

In a transport context, joint development refers to the development of a transport facility and adjacent private real estate, often based around a railway station, where higher-density development might accompany station redevelopment (e.g. transit-oriented development). This might involve a partnership between a public land development agency or transport authority and a private sector developer. There are a number of possible joint development models, with varying equity, risk allocation and revenue and cost treatments. Joint development may include air rights development (see the next section), such as above a railway station. Such proposals are unlikely to generate sufficient funding to facilitate developments beyond those covered by the particular joint development arrangements, but they can be significant for a small number of particular major development opportunities.

Air rights

Major new transport projects, or urban development projects, may add value to the space above (or below) a transport facility. For example, air rights above Wurundjeri Way in Docklands (Melbourne) were part of a development proposal before the market for bidding in early 2012. Air rights agreements establish the right to develop above (or below) a facility, in exchange for a financial contribution or future additional property and/or income taxes (depending on jurisdictional income-raising opportunities). Revenue from such an initiative may be used for a range of public purposes, such as place making, but is most likely to be retained within the development site.

In Australian cities, development above railway stations usually has a high cost for podium development, relative to surrounding land prices. This typically means high-density development will be needed to establish a financially feasible opportunity.

12.7 Increased borrowings

A traditional way of financing investment in land transport infrastructure has been government borrowings. A major advantage of this approach is that it enables the funding of these financing costs to be spread over the life, or part thereof, of the asset, so that the generation(s) that benefit(s) can meet the financing costs. A disadvantage, however, is that these costs are not specifically financed by users (unless explicitly levied as a user charge for the facility developed with the borrowings, like a toll) but by taxpayers more broadly. In the current fiscal environment, where governments are seeking to keep a tight rein on spending, increased borrowings for infrastructure are not popular politically. However, if the relevant investment generates significant public value, then borrowings may be an effective means of assuring implementation.

12.8 Public–private partnerships

Public–private partnerships (PPPs) have played a major role in development of significant transport infrastructure investments over the past couple of decades, particularly urban toll roads, where private equity and borrowings for infrastructure financing are rewarded through associated user pays (toll) funding. As an investment vehicle, PPPs have lost some of their gloss in recent years, with (for example) concern over high and escalating bidding costs and some significant shareholder losses being associated with some poorly bid projects (some PPPs seem to have been particularly vulnerable to ‘optimism bias’). Private borrowing costs are typically well above public costs, suggesting that PPPs need to play a role of complementing publicly funded infrastructure, such that the total level of investment is higher than would otherwise be possible. Higher private sector borrowing costs (and profit expectations, in a risky environment) mean that careful selection of major projects to be the subject of PPPs is vital. From a public sector viewpoint, it is critical that the granting of a major transport PPP, with its associated long-term operating rights, does not entail significant loss of transport network control.

In view of the losses on some major recent projects, a greater reliance on the public sector taking more of the construction-stage risk, with the

operational stage being contracted out once traffic flows have settled down (essentially as a management contract), might be worthwhile for some project PPPs.

12.9 Asset sales

The sale or lease of government assets, such as publicly owned ports, is a means of freeing up funds for new infrastructure. Sale of existing freeways is another possibility for raising substantial sums. This could take the form of an outright sale or a long-term management lease or perhaps the narrower form of the sale of a lane on a freeway, for use as a high-occupancy toll lane.

12.10 Conclusion on sustainable pricing and funding

Infrastructure backlogs and governmental focus on reining in spending, together with the emerging trend in a number of countries of declining fuel tax revenues, are highlighting the urgency of finding new ways to fund transport infrastructure. At the same time, the unpriced external costs of road use are increasing the focus on user pays charging, for more efficient resource allocation. Arguments of efficiency and fairness support a greater reliance on user pays and beneficiary pays pricing systems. User pays systems have the allied benefit of reducing the size of the apparent investment backlog (by encouraging behaviour change), provided equity concerns are handled. There are a number of ways user charging can be improved, from better aligning fuel taxes through to comprehensive road pricing reform.

The increased focus on land use/transport integration as a policy direction focuses attention on how the benefits of transport infrastructure are transmitted through the urban system. Much benefit will ultimately accrue to landowners or developers, who should contribute to the costs of the initiatives that increase the value of their assets. This land value capture (beneficiary pays) approach should be used more widely, and there is a range of value capture mechanisms available to this end.

Optimising funding opportunities across user pays mechanisms (including those that are associated with PPPs) and various beneficiary pays mechanisms, together with direct government funding, requires careful balancing of the funds raised from each mechanism, to ensure the totality is effective and equitable.

NOTES

- 1 These two schemes approach their cost recovery purposes differently, the European scheme adopting a design lifetime costing framework, including a return on capital, whereas the Australian scheme is a pay-as-you-go scheme.
- 2 Chapter 4 set out some definitions of costs that are relevant to the following discussion.
- 3 A very useful reference on the complexities of congestion pricing, and the technologies used for implementation, is de Palma and Lindsey (2011).
- 4 UK research suggests that reductions in road traffic levels of 4 per cent can cut congestion costs by about 40 per cent (DfT 2004, Table B).



REFERENCES

- Anas, A. and T. Hiramatsu (2013), 'The economics of cordon tolling: general equilibrium and welfare analysis', *Economics of Transportation*, **2** (1), 18–37.
- Anas, A. and R. Lindsey (2011), 'Reducing urban road transportation externalities: road pricing in theory and practice', *Review of Environmental Economics and Policy*, **5** (1), 66–88.
- Bayliss, D. (2011), 'A speculative estimation of direct road user charging impacts', Appendix in S. Glaister, L. Lytton and D. Bayliss (eds), *Funding Strategic Roads*, London: RAC Foundation.
- Button, K. (2010), *Transport Economics*, 3rd edn, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- CEC (Commission of the European Communities) (2008), *Directive of the European Parliament and of the Council: Amending Directive 1999/62/EC on the Charging of Heavy Goods Vehicles for the Use of Certain Infrastructures*, COM/2008/0436 final, Brussels: Commission of the European Communities, available at: <http://eur-lex.europa.eu/LexUriServ.do?uri=COM:2008:0436:FIN:EN:PDF> (accessed 24 December 2012).
- Clarke, H. and D. Prentice (2009), 'A conceptual framework for the reform of taxes related to roads and transport', report prepared for the Treasury, Canberra, June.
- Committee for Melbourne (2012), *Moving Melbourne*, Taskforce Report, September, Melbourne, Victoria: Committee for Melbourne.
- CTS (2009), 'Harnessing value for transportation investment: a summary of the study. Value capture for transportation finance', Centre for Transportation Studies, University of Minnesota.
- DfT (2004), *Feasibility Study for Road Pricing in the UK: Report*, Appendix B: 'Modelling results and analysis', report to the Secretary of State, London: Department for Transport.
- DfT (2010), *Transport Analysis Guidance (TAG) Unit 3.9.5: MSA Major Schemes Appraisal Road Decongestion Benefits*, London: Department for Transport.
- Dye, R. and D. Merriman (2008), *Tax Increment Financing: A Tool for Local Economic Development*, Cambridge, MA: Lincoln Institute of Land Policy, available at: http://www.lincolninstitute.edu/pubs/1078_Tax-Increment_Financing (accessed 10 May 2012).
- Graham, D.J. and S. Glaister (2004), 'Road traffic demand elasticity estimates: a review', *Transport Reviews*, **24** (3), 261–74.
- Johnson, P., A. Leicester and G. Stoye (2012), *Fuel for Thought: The What, Why and How of Motoring Taxation*, report prepared by the Institute for Fiscal Studies for the RAC Foundation, May, London: RAC Foundation.
- Maibach, M., C. Schreyer, D. Sutter, H.P. van Essen, B.H. Boon, R. Smokers, A. Schroten, C. Doll, B. Pawlowska and M. Bak (2007), 'Handbook on estimation of external cost in the transport sector', produced within the study Internalisation Measures and Policies for All External Cost of Transport, CE Delft, 19 December.
- May, A. (2010), 'Road user charging and implications for transport policy: findings from the

- CURACAO Project', paper presented to World Conference on Transport Research, Lisbon, July.
- National Economics (2010), *State of the Regions 2010–11*, Canberra: Australian Local Government Association.
- NSTIFC (2009), *Paying Our Way: Report of the National Surface Transportation Infrastructure Financing Commission*, Washington, DC: National Surface Transportation Infrastructure Financing Commission.
- NTC (2007), *2007 Heavy Vehicles Charges Determination: Regulatory Impact Statement*, Vols 1 and 2, December, Melbourne: National Transport Commission.
- Palma, A. de and R. Lindsey (2011), 'Traffic congestion pricing methodologies and technologies', *Transportation Research Part C*, **19** (6), 1377–99.
- Parry, I.W.H. (2009), *How Much Should Highway Fuels Be Taxed?*, Discussion Paper RFF DP 09-52, Washington, DC: Resources for the Future.
- Parry, I.W.H. (2012), 'Reforming the tax system to promote environmental objectives: an application to Mauritius', *Ecological Economics*, **77**, 103–12.
- Parry, I.W.H. and K. Small (2005), 'Does Britain or the United States have the right gasoline tax?', *American Economic Review*, **95** (4), 1276–89.
- Ramsey, F. (1927), 'A contribution to the theory of taxation', *Economic Journal*, **37**, 47–61.
- Safirova, E., S. Houde and W. Harrington (2007), *Marginal Social Cost Pricing on a Transportation Network: A Comparison of Second Best Policies*, Discussion Paper RFF DP 07-52, Washington, DC: Resources for the Future.
- Sansom, T., C. Nash, P. Mackie, J. Shires and P. Watkiss (2001), 'Surface transport costs and charges', report prepared for the Department of the Environment, Transport and the Regions, Institute of Transport Studies, University of Leeds, April.
- Stanley, J.K. and D.A. Hensher (2011), 'Environmental and social taxes: reforming road pricing in Australia', *Road and Transport Research*, **20** (4), 71–83.
- Stopher, P.R. (1993), 'Theory and practice in financing urban rail projects: the case of Los Angeles', *Transportation*, **20** (3), 229–50.

13

Potential solutions – public transport investment and technology



CHAPTER OVERVIEW

This chapter presents some broad background information on the role played by public transport (PT) and then reviews a range of key policy issues associated with its provision. These issues encompass why PT is needed, how PT might be best provided and the future role that PT might play. The discussion about role notes that a small number of PT services operate commercially but that the main transport policy argument for urban PT is the savings in the external costs of road use that it can deliver. A case study of Melbourne illustrates this proposition. The potential roles of different PT modes are briefly considered. Where PT is to be provided by the private sector, the merits of competitive tendering and negotiated performance-based contracts as alternative awarding mechanisms are considered.

13.1 Land transport policy context

Public passenger transport is transport that is available to the general public for purposes of undertaking trips. The usual modes of PT travel are rail (including metros, heavy rail, high-speed long-distance rail, light rail and trams), bus (including bus rapid transit (BRT), conventional bus, minibuses, and bus taxi services), ferries and taxis (from the famous London variety through to Bangkok's equally famous tuk tuks and the minibuses of the South African 'informal' sector).

A close relation of PT is what is sometimes called 'community transport'. Whereas PT is characterised by universal availability, particularly in jurisdictions that use disability discrimination provisions, 'community transport' services have more restricted availability but are also important in many jurisdictions. These are typically limited services that are targeted to particular

groups with special needs, defined (for example) by age, physical or other capacities. Some reference to community transport is included in Chapter 6. It is not part of the subsequent discussion in the present chapter. We note, however, that bridging the divide between PT and community transport, to provide a more seamless set of publicly available passenger transport options, is a matter of emerging policy interest.

The growing international concern with traffic congestion and with a number of other external costs associated with motor vehicle use that are discussed in various chapters of this book (such as air pollution, greenhouse gas emissions, energy insecurity, adverse safety and health outcomes, and social exclusion), together with the growing understanding of the economic importance of agglomeration economies, has led to an increased policy interest in a greater role for PT. This increased interest recognises the *potential* PT has to improve outcomes with respect to these areas of policy concern, in terms of moving people. This interest is not confined to urban areas, although it is vital in such locations. The strong interest in high-speed rail that is apparent worldwide at present, for example, is also partly a response to external costs of current transport choices, in this case related (for example) to greenhouse gas emissions from aviation. A critical land transport policy challenge is to plan and deliver future PT services in such a way that the potential to contribute to valuable economic, social and environmental outcomes is realised cost-effectively.

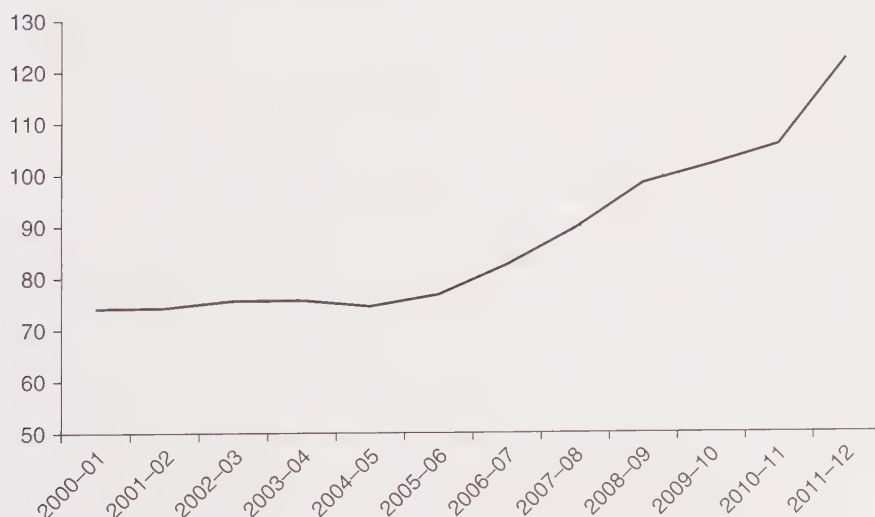
The proportion of the passenger transport task performed by PT varies substantially between countries and also within particular countries. In general, the PT mode share increases as population or job densities increase and service levels improve but has tended to decline as income levels rise. High-income cities that are characterised by low-density settlement patterns seldom exceed a 10–15 per cent PT mode share, Sydney (for example) having Australia's highest PT mode share for urban travel at 12.1 per cent (BTS 2012). However, such cities may exceed 50 per cent PT mode share for peak-hour journeys to work in their central business districts, Sydney PT (for example) carrying about three in every four peak CBD commuters who use motorised transport (BTS 2012). On the urban fringe, however, PT mode shares nearer to 5 per cent are common in developed low-density cities.

High-income countries with high population or job densities typically have higher PT mode shares. For example, the 27 European Union countries as a whole had 18.1 per cent PT mode share for *motorised trips* (calculated as a proportion of passenger kilometres), with a 22.7 per cent share for the core EU 12 countries (EC 2010). Hungary had a high 40.1 per cent of

passenger kilometres by motorised modes undertaken on PT. These shares are for all passenger travel in the EU countries; the urban PT mode shares will be higher, and they are typically higher in (for example) Asian and South American countries.

Over the period from 1995 to 2008, the 27 EU countries' PT mode share declined by 1 percentage point. In contrast, it has grown strongly in some Australian cities (e.g. Melbourne, Perth and Brisbane). Such contrasting results are often a reflection of investment in PT system improvement over the period in question. For example, Melbourne invested in a substantial upgrade of its bus services over the period commencing 2005–06, with strong patronage growth resulting (Figure 13.1). Brisbane has shown similar high growth rates in bus patronage, linked particularly to the roll-out of its world-class bus rapid transit network and implementation of high-frequency bus upgrade zone (BUZ) routes, some of which use the busways. Brisbane Transport bus passenger numbers increased from about 45 million in 2000 to 80 million in 2011. The lesson from both Melbourne and Brisbane is that investing in high-quality bus services increases patronage. Service frequency and the span of service hours and days are central to patronage growth outcomes.

Historically, PT has often been thought of as an 'inferior good', in the sense that demand declines as income levels rise (implying a negative income



Note: In million boardings per annum.

Figure 13.1 Patronage on Melbourne's route buses

elasticity of demand). Frank (2008), for example, cites an income elasticity of demand of -0.36 for public transport, suggesting that a doubling of income would reduce PT demand by 36 per cent. The tendency for PT mode shares to decline in most advanced countries from the 1950s, as car ownership levels increased, supported this view. Holmgren (2007), in contrast, found a small positive income elasticity of demand for bus travel (0.17), and Asquith (2011), in reviewing a range of studies on income elasticity of demand, suggests that the results are rather ambiguous. The tendency for urban PT mode shares to be highest for rail commuting trips to central business districts, where incomes are typically highest, reflects this ambiguity and highlights the difficulty of disentangling causal relationships. The major turnaround in PT patronage growth being experienced in some high-income cities, such as Melbourne, Brisbane and Perth in Australia, suggests that the 'inferior good' label may be losing its relevance, in the face of increasing traffic congestion and improved PT services. These are issues that need to be understood in specific jurisdictional contexts if PT is to be able to realise its future local potential.

The focus in this chapter is on policy issues that relate to various aspects of the provision of passenger PT, particularly urban PT. We do not seek to provide detailed descriptions of the characteristics of individual modes, such as operating performance, cost structures and the like. Texts such as Vuchic (2005) do this more than admirably. Our focus instead is on matters that are likely to be of concern to the policy analyst whose interests include public transport and its potentialities, either as a dominant focus or perhaps as one within a wider policy frame of reference.

The key policy issues on which we focus are:

1. the need for public transport – where commercial and wider social benefit issues are particularly important;
2. the provision of public transport – which includes questions such as whether provision should be by government (either directly or through some form of contracting arrangement), the merits of different forms of contracting, the need for regulations to help ensure particular operating standards, and a range of distributional policy matters; and
3. the future role public transport might play – where we focus not on potential modal innovation so much as on the potential for public transport to perform a role of increasing policy importance to society, particularly in terms of triple-bottom-line outcomes.

In exploring these matters, we do not start from any particular presumptions about PT having particular intrinsic value. We believe that PT value is

context specific and, in our view, PT needs to be able to clearly demonstrate this value in relevant policy contexts. The chapter provides some examples to illustrate how this might be done.

13.2 The need for public transport

User benefits

The answer to the question of the need for PT is different in different settings, but a core element of the answer is common: PT provides benefits to users and potential users. At the simplest level, PT provides a means of undertaking trips that is an alternative to the private car. Some who have a private car may choose to use PT, because they find its generalised cost is less than that of the car for particular trip purposes. Generalised cost is sometimes expressed as follows:

$$G = g(C_1, C_2, C_3, \dots, C_n) \quad (13.1)$$

where G is generalised cost and C_1, C_2 and so on are the various time, money and other components that constitute the 'costs' of travel. The various cost arguments each have a quantity dimension (e.g. minutes of in-vehicle time and waiting time, minutes of walking time, litres of fuel) and a value dimension (e.g. the unit values of walking time and waiting time, price of fuel per litre, and so on), which enables an estimate of the overall user generalised cost of selecting a particular mode for a particular trip.

Some people may choose, for whatever reason, not to own a car, or they may not have access to a car. PT may provide a very important means of mobility and accessibility for these people. The high user benefit of mobility (trip making) for many people in this situation is discussed in Chapter 6.

Another group of people, who use their car for a particular trip, may still value the *option* of using PT and/or value its availability for others, and they may be willing to pay for this option. The concept of option value is explained in Chapter 4. These people do not receive a direct user benefit, but their option value is relevant to consideration of the value of PT.

External benefits

In transport policy terms more broadly, it is useful to think of PT as performing two major roles: a *mass transit* role and a *social transit* role, which largely associate with trunk movement and local movement respectively.

Users benefit from both these roles, but the roles are differentiated by the primary external benefits that are associated with each, as illustrated in Figure 13.2. The trunk movement (mass transit) role performed by PT is essentially about aggregating large numbers of passengers in time and space. Destination density is an important driver of load possibilities and of the market share possibilities of PT (Ewing and Cervero 2010); the peak is particularly important, and peak trip purposes are primarily work and education. Capital costs of meeting peak movements are usually high, such that peak-shifting strategies are increasingly common in PT systems where loads are on the increase, to improve asset utilisation.



Figure 13.2 The benefits to society from mass transit and social transit

The mass transit role can realise substantial urban PT benefits in the realm of congestion mitigation, particularly for rail and BRT modes. The capacity of a single rail line is equivalent to about ten lanes on a freeway, such that rail is a very efficient way to cater for highly concentrated peak movements. The mass transit role is thus also crucial in the sustaining of urban agglomerations in central cities and major urban nodes, as discussed in Chapter 11. Figure 13.2 also lists a number of other potential external benefits from the operation of PT in mass transit (trunk) mode, with some being relevant in regional settings as well as urban. Chapters 7 and 8 are relevant to such matters.

The social transit role is primarily about reducing the risks of social exclusion that are associated with poor mobility opportunities (Chapter 6). This is mainly about local accessibility rather than with mass movement on trunk routes, although the local movement may feed a mass transit service. Australian research in this area has been notable in developing the idea of urban PT as a social safety net service, to enable most people to undertake most of the activities they wish to undertake, most of the time. Low-density settlement patterns (urban and regional) are the most difficult locations in which to accomplish this social safety net role, because of the costs of service provision in such settings.

Melbourne has a population of 4 million people. Its route bus services operate primarily in middle and outer suburban areas, with train and tram services providing the major PT service to the CBD and in the inner suburbs. There are only a small number of trunk radial bus routes. Table 13.1 sets down the indicative 'social' value estimated to be provided annually by Melbourne's route bus services, focusing particularly on lowering the estimated external costs of motor vehicle use. These values were estimated against a hypothetical base situation in which it was assumed that the route bus services ceased to exist. Estimates were made of alternative travel choices that were likely to be taken, including trip cancellation. No attempt was made to ascribe agglomeration benefits, on the argument that train and tram services provide the bulk of the Melbourne concentrated commuter PT task, or to estimate option value (because there is no local evidence on this matter).

The analysis suggested that social inclusion benefits are the largest single benefit from Melbourne's route bus services, drawing on the analysis set out in Chapter 6, substantially greater than the (also very significant) congestion

Table 13.1 Indicative annual value of Melbourne's route bus services, 2010

Value of route bus services in Metro Melbourne	\$ million p.a.
Congestion time (\$518m) and fuel (\$70m) benefits	588
GHG (\$7.5m), local pollution (\$12.2m) and energy security (\$1.6m)	21
Accident savings	15
Bus user benefits of social inclusion = 33m trips @ \$23.25 per trip	767
User benefits for other bus users = 67m trips @ \$5 per trip	335
Total value (externality + user benefits)	1726
Gross cost to budget	486
Benefit cost ratio (BCR)	c.3.5

Source: Updated by the authors from Stanley (2008).

cost savings. These relativities reflect the focus of route bus operations in middle and outer suburbs. Melbourne's train and tram services would score highest on congestion mitigation and agglomeration economies, in terms of external benefits. Environmental and safety benefits of the route bus services were assessed as having only minor value, partly because of the low loading rates on some local bus services. This reflects low settlement densities, although the social exclusion research suggests that boarding rates of about eight passengers per hour are a break-even loading in the Melbourne setting for social inclusion benefits (see Chapter 6). User benefits were much smaller than the external benefits estimated to be attributable to service provision, although the user benefit estimation process was crude. Overall, the services were found to deliver annual benefits of over three times the annual cost of service provision. Financial cost recovery is closer to 30 per cent of annual costs, or less than one-tenth of the estimated service benefit/cost ratio, showing the importance of taking both financial and wider triple-bottom-line perspectives for informed PT policy making. Looking solely at the financial cost recovery rate one might well conclude that the Melbourne route bus services are not viable. Taking account of their wider external benefits suggests they are very worthwhile for the city and its residents, particularly those living in the middle and outer suburbs.

Table 13.1 sets out the results of an *overall* assessment of the value of a set of PT services. It is more usual for policy assessments to deal with the value of *changes* in the level of service that is provided. In this case, marginal analysis is relevant, looking at how changes in service levels will change user benefits and relevant external costs.

Commercial/non-commercial operation

If PT provides benefits to users, then a fare can be charged to extract some of the consumers' surplus (as defined in Chapter 4) and help pay for the service. Thus, for example, most UK bus services outside London rely on the fare box for their survival, with some tax concessions, as does the informal minibus/taxi sector in South Africa. Long-distance coach services in many countries also operate solely on the basis of fare-box revenue. Asian rail operators often capture some of the value created by a rail line through property development. In consequence, a number of PT services are able to operate fully commercially, from a combination of fares and value capture and other minor revenue sources (Chapter 12 discusses value capture in more detail).

In many situations, however, commercial revenue opportunities are not sufficient to meet the costs of public transport. A financial deficit is very

common. The Melbourne route bus cost recovery rate of about 30 per cent was noted above. In 2010–11, Sydney's PT fares recovered 22 per cent of city rail operating costs, 32 per cent of Metro Bus operating costs and only 11 per cent of outer Metro Bus operating costs (Australian Government 2012). Canadian PT systems average about 60 per cent recovery of operating costs from the fare box. In both countries the cost recovery rate is even lower on total costs.

When PT revenue streams fall short of meeting the costs of service provision, governments need to decide if some level of financial support is justified. Valuation of the relevant external benefits and costs of PT, along the lines illustrated in Table 13.1, is then important to shed light on policy choices. In constrained funding situations, governments may seek evidence that benefits are proportionately higher than costs, to reflect forgone opportunities in other sectors of funding that is committed to PT, the implication being a minimum acceptable benefit/cost ratio (or shadow price of public funding) greater than 1.

A further set of considerations relating to the commercial/non-commercial operation of PT services relate to the question of declining long-run costs, such as may characterise some rail systems. Chapter 3 pointed out that maximising the social benefits from particular goods and services generally requires pricing at marginal cost. However, in a situation of declining long-run costs, marginal cost pricing will incur financial losses, marginal costs being less than average costs. A related concept is the Mohring effect, whereby increased frequency of PT services reduces waiting times and costs, having the effect of producing increasing returns to scale (or decreasing costs) for urban PT services (Mohring 1972). Governments in such situations may choose to follow a marginal cost pricing strategy and fund the resulting 'deficit' by other means (e.g. from general revenue or from charges on the value of land, as a form of beneficiary pays, as discussed in Chapter 12).

Choosing the modes, or horses for courses

The various potential external benefits that may be created by PT services will be met, to a greater or lesser extent, by all public transport modes, but the efficiency and effectiveness with which different modes might contribute to meeting these desired urban outcomes will usually differ. A frequent response of politicians to transport problems, especially congestion, is to adopt a policy of investing in public transport, with the preference usually being some form of rail, regardless of whether it is the most appropriate. It

is, therefore, worth examining briefly the characteristics of various public transport modes, to understand their applicability in various situations. For example, while metros, heavy rail and BRT are probably the most effective and efficient way to promote urban agglomeration economies and reduce urban congestion costs, in many settings they will not be the best way to promote social inclusion as it is affected by mobility.

Different modes are better at some tasks than others. Table 13.2 sets out indicative passenger-loading capacities of particular modes, to help position their potential contribution to urban passenger transport. Development of a city's most effective and efficient transport system requires attention to such technological and economic characteristics of various options and how they align with user needs, wider economic, social and environmental opportunities and constraints and the geography of place in that location.

Table 13.2 Maximum offered line capacities of some urban public transport modes

Mode	Vehicle dimensions (metres, L × W)	Maximum offered line capacity per hour
Standard bus, single stops	12 × 2.5	3800–5400
Articulated bus, single stops	18 × 2.5	5400–7200
50% standard/50% articulated bus, multiple stops	12 × 2.5; 18 × 2.5	8800–17 500
High-capacity bus	12 × 2.5	9000–30 000
Streetcar, surface street, mixed traffic	14 × 2.2	13 200–26 400
Light rail transit, separated right of way, at-grade crossing	24 × 2.65	12 200–26 900
Automated guided transit – Westinghouse, VAL systems	12.5 × 2.05	13 800–16 500
Automated guided transit – rail	12.7 × 2.5	21 600–24 000
Rapid transit (large city, developed country)	18 × 2.9	51 800–57 600
Rapid transit (new systems to maximum capacity)	21 × 3.15	67 200–72 000

Source: Based on Vuchic (2005, Table 2.4)

Rail

In many countries, the idea of investing in rail is seen as politically desirable and, in some cases, has been for some years. The rail starts programme in the US is a manifestation of this. In comparison, bus is often seen as being much less politically desirable.

Rail comes in a number of different forms, from heavy rail to light rail, and with variants on each of these, from a true heavy rail system, such as British Rail in the UK or Amtrak in the US, to trams or streetcars, such as those to be found in cities like Melbourne, Adelaide, Budapest, Vienna, and many others around the world. True heavy rail is usually a form of intercity public transport, built at grade or in cuttings and on embankments, utilising trains of substantial length and carrying capacity, operating with diesel engines or overhead electric supply, and usually entailing station spacings of 3 to 5 kilometres or more, except in downtown areas, where station spacings may be as close as 1.5 kilometres. Next is metro rail, which is often built underground, although may also be on the surface, is usually electrically powered, either from overhead wires or a third rail (occasionally even a four-rail system), has moderate carrying capacities, and usually has station spacing of 1 to 3 kilometres, but with most in the range of 1.5 to 2.5 kilometres.

Light rail also comes in two or more forms. It may be provided on its own right of way, or it may operate within the streetscape, as most tram systems do. It is usually at grade, although it may be grade separated at times. Light rail trains are much lower in carrying capacity than heavy rail, and station spacings can be as low as 0.5 kilometres. Systems are usually electrically powered, using an overhead wire and a trolley bar or catenary. Light rail trains may consist of as many as four carriages, but are often two or even one, especially in the lightest of light rail, which is often called a tram or streetcar.

With the exception of the lightest versions of light rail, all rail has in common that it requires right of way, plus the investment of track, wayside controls, and electrical supply in most cases, so that it is least flexible in routeing (once the decision is made to lay a particular route, it will be there for the long term). Rail is also limited in the horizontal and vertical curves that can be negotiated, with gradients over 3 per cent normally not being desirable and large horizontal radius curves required for the long wheelbases of the carriages. In addition, rail best serves a many-to-one transport function, meaning that it operates best in moving people from residential areas to one or two concentrated employment centres, so a traditional city with a strong downtown and dense residential concentrations on radial routes out of the city is best suited to rail. Because of the total carrying capacity of rail, heavy rail and metro require high-density residential development around urban and suburban stations, while light rail and trams or streetcars can serve medium- to high-density areas. Rail is not suited generally to low-density areas, unless there are periodic high-density developments close to stations. High-speed operation is also possible, but only where station spacing is substantial, so that trains can accelerate to a cruising speed, maintain the cruising speed for

some distance, and then decelerate smoothly to the next station. Urban or suburban station spacings are not appropriate for high-speed rail operations.

New technologies, such as magnetic levitation (maglev), are also feasible for train operations, but, as with all high-speed rail services, require long distances for service and lengthy station spacing, so that the speeds can be utilised for significant periods of time and distance. Stations spacing of less than about 50 kilometres is generally inappropriate for high-speed services, whether by conventional high-speed trains or maglev trains.

Monorail, which still seems to have some appeal to politicians as a supposedly more modern rail system, is generally not applicable to urban or suburban services. It is, in fact, almost as old as conventional duo-rail, having been invented originally in the 1890s, but lacks wayside for signals and emergency evacuation of passengers, and is difficult to operate in anything other than an out-and-back, a loop or a figure-of-eight configuration. Switches are extremely difficult to provide for monorails. Sydney, Australia is perhaps one of the only cities to try installing a monorail in recent times, and it has ended up operating almost solely as a tourist attraction and is now scheduled for demolition.

Bus

Buses come in more varieties than rail. Buses can range in size from mini-buses with 8- to 12-passenger capacities to standard buses that usually have a seating capacity of 40–45 passengers, to articulated buses that may have capacities as high as 200 passengers, while double-decker buses can carry somewhere between about 80 and 125 passengers. In addition to this variety of vehicles, there is also a wide range of operating schemes for buses. The most common is the fixed-route bus with stops that are usually spaced between 200 and 500 metres apart, with somewhat longer spacing in less dense suburban areas. These buses may be required to stop at every stop, or, as is common in the UK, some stops may be mandatory while others are at the request of a boarding or alighting passenger. A common variant on the fixed-route operation is limited-stop or express operation. In these operations, buses may stop like fixed-route buses in a suburban area and again in a business district area, but may stop at only a few or no stops in travelling between these two areas. An interesting example of this type of operation is the O-Bahn in Adelaide, which operates as a normal fixed-route bus in downtown Adelaide, and then enters a dedicated guideway with only three stops along it, with buses exiting the guideway and travelling as conventional fixed-route buses from two of the guideway stations (Paradise Interchange

and Tea Tree Plaza). Express and limited-stop services may also, as in the case of the Adelaide O-Bahn, use a dedicated right of way for the line-haul portion of service. This type of operation then leads directly into bus rapid transit, which usually involves higher-capacity buses operating on extensive segments of dedicated rights of way, with station stops that may involve rather more infrastructure in the form of shelters, ticketing machinery, and so on than is normal for conventional services. BRT may sometimes be the only appropriate service to provide in medium-density areas, or may be considered as a precursor to a rail system, to build patronage and encourage denser development in station areas. There are many examples of excellent BRT systems around the world, with the system in Curitiba, Brazil being one of the best-known examples of such systems.

At the other end of the spectrum, buses may be operated as on-demand services or as one-to-one services to ferry particular patrons from one location to another. A common example of such services is that of community buses, which may provide service from residential locations for seniors to shopping or to medical services. Another example of such services are dial-a-ride or pre-booked services, where patrons can contact the operator and request a pick-up at a particular time to travel to a specific location. The operator may pick up a number of patrons within a limited area before travelling to either one or a few destinations to drop off the passengers. An example of such a service is the many airport shuttle buses that operate in cities around the world.

A key characteristic of all bus services, except those that use a dedicated right of way, is flexibility. Unlike rail, these services can be routed differently in a very short period of time and may require little more than training of drivers on new routes and possible decommissioning of some bus stops and installation of new ones. On-demand and community bus services need no such investment, because they are designed for fully flexible routing.

Unlike rail, buses are readily able to serve many-to-many demand patterns and are, therefore, well adapted to providing services in low- to medium-density residential areas and to relatively modest-density employment areas. By careful choice of the type of service and vehicle, bus can also provide service that is competitive with the car.

The role of public transport services

In general, public transport can serve one of two roles. On the one hand, it can provide service to specific markets, sometimes termed niche markets,

where public transport service is clearly the obvious, if not the only, logical way of providing transport for the majority of the market. On the other hand, public transport can be offered as a competitor to private transport (i.e. the car) in more or less any market.

In the former role, public transport service must be looked at more in terms of the overall level of service that needs to be provided, with price, convenience and comfort probably being the major characteristics of concern in the delivery of service. For example, for people commuting into major central cities such as New York, London, Mexico City and even Sydney, car is not the mode of choice for the majority of such commuters, and public transport offers by far the best means of travel, especially given the road congestion, lack of parking, high price of parking, and other deterrents to the use of a car. This is clearly a niche market, where frequency of service, system capacity (which translates directly into passenger comfort) and convenience of the service are probably of greater importance than the actual travel speeds.

In the latter role, where public transport is provided as a competitor with the car and with the desire to reduce levels of car use, it becomes much more important to focus on speed of service. Public transport, because of its need to make stops to pick up and set down passengers, is at an immediate disadvantage to the car in terms of speed, so that line-haul speeds need to be higher than those of cars if public transport is to compete. At the same time, costs must also be considered as extremely important, as is also the convenience of accessing public transport and egress from it. It is also well known among transport planners that people dislike being required to make transfers in a public transport journey, so that a design that minimises the need for transfers is imperative for competition with the car. Much more could be said on this subject, but it is not appropriate for this book. There are excellent treatments of such matters elsewhere. However, in a policy context, it is important to make choices of the type of technology and operating system that keep in mind the limitations and the benefits of each type of public transport system available, and not to make decisions blindly on the basis of the political appeal of a particular public transport mode.

A cautionary example of doing the latter is provided by Miami, Florida. In the late 1970s, a political decision was made to invest in a metro rail line (albeit an above-ground system), to serve a more or less north–south corridor of about 40 kilometres in length, centred on the central business district. The decision was made partly on the basis of newly available federal funding for new rail lines and partly on a political belief that a city without a rail service was not an international city of any repute. The line was built at a

cost of some \$US700 million in 1980s currency, but, after opening, struggled to carry more than about 20 000–30 000 passengers per day. By the 1990s, when it had been predicted that the system would be carrying 170 000 passengers per day, it was carrying about 45 000, and was carrying about 66 000 in 2012. One commentator observed in the mid-1980s that it would have been cheaper to purchase a chauffeur-driven limousine for every passenger on the recently opened line than to have built the metro rail line. The decision to build the line was a purely political one, and was not based on determining what the most suitable public transport service might be for the city. The majority of the patronage was drawn from existing bus services, with little impact on the levels of car use into the downtown area of Miami.

13.3 Providing public transport

Internationally, the means of delivering PT services is highly variable. In some jurisdictions, services are delivered by government, often through a government business enterprise (GBE). This model of operation is intended to introduce private sector practices into public sector operation, to improve service efficiencies. In others, services may be primarily left to the private sector, where the profit motive will mean a focus by the provider on high-revenue-yielding routes. Little attention may be devoted to routes where social need is a stronger driver for service. Alternatively, government may rely on the private sector to initiate and deliver most services but provide funding support to (for example) social safety net services, which may also be delivered by the private sector under a contract to a government agency (largely the UK model outside London). For rail, where capital costs of track and rolling stock are high, government may do the lot itself, may set up a government entity to have responsibility for the track and encourage above-ground competition for service provision, or may contract out both the track and service sides to the private sector. These are just some of the possibilities. Gwilliam (2008) notes the possible long-term existence of a regulatory cycle for route bus services in developing and transitional economies, under which PT delivery models change over time through a cycle of public or municipal monopoly to fragmented informal supply, an informal sector cartel and then private regulated monopoly, with a subsequent return to public or municipal monopoly, with different jurisdictions starting the cycle at different points.

The Thredbo International Conference Series on Competition and Regulation in Public Transport is a vast source of case study material on such possibilities and experiences with their implementation. At present, there is a general tendency for increased liberalisation of public transport service provision, particularly route bus services, but with growing focus on passenger

rail. Stanley (2011) suggests that this has sometimes over-promised and under-delivered.

Some general conclusions from the Thredbo Conference Series are relevant to the question of how public transport services might be best provided, particularly where private provision is being pursued (Hensher and Stanley 2010):

- Competitively tendering a government monopoly urban route bus service to the private sector usually leads to cost savings of 10 per cent and sometimes substantially more (depending on the efficiency of the prior government monopoly).
- These savings tend to be one-off, with second- and third-round tenders not delivering significant cost savings.
- The evidence for cost savings from the contracting out of passenger rail services is more limited and more mixed (Stanley and Smith 2013).
- Where bus services have traditionally been provided by the private sector, contracting out by competitive tender usually delivers only small, one-off savings.
- Designing and managing public transport contracts to deliver efficient and effective services are difficult because of the combination of service cost and multiple service quality variables that enter the service bidding equation. Specifying multidimensional quality is difficult, which tends to see heavy reliance placed on lowest cost in the award of tenders, which may not deliver best value for money.
- Linked to the preceding point, PT service contracts are inevitably incomplete (bounded rationality), and it is better for the authority that is managing services and the service provider to have a strong working relationship during the course of a contract to deal with this, rather than to seek to specify every last point in complete detail, which can be a distraction from performance.
- Negotiating the allocation of the rights to provide a bus service to an existing efficient private bus operator, through a negotiated performance-based contract (NPBC), is likely to be at least as efficient as putting these services to open tender, particularly because of the inevitable uncertainty in contracting for service quality. An NPBC can use an adaptive awarding mechanism that sustains operator performance pressure *during the course of the contract*.
- In NPBC regimes, competitive pressure can be sustained by performance benchmarking, an open-book approach to costs, use of a probity auditor, public disclosure of the contract, and contract termination in the event of poor operator performance.

The Thredbo Conferences have developed the idea of a *trusting partnership* between government PT purchaser and private sector provider, as a way of maximising effective PT service planning and delivery, grounded in (Hensher and Stanley 2010):

1. **common** core objectives tied to public policy purposes;
2. **consistency** of behaviour and direction;
3. **confidence** in a partner's capacity to deliver;
4. respect for each other's **competencies**; and
5. demonstrated **commitment** to good faith in making and keeping arrangements and in principled behaviour.

Hensher and Stanley (2010, p. 143) argue that these five Cs support contract clarity (before signing the contract) and clarity of obligations once the contract is signed.

A useful recent test of some of the Thredbo ideas has been provided by bus service contracting in Adelaide in recent years. Wallis et al. (2009) reviewed Adelaide's experience with three rounds of tendering bus services, concluding that there was little to gain in terms of cost efficiency and quality enhancement by going to a fourth round of tendering. They argued for a move to negotiated performance-based contracts, to lower transaction costs (associated with tendering), but also to offer the opportunity for the authority to work more closely with efficient incumbents to grow trust and build patronage. The authority decided instead to go to open tender, with the subsequent result that tender costs were higher and service quality lower than expected.

Hensher (2013) has subsequently benchmarked a number of Australian private route bus operations that were tendered against others that were awarded by negotiation with an existing operator (NPBCs). Australia is probably the only country in which this comparative assessment could be undertaken, because of its leading role in implementing NPBCs (especially in Melbourne and Sydney) and its long involvement in competitive tendering (e.g. Adelaide, Perth, and parts of Sydney and Melbourne). He found that NPBCs deliver slightly better cost-efficiency outcomes than competitive tendering, when the incumbents are private operators. This strongly challenges the economic textbook orthodoxy of putting contracts to the market for bidding to maximise cost efficiency.

If public transport is provided by a private operator under contract to a government authority, then the service contract is an opportunity to set down some of the service requirements that are expected, taking account of the

difficulties noted above about not seeking to go too far in this regard. If services are the prerogative of the private sector alone (competition *in* the market), then more general legislative and regulatory provisions, with related compliance mechanisms, are the main way of assuring performance. These legislative and regulatory provisions are essentially safety, environmental and asset protection requirements (e.g. working hours regulations, vehicle emissions control requirements, driver licensing requirements, driver blood and alcohol limits, vehicle mass limits, and vehicle swept paths). They may also include an operator accreditation regime, intended to help ensure that the operator is able to provide a service of a certain standard. Such requirements, in a regime of competition *in* the market, will usually be no different to those that must be met by any contracted service providers, if there are such services. This makes competition *in* the market a lighter-touch operating environment, recognising the market risks taken by service providers. The lack of a contractual opportunity to assure operator performance places emphasis on the importance of an effective legislative and regulatory environment to achieve environmental, social (including safety), and asset protection goals, together with basic service standards.

There is no one answer to the question of what is the ‘best’ way to provide PT services. What is ‘best’ in any situation will depend on such factors as the history of service provision in the jurisdiction, the dominant ideology of the electorate (especially if it is pro-government provision or pro-private sector), the quality of the government authority that is providing or purchasing the PT service, the size and complexity of the PT market(s) and the related capacity to attract and retain interest from a range of service providers. There is a large and growing body of accumulated experience to assist policy makers and land transport planners through this maze, including the Thredbo Conferences and agencies such as the International Association of Public Transport (UITP), Canadian Urban Transit Association, American Public Transportation Association, and Australia’s Bus Industry Confederation.

13.4 The future role that public transport might play

It is now very common for governments to be seeking to increase the passenger transport task that is performed by public transport. In Europe, for example, policy to mitigate climate change pressures is seeing a greater role expected for rail and water for medium-distance passenger (and freight) movement, with a target of a 50 per cent shift outlined in the EU White Paper (EC 2011). The Victorian government in Australia set a target back in 2000 for 20 per cent of Melbourne’s motorised trips to be made by PT by

2020. At the time, the mode share was about 8–9 per cent but has currently reached 13–14 per cent and is on track to meet the ambitious target, showing the benefit of strong system planning and effective delivery capabilities.

Decisions about the future role for public transport in any particular setting should flow from a careful examination of desired land use futures and the economic, social and environmental costs and benefits associated with change. Chapter 16 explores some of the wider land use considerations. The current section seeks to put some context around the idea of scale of change.

Public transport mode shares typically do not change quickly, because of the influence of long-term land use decisions on user PT choices. However, the Melbourne and Brisbane experiences cited above suggest that it is possible to *grow* the PT mode share by a significant relative proportion and quite quickly, if there is investment in service improvements. These investments need to target the ‘demand drivers’ such as service frequency, operating hours, reliability, connectivity, information provision, fare levels, safety at stops, ease of access to and from services, and the like.

The importance of particular demand drivers will vary from place to place. For example, if service frequencies are high, then reliability may be the most important improvement priority. Low frequencies are likely to mean that improving frequency is a high priority. For example, Melbourne bus frequencies were low a decade ago. Service improvements primarily aimed at improving frequencies were introduced, and the resulting patronage growth rate over a three-year period to 2008–09 was very high (see Figure 13.1). The implied service elasticity was marginally above 1 (a service elasticity being defined as the relative increase in patronage divided by the relative increase in service kilometres), which is a strong result in international terms. This has been continued as further service increases have been provided.

More broadly, achieving PT patronage increases requires *packages* of complementary policy initiatives tailored to local circumstances, because any single demand driver is likely to have only a small impact. This emphasises the importance of detailed local market knowledge and a deep understanding of how to increase local patronage cost effectively, if that is a policy or operational goal. Complementary policy packages may include measures to increase the relative generalised cost of car travel, through pricing initiatives of the kind discussed in Chapter 12, and (for example) measures to improve walkability in the urban realm, which support public transport use (Ewing and Cervero 2010).

It is vital to understand that relatively small changes in travel mode shares in favour of PT can deliver substantial public benefits. This is particularly the case in the peak period, when marginal road congestion costs are high (see, for example, Figure 5.9 in Chapter 5). One implication of the shape of the speed–flow curve in congested road conditions is that small reductions in congestion levels can generate large savings (benefits). For example, Levinson (1995) calculated that marginal US peak period congestion costs for an urban freeway average 6–9 cents per vehicle mile when traffic travels faster than 50 mph, and up to 37 cents per vehicle mile when traffic flows at less than 40 mph, which would be much higher if updated to current prices. Figure 5.9 in Chapter 5 showed even higher peak current congestion costs in an Australian context. UK research has suggested that urban congestion costs (in the UK) can be cut by over 40 per cent if congestion pricing reduces urban traffic volumes by about 4 per cent (DfT 2004, Table B3). School holiday traffic levels in Australia typically involve slightly larger traffic reductions, illustrating the significant congestion gains to be achieved from small reductions in volumes. Thus, if public transport improvements over a few years can sustain a reduction in the car mode share by 4 to 5 percentage points, very substantial congestion benefits result. If the benefits of such a reduction in traffic volumes and associated congestion costs are to be sustained, measures are needed to limit any subsequent traffic generation caused by lower congestion costs (the ‘bounce-back’ effect). Pricing solutions and capacity reductions are a way to achieve this outcome.

It is also important to note that increases in the PT mode share usually mean important *complementary benefits* in terms of a lower accident rate, reduced greenhouse gas emissions, improved air quality, better health, improved energy security and greater social inclusion. A higher PT mode share may also be an indicator of agglomeration benefits. Realising such benefits involves costs and forgone investment opportunities, underlining the importance of careful analytical approaches to prioritisation, along the lines discussed in Chapter 4.

13.5 Conclusions on public transport

In land transport policy terms, public transport is at a turning point. There is widespread recognition that it has the potential to play a larger role in moving people in cities and regions. However, the realisation of this potential requires detailed understanding of the drivers of PT demand, service costs, the most effective ways to deliver service and the potential benefits from growing market share. Increasing the PT mode share in the medium term by 4 to 5 percentage points can deliver major benefits. Longer-term,

much greater increases in market share are likely to be needed in many jurisdictions, for example to mitigate increasing congestion costs and help meet greenhouse gas reduction targets. For the achievement of such outcomes to be more than aspirational, P/T service standards will increasingly need to mirror those offered by the private car and be delivered in a highly efficient manner.



REFERENCES

- Asquith, B.J. (2011), 'Income elasticity of demand for large, modern rapid transit rail networks', *Undergraduate Economic Review*, 7 (1), Article 20, available at: <http://digitalcommons.iwu.edu/uer/vol7/iss1/20>.
- Australian Government (2012), *State of Australian Cities 2012*, Canberra: Department of Infrastructure and Transport.
- BTS (2012), *NSW and Sydney Transport Facts*, Sydney: New South Wales Government, Bureau of Transport Statistics.
- DfT (2004), *Feasibility Study of Road Pricing in the UK: Report*, report to the Secretary of State, Appendix B: 'Modelling results and analysis', London: Department for Transport.
- EC (2010), *EU Energy and Transport – in Figures: Statistical Pocketbook*, Luxembourg: Publications Office of the European Union.
- EC (2011), *Roadmap to a Single European Transport Area: Towards a Competitive and Resource Efficient Transport System*, COM/2011/0144 final, Brussels: European Commission.
- Ewing, R. and R. Cervero (2010), 'Travel and the built environment: a meta-analysis', *Journal of the American Planning Association*, 76 (3), 265–94.
- Frank, R. (2008), *Microeconomics and Behavior*, 7th edn, New York: McGraw-Hill.
- Gwilliam, K. (2008), 'Bus transport: is there a regulatory cycle?', *Transportation Research Part A*, 42 (9), 1183–94.
- Hensher, D. (2013), 'Cost efficiency under negotiated performance-based contracts and benchmarking: are there gains through competitive tendering in the absence of an incumbent public monopolist? The Australian experience and evidence', keynote address delivered to the South African Bus Operators Association Annual Conference, Pretoria, 28 February.
- Hensher, D. and J. Stanley (2010), 'Contracting regimes for bus services: what have we learnt after 20 years?', *Research in Transportation Economics*, 29 (1), 140–44.
- Holmgren, J. (2007), 'Meta-analysis of public transport demand', *Transportation Research Part A*, 41 (10), 1021–35.
- Levinson, H. (1995), 'Freeway congestion pricing: another look', *Transportation Research Record*, 1450, 8–12.
- Mohring, H. (1972), 'Optimization and scale economies in urban bus transportation', *American Economic Review*, 62 (4), 591–604.
- Stanley, J. (2008), 'Costing transport modes for externalities', in Metropolitan Transport Forum, *The Economics of Transport: Smarter Transport – Better Cities*, Melbourne: Metropolitan Transport Forum.
- Stanley, J. (2011), 'Public transport liberalisation: achievements and future directions', in M. Finger and R.W. Kunneke (eds), *International Handbook of Network Industries*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Stanley, J. and A. Smith (2013), 'Workshop 3A: governance, contracting, ownership and competition issues in public transport: looking up not down', *Research in Transportation Economics*, 39 (1), 167–74.

- Vuchic, V. (2005), *Urban Transit: Operations, Planning and Economics*, Hoboken, NJ: John Wiley & Sons.
- Wallis, I., D. Bray and H. Webster (2009), 'To competitively tender or to negotiate: weighing up the choices in a mature market', *Research in Transportation Economics*, **29** (1), 89–98.

14

Potential solutions – TSM, TDM and VTBC



CHAPTER OVERVIEW

Continuing the exploration of potential solutions to transport issues, this chapter deals with the low-capital strategies of transport system management (TSM), travel demand management (TDM) and voluntary travel behaviour change (VTBC). The first two of these – TSM and TDM – are largely coercive in nature, taking the form of actual supply changes, reconfigurations of the transport facilities, and the introduction of various rules, regulations and top-down strategies. The various available techniques for TSM and TDM are discussed. The third potential solution – VTBC – is a non-coercive policy direction that seeks to motivate people to change their behaviour voluntarily either through appeals to altruism or by helping people to solve their own problems. Techniques to do this are discussed in this chapter.

14.1 Introduction

Transport system management, travel demand management and voluntary travel behaviour change are three potential low-capital strategies for tackling the problems of transport, especially those related to air pollution, noise, congestion, and a number of other related problem areas, such as improving flow of traffic, increasing the use of public transport and active transport modes, and so on. The first two of these strategic directions emerged initially in the 1970s, when it was realised in a number of countries that the funds for continuing to expend on large capital-intensive projects were no longer available. These two strategies were put forward – TSM first, with TDM then being added – as low-capital options to increase capacity through better management of the supply and demand for transport. VTBC emerged more recently, mainly in the 1990s, as a new set of strategies for increasing capacity through non-coercive behaviour changes (in contrast to TSM and TDM, which may often be regarded as somewhat or completely coercive).

The various tools discussed in this chapter are low capital cost and generally fast implementation strategies. They may be introduced on their own, especially when there are limited funds available for implementing solutions to traffic problems, or they may be introduced as part of a policy suite that may also include some high-capital and/or slow implementation policies and strategies. As is discussed later in the chapter, it is often important that both TSM and TDM strategies are implemented together, so that the full potential benefits of both strategies can be achieved. It is also important to note that the analytical tools provided in Chapters 4 and 5 will need to be used to develop policy packages that will best meet the social goals, as well as the decision makers' views on acceptable trade-offs.

In this chapter, the options that are included within these strategies are described, together with the potential effect that these strategies may have on the performance of the transport system. Specific issues that the strategies may contribute to solving are also described.

14.2 Transport system management

As its name implies, this group of strategies cover those that can be implemented to manage the supply side of the transport equation, generally by providing capacity increases or better use of existing capacity (e.g. by changing the demand pattern over time, to utilise capacity that exists at other times of the day) without the need for major capital investment.

Intersections

On a freeway or motorway, capacity is mainly determined by the geometrics of the lanes, the mix of traffic, and the driver population, as is discussed in Chapter 5. However, the capacity of streets and roadways with at-grade intersections is largely determined by the characteristics of the intersection. For example, if one has two roadways that intersect, with roughly equal volumes approaching the intersection on each of the four legs, then the capacity of the roadways leading to the intersection is determined entirely by the maximum flow that can be achieved through the intersection, which would, with roughly equal approach volumes, be about 50 per cent of the capacity in the middle of each approach roadway. This is because vehicles can safely move through the intersection from each approach for a maximum of 50 per cent of the time, thereby reducing the effective capacity to 50 per cent of the unimpeded flow on the approach roads. If a lane has a nominal capacity of 2000 vehicles per hour (TRB 2010), then the intersection will reduce this to no more than 1000 vehicles per lane per hour. This capacity is further reduced if there is a

substantial volume of turning traffic at the intersection. In the following subsections, the various means to increase intersection capacity are examined, apart from that of widening the approach roads to the intersection.

Dedicated turn lanes

In a country where people drive on the left, right turns will have the biggest impact on capacity, while in countries that drive on the right it will be left turns that have the biggest impact. The reason for these impacts is that such vehicles must wait at the approach to the intersection until there is an acceptable gap in the opposing traffic stream before being able to complete the turn in safety. If turning vehicles are mixed into the same lane or lanes as through traffic, then these turning vehicles will reduce even further the total possible flow through the intersection.

Left turns (right turns in countries that drive on the right), will have a lesser effect, because there is not a need to wait for opposing vehicles. There may, however, be a need to wait for pedestrians to cross, and the vehicles turning left will have to slow down much more than through traffic to negotiate the turn.

It follows, then, that a potentially successful strategy for intersections that have significant turning movements would be to reconfigure the approaches to the intersection to allow construction of dedicated turn lanes. In countries driving on the left, this would be most important for significant volumes of right-turning traffic, and vice versa for countries that drive on the right. Such dedicated turn lanes can often be constructed by narrowing the lanes as they approach the intersection, and a modest widening of the total right of way in the approaches to the intersection, as shown in Figure 14.1. Hence, this is a low-capital strategy. When volumes of pedestrians crossing are large enough, or when left-turning traffic is of a high enough volume, a dedicated left-turn lane may also be worthwhile. Dedicated turn lanes can work where there is minimal intersection control (e.g. stop and yield signs), but work much more effectively where there are traffic signals or roundabouts.

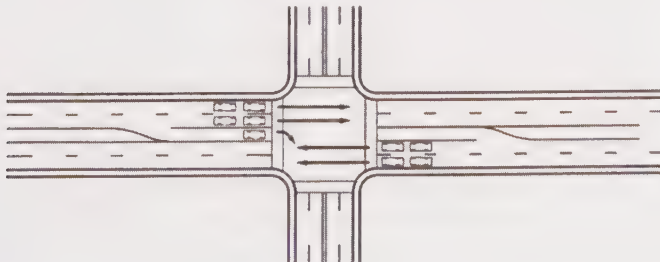


Figure 14.1 Restriping to create a right-turn lane

Roundabouts and traffic lights

Another TSM strategy that relates to the capacity of urban streets and roads has to do with the control of intersections. Uncontrolled intersections probably have the lowest capacity, because drivers must slow down at the intersection to determine if it is safe to proceed. Stop and yield signs represent the next level of control. If two movements through an intersection are much more major, then the stop or yield signs will be on the minor approaches, thus increasing the capacity of the major road, but decreasing the capacity of the minor road. However, the minor road capacity is probably not an issue. Therefore, this will result in an increase in capacity for the major road. In some countries, when the volumes on the minor road become higher, a four-way stop may be introduced. This will usually decrease capacity, although it is necessary from a safety viewpoint. In other countries, a roundabout may be introduced. This will normally have benefits in both increased safety and capacity increase for the previous minor road. It may decrease the capacity of the major road.

When volumes become high enough for four-way (or all-way) stop intersections or roundabouts, the stop signs or the roundabout is replaced by traffic lights. Traffic lights offer various levels of sophistication and capacity benefits. The simplest traffic lights are fixed-time signals that are preset to provide certain lengths of green and red time to each approach at the intersection. Fixed-time signals can also include right- or left-turn phases when turning volumes warrant it. The next level of sophistication is to have fixed-time signals whose fixed times may vary by different periods of the day. These signals may be used where the various flows at the intersection vary with different time periods of the day. Fixed-time signals can also be linked together in a sequence that produces a 'green wave', such that a vehicle that passes through the first green traffic light will, if it proceeds at the design speed for the green wave, meet a series of green lights in travelling along the roadway. This is done by calculating the length of time required at the design speed to travel from one traffic light to the next and offsetting the start of the green phase on the second light by the amount of the travel time from the first to the second light. This is repeated for successive lights along the road. Any number of traffic lights can be linked in this way. Other linked signal phases can also be implemented with fixed-time signals.

The next level of sophistication is to introduce traffic-actuated signals. These are signals that are controlled through a controller box and loop detectors in the roads at the approaches to the intersection. These signals can vary

the green times within certain limits on a cycle-by-cycle basis, according to the approach volumes of traffic. When a minor road intersects a major road, and the minor road normally has a very low volume, but a traffic signal is required, traffic actuation may be used so that the traffic light for the minor road turns green only when there is a minimum of one vehicle waiting, or when the queue exceeds a certain length or a vehicle has been waiting a certain amount of time. Because traffic actuation responds to the volumes of traffic on the approaches, and provided that overall volumes are not too high, the capacity of the intersection can be increased significantly by using such traffic signals. The cost of adding a controller for traffic actuation is not high, and the controller simply needs to be set for different periods of the day to effect this.

The final significant step in traffic signal design is to move to area-wide computer control of the traffic signals. In this case, information on the approach flows to all intersections in an area is fed into a computer and the computer uses various algorithms to determine optimal signal timings in the network of signals that cover the area. A good example of such a system is the Sydney Coordinated Adaptive Traffic System (SCATS) (RMS, n.d.).¹ In such systems, there are usually mid-block detectors as well as the stop-line detectors that provide the computer system with more detailed information about the volumes of traffic flowing towards the intersections. As a result, traffic signal phases can be adjusted to the flows, and delays can be minimised. Computer control of traffic signals probably represents the best that modern technology can do to improve the capacity of intersections within urban road systems. However, even computerised traffic control cannot remove congestion or delays if volumes in the system are too high.

Traffic lanes

A number of TSM strategies can also be used to increase the capacity of roadway segments between intersections. While these may not always be the controlling issue for capacity, if improvements are made to the intersections then additional capacity may need to be added between the intersections.

Removing parking

The simplest strategy that can be applied is to remove kerbside parking. Removing parked vehicles may provide sufficient road space to add an additional lane of moving traffic in each direction on a roadway. Removing parking can be done on a periodic basis (e.g. in the peak periods in the peak

direction only) or may be permanent by time and location. The former case is often referred to as the operation of 'clearways', where parking may be prohibited on one side of the street between certain hours only.

As well as the fact that parked vehicles remove road space that may be sufficient to add another lane to the roadway, the presence of parked vehicles also decreases capacity of the existing lane nearest to the parked vehicles. As noted in Chapter 5, the presence of lateral obstructions has the effect of depressing speeds and volumes in a traffic lane. Parked vehicles represent such an obstruction. Therefore, removal of parked vehicles will increase the effective capacity of the lane that was closest to the parked vehicles, as well as offering the opportunity to add an additional lane of traffic.

Removal of parking may meet with local opposition, however. If the parking removal occurs on a shopping street, for example, the tenants of the shops and stores may oppose this on the grounds that it reduces potential patronage of their shops and stores. In a residential area, especially where residences do not have off-street parking available, residents may raise objections if the parking ban operates all day and night.

Tidal flow lanes

Another strategy that can be used works well when there is no median in the roadway and where peak flows are substantially out of balance in the two directions. In this case, one or more lanes may have their flow direction reversed between the morning peak and the evening peak. This must, of course, be done on those lanes that are nearest to the centre of the roadway. Removable barriers are normally used to indicate the boundary between the two directions of flow. These removable barriers may be in the form of manually placed road cones or similar markers, automatic pop-up posts that delineate the division between the directions, and even movable concrete barriers between the two directions of flow. Tidal flow lanes are useful only in cities that have strong flows in one direction in the morning peak, followed by strong flows in the opposite direction in the evening peak. Tidal flow lanes, for example, are used in numerous places in Sydney, Australia, where flows into the city in the morning are much higher than outward flows, and the reverse occurs in the evening. The Sydney Harbour Bridge is an excellent example of tidal flow lanes, where the directions of flow are shown by a red X or green downward arrow on gantries across the roadway. (The red X indicates that vehicles seeing that sign may not travel in that direction in a specific lane, while the green downward-pointing arrow indicates that travel is permitted in

Figure 14.2 Lane-marking gantries on the Sydney Harbour Bridge



that direction in that lane.) Figure 14.2 shows an example of the gantries that indicate the lane directions of use. There are no removable barriers between the flow directions in this case.

In some instances, especially in North America, a road may have an odd number of lanes, with the centre lane being designated as a turning lane outside the peaks. However, this turning lane may be utilised for traffic flow in the peak direction during peak periods.

Lane widths

Another TSM action that can increase the number of lanes on a roadway is to restripe the lanes to a narrower width. For example, a road may have been designed and built originally with four 3.6-metre-wide lanes. These lanes could be narrowed to 2.9 metres, thereby adding 2.8 additional metres that could permit one additional lane to be created without any road widening. The amount by which lanes can be narrowed will depend on the legal allowable vehicle widths, however. Such an additional lane could be allocated to one particular direction of traffic or used as a tidal flow lane, as described in the previous subsection. The extra lane could also be used as a bus lane. Similarly, for roads designed with shoulders both on the outside edge and beside the central median, the shoulders may be incorporated into the travelled way and, with restriping of the existing lanes, provide sufficient pavement width to allow creation of an additional lane.

A closely related strategy is to permit a wide offside shoulder to be used as a travelling lane only in peak periods or in the peak direction in peak

periods. In this case, there is no restriping of lanes, but signs are used to indicate that vehicles can use the shoulder as a travel lane at certain times. The disadvantage of this strategy is that there is nowhere for a disabled vehicle to be moved out of the moving traffic lanes. This strategy has some negative safety implications, because a disabled vehicle in a moving lane represents a hazard to traffic in that lane and in the lane beside it. Clearly, the decision to use this strategy must be based on a trade-off between the frequency with which disabled vehicles use the shoulders and the extent of the traffic congestion that can be relieved by using the shoulder.

High-occupancy-vehicle lanes

A further TSM strategy that is aimed at increasing capacity through increased vehicle occupancy is to designate specific lanes as high-occupancy-vehicle (HOV) lanes, carpool lanes, or T2 or T3 lanes. These are lanes that are reserved for the use of vehicles with two or more occupants, taxis and buses. This TSM strategy can be effective at moving more passengers per hour through the system, if people respond to these lanes by abandoning the use of single-occupant vehicles and sharing rides. Usually such lanes, if created by diverting an existing lane of traffic from mixed use to high-occupancy use, will actually decrease vehicle capacity, but they can increase person movement capacity, which is increasingly the policy focus in most jurisdictions. This strategy is discussed further in conjunction with TDM strategies that are aimed at increasing vehicle occupancy. If HOV lanes are not adequately policed, however, then violations of the high occupancy will occur and the value of the lane will be greatly reduced.

A special case of HOV lanes can be considered where freeways have ramp metering. Where ramp metering is used to regulate the flow of vehicles into a heavily used freeway, carpool vehicles may be assigned a bypass lane that allows these vehicles to merge directly into the freeway traffic without having to join a queue and wait at the ramp meter signal. Currently, this type of bypass lane is seen in a substantial number of cases in Los Angeles, among other places. Also, in Los Angeles and some other cities where carpool lanes have been designated on freeways, at some freeway-to-freeway interchanges there may be separate ramps for carpools that connect directly from the carpool lane on one freeway to the carpool lane on the other freeway. An illustration of this is shown in Figure 14.3, which shows the interchange between the I-110 and I-105 freeways in Los Angeles, with separate HOV ramps.



Figure 14.3 I-110 to I-105 interchange in Los Angeles, showing dedicated HOV ramps

Bicycle lanes and paths

Another strategy that affects traffic lanes is to provide a special lane for bicycles, as shown in Figure 14.4. This may also be associated with separated bicycle paths. Because bicycles are often seen as being vulnerable in mixed traffic, especially to large vehicles, to the opening doors of stationary vehicles, and at intersections, provision of bicycle lanes and bicycle paths may offer a safer alternative for cyclists from that of using a mixed traffic roadway. In some countries, such lanes and paths may also have their



Figure 14.4
A kerbside bicycle lane

own separate traffic light phase at major intersections, or a joint phase with pedestrians crossing in the same movement as the bicycles. However, this is normally only warranted where the volumes of bicycle use are particularly high.

Two observations are important in relation to bicycle lanes and bicycling in general. The first relates to the cyclists themselves. It is important that cyclists are educated in how traffic rules apply to them and that the riders obey these traffic rules. All too frequently, cyclists in countries such as Australia and the USA seem to feel that they do not have to obey street signs and signals and that they can use the road and the footpath as they see fit. Cyclists also often do not observe the requirement to ride in the same direction as motorised vehicle traffic. This disregard for the rules of the road leads to increased accident exposure and a perception that bicycling is not safe.

Second, it is important that bicycle lanes are an actual reservation of road space for bicycles only. Placing these lanes alongside kerb parking is also not useful, because one of the major hazards for cyclists is the opening door of a parked vehicle. This can be avoided by running the bicycle lane on the inside of parked cars, although this moderates but does not eliminate the problem (because car passengers may now open doors into the lane). Simply painting a bicycle symbol on the road, in the outside edge of the vehicular lane, does not represent a bicycle lane and is not a contributor to safety for the cyclist. It is probably safest and ideal if kerbside parking can be removed and the bicycle lane placed alongside the kerb, still leaving a full lane width for car and truck traffic. A barrier that separates the bicycle lane from the moving vehicle lanes is also ideal from a safety perspective and as encouragement to cyclists to use the special lanes. Special markings should also be placed at intersections, indicating the turn and through lanes for the bicycle. Of particular importance in this context are the marking of lanes and the provision of safety for bicycles that need to turn across the motorised vehicle traffic lanes (e.g. for right turns in a country where people ride or drive on the left). Care is also needed in the design of such lanes along roads that have frequent driveway openings from commercial and other non-residential land uses, where there may be frequent vehicle crossings.

Bus priority

The final TSM strategy discussed in this chapter is that of bus priority. Again, this is a strategy that is aimed at reducing single-occupant vehicle use and increasing the passenger flow on a roadway, but it may reduce the total vehicle flow. Bus priority can be achieved with several strategies. First, as

with the high-occupancy-vehicle lanes described earlier, some lanes can be designated as bus lanes only. When bus volumes are high enough, bus lanes may represent an effective way of permitting buses to compete on journey times with cars, because the buses are usually able to operate at higher speeds than the remaining, and now more congested, mixed traffic lanes.

Buses travelling along bus lanes may also be given separate bus phases on traffic signals. This is usually done in response to as few as one bus being detected in the bus lane approach to the traffic signals. In other cases, bus pre-emption of traffic signals may be implemented, whereby a transponder on the bus interacts with the controller on the traffic signal, causing the green phase to be ended for the traffic currently with a green phase, and then providing a green light specifically for the bus. Signal pre-emption for buses does not work very well in computer-controlled area-wide traffic signal systems, where the separate bus phases will work better. Signal pre-emption is applied best in systems where traffic-actuated and fixed-time signals are in use.

Summary of TSM actions

The preceding subsections provide a number of examples of the types of actions and strategies that can be applied to obtain more capacity out of the existing system. As can be deduced from these examples, the capital costs are generally much lower than the capital costs of adding more road space, and also are potentially more sustainable than the addition of road space. Adding road space in an Australian city, for example, may cost in the order of tens if not hundreds of millions of dollars per kilometre, whereas most TSM actions cost in the hundreds of thousands to a few million dollars. Some TSM strategies and actions may have quite marked effects on system performance, while others may have minimal effect. However, on balance, TSM strategies are usually highly cost-effective.

To produce the greatest benefits, TSM strategies should be only a part of the overall strategy, and should be coupled with complementary TDM strategies. If there are no demand-side strategies, then this will probably limit the usefulness of the supply-side strategies of TSM.

14.3 Travel demand management

Travel demand management is the other side of the equation from TSM. Where TSM strives to find ways to get more use out of the existing supply of road space, TDM looks at possible ways to manage the demand for road space, particularly with regard to reducing the height of the peaks in traffic

flows, and extending the usefulness of the road system through time, by diverting some peak traffic into less heavily used time periods. As with TSM, most TDM strategies are also low-capital strategies. However, there are two cautionary remarks that should be made about the majority of TDM strategies. First, if no complementary TSM strategies are implemented alongside the TDM strategies, the latter will usually fail to bring about significant benefits. Second, TDM is all about changing human behaviours, which is generally much harder to do than the type of engineering changes described as TSM strategies.

TDM involves managing the incidence of demand in terms of:

- the amount of travel;
- the timing of travel; and
- the means by which people travel.

Most TDM strategies are 'carrot-and-stick' strategies, or coercive strategies. In other words, TDM usually attempts to persuade or compel people to change their behaviour, and usually offers some sort of reward to those who are persuaded.

Strategies affecting working hours

A number of TDM strategies affect working hours, by attempting to remove the requirement for people to travel to and from work at the same time as one another. Some of these strategies can be applied only to certain types of jobs. Many of these strategies are favoured by workers, but disliked by employers, for a variety of reasons.

Flexible work hours

The first of these working hour strategies is that of flexible work hours. Under this strategy, employers define a core time period in the day, during which employees must be present and also during which important meetings that all must attend would normally be scheduled. The core period might, for example, be defined as from 10 a.m. until 3 p.m. Outside this core time, employees are free to choose their time of arrival at work and departure from work, provided only that they work the prescribed number of hours each day or other agreed time period. So, for example, one employee might still work from 9 until 5 each day. Another employee might work from 7 until 3 each day, while another one would work from 10 until 6. The effect of this strategy is to spread the peak period, because not all employees have to arrive at work

around 9 and leave around 5. The shorter the core period, the greater flexibility is possible, while conversely the longer the core period the less is the flexibility offered by this strategy. For such a strategy to have a noticeable effect on the system performance, either a very large employer or many employers in the same city need to adopt the strategy. Otherwise, effects are not likely to be measurable.

Flexible work hours will usually work only in certain types of jobs, where the full complement of staff is not required to be present through the entire working day. It is not possible for shift workers, or in such businesses as retail sales, in general. However, it may be feasible in government offices and a wide variety of private business offices.

A policy difficulty with flexible work hours is that the individual employer is not able to capture the wider external benefits (e.g. reduced congestion costs, cleaner air) that flow from the effect of flexible hours on reducing peak traffic flows. This will mean that the individual firm-level incentives to adopt flexible hours are less than the social incentives and there is underinvestment in flexible hours. This is an argument supporting government leadership in implementation, since government can champion wider community benefits by example. Alternatively, governments may seek to provide a benefit of some type to employers that pursue flexible hours, to encourage wider adoption.

Telecommuting

Telecommuting usually means employees being permitted on one or more days per week or per month to work from home. It is probably best suited to the same types of jobs as flexible work hours. Telecommuting usually requires at a minimum that the employee is contactable throughout the working day by telephone. It may also require that the employee has email and internet access from home, so that work can be done at home that would normally be done at the workplace, and that contact can be made through a variety of mechanisms throughout the working day. Mobile communications devices are a great support in this regard, reducing the need for a home base.

Telecommuting may present difficulties to employees who have young children at home, who may be distracted or expected by their spouse or partner to share in childminding chores during the telecommuting workday. Apart from this, employees generally are favourably inclined to telecommuting, while employers are frequently unfavourably inclined. Ensuring that productivity does not suffer and ensuring that employees take seriously the

responsibility to accomplish work from home are usually the two major, and related, hurdles that must be surmounted. However, each day that an employee is able to telecommute reduces that person's contribution to the peaks by 20 per cent per week. If all employees in a city were able to telecommute for one day per week, then the peak would be reduced by 20 per cent overall. If this could be achieved in the majority of employment (which is not possible), then congestion would all but disappear.

Compressed work weeks

There are two models of compressed work week: four days per week or nine days per fortnight. In each case, employees work fewer days, but work longer hours on the days on which they work. Thus, if an employee is supposed to work, say, a 40-hour working week and is on a four-day work week, then, instead of working eight hours each day, the employee would work ten hours on each of four days. Similarly, for the same employee to work a nine-day fortnight, instead of working eight-hour days this would involve working just under nine hours per day. This strategy has two effects on the demand for travel. First, it reduces the overall amount of travel by 20 per cent per week for the four-day work week, or 10 per cent for the nine-day work fortnight. Second, it also spreads the peak, because the longer working days for these employees now necessitate that they come earlier to work and/or leave later. If the day off in a week or fortnight is rotated throughout the week, then the employer can maintain the same number of staff on each day, but also retain the same number of total working hours as with a fixed five-day work schedule.

Again, this strategy is best suited to certain types of occupations. Also, employees generally like compressed work weeks or fortnights if they can take a Monday or a Friday as their non-working weekday, but are less excited by the strategy if required to take a midweek day off. Some employers like compressed work weeks, while others find them difficult to manage. As with flexible work hours and telecommuting, the effects will generally only be evident when a large employer or a large number of smaller employers implement the strategy.

Working time strategies

In summary, employees in certain types of occupations can be offered one or more of three different strategies with respect to working hours, two of which (telecommuting and compressed work weeks) have the potential to reduce the number of commute trips made by each employee, and all three of which

have the potential to spread the peak periods, thereby flattening the highest part of the peak and moving some of the peak traffic flows into the shoulder periods. This offers the potential of greater efficiency in the use of existing transport capacity. It is potentially beneficial not only to road-based traffic, but also to rail. However, if current passenger volumes on bus or rail are not high, these strategies can lead to potential revenue loss for public transport operators. These losses may be offset by a lower requirement for peak vehicles, and less congestion, leading to higher productivity in the public transport industry.

It should also be noted that there is no reason why these strategies may not be offered in combination to employees. Thus, for example, some employees could be allowed both flexible working hours and telecommuting. Yet others may be offered telecommuting and compressed work weeks. In fact any combination of two of the strategies, and even all three together, may be offered. When the strategies are combined, the effects are additive from these strategies.

Encouraging ride sharing

As is noted under the TSM strategies, one of the inefficiencies of urban transport is the number of single-occupant vehicles travelling on the roads. While TSM can include the provision of dedicated lanes for those who are sharing rides, TDM strategies are needed to encourage people to change their behaviour of driving alone and sharing rides instead. To encourage ride sharing, most or all of the following strategies may be necessary, together with the TSM provision of carpool lanes and other supply-side strategies.

Carpool-matching programmes

Probably one of the simplest, easiest and lowest-cost actions that can be taken to encourage carpooling is to use a carpool-matching programme. Ideally, such a programme would be run by a government or service agency, which would offer the service to multiple employers in an area. The more employers included, the more likely the programme is to provide feasible matches. The principle of the programme is a bit like that of a dating service. In this case, however, the match is made by determining the work start and work end times and the locations of home and work, and then finding two or more employees who live and work fairly close to one another and have similar start and end times. If required, the matching programme can also be gender specific. Once matches are found, the matched employees are provided with contact information and encouraged to contact one another

to see if a carpool can be formed. A large employer may run such a programme just for its employees, but, as noted earlier, the more employers that utilise a single programme, the greater is the chance for matches to be found.

Carpool parking spaces

Similar to the idea of providing disabled parking spaces closest to building entrances, this strategy designates the next nearest parking spaces for carpool and vanpool vehicles. The idea here is to give preference and shortest walks into the workplace for those who ride-share to and from work. Only registered carpools can take advantage of these spaces. The existence of such spaces will also require those who continue to use single-occupancy vehicles to park increasingly far from the building entrances.

Guaranteed ride home programmes

One of the biggest objections people raise to depending on carpools is how to handle personal and family emergencies that might arise during the working day. For the driver in the carpool, this could mean a requirement to drive from work to home or elsewhere during the day, thus depriving members of the carpool of their ride back home at the end of the working day. For a passenger in the carpool, it could mean not having a means to travel home or elsewhere in the middle of the day to deal with the emergency situation. To counteract these problems, employers offer a guaranteed ride home. This may mean, in the case that the carpool driver has to leave work before the end of the working day, providing a company vehicle that can be used to transport carpool members home at the end of the day, or taxi fares to get them home. In the case of a passenger in the carpool who needs to attend an emergency, it may mean providing transport or taxi fare for that person to be able to travel to the required location.

Where such programmes have been put in place by employers, the usual experience is that there is very little call on the programme. However, its existence is of considerable help in encouraging employees to sign up and make use of carpools.

Encouraging public transport use and walking or bicycling

Another group of TDM strategies have to do with encouraging employees to use public transport, or to walk or bicycle to work. Several strategies are possible to encourage such practices.

Public transport information

A simple and very low-cost strategy is to provide at the workplace information on nearby public transport stops and stations, together with timetables and maps that show where the services operate. It is not uncommon that one of the reasons that people do not use public transport is simply ignorance of what services are available and which services operate from near home to near the workplace. Providing information in the form of timetables and maps, and identifying the locations of nearby bus stops and train stations may encourage employees to examine these alternatives.

Subsidising public transport fares

Another strategy that can be used is for employers to subsidise public transport fares. For employers that provide free or low-cost parking, the provision of public transport fare subsidies could be seen as equitable treatment for those who do not drive. However, there may be tax implications of such subsidies in different countries, so care must be taken that provision of subsidies does not incur some new tax burden for those so benefited. Subsidies for season tickets and multi-trip tickets are usually the easiest to provide, and encourage habitual use of public transport. In countries, like Australia, where tax incentives are offered for car use, the abolition of these incentives and possible replacement of them with public transport tax incentives would be similarly useful (in the absence of marginal social cost pricing of road use).

Secure bicycle storage

To encourage those who are within an appropriate distance of the workplace to use bicycles, one of the most obvious strategies that can be used is to provide secure storage at the workplace for bicycles. This is a relatively low-cost option for the encouragement of bicycling to and from work. It is important, in many climates, that the bicycle storage protects bicycles not only from theft, but also from inclement weather. An appropriate form of storage is to provide bicycle lockers with combination locks.

Showers and clothes lockers at work

One of the major complaints made by those who bicycle to work is that they may arrive hot and sweaty, and also that they may need to change into smarter clothing at the workplace than is consistent with bicycling to work. Provision of showers and lockers that will allow bicyclists to shower and to change into more appropriate clothing for work is a further significant

incentive to bicycle to work. Those who walk to work, especially in climates where very warm and humid weather occurs, may also desire the use of such showers and lockers.

Parking

There are two potential parking strategies that are usually included in the list of potential TDM actions. The first of these is aimed at discouraging use of the car, while the second is aimed at increasing public transport use through car access and egress.

Parking cost changes

Many employers provide free or subsidised parking to their employees. This is a direct incentive, in many cases, to use the car for the commute. In the past, building permits often required new commercial buildings to provide certain minimum amounts of parking. In this case, employers may decide that the marginal cost to them of the parking spaces is very low, and providing free or reduced-cost parking to employees is a useful employment benefit. If it is desired to divert car users to other means of transport, including car-pools and vanpools, then introduction of parking charges for commuters or increasing the price of parking to commercial rates is an obvious strategy worth pursuing, in addition to taxing employees on the value of their parking benefit. In addition, there are many parking garages and car parks that are run commercially that offer a reduced parking rate for all-day parking, especially if the vehicle is parked before a certain hour in the morning. Such deals need to be reversed, so that those parking all day are charged at a higher rate than the short-term parker. Another related option that has been considered in a few cases in the USA, at least, is to offer to employees the option to cash out their privilege to park at the workplace at a low or no parking charge. The cash-out option may be available for a year or for the duration of the time that the employee works for the employer (with safeguards against employees who may stay with the employer for too short a time). Once cashed out, the employee may no longer make use of free or reduced-cost parking, but would have to pay full commercial rates, possibly at a commercial parking facility.

Another option is to increase parking charges within a defined area, such as a CBD, by having government impose a parking tax on all parking in the area. Such a parking tax might be used, for example, to invest in improvements to public transport serving the area. Provided that the tax is large enough to make a significant difference to daily parking costs, this can work as a strong

disincentive to drive into the area so taxed and divert car users to public transport. Higher taxes could be levied if the car park operator sets low prices for all-day parking. However, there needs to be a good public transport service to the area in question, or the ultimate effect could be to drive employment out of the taxed area, or incentivise employers to add further subsidies either directly or through increased pay levels.

Park-and-ride and kiss-and-ride parking spaces

As a means to encourage public transport travel more directly than through parking charges and other arrangements, provision of parking spaces and drop-off and pick-up spaces close to train stations and major bus stops may provide an incentive for people to use public transport where they live too far from the stops and stations to be able to access them by walking. Safe and secure park-and-ride spaces should be provided free or at a very low cost for all-day parking, possibly with the public transport fare included in the parking charge. Similarly, there should be provision of safe spaces for kiss-and-ride public transport users (i.e. public transport users who are provided with a car ride to and from the stop or station). It is important that kiss-and-ride spaces be close to the station entrances for train stations and be in close proximity to the bus stop for bus services, so that kiss-and-ride patrons are encouraged to use these options.

Differential pricing

The final set of options that can be adopted as TDM strategies relate to pricing of different elements of the transport system. This may affect tolls and other road user charges and also public transport fares.

Tolls and road user charges

The goal of TDM actions is usually to reduce the amount of car use. In particular, these actions aim at reducing peak use of the car. However, too often in the past, pricing has operated in the reverse direction to that required to discourage frequent car use. As is also discussed earlier, in the section on parking cost changes, parking has also frequently been charged in a reverse pattern to what makes sense for discouraging car use and encouraging other forms of transport use. In the case of tolls and other road user charges, as with parking, higher charges should be imposed during peak periods and lower charges outside the peak. Such charges may help to divert some users to public transport, ride sharing, walking and bicycling, while they will motivate other users to travel outside the peak, thereby reducing peak loads on

the system. This, of course, flies in the face of other strategies that are used outside the transport arena, such as frequent-user rewards and ability to purchase multiples of some commodity at a lower price than for single units. However, when one is trying to change travel demand to lower external costs, especially to flatten the peaks and divert some users to other options, such as public transport, the typical frequent-user rewards are not the correct strategy to adopt (other than for frequent use of public transport).

Public transport fares

In direct contrast to the situation with road user charges, the argument should be to reduce the unit cost of public transport fares for frequent users. In this case, as with the more common frequent-user programmes, the idea is to encourage further use, which can best be done by reducing the cost for public transport users. Bearing in mind that there are a number of TDM strategies that are aimed at reducing the number of times people travel to and from work, without reducing their work effort, it is also important that public transport operators recognise that such fare reductions should not be applied on the basis of uses in a week or a month, but should rather be based on the number of times of use within some specified time period. Many public transport operators offer weekly or monthly tickets where the cost reductions compared to purchasing fares on a daily basis only arise for the fifth round trip in a week or the 18th or 19th round trip in a month. Such breaks in price are ineffective if employers are also offering workers the opportunity to telecommute or compress work weeks. A better strategy is to offer, for example, a ten-ride ticket that may have a validity period of two to four weeks. This still allows the frequent user who is travelling less than daily to obtain a reduction in fare. In addition, the availability of regional unlimited ride tickets, permitting multiple rides within a fixed time period, can be very attractive to public transport users and may result in increased use of public transport. In this case, once patrons have purchased such a ticket, they may be permitted unlimited free rides within a specified time period and geographic region, which will be likely to have the effect of encouraging use of public transport for more than just commuting trips.

Summary of TDM actions

As evidenced in this section, there are many TDM actions that can be taken. As noted at the outset, these actions are generally aimed at changing the demand behaviour of individuals, which is much harder than simply adding capacity into the transport system by physical, regulatory and management strategies applied to the network itself. Because TDM actions are aimed at

changing behaviour, they have generally been somewhat less effective than TSM actions in reducing peak traffic flows and diverting people to public transport. TDM strategies are usually most effective in changing behaviour when they confer other benefits on the people who adopt the strategies. For example, changes to work schedules are often seen as being beneficial when they provide more time at home for workers, although these changes are also often those considered less attractive by employers.

In the late 1980s, the Los Angeles Region Air Quality Management District introduced a regulation called Regulation XV, which required employers with more than 100 employees to introduce a range of strategies that would increase average vehicle ridership (AVR) to the workplace. Goals were set for average vehicle ridership that depended on the geographic location of the workplace, with the highest goals being within downtown Los Angeles (AVR of 1.75), a lower goal in the areas surrounding the CBD (AVR of 1.5) and a yet lower goal in the suburban areas (AVR of 1.25). Over a period of some two or three years, many employers introduced a variety of these TDM strategies, such as carpool matching, guaranteed ride home programmes, subsidies for public transport use, and so on. However, the average vehicle ridership of around 1.1 in the peak period was increased only slightly to about 1.2, far short of the goals. Some further gains arose as TSM strategies that favoured carpools and on-road public transport were introduced, but the overall effect was nowhere near what had been anticipated and hoped for. Employers of more than 100 employees were required to undertake a range of actions to change the average vehicle ridership, and the consequences of not doing so were some rather substantial fines. Employers were greatly opposed to the regulation and pressured the state government of California to repeal the regulation, an action that was successful after a few years of implementation of the regulation. Since the repeal of Regulation XV, carpool lanes have been expanded considerably in the Los Angeles region, but overall vehicle occupancy has not increased noticeably.

TDM strategies are introduced mainly by employers, although some are the responsibility of public transport operators and others of the roads authorities. However, the effectiveness of these strategies has largely been unclear, because evaluation has been almost non-existent, and introduction by only a few scattered employers is often not optimal to achieve real gains in shifting behaviour. Implementation across a large group of employers seems to be needed, together with complementary introduction of TSM strategies that will support and strengthen the TDM actions. For example, if carpooling is to be encouraged, then carpool lanes, carpool-matching services, carpool parking spaces, and guaranteed ride home programmes may all need to be

implemented. It is also important to note that, for carpool lanes, bicycle lanes and other similar strategies to be effective, they must provide preferential services that will match the commute trips of many commuters. Thus, the demand pattern of travel must be known and these preferential treatments provided for a significant proportion of the average commuter's work trip.

4.4 Non-coercive behaviour change strategies

TSM and TDM strategies are often coercive in the sense that they involve regulation, control, or pricing mechanisms to change behaviour. While such strategies may offer benefits to the user, they represent a 'carrot-and-stick' approach, in that there is a penalty attached to not changing behaviour, although a reward is offered for changing behaviour. In addition, most TSM and TDM strategies are evaluated as being beneficial mostly from the viewpoint of government, but not necessarily from the viewpoint of individual travellers. There are, however, non-coercive strategies that focus on the benefits to individual travellers, rather than benefits to government, and that remove the 'stick' and provide only the 'carrot'. In Australia, in particular, these strategies have been labelled voluntary travel behaviour change, but may also be known as TravelSmart or Smarter Travel.

There are two distinct ways of going about achieving VTBC. The first is a social marketing approach, and the second is a community development approach. Both approaches are applied at an individual level. They are described in the following subsections.

The social marketing approach

This approach is based on the assumption that one of the reasons that people do not choose to use alternatives to driving their own car (often alone) is because they do not know about alternatives to the car or are otherwise not confident or knowledgeable enough to choose other alternatives. This perception is closely related to the microeconomic view of the world that assumes that people make rational choices when they have perfect knowledge of the alternatives available. Clearly, under this assumption, if people lack perfect knowledge, they are unlikely to be able to make rational choices. In addition, the social marketing approach assumes that, if people are made aware of a problem or issue and of how they might affect that problem or issue by a change in behaviour, given awareness of the alternative behaviours, a substantial number of people will change behaviour to reduce or eliminate the problem or issue.

One could, therefore, describe the social marketing approach as being one in which a problem or issue is first identified from the point of view of government or the community. In a transport context, this could be congestion, pollution, greenhouse gases or other similar issues. The next step is to inform the community of these problems and how it is affecting the community in which they live. Having increased awareness of the problems (as with the way in which government has made people aware of the issues of recycling, water conservation, and so on), individuals are offered information and materials, possibly with incentives, to change their behaviour, so as to solve a problem that has been identified by the government or the community. Examples of the information and materials provided include:

- bus, train or tram timetables;
- public transport route maps customised to the location where the individual lives;
- free tickets to travel on public transport for a limited time;
- bicycle-riding promotional material, such as identification of nearby bicycle shops; and
- opportunities to have a bus driver visit to talk about how to use the bus system.

People are then encouraged to assist government to solve the community problems. However, this strategy may require the subsequent use of incentives to keep people motivated to maintain the changed behaviours.

The social marketing approach is positive, in that it increases knowledge about alternatives to the car, especially driving alone. It may also result in increased use of public transport, walking and bicycling. However, the disadvantages of the approach are twofold. First, it is a top-down approach, in that government (or the community) identifies the problem and enlists the assistance of people to solve the problem. This can give rise to the economist's 'drop in the bucket' attitude (i.e. that the behaviour change of one person will have too little effect on the problem). Second, the approach does not work from a consideration of the individual's lifestyle and ability or willingness to change, so that the change may not be adopted wholeheartedly and may require continued reinforcement in the form of incentives and promotions for a longer-term or permanent behaviour change.

The community development approach

The second approach to non-coercive behaviour change is a bottom-up approach that seeks to help solve a problem perceived by an individual

person and, in so doing, also helps to solve a government or community issue, such as congestion, air pollution, greenhouse gas emissions, and so on. In this case, the approach is based on attempting to understand what motivates people to change their behaviours. There are probably four primary reasons why a person may change his or her behaviour. First, a behaviour change will take place when the effects of a current behaviour reach a certain level that the person finds no longer acceptable and therefore is motivated to change. Second, there has to be recognition that it is possible to change. This recognition usually comes from one or more of at least three sources – the behaviour of trusted others, the occurrence in a person's life of a change moment (e.g. moving to a new home), and fashion or the perception of what has now become a popular thing to do. Third, a behaviour change may come about when a person recognises that he or she will receive significant benefits from the change. Finally, there may be other unexpected reasons why a person is motivated to change, apart from the preceding three. This may be a highly individual motive for change.

Behaviour change can also be motivated in different ways. Money is sometimes a motivator. However, it must be emphasised that it is only sometimes a motivator. Money is not important to everyone, and social and cultural values may be more important to some people. Financial incentives may also be lost to 'free riders', that is, those who are able to receive the financial incentive without actually making a behaviour change (perhaps because their behaviour already aligns with what is desired). Information also sometimes motivates behaviour change. However, information is often more effective at changing attitudes than changing behaviour (Darnton 2008). Both financial incentives and information may also produce non-desired behaviours. For example, providing free rides on public transport may induce some people to try the public transport system, without changing their car use, simply by undertaking additional travel by public transport. Attitudes may motivate change in behaviour. However, attitudes and behaviour are not necessarily aligned with each other (Kollmuss and Agyeman 2002). Changing attitudes is usually based on the belief that attitudes cause behaviour. The causal link is not clear, however, and attitudes can change without a commensurate change in behaviour.

The community development approach builds on these concepts and ideas and seeks to initiate behaviour change by approaching individuals in the community and inviting them to express something that they see as a transport problem in their lives. For example, an opening that is used in this approach is for a trained conversationalist from a specialist consultancy or government agency to engage a person in a conversation in which the person is asked to

say when he or she was last in a car and wished that he or she was not in the car. This is expected to result in the person identifying an issue relating to use of the car. An example might be non-working parents who find themselves acting as taxi drivers for their children to various activities. The next step is to ask the people what they think they could do about this. As the alternatives are explored, it may become apparent that some tools are needed to assist people to change their behaviour. These tools might include bus timetables and maps, discovery of a safe walking route for children to take between home and school, and so on. However, the emphasis remains on having people solve their own problems. Thus people buy into the solution and are motivated to maintain a change in behaviour, because it provides direct benefit to them. In addition, the change needs to be comparatively easy to make, must fit people's core values and must provide positive recognition or reinforcement for the new behaviour.

In contrast to the social marketing approach, this approach has individuals identify a problem and then find their own solution to the problem, which might possibly require one or more tools to help them change their behaviour. The change becomes self-sustaining as the individuals recognise the benefits that stem from the behaviour change. At the same time, these changes, motivated at the individual level, come together to help solve the community or government problem, such as congestion, air pollution, greenhouse gas emissions, and so on. The positive aspects of this approach are that the changes fit the lifestyle of each person concerned, the changes are more likely to be self-sustaining because they solve the individual's own problem, and people will be likely to diffuse information about the changes each has made to friends and relatives, thereby bringing about potential further change, much in the way that someone who embarks on a successful weight loss programme may tell friends about it, inducing some of those friends to try the same programme. The negative aspect is that this is a more costly and time-consuming approach in the short range, although it may provide more sustained improvements to the use of the transport system in the longer run.

Summary of non-coercive approaches

Non-coercive approaches have been subject to some of the most rigorous evaluation of any of the strategies discussed in this chapter (e.g. Richardson et al. 2005; Brög et al. 2009; SA Government 2009; Stopher et al. 2009). These evaluations have concluded that short-term changes of 4 to 20 per cent in the amount of drive-alone car use can be achieved among participants by means of these approaches. Differences in efficacy of the two approaches described here have not been subject to rigorous testing, so it is not known

at this time which of the two approaches – social marketing and community development – may be better. However, based on the results found for evaluating non-coercive approaches as stand-alone projects, the results are impressive and often greater than TDM or TSM strategies. Again, however, it is important that other strategies be implemented in concert with these VTBC strategies. It is unlikely that a VTBC programme will be very effective if public transport is already operating at or near capacity. If this is the case, then efforts to change behaviour are likely to fail to be sustained. Similarly, if VTBC is aimed at making ride sharing more attractive, then some of the TSM and TDM actions discussed in earlier sections of this chapter should be implemented also.

NOTE

- 1 SCATS is now used in 263 cities in 27 countries around the world. It is probably the most widely used computer traffic control system (RMS, n.d.).



REFERENCES

- Brög, M., E. Erl, I. Ker, J. Ryle and R. Wall (2009), 'Evaluation of voluntary travel behaviour change: experiences from three continents', *Transport Policy*, **16** (6), 281–92.
- Darnton, A. (2008), 'Reference report: an overview of behaviour change models and their uses', GSR Behaviour Change Knowledge Review, University of Westminster, July.
- Kollmuss, A. and J. Agyeman (2002), 'Mind the gap', *Environmental Education Research*, **8** (3), 239–60.
- Richardson, A.J., S. Roddis, D. Arblaster, D. Attwood and J. Newman (2005), 'The role of trend analysis in the evaluation of a TravelSmart program', paper submitted for presentation to the 28th ATRF meeting.
- RMS (n.d.), *History of SCATS Development*, Sydney: Transport for NSW, Roads and Maritime Services, available at: <http://www.scats.com.au/product/history.html> (accessed 28 November 2012).
- SA Government (2009), *TravelSmart Households in the West Project Report*, Adelaide: SA Department for Transport, Energy and Infrastructure, January, available at: http://www.transport.sa.gov.au/pdfs/environment/travelsmart_sa/Households_in_the_West_Final_Report.pdf (accessed 25 June 2009).
- Stopher, P., Y. Zhang and B. Halling (2009), 'Results of an evaluation of TravelSmart in South Australia', paper presented to the 32nd ATRF Meeting, Auckland, New Zealand, September.
- TRB (2010), *Highway Capacity Manual*, Washington, DC: National Academies of Science and Engineering, Transportation Research Board.

15

Goods movement



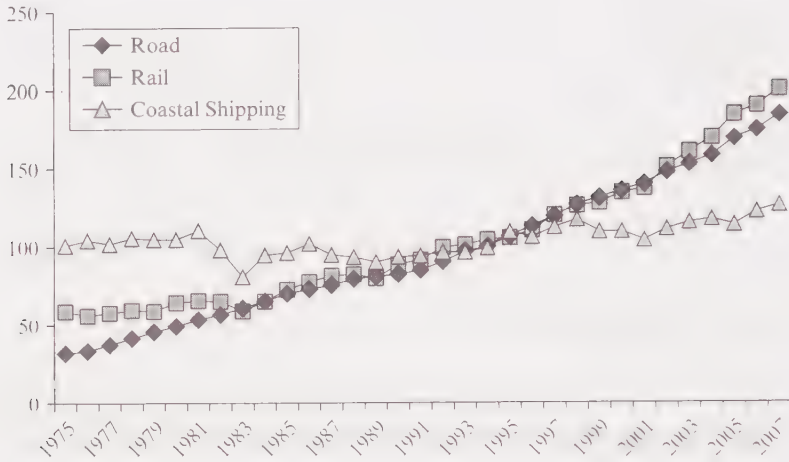
CHAPTER OVERVIEW

Internationally, trade volumes are growing more strongly than gross domestic product. With road movement being generally so dominant in the land-side movement of freight, transport policy is taking increasing interest in this sector. The chapter overviews some of the key land transport policy issues associated with goods movement. These issues encompass congestion, network access considerations, encroachment, energy use, supply chain coordination, safety, and environmental outcomes (climate, air pollution and noise), together with the regulatory framework. The way goods movements have been incorporated in a number of urban transport plans is summarised. The chapter highlights the tensions between improving the productivity of freight movement and making the sector accountable for its external costs.

15.1 Context

It is common for the growth rate in freight (or goods) movement to be higher than that of gross domestic product in real terms. For example, OECD trade data show that overall OECD imports, in volume terms, have grown at an average annual rate of 5.6 per cent per annum over the past 20 years (OECD 2012). Highest individual country growth rates were Poland (10.2 per cent) and Korea (9.5 per cent). Australia's imports grew 8.6 per cent annually and those of the US 6.0 per cent, while the UK lagged somewhat at 4.6 per cent and Greece brought up the rear at 3.0 per cent. These figures are generally well above comparable country real GDP growth rates. They help illustrate why the growth in freight or goods movement has been so strong.

At a national level, Figure 15.1 shows data for Australian freight movement by mode. Road, rail and coastal shipping are all important in tonne kilometre (tkm) terms, with rail and road both moving about 200 billion tkms annually, this movement task growing strongly. Road is more than three times as important as rail in terms of tonnes moved, the greater distances freight



Note: In billion tkms.

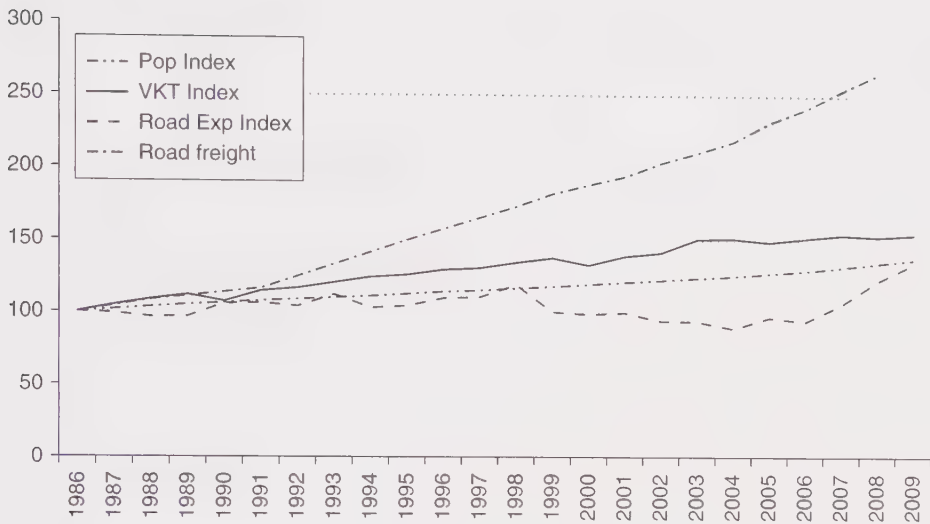
Source: BITRE (2009)

Figure 15.1 Australian domestic freight task

moves by rail and coastal shipping explaining the more significant roles for rail and coastal shipping in tkms.

Figure 15.2, again for Australia, shows road traffic growth for freight and passenger movement separately, compared to real GDP growth and to growth in road expenditure. Road freight traffic (in billion tkms) grew almost twice as quickly as real GDP, whereas person movement (in billion passenger kilometres, or pkms) grew about in line with real GDP and appears to have levelled off in recent years. This Australian road freight growth rate was relatively high in international terms. For example, from 1988 to 2008, ITF data show that the Australian road freight task increased by almost 150 per cent, compared to 65 per cent growth for the US, 40 per cent for Japan and only 31 per cent for the UK (ITF 2012). The Japanese and UK rail freight tasks did not show much growth over this period, whereas Australia's rail freight task increased by 168 per cent, reflecting (for example) strong growth in mineral movements.

This tendency for the growth in freight or goods movement to exceed GDP growth reflects increasing globalisation and competitive advantage, countries and regions specialising in what they do best and trading with others to improve their overall competitiveness. It arguably also reflects a tendency not to price adequately the external costs of freight movement, which in many countries would be far in excess of any taxation/road user charge payments (discussed in Chapter 12).



Note: 1985–86 = 100.

Source: BITRE (2009).

Figure 15.2 Growth in road traffic task and road expenditure, relative to population growth

Table 15.1 Typical freight movement patterns

Task	Location	Main modes
Bulk minerals	Mines to ports	Rail, shipping, pipeline, inland waterways
Agricultural products and livestock	Rural areas to urban areas or ports, perhaps via a processing stage	Road and rail as complements, and, as substitutes, shipping, inland waterways and some aviation
General long-distance land transport freight	Between urban areas; ports to/from distribution centres (for example)	Road and rail as complements and, as substitutes, some inland waterways
General freight urban distribution	Urban centres	Road predominantly
Industrial intermediate freight	Between or within industrial areas	Road and rail as complements and as substitutes

The freight movement task is often highly complex. Table 15.1 provides a number of simple examples of a country's typical flows. With the private sector dominating freight movement, the resulting allocation of resources is substantially influenced by individual corporate decisions, where the firm's

profitability will be a key influence but may account for only one step of many in a logistics chain.

In recognition of the many externalities associated with freight movement, the sector is usually subject to considerable regulatory oversight, and it is here that the main land transport policy issues typically (but not solely) arise. In designing policy for land freight movement, the same high-level economic, social and environmental goals (or specific jurisdictional variants thereof) that are used for person movement are again relevant. These might be reiterated as (for example):

- economic competitiveness;
- social inclusion;
- environmental sustainability; and
- safety and health.

With the commodities being moved being part of a jurisdiction's competitive offer, a more targeted objective for freight transport might be to improve productivity while respecting social and environmental constraints. Regional development might also be important to some jurisdictions, with land transport of freight being a contributory element to regional economic opportunity.

The productivity focus reflects the high-level economic goal, with social and environmental constraints being included to ensure due recognition is given to potential externalities associated with freight movement. In setting this book's focus on land freight, it should be noted that many freight movements involve sea (domestic and/or international), and productivity should be considered on a full movement basis, together with social and environmental considerations (a value position on our part is that it is, for example, unacceptable to export environmental costs to another country without adequate and fair compensation, or to exploit workers in another country).

A closely related approach has been taken in the *London Freight Plan*, where the language of sustainability has been used to reflect triple-bottom-line outcomes, including a focus on human rights. The sustainability definition used by Transport for London in its freight strategy is as follows (TfL 2008, p. 3):

Definition – sustainable freight distribution

This Plan defines sustainable freight distribution as 'the balanced management and control of the economic, social and environmental issues affecting freight transport that:

- Complies with or exceeds environmental standards, regulations or targets aimed at reducing emissions of climate change gases, improving air quality and minimising impacts from accidents, spillages or wastes
- Ensures freight is run efficiently, reduces unnecessary journeys, minimises journey distances and maximises loads with effective planning
- Complies with labour, transport and human rights standards and regulations ensuring that employees and communities affected by freight can function in a healthy and safe environment
- Minimises the negative impacts of freight activities on local communities.'

Taking a freight land transport policy frame of reference that is based on productivity, modified by social and environmental constraints, suggests there are a number of critical issues that will need to be considered by policy makers, such as:

- traffic congestion;
- network access;
- encroachment;
- energy use and outlook;
- coordination;
- safety;
- environmental outcomes:
 - climate policy;
 - air pollution; and
 - noise; and
- the regulatory framework.

Relevant policy instruments available to governments to respond to such issues include investment, pricing, regulation or legislation, and behaviour change. It is common for governments to use regulatory impact statements (RISs) as a framework for gathering information to inform decisions about policy and the use of these instruments (particularly regulations).

15.2 Congestion

Traffic congestion adds to the costs of moving goods and, to that extent, reduces economic competitiveness. Land transport costs are typically minimal for many high-value goods (such as Rolex watches or gold) but constitute a high proportion of product costs for bulky goods, such as construction materials. In some countries, the fast growth rate in road freight, compared to growth in person movement by car, suggests that freight may be an increasingly important contributor to congestion costs, even though

these costs provide an incentive to switch some freight tasks to less congested times of day. Much of the fast growth is in the lighter end of the commercial vehicle range, linked to the need for timely deliveries. Mass/distance/location (MDL) charging, discussed in Chapter 12, has much to recommend it in terms of improving the efficiency of freight movement.

Shrank et al. (2011) estimate US congestion cost in 2010 at \$US101 billion. They suggest that \$23 billion of this total is imposed on freight movement, even though trucks account for only 6 per cent of distance travelled in urban areas. The high unit values ascribed to freight movement (\$US88 per hour truck time for congestion costing) are an important reason for this scale of impact. A significant share of the increased freight cost is subsequently passed to consumers in higher prices. Chicago (\$2.32 billion), Los Angeles–Long Beach–Santa Ana (\$2.25 billion) and New York–Newark (\$2.2 billion) had the highest estimated freight congestion costs. Shrank et al. (2011, p. 16) argue:

Urban and rural corridors, ports, intermodal terminals, warehouse districts and manufacturing plants are all locations where truck congestion is a particular problem. Some of the solutions to these problems look like those deployed for person travel – new roads and rail lines, new lanes on existing roads, lanes dedicated to trucks, additional lanes and docking facilities at warehouses and distribution centers. New capacity to handle freight movement might be an even larger need in coming years than passenger travel capacity.

While freight movement experiences significant congestion cost, the size and operating characteristics of trucks mean that they also impose substantial congestion costs on other road users. As Parry (2009) notes, trucks take up more road space and drive more slowly than cars. These impacts on other road users are magnified in hilly conditions. It is for such reasons that congestion costing analyses typically use a passenger car equivalent (pce) value of 2 to 3 for trucks.

Pces for trucks are determined from several factors – the gradient of the road, the percentage of trucks in the traffic flow, and the length of the gradient. Separate pces for different sizes and types of trucks have not been published, and the published values are based on an 'average' truck. On level terrain, the pce for a truck is estimated as 1.5, irrespective of the distance or the percentage of trucks in the traffic flow. For rolling or mountainous terrain (rolling is gradients in excess of 2 per cent but less than about 4 per cent, and gradient lengths of less than about 2 kilometres), the average pce for a truck is about 2.5. For steeper and long gradients, and especially when trucks constitute a relatively low percentage of traffic, the pce can be as high as 7.0, although the

average is around 4.0. These equivalences reflect the effect of trucks on both the carrying capacity of the road and the speed of traffic. For example, if the capacity of a lane is 2000 cars per hour and there are 10 per cent of trucks in the traffic on a steep and long upgrade, where each truck is equivalent to four passenger cars, the capacity is effectively reduced to 1538 vehicles per hour, made up of 154 trucks and 1384 cars. This represents a drop of 30 per cent in the effective capacity, or equivalently it means that the 10 per cent of trucks take up almost 40 per cent of the road capacity.

15.3 Network access

Road

The road and bridge system (toll roads and some gated communities excepted) operates like a commons, with almost universal vehicular access potentially available. However, roads and bridges are not constructed or maintained necessarily to permit access by all who might wish to use the network. Considerations such as the land use in an area (e.g. whether it is a historic town centre with narrow winding streets), the wealth of a country and the priorities it attaches to roads and the nature of the demand for road use in an area (e.g. whether there are local quarries) all exert an influence on how the road and bridge system might develop. Jurisdictions make decisions about vehicle configurations that will be provided with unlimited access to the road and bridge system, configurations that will be provided with access under certain conditions and configurations that will be prohibited. Vehicle axle mass loading (e.g. which causes road and bridge damage), width (e.g. which affects the safe passage of other vehicles on any given road or bridge) and length (e.g. which affects the feasibility or safety of manoeuvrability or turning on a particular road geometry and bridge damage) are critical influences on access conditions that are likely to be allowed.

A vehicle's environmental performance is also taken into account in the decision of whether or not it can operate on the road network (e.g. black smoke emissions frequently have limits). Environmental aspects are discussed in section 15.8. Similarly, drivers must meet certain requirements to be able to drive a motor vehicle, managed through various licensing regimes and related areas like alcohol and drug regulations. Driver aspects of network access are considered in section 15.7. The current section focuses on vehicular network access and its productivity implications.

Prescriptive standards are usually set down in regulations, to govern access opportunities that will be available for any given vehicular configuration.

These encompass such matters as vehicle mass, length, width, height, swept paths, lighting requirements, speed limits, seat belt requirements, a range of other safety requirements (e.g. rear underrun protection, rollover requirements, load restraint requirements, seat strengths for bus seats, and so on) and environmental requirements. Different access impacts may be connected to particular prescribed standards. Thus some regulations determine which parts of the road and bridge network a particular vehicle configuration (such as a B-triple, which is a prime mover with three trailers) may be able to access, while others affect the nature of such access (for example, whether different speed limits might apply).

It has long been recognised, however, that prescriptive access standards can inhibit innovation in vehicle design and, as a consequence, hold back productivity growth. The evolution of huge trucks that shift minerals in places like Western Australia's Pilbara, on company-controlled roads, shows how freeing up design constraints can drive productivity gains, where the road and bridge design opportunities are permissive. The key transport policy question associated with network access concerns how to balance the productivity gains that can flow from freeing up vehicular network access with the road and bridge damage, safety and environmental impacts that might result.

In the past two decades, there has been increasing interest in the idea of performance-based standards as an alternative way of managing access to the road and bridge system. Performance-based standards provide scope for vehicular configurations that deviate from those prescribed by a jurisdiction's regulations and set down the associated performance standards or requirements that must be met and the process for meeting such requirements. For example, rather than setting down precise regulatory requirements for a vehicle's swept path, which determines its turning capabilities, some alternative possibility might be allowed if the proponent can demonstrate that this alternative can operate safely on the routes on which it is to be used and provide a means of assuring that these are the only routes that will be used (e.g. via some on-board monitoring system that is accessible by the regulatory authority).

In a road user charging system that seeks to fully reflect the external costs of road use in user charges, charging for specific cost external consequences (on a marginal social cost basis) flowing from particular decisions to allow performance-based vehicular operation should be a part of the charging regime.

Performance-based standards systems typically need their own set of authorising arrangements that sit alongside the prescriptive regime.

Rail

Rail access also raises important policy questions, with similar access matters being taken into account as for road (e.g. track capacity to carry different loads and the speed at which a train can operate). Rather than focus on the engineering, safety and environmental issues associated with access regimes, as in the preceding subsection on road, we consider some broader access policy questions for rail. In an urban setting, for example, how should competing demands for freight and passenger movement be managed in a context of limited track capacity? If rail track is owned and/or managed by a private operator, how can public interest issues associated with competing demands be brought to bear on the decision about capacity allocation? If a private owner has built its own rail facility to transport its own products (e.g. as is common in bulk haulage of minerals from mine to port for export or coastal shipping), what access conditions (if any) should govern third-party access? Similar issues arise, to a greater or lesser extent, with respect to access to the road system.

15.4 Encroachment

By encroachment, we mean the conflict that arises between freight movement and abutting land uses that are adversely affected by that movement. This commonly arises, for example, in areas close to major feeder routes to ports, where growing numbers of heavy trucks can significantly reduce the amenity of neighbourhoods through which vehicles move. For example, a narrow residential street (Francis Street, Yarraville, Victoria, Australia) is used by about 9000 trucks a day, mainly going to or from the Port of Melbourne, with about one in ten of these trucks travelling during a (rarely policed) curfew period. This is likely to lead to higher levels of air pollution, with adverse health impacts, severe noise nuisance, increased safety risks and suchlike. Media coverage suggests higher rates of patients reporting to hospital with respiratory disease from this area than is usual.

Encroachment problems typically arise as land use/transport problems, when the intensity of land use changes over time. The strong growth in trade flows reported at the beginning of this chapter, for example, will be associated with strong growth in freight movements through ports. If those ports are located near markets, they are likely to be in or near major cities. Impacts of increasing freight movements on abutting land uses will thus become more pronounced over time.

These problems are associated not only with ports. For example, factories may have been long located near residential areas and be associated with

similar problems as business grows, particularly when access is via local streets (a 'last mile' issue). Again the problem is one of land use/transport interaction as intensity of use changes.

Encroachment problems may also reflect changing social values over time. As incomes increase, people tend to place a higher value on protecting their environment. Conflicts between freight movement and abutting land uses may thus reflect changes in people's preparedness to put up with noise, even if the noise level itself is not changing.

If encroachment problems are mainly due to land use/transport interactions, then that is where solutions will primarily need to be found. Long-term strategic land use plans need to ensure that adequate provision is made for growth in freight volumes, within constraints set by community values. For example, if a port is located close to a city's CBD and government is trying to increase residential and job densities in abutting areas, land values will be very high and there may be pressure to limit growth in port throughput. This may require designating a new location to accommodate port growth. Also, policy measures can be implemented to improve the way freight moves to and from the area of concern. For example, inland freight terminals or hubs (which could be on a city's fringe) can be used to consolidate freight, with movement to and from the port relying more on rail than road.

Land zoning more generally can be used to contain uses with adverse impacts on residential and other sensitive environments, with buffers to limit external costs. Access restrictions (e.g. truck bans) can direct freight traffic to particular routes, where impacts are lower, provided compliance with these restrictions is adequately enforced. Measures can be put in place to mitigate some of the more severe impacts (e.g. freeway noise barriers). Modal restrictions or incentives can be used to achieve movement by modes with the lowest impacts, such as development of intermodal freight hubs. These incentives would be justified by the inability of a private operator to capture all the external benefits that might flow from such an initiative (e.g. lower levels of air pollution, noise and traffic congestion). Marginal social cost pricing of the external costs of transport is well suited to such circumstances.

Similar issues arise with rail. For example, freight trains moving through urban areas frequently create noise problems for people living in the high-density residential environments often found alongside railway lines, where those lines are shared by freight and passenger traffic (transit-oriented development).

15.5 Energy use/security

With road freight movement being heavily dependent on fossil fuels (diesel), certainty of supply is a fundamental requirement for an efficient freight sector. Issues such as peak oil suggest that high and rising fuel prices are likely over time, and carbon pricing can be expected to accentuate price pressures on transport fuels over time. Taking these two sets of pressures in tandem suggests increasing fuel risks over time, from the perspectives of both availability and cost. Countries that are relatively highly dependent on imported supplies can be expected to be particularly sensitive to such issues, where increasing balance-of-payments pressures add a further complexity, particularly if the degree of import dependence is on the rise (e.g. as for Australia).

Policy responses to transport fuel availability and cost concerns, for freight and passenger movement, are generally similar and, in terms of transport policy responses, include (for example) a range of measures to reduce dependence on fossil fuels and diversify sources of supply (which will not remove all the cost risks). Typical policy responses, as they relate to road freight movement, could include:

- pricing or incentives and investment to encourage modal switching away from road to rail;
- improving freight movement efficiency, to reduce the fuel use per tonne kilometre (e.g. through improved driving techniques, use of truck-only lanes, or better vehicle design); and
- seeking alternative fuel sources (e.g. many buses use ethanol or compressed natural gas).

15.6 Coordination

It is common for freight movements to involve use of multiple modes, such as when shipping is involved. This means that a range of stakeholders impact on the efficiency with which the movement task is accomplished. Each stakeholder is usually concerned about the efficiency of the movement task only in so far as it affects their link in the chain, which does not assure a maximally efficient movement task from society's viewpoint. This can deliver suboptimal outcomes to the overall logistics task.

Coal is a major export from New South Wales, with the Hunter Valley being a major source of supply. The Hunter Valley Coal Chain handles product valued at about \$A15 billion (variable with price, in particular). Concern about the inefficient movement process from mines to customers led to the

formation of the Hunter Valley Coal Chain Logistics Team (HVCCLT) in 2003. HVCCLT effectively integrates the operation of port operators, railway operators and infrastructure managers. About 120 million tonnes per annum of coal are moved through this process. Over 24 coal points are coordinated. Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO) has developed software to optimise the movement task. It is important that competition legislation does not arbitrarily limit the scope to coordinate stages in the freight task in this way.

15.7 Safety

Freight movement is undertaken in vehicles that are potential killers, and many countries have a poor safety record for truck movement in particular. Rail safety is also an important transport policy concern. If the vehicles used to move freight are potential killers, then those in control of those vehicles are also potential killers. Those in control are not limited to vehicle drivers. 'Chain of responsibility' legislation was pioneered in Australia by the National Road Transport Commission (NRTC) in the 1990s, to hold all who have an influence on heavy vehicle-related safety outcomes accountable for their performance. This includes drivers, but also includes, for example, those who schedule driving tasks and customers who place demands on trucking companies that require particular delivery schedules (which may involve, for example, exceeding legislated working hour limits).

Road safety requirements cover the vehicle, the driver and the road environment. Examples of relevant policy approaches in each area include:

- the vehicle:
 - mass and dimension requirements (including swept path), which are partly safety related;
 - lighting;
 - rear underrun barriers;
 - dangerous goods movement requirements; and
 - restrictions on operation of 'over-mass' or 'over-dimension' vehicles;
- the driver:
 - licensing requirements related to the vehicle class;
 - blood alcohol and drug-taking restrictions; and
 - driving or working hour and rest requirements; and
- the road environment:
 - road curvature, width and condition;
 - allowable speeds; and

- lane restrictions on heavy vehicle operation (e.g. some jurisdictions ban heavy trucks in one lane of a three-lane dual carriageway freeway).

Valuing the health and safety impacts of heavy vehicle operation is a critical part of the preparation of many regulatory impact statements, whether these RISs are prompted by an interest in improving vehicle productivity, concern about some specific safety questions, or environmental regulation, where health benefits and costs are usually central. The Australian National Transport Commission has long been a leader in this area.

15.8 Environmental outcomes

Freight movement is a source of several significant environmental concerns. On the road side, these concerns include emissions (both local, such as particulate emissions, and global, particularly greenhouse gas emissions), noise and amenity concerns (as outlined in section 15.4). This discussion concentrates on air pollution concerns from heavy vehicle operation.

As pointed out in Chapter 7, diesel fuel is a source of fine particulate emissions, which are known carcinogens. Nitrogen oxide and hydrocarbon emissions from heavy vehicles are also of concern. Emission standards have regularly tightened the level of allowable emissions from new vehicles, and many countries also monitor in-service emission performance of a vehicle during its working life.

European emission standards are defined in a series of European Union directives, with the Euro VI standards for heavy-duty vehicles applying from 2013. In the executive summary of the impact statement presenting its recommendations for the Euro VI standards, the Commission Staff argued the case for tighter emission standards on heavy-duty vehicles (CEC 2007, p. 3):

By 2020 the EU will still be a long way from achieving the objectives of the 6th Environmental Action Programme. Significant negative health and environmental impacts will continue to persist even with effective implementation of current legislation . . .

Particulate matter (PM) pollution is of increasing scientific concern and harmonised emission limits at European Union level are necessary to prevent barriers to the distribution and circulation of heavy duty vehicles and to achieve substantial reductions in particulate emissions from those vehicles across the EU. A 66% reduction in the limit value for particulate matter is proposed with

an emission limit of 0.01 g/kWh, which, given current technology, will require particulate filters to be fitted to all diesel vehicles.

Additional action on nitrogen oxide (NO_x) and hydrocarbon (HC) emissions is also justified . . . An 80% reduction in NO_x to an emission limit of 0.4 g/kWh is proposed. Increased use of engine technology and after-treatment technology shall be needed to fulfil the required NO_x emission limits.

Detailed argument supporting the tighter new standards is contained in the body of the impact statement. The impact statement for Euro VI comprehensively assessed alternative options for lowering emissions, in terms of:

- changes in aggregated utility levels of households;
- changes in aggregated production costs of firms;
- welfare changes stemming from changes in government tax revenues; and
- changes in external environmental costs.

The emissions evaluation processes followed by the EC are recommended reading for policy analysts interested in best practice.

15.9 The regulatory framework

Our particular concern in this section is the approach taken by neighbouring jurisdictions to harmonise their regulations to facilitate smooth operation of markets and avoid, for example, a vehicle's operation suddenly becoming illegal because it has crossed a border, be that an international border or a border between two neighbouring states of the same country. Relevant examples are the approaches taken by the European Union and Australia. In both cases, the operating environment across neighbouring jurisdictions is decided by a negotiated outcome, where representatives of the relevant jurisdictions are the parties to the negotiation.

In Australia, for example, the National Road Transport Commission was established in 1991 to harmonise road transport legislation and regulations across all Australian states and territories, which had primary responsibility for managing the road use. State and territory arrangements led to conflicting regimes, which inhibited innovation and were frequently difficult for business to understand, much less comply with, adding to costs. The NRTC worked with the states and territories to develop the harmonised regulatory regime. This task was subsequently taken over by the National

Transport Commission, when rail was added to the scope of the agency's work. Ministers from all jurisdictions meet as a ministerial council to sign off relevant legislative and regulatory provisions that govern how (for example) the road system will be used to meet productivity, safety and environmental goals, with national uniformity the target. This covers the full range of matters that take place on the road and bridge network, including the many issues related to vehicles and drivers discussed above. Road construction and maintenance are handled elsewhere.

In developing harmonised operating regimes, the greater the number of affected jurisdictions the stronger the case for one entity that sets the common standards and rules. The risk otherwise is a race to the bottom, with the lowest common denominator being the option most likely to get widespread support, if a negotiated outcome is required.

15.10 Freight in land transport plans

Because of the significance of freight movement to a productive economy, and the economic costs of congestion for goods movement, it is common for national and metropolitan transport plans to devote considerable attention to this sector. In metropolitan transport plans, this attention typically concentrates on the issues discussed above, such as port access, encroachment more broadly, and environmental outcomes. For example, the 2012 *Draft New South Wales Long Term Transport Master Plan* sets out five major policy directions for land transport, one of which is 'Supporting efficient and productive freight' (NSW Government 2012). That policy direction sets out seven major action areas:

- targeted network upgrades (to improve freight movement);
- port growth plans;
- ports action plan;
- higher productivity vehicle pilot programme;
- rail freight infrastructure enhancement;
- protecting strategic freight corridors; and
- development of a package of measures to grow off-peak freight movement.

Environmental impacts are not singled out in this set of measures, probably because much of the environmental regulation of freight movement in Australia is set at the national level, whereas the *Master Plan* is a state government policy. However, several of the measures indicated would have positive environmental outcomes from better separation of freight and other traffic.

The London Freight Plan (TfL 2008), framed against a much slower expected growth in freight movement than Australian cities have been experiencing (London's plan was forecasting only 15 per cent freight growth to 2025, from 2007), places much more emphasis on operational aspects of the freight task than the New South Wales plan and more focus on modal shift towards rail. It includes:

- a freight operator accreditation and recognition system, to encourage improved operator performance across a range of sustainability dimensions;
- delivery and servicing plans, to increase the building of operational efficiency;
- construction logistics plans (similar to delivery and servicing plans, with application to the design and construction phase of premises); and
- a freight information portal.

The London approach looks to partnerships across the logistics chain to deliver improved outcomes. Reducing CO₂ emissions is a strong motivation behind the plan, much more so than for Sydney. The existence of the London congestion charging scheme should have contributed to improving the efficiency of freight movements in the city.

The Oregon 2035 *Regional Transportation Plan's* freight focus is much closer to Sydney's than to London's (Metro Council 2009). It identifies six issue areas:

- congestion and hotspots – which impede freight operation;
- reliability – improving the predictability of travel times;
- capacity constraints – due to physical and operational issues or lack of capacity in key corridors;
- network barriers – associated with issues such as over-mass or over-dimensional operation;
- land use (protecting land for future industrial uses); and
- impacts – managing adverse impacts (e.g. environmental, land use).

At the national level, the 2012 US Moving Ahead for Progress in the 21st Century Act (MAP-21) includes a focus on freight, where the intention is to improve the national freight network, strengthening the ability of rural and regional communities to access national and international trade markets and supporting regional development (US DOT 2012). Road improvement is a key target, with funds being provided to the states by formula for projects that improve regional and national freight movements on highways,

including intermodal freight connectors. A primary freight network of up to 30 000 miles (based on freight volume) is to be identified (by the Federal Secretary of Transportation) within the first year of the legislation, and there is a commitment to development of a national freight strategic plan, with related state plans. Some \$US2 billion annually is available over the two years of this legislation for improvements to the primary freight network and/or other key freight links, which are part of a wider designated national freight network. The focus on the highway network reflects a growing concern in the US about underinvestment in the nation's infrastructure base.

15.11 Concluding comments

The fast growth in land freight traffic in many countries, and pervasive externalities associated with freight movement, is leading to an increased policy focus on this sector. This has been heightened by a slower growth in car traffic in many developed economies. Land use/transport integration is critical to resolving many of the issues outlined in this chapter, but others can be resolved by tighter regulation of the freight-operating environment. This is likely to impede innovation, such that performance-based standards are likely to play a stronger role in coming years. Charging for freight movement that recognises variations for mass, distance and location (MDL charging) is particularly important for more effective resource allocation for freight movement, as discussed in Chapter 12.



REFERENCES

- BITRE (2009), *Australian Transport Statistics: Yearbook 2009*, Canberra: Commonwealth of Australia, Bureau of Transport, Infrastructure and Regional Economics.
- CEC (2007), *Annex to the Proposal for a Regulation of the European Parliament and of the Council on the Approximation of the Laws of the Member States with Respect to Emissions from On-Road Heavy Duty Vehicles and on Access to Vehicle Repair Information: Impact Assessment*, Commission Staff Working Document COM(2007) 851 final SEC(2007) 1720, 21 December, Brussels: Commission of the European Community, available at: http://ec.europa.eu/enterprise/sectors/automotive/files/environment/sec_2007_1718_en.pdf (accessed 13 October 2012).
- ITF (2012), *Trends in the Transport Sector: 1970–2010*, Paris: OECD, International Transport Forum.
- Metro Council (2009), *2035 Regional Transportation Plan*, Portland, OR: Metro Council, available at: http://library.oregonmetro.gov/files/freighplan_fact_sheetpdf.pdf (accessed 25 October 2012).
- NSW Government (2012), *Draft New South Wales Long Term Transport Master Plan*, September, Sydney: Transport for New South Wales.
- OECD (2012), *OECD Economic Outlook*, Vol. 2012, Issue 1, No.91: *Statistical Annex*, Paris: OECD.
- Parry, I.W.H. (2009), *How Much Should Highway Fuels Be Taxed?*, Discussion Paper RFF DP 09-52, Washington, DC: Resources for the Future.

- Shrank, D., T. Lomax and B. Eisele (2011), 'TTI's 2011 urban mobility report', Texas Transportation Institute, Texas A&M University System, September, available at: <http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/mobility-report-2011.pdf> (accessed 25 October 2012).
- TfL (2008), 'Executive summary', in *London Freight Plan: Sustainable Freight Distribution – A Plan for London*, June, London: Mayor of London, Transport for London, available at: <http://www.tfl.gov.uk/microsites/freight/documents/publications/London-Freight-Plan-Executive-Summary.pdf> (accessed 25 October 2012).
- US DOT (2012), *Moving Ahead for Progress in the 21st Century: Public Law 112-141*, Washington, DC: US Department of Transportation, Federal Motor Carrier Safety Administration, available at: www.fmcsa.dot.gov/about/what-we-do/MAP21-21/Map21.aspx (accessed 21 March 2013).

16

An integrated land use/transport policy



CHAPTER OVERVIEW

This chapter brings together much of the material from preceding chapters to suggest how a national land transport policy and also a city's transport policy might be framed. It emphasises the importance of taking an integrated land use/transport approach, with transport helping to achieve intended land use policy outcomes over the longer term. National and city-level approaches from various jurisdictions are observed to have much in common, primarily because of the importance usually attached to dealing with the external costs of motor vehicle use and the policy directions that this implies. Case studies are used to illustrate top-down and bottom-up approaches to city-level transport policy. The chapter ends by recalling some of the funding issues discussed in Chapter 12, funding being seen as perhaps the biggest single constraint on improving integrated transport policy outcomes.

Today the challenge is to take transport out of its box in order to ensure the health, vitality and sustainability of our metropolitan areas. (Brookings Institution 2008, p.9)

16.1 Context

This chapter brings together much of the discussion from previous chapters, in a context of developing an integrated land use/transport policy. This is usually done at the city level, or for a part thereof, since this is where the interactions between land use and transport are usually most visible. The KonSULT tools, which were included in Chapter 3, are applicable at either level. The potentially strong linkages between transport infrastructure improvement and development of agricultural and/or mining activities should also be noted, particularly where prior infrastructure condition is poor. The focus in this chapter, however, is city-wide, within a national transport policy context.

National and city-level policy alignment cannot be taken for granted but, with half the world's population now living in cities and this share growing, national economic and social success is intimately linked to city performance. Jurisdictional responsibilities for land use typically sit below national level, at local government level or sometimes at a regional or state/provincial governmental level. However, national policy settings can contribute to enhanced city-level outcomes, and policy integration between national and city-wide levels is likely to benefit both.

The concentration of people in cities improves productivity and many aspects of liveability, particularly through scale economies and agglomeration effects in production and consumption. The origins of agglomeration economies in production have been well understood for many years. They include such sources as improved access to inter-industry information flows (information spillovers), thick labour markets, and better access to specialised services (e.g. legal services, design and testing, financial services) and to public infrastructure.

Agglomeration effects in consumption, an important element of liveability, are a relatively new area of quantitative research, but Chapter 11 pointed to Borck's (2007) research (for example), which provides evidence of agglomeration externalities in consumption from a German case study. Bigger cities (in population terms) showed benefits for residents from a larger range of service choices, across areas such as restaurants and bars, concerts, dancing, theatres and museums.

There is a trade-off in city size between agglomeration benefits and the external costs of (for example) traffic congestion, pollution and noise. Size carries risks: bigger cities tend to have greater agglomeration benefits but larger external costs. Getting the balance right is a key task of urban policy. The tension between agglomeration benefits and external costs is not only an issue of absolute city size. It frequently also arises when cities grow quickly, with infrastructure and services lagging behind population growth.

Transport is a major contributor to agglomeration benefits, particularly through its contribution to accessibility and the related concept of effective density, and is a vital requirement in growing suburbs. It is also a source of some of the major external costs of cities, such as congestion, air pollution, social exclusion and noise. Getting the land use/transport balance right is thus a vital element in having a great city and a motivating reason for taking an integrated approach to land use and transport.

Noted urban scholar Cervero (2001) argues that large cities that are compact and enjoy good accessibility, matched by efficient transport infrastructure, are among the most efficient urban settlements. These cities do not arise by chance: they require decades of careful management and guidance.

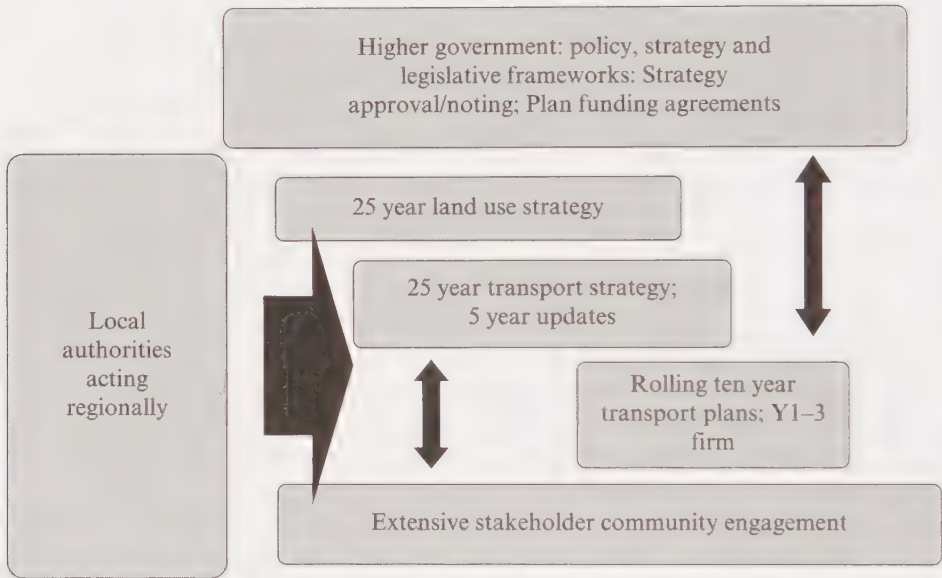
16.2 The need for integration

The agglomeration/external costs trade-off illustrates why an integrated approach is important in land use/transport policy and planning. People and goods do not usually move around simply for the sake of it. For people movement, for example, the demand for transport is essentially derived from the demand for activities that people wish to undertake. In consequence, understanding what drives demand typically requires some understanding of the way spatial systems work. This includes understanding how different land use configurations impact on travel demands and, in turn, how transport systems influence land use. Major transport improvements, in particular, can play important city-shaping roles. For example, urban rail capacity expansions can support policy intent to foster growth in a city's central business district. Radial urban freeways constructed at a city's fringe can lead to urban sprawl in the absence of strong land use controls. It is better that such potential outcomes be planned, to achieve intended societal outcomes, rather than arising as an unintended consequence of seeking to solve transport problems in siloed isolation. This requires an integrated approach.

The importance of integrating land use and transport policy and planning has been widely recognised for a long time, but the practice of implementation generally falls short of expectations. Institutional design for land use/transport integration frequently concentrates on integration within one particular level of government. However, if service impacts (benefits and/or costs), service delivery responsibilities and/or funding obligations cross jurisdictional boundaries between levels of government, then institutional arrangements also need to facilitate and manage this cross-governmental involvement for effective integration, even if service delivery responsibilities lie largely (or entirely) at one particular level of government (as is common).

Figure 16.1 sets out a broad framework for land use/transport integration at city level, which recognises the importance of structuring arrangements, both within and between levels of government, to support an integrated approach. For example, national-level governments (called the federal government in many countries) typically have the strongest revenue-raising powers and often make funding available to lower levels of government to pursue initiatives that advance a national policy agenda. They can also use other national

policy or programme measures to support city-level initiatives (e.g. locating government offices and/or other infrastructure facilities in places that are supportive of city development intentions, or implementing supportive environmental regulatory regimes for issues such as air quality). These are practical reasons for approaching city land use/transport in an integrated way across levels of government (in the examples given, by recognising national policy imperatives).



Source: Based on Stanley and Smith (2012)

Figure 16.1 A framework for land use/transport integration

Alignment of goals between layers of government becomes an important process in such cases, together with the use of evaluation and monitoring processes that help to assure accountability for outcomes. Funding agreements can help to cement cross-governmental integration, as has been practised (for example) in Canada with respect to major transport infrastructure projects.

Internationally, the most common institutional arrangement for land use/transport is that local government, acting at a regional level, has primary responsibility to drive the process, based on the principle of alignment of decision-taking responsibility with the jurisdiction in which the consequences (costs and benefits) of those decisions are most concentrated. In some cases this may mean a single local authority, if that authority covers

the entire geographic space of interest. In others, it may require some means of local authorities working together to act regionally, as in Vancouver (for example). Figure 16.1 shows the relationships with local government exercising this driving role. It is recognised, however, that in some jurisdictions a higher level of government (the state government in Australia, for example) may have land use/transport responsibility at city level (which indicates inclusion in the left-side box).

Higher levels of government (national/federal, state/provincial) have legitimate interests in land use/transport integration, particularly because of impacts on high-level social goals (e.g. the national economic significance of city economic performance in the knowledge economy and the impacts of poor land use/transport integration on this performance; the national economic costs of traffic congestion, which has adverse productivity impacts; the national and international environmental consequences of city performance, especially with respect to greenhouse gas emissions associated with transport and land use choices). It is such impacts on high-level social goals that justify national or federal funding flows to support implementation and outcome achievement at subsidiary levels of government (and also state or provincial funding flows, if local government has primary land use/transport planning and delivery responsibilities). Legislative frameworks set at higher levels of government may also impact directly on what is possible at local government level in terms of implementation.

The framework in Figure 16.1 suggests a long-term land use strategy of about 25 years, although there are good arguments for taking a 40- to 50-year perspective, because of the slow rate of change in land use in many cities, and some US cities use both high-level 40- to 50-year plans and more detailed 20- to 25-year plans. The 25-year land use plan is shown as marginally leading preparation of an associated and integrated long-term transport strategy. This phasing is because a policy view on broad land use issues, such as (for example) the balance between growth on the fringe and more compact infill settlement patterns, and whether or not to protect agricultural lands or scarce natural areas on urban fringes, should help to guide strategic transport directions, because of the way major transport initiatives can shape urban form (support, or detract from, intended outcomes). The 25-year transport strategy is then seen as being given more detail in a short-term implementation/transport plan (a five-year or a ten-year transport plan), which should include a related funding plan. The 25-year strategy should be updated about every five years and the short-term plan annually. The first three to five years or so of the short-term plan need specifics on policies, projects and funding, with the subsequent years being more indicative.

16.3 National policy directions for land transport

In this section, we address the kinds of national or federal transport policy directions that might be common and relevant to integrated land use/transport policy and planning at city level.

Effective transport systems are a vital foundation of competitive economies and of liveable, inclusive communities. They enable the efficient and safe movement of people and goods that is critical to quality of life. Roads, public transport systems, footpaths and cycle ways provide people with opportunities to access family and friends, jobs, recreation, education, health care, and the many other activities that contribute to individual and community well-being. Many of the same roads and rail lines are used as part of the logistics systems that support goods movement, with ports also being critical components of the relevant infrastructure sets. An efficient transport network supports a healthy and vibrant national economy and society.

Land transport systems, however, are also associated with a number of serious economic, environmental and social problems (externalities), such as congestion, greenhouse gas emissions, air pollution, social exclusion for some because of a lack of mobility options, energy security issues, and health (e.g. obesity) and safety concerns. While most of these issues arise on a place basis, particularly an urban place, the economic and social significance of cities in most countries makes them issues of national importance.

Countries from Europe to North America, and elsewhere, have recognised the pervasive influences of transport and the importance of a national approach to transport policy, to maximise the potential benefits from an effective and efficient transport sector. As a result, transport has typically become a significant element of national policy agendas.

This increasing policy interest is partly a reflection of inadequate spending on transport infrastructure during the 1970s, 1980s and, in some cases, 1990s, as transport investment fell as a proportion of gross domestic product (GDP) in many countries around the world (Stanley and Barrett 2010). In an increasingly globalised business world, the impacts of declining transport infrastructure spending on productivity levels and economic competitiveness have rung alarm bells in many countries (e.g. the US, Canada, many European countries, Australia). This concern has been compounded by declining revenue flows from fuel taxes in countries such as the US and the UK, as per capita motor vehicle use has been in decline, vehicle fuel economy has improved and governments have been unwilling to increase fuel tax rates.

Such national concerns have led to growing interest in the preparation of national land transport policies. For example, Chapter 3 quoted the Ministerial Foreword to Norway's *National Transport Plan 2010–2019*: 'It is about making everyday life easier; about sustainable development; making society more inclusive and universally accessible; and strengthening the competitiveness of Norwegian commerce and industry' (NMTC 2011, p.3). The overall objective of the plan is 'to provide an effective, universally accessible, safe and environmentally friendly transport system that covers the Norwegian society's transport requirements and advances regional development' (NMTC 2011, p.6). The important role played by land use policy is acknowledged.

The major policy directions for improvement in the Norwegian Plan are (NMTC 2011):

- improving standards in infrastructure;
- reducing the number of fatalities and serious injuries in road accidents (targeting a reduction of one-third by 2020 and with a zero fatalities or long-term injuries target for the long term);
- a public transport system that is suitable for all users (including stimulating regional and municipal authorities to implement schemes that restrict use of private vehicles, including local initiatives on congestion charges, and encourage use of environmentally friendly modes);
- better accessibility and reliability;
- reducing climate and environmental impact (including provision for a carbon price on fuel equivalent to the price of international emission quotas);
- reducing the risk of ice and rockslides on large parts of the road and rail network;
- improving safety and navigability in Norwegian waters; and
- improving efficiency in ports.

With the exception of problems of ice and rockslides, these policy directions for improvement are very similar (for example) to Australian directions. The Norwegian Plan emphasises improving public transport in urban areas and improving roads in other areas, but recognises the need also to improve public transport in regional areas. Toll finance is expected to fund a significant part of the growth in spending. Upgrading major international and national transport corridors is a significant priority in the Norwegian Plan, reflecting broader European policy concerns. These national policy directions are also common elsewhere.

The European Commission's 2011 *White Paper on Transport* (EC 2011) sets out a comprehensive set of policy directions intended to complete the internal market for transport, to promote European economic growth and integration, while anticipating resource and environmental constraints. It sets a target of reducing transport greenhouse gas emissions by at least 60 per cent on 1990 levels by 2050, within wider EU GHG emissions reduction targets of 80–95 per cent. As well as decarbonisation of transport, the plan also aims to reduce oil dependence.

Congestion is recognised as a growing European problem, as is funding the required infrastructure (including public transport infrastructure). The White Paper supports charges that internalise external costs of road use. It proposes moving towards reformed pricing by first broadening heavy vehicle charging to incorporate externalities and then extending this to passenger cars, recognising that this will deliver benefits in terms of both behaviour change (away from car use towards lower-impact modes) and revenue raising, which can be used to improve infrastructure. Pricing reform accounts for a significant proportion of the White Paper, reflecting the importance attached to this reform by the Commission.

Improving public transport quality, accessibility and reliability is central to the policy directions in the White Paper, because of the multiple benefits flowing from a shift towards PT use, within the context of an integrated approach to urban mobility: 'In the urban context, a mixed strategy involving land use planning and pricing schemes, efficient public transport services and infrastructure for non-motorised modes and charging/refuelling of clean vehicles is needed to reduce congestion and emissions' (EC 2011, p. 14). There is a strong focus on transport corridors in the White Paper, mainly for long-distance movement (including freight) but with links to major nodes such as capital and other main cities, ports, airports, and key border crossings. Improvement priorities are proposed to target missing links, upgrading existing infrastructure and development of multi-modal terminals.

These examples illustrate that, because the external costs of urban land transport have an adverse impact on achievement of high-level economic, social and/or environmental goals, it is common for national or federal land transport policy (as it relates to cities) to be directed at:

- managing congestion costs and improving economic competitiveness and liveability as it is affected by land transport (including pursuing the potential external benefits of agglomeration economies);

- achieving substantial cuts in transport greenhouse gas emissions;
- ensuring adequate mobility options are available for all;
- making the transport system safer;
- encouraging healthier transport choices; and
- increasing energy security, by reducing reliance on (imported) fossil fuels.

In consequence, improving the sustainability of most city land transport systems is essentially about policies aimed at:

1. getting the modal balance right for transport of people and goods, which in most developed countries means a shift away from high dependence on motor vehicles towards walking, cycling and public transport, including rail for freight, and, in developing countries, avoiding problems of excessive reliance on motor vehicles as incomes rise;
2. improving the environmental performance of all transport modes but particularly of cars and trucks, because of their dominant roles; and
3. ensuring that travel opportunities are available to all, irrespective of personal circumstances.

These three policy objectives can usually be translated into six major programme directions, with indicative actions of the type shown below:

1. **reducing the demand for travel:**
 - land use planning (increased density, and co-location, while pursuing agglomeration economies);
 - maximising opportunities for walking and cycling;
2. **achieving a shift to lower-carbon or lower-air-polluting transport modes:**
 - cars to public transport, walking and cycling;
 - trucks to rail;
3. **improving vehicle utilisation:**
 - higher car occupancy;
 - more efficient freight movements;
4. **reducing vehicle emission intensity:**
 - more efficient vehicles;
 - smaller passenger vehicles;
 - alternative fuels;
 - intelligent transport systems;
 - better driving practices;
5. **increasing mobility opportunities:**
 - provision of reasonable base public transport service levels;

- using existing public transport opportunities (e.g. school and community buses) more effectively;
- 6. **creating a more sustainable freight network:**
 - focus on freight movement to ports and hubs and to connect key manufacturing or distribution centres.

Table 16.1 shows how most of these programme directions can contribute to reducing a number of the major national land transport issues, again underlining the importance of taking an integrated approach. These policy directions then need a sustainable funding package to support implementation. In any particular city, these generic policies need to be nuanced by the particulars of place in that city, as we illustrate in sections 16.5 and 16.6.

16.4 Some land use/transport research findings

If an integrated land use/transport approach is to be pursued, what does the available evidence tell us about key relationships that might inform policy and planning thinking? Ewing and Cervero (2010) have produced a very useful meta-analysis of studies on relationships between travel and the built environment, drawing on statistical analysis from about 50 published studies. They examine the influence of density, diversity (of land uses), design (particularly street network characteristics), destination accessibility (ease of access to trip destinations) and distance to transit. While the subject matter is almost all from the US, the findings provide some insights that are likely to have wider application. They report impact elasticities, which show the sensitivity of various response variables (such as vehicle miles travelled, or VMT) to changes in a range of potential causal influences. While most elasticities are quite small, the combined effect of a number can be significant, emphasising yet again that integrated policy packages need to be important in the land use/transport space. McKibbin (2011) has examined similar factors for Sydney, with the central zones excluded from analysis because they skewed the data set. The main Ewing and Cervero findings are summarised below, with relevant McKibbin findings presented for comparison.

Ewing and Cervero (2010) find that, of the various land use variables considered, destination accessibility is the most important factor in determining a household or person's amount of driving and walking. The more accessible a centre, the lower the VMT, probably because of lower car ownership rates and less car dependence in highly accessible central or inner areas, and the higher the walking mode share. Distance to downtown is also highly

Table 16.1 Alignment of measures and their expected benefits

Critical national land transport issue	Land transport major programme directions					
	1. Reducing the demand for travel	2. Achieving a shift to lower-carbon or lower- air-polluting transport modes	3. Improving vehicle utilisation	4. Reducing vehicle emission intensity	5. Increasing mobility opportu- nities	6. Creating a more sustainable freight network
Congestion	Yes	Yes, with suitable infrastructure provision, to ensure congestion is not shifted to another mode	Yes, provided traffic generation is managed	Neutral	Neutral	Needs to minimise road traffic generation
Greenhouse gas emissions	Yes	Yes	Yes	Yes	Depends on how provision is made. Low emission modes best	Focus on fuel-efficient flows
Social exclusion	Should target shortening trip lengths, not eliminating activities	Yes. Because these means of travel are low- or no-cost, they are inherently relatively inclusive	Yes. This is a common way people at risk of social exclusion improve their mobility options	Price increases may have negative impacts	Yes	Neutral
Energy security	Yes	Yes	Yes	Yes	Neutral	Needs fuel-efficient modes
Safety and health	Yes	Yes	Yes	Neutral	Yes	Risk area

related to vehicle miles travelled, VMT increasing with distance. Personal characteristics, such as household income, have a larger influence on VMT than such land use variables.

McKibbin (2011) finds that destination accessibility by public transport is an important influence on mode share of non-car modes (PT, walking and bicycling) for the journey to work in Sydney. The relative accessibility performance of PT compared to the car is significant in the Sydney study, such that improving car accessibility without improving PT access will increase the car mode share.

An important inference from this study on destination accessibility is that, if governments want to promote locational agglomeration economies (i.e. lift urban productivity), while reducing the external costs of road use, a major focus should be on opportunities for redevelopment and revitalisation of central locations, supported by improvements in accessibility by public transport. The growth in the knowledge economy is supportive of this direction. Ewing and Cervero (2010) suggest that 'Almost any development in a central location is likely to generate less automobile travel than the best designed, compact, mixed-use development in a remote location.'

Ewing and Cervero (2010) found that the design metrics of intersection density and street connectivity were also important influences on VMT, particularly through their impact on cycling and, more particularly, walking. Short blocks and many intersections seem to shorten travel distances, with higher intersection density seeming to be strongly linked to increased walking. They point out that linking where people live and work (the jobs-work balance) allows more walking, particularly if intersection density is supportive. This is an important design insight for promotion of activity centres and urban villages, more important than mixed use in the reported findings. McKibbin (2011) did not find street density had a significant influence on PT mode share in Sydney, suggesting that this could be because of the difficulty of finding suitable metrics.

Transit accessibility is related to VMT and to walking (greater accessibility reducing VMT and increasing walking), while transit use is most closely correlated with distance from a transit stop and the shape of the street network. While Ewing and Cervero (2010) identified these aspects as significant, their finding was based on a small number of studies. McKibbin (2011) found distance to transit to be a significant but minor influence on mode choice for the journey to work in Sydney, suggesting that this effect may perhaps be being obscured by destination accessibility that the train station provides.

Perhaps surprisingly, neighbourhood population and job densities were not as strong influences on VMT as some other factors (such as location) in the Ewing and Cervero work. This may be because of problems of multi-collinearity, since dense settings usually have mixed uses, short blocks and central locations, all of which Ewing and Cervero (2010) suggest shorten trips and encourage walking. McKibbin (2011) found that density had only a moderate influence on transport mode share for the journey to work in Sydney, reiterating concerns about multi-collinearity.

Ewing and Cervero went to pains to control for problems of self-selection: for example, people walk more in places with a good walking environment because people who like to walk choose to live in such places! However, as Benfield (2010) of the US Natural Resources Defense Council points out on his blog:

All indications in the market suggest that we have a large, growing, unmet demand for close-in, walkable neighborhoods and an emerging surplus of automobile-dependent environments; research consistently shows that, where walkable neighborhoods in smart locations exist, walking goes up and driving goes down . . . The environment doesn't care what the psychological motive is.

McKibbin (2011) highlighted the relatively small role of the various factors analysed on mode share of non-car modes for the Sydney journey to work. Car ownership, income and workplace location were far more important influences, suggesting that efforts to promote transit-oriented development (TOD) should target reducing car ownership (e.g. low parking availability). The author suggests that TOD should mainly focus on building population growth in highly accessible locations, to drive use of non-car modes (car ownership declining in higher-density, more PT-accessible locations).

Bento et al. (2005) examined the effects of urban form and public transport supply on travel mode choices and annual vehicle travel in 114 US cities. This was one of the studies included in the Ewing and Cervero (2010) analysis. Bento et al. found that population centrality, the jobs–housing balance, city shape and density, in combination, had a significant effect on the amount of vehicle travel, generally mirroring the Ewing and Cervero conclusions. The effect of moving a sample of households from a city like Atlanta (733 persons per square kilometre, 7000 rail miles of service per square kilometre, 10 000 bus miles of service per square kilometre) to a city with the characteristics of Boston (1202 persons per square kilometre, 18 000 rail miles of service per square kilometre, 13 000 bus miles of service per square kilometre) was a projected reduction in annual vehicle travel of 25 per cent. The

result underlines the important role of urban renewal and the supportive role of a good public transport system, to increase the scope to generate locational agglomeration economies and reduce the external costs of road use.

Similar conclusions emerge from a comparison by Buehler et al. (2009) between Germany and the US. It was found that Americans travel by car about twice as much as Germans. The analysis suggests that transportation policies and spatial development (German cities are more compact than US cities) each account for about 25 per cent of the explained variability in travel behaviour.

Compact, pedestrian- and bicycle-friendly, mixed-use developments, containing medium- to high-density residential, office and retail uses within walking distances of rail stations (or tram or bus rapid transit routes), are sometimes called TODs. A number of studies have shown how such developments can reduce car use by 20 per cent or more. For example, a study in Seattle, Washington found that in mixed-use TODs car use was reduced by about one-third, with public transport, walking and cycling playing correspondingly larger roles. Residents of TOD-like neighbourhoods in the San Francisco Bay Area had almost half the vehicle miles travelled of those in new suburban developments (Sydec 2007).

In the Australian context, more compact urban development is a goal for most major cities. This is likely to require a much greater focus on building activity levels (including residential populations) in CBDs, increasing jobs and population in key urban nodes and increasing development densities along principal public transport corridors (higher-density, low-rise) and at major public transport nodes (e.g. around rail stations). Public transport investment will be needed along the major corridors to cater for the higher-density, mixed-use developments, and capacity expansion will be needed in several CBD-oriented public transport services, to cater for patronage growth. Similar policy directions will be relevant for large cities in other developed economies where density patterns are low.

16.5 Land use/transport integration in Melbourne

Background

The translation of some of the preceding generic material into a specific urban transport policy is perhaps best illustrated by a couple of case studies. Melbourne (Victoria, Australia) and New York (USA) are used for this purpose.

Melbourne is in the course of updating its land use/transport strategy at the time of writing this book (2012–13), with one of us a member of a small group that is advising the responsible Victorian Minister for Planning on this update. The discussion is at a strategic level, with particular initiatives that are identified requiring detailed evaluation to test their performance effectiveness. The approaches outlined in Chapter 4 are applicable to those assessments. The subsequent New York discussion is more specific on projects.

Melbourne (4 million people) is consistently rated as one of the world's most liveable cities, which is a major competitive advantage in the knowledge economy. The city was listed at number 1 in both the 2011 and 2012 rankings by the Economist Intelligence Unit (*Economist* 2012). The city's high liveability reflects its many distinctive qualities, qualities whose continued nurturing are front and centre in the Metropolitan Planning Strategy, to protect liveability. The city's population will pass 5 million in the next two decades, a population level where the balance between agglomeration benefits from size and offsetting external costs is questioned by some (see, for example, Cervero 2001). Melbourne's new Metropolitan Planning Strategy will include a plan for Melbourne's transport system. The transport plan is being assembled to support the desired land use planning directions for the city.

A ministerial advisory council (MAC) was appointed to consult widely and assist the state government with developing the Metropolitan Planning Strategy. The MAC has taken the view that, unlike previous land use/transport strategies for the city, the new strategy will emphasise:

- working with (not against) markets in terms of shaping the future of the city (e.g. the previous Melbourne strategy selected some major urban nodes for development, which have languished because of a lack of private sector interest);
- seeking to enhance individual choices (recognising that this can improve individual well-being);
- supporting people's capabilities to exercise choice; and
- ensuring a focus on implementation (which means actively pursuing support across multiple stakeholders).

The MAC has identified five principles or outcomes it thought people would value about Melbourne in the future. These are a Melbourne that (1) is distinctive, (2) is globally connected and competitive, (3) enables social and economic participation, (4) has strong communities and (5) is environmentally resilient. This is a local expression of the usual triple-bottom-line

goal formulation, with the focus on distinctiveness particularly important to 'brand Melbourne' (the city's liveability).

Two further principles relate to the shape of the city thought most likely to support achievement of these outcomes: Melbourne should be (6) a poly-centric city and (7) a 20-minute city. These are central land use development directions. The first of these spatial principles requires a number of suburban activity nodes that support the primary role of the CBD, some of which will be strongly connected to the global economy (e.g. Monash–Clayton, Melbourne Airport, Dandenong). The second reflects a desire for people to be able to undertake more activities closer to where they live.

What does this mean for future transport directions? It suggests that Melbourne's transport plan should generally seek:

- to emphasise elements such as trams, boulevards and main streets which contribute to Melbourne's distinctiveness;
- to facilitate global competitiveness, through supporting the growth of key economic nodes (e.g. access for skilled labour to support concentrations of knowledge industries) and efficient freight and logistics operations (a Melbourne national competitive strength);
- to improve accessibility to jobs, services and people;
- to contribute to affordable living and a sense of place; and
- to contribute to environmental resilience through respect for the natural and built environments, and a focus on low-impact transport.

These general directions need to be worked up on a place basis. Some relevant considerations are set out below.

Central and inner area

Central Melbourne employment has grown by about 200 000 over the past 20 years, reflecting the city's competitive advantage in high-end knowledge-based industries. No other Australian central city has experienced such growth over that time. Land available immediately adjacent to the current central area (CBD) can support future employment growth of a similar scale. Many of these additional jobs will be filled by people living outside the central area, and rail, which is already the major mode for accessing the CBD, will need to continue to be the major means of access (there is no effective alternative to carry the numbers). However, train use has increased over 80 per cent in the past decade, and the core of Melbourne's metropolitan rail network is close to capacity. This constrains future employment opportuni-

ties. The \$10 billion Melbourne Metro (underground) rail project aims to resolve this network capacity constraint, to unlock future central area growth potential and associated agglomeration economies. This will be a major city-shaping project, implemented to drive a particular pattern of future central area land use.

Walking, cycling and tram service upgrades are also needed to support central city growth. This transport policy direction will support use by central and inner urban residents and by others travelling to and from this area and will support greater dwelling density in the centre. For example, development of the expanded central area will require tram network extensions into the CBD growth areas. Some major central and inner area arterial roads could be developed as major boulevards, to complement one of the city's current sources of distinctiveness. The inner public transport corridors can then be targeted to support higher development densities, with suitable public realm improvements.

Freight and logistics are important areas of competitive advantage for Melbourne. However, the city is over-dependent on one major freeway corridor. A major new road tunnel (\$10 billion) is one way to help ease these pressures and improve the city's economic resilience. Project emphasis should be on traffic bypassing the city and traffic to and from the port, to allow a continuing high quality of life in central and inner areas, to minimise risks of significant generation of new road use (and the attendant external costs). Removing some through traffic will enable surface public transport improvements and walking and cycling improvements and free up opportunities for improvements to the public realm in inner areas.

Middle and outer suburbs

Victorian government research suggests that the city's economic, social and environmental performance is better served with a small number of major activity centres (polycentric shape) and constrained fringe growth than with either more centres and/or more dispersed growth. Because of the high costs (including external costs) of fringe growth, probably about 70 per cent of Melbourne's housing growth over the next 40 years, and a higher proportion of employment growth, will need to be in established suburbs, where it is difficult to add road space.

Improving accessibility within existing middle and outer urban areas will facilitate population and activity growth. This requires better circumferential movement opportunities for people and goods along major trunk roads.

Better accessibility is also required for employment and activity nodes, particularly by public transport, walking and cycling, and also for commercial traffic. Smarter management of the road network is needed to cope with competing demands, with priority for use of limited road space given to high-value road uses (which can vary by time of day).

Improved trunk bus services (primarily circumferential) and local bus services (e.g. increased frequencies and longer operating hours, bus priority treatments) are central to improving access opportunities in existing suburbs, including supporting growth of the selected activity nodes (which tend to be located at junctions of radial rail and circumferential trunk bus services). Patronage growth on Melbourne's often-derided bus services has been very strong in recent years (70 per cent growth since 2005–06), as services have been improved. Improving accessibility should assist in promoting job growth in activity centres, as part of an integrated set of measures to promote activity centre development (e.g. selective increases in density, mixed-use development, focus on growing higher-order service employment).

These improvements should help to increase the PT, walking and bicycling mode share for work and non-work journeys, most of which are not to the CBD but are within the home local government area (LGA) or to a neighbouring LGA. Target PT service frequencies on the trunk circumferential middle and outer corridors should be 15 minutes or better for most service hours (which should be from about 5.00 a.m. to midnight in most cases, with a lesser frequency being acceptable late in this operating span).

High priority should be given to walkability and cyclability within and to and from local centres, to support greater use of more sustainable travel modes and also assist development of more compact settlement forms, particularly in the middle and outer suburbs (inner suburbs are already typically characterised in this way).

Rail grade-separations are a major issue in Melbourne. The city has about 170 level crossings. Traffic delays at these locations will get worse in future, as rail volumes continue to increase, adding significant costs to industry and delays to travellers. Level crossing abolition is a high priority.

Growth areas

Melbourne's historical outward suburban expansion is at a tipping point, with commuting times and distances in danger of blowing out as a result of

disconnection between housing and jobs (partly reflecting years of underinvestment in transport infrastructure and services in growth suburbs, as the city has grown quickly). House prices are generally lowest in growth areas, household incomes also tend to be lower, and risks of high interest rates and high fuel prices are highest for people in these areas.

Strategies are needed to make sure residents in growth areas have better job availability and better access to jobs elsewhere. This requires convenient public transport services when new housing estates are opening (with a target 30-minute headway and at least 15 hours of daily operation) and safe walking and cycling opportunities. It also requires the upgrading of some old rural roads to urban arterial status, while seeking to limit further urban sprawl.

Availability of jobs in growth suburbs is relatively low, and the manufacturing sector, which has traditionally provided many job opportunities for people from these areas, is generally contracting, under cost pressures and the high Australian exchange rate. Providing high-quality trunk PT services between outer growth suburbs and the most proximate employment hubs, ensuring that road capacity is sufficient to meet these PT service needs (if rail is not available), is one way to enhance access to employment opportunities and promote social and economic participation. In many cases this will mean improving trunk PT services between the outer suburbs and the middle suburbs, where jobs are more readily available, while also seeking to increase the availability of local (non-transport) services and of jobs in growing outer suburbs, to reduce the need to travel.

Freight

In addition to many initiatives included above, transport improvements to support growth in freight and logistics will include:

- developing new freight precincts and gateways as part of a more decentralised network;
- better rail connections to ports, including new longer-term port development options;
- preservation of a corridor for an Outer Metropolitan Ring Road in the city's west, which is the area of highest freight and logistics specialisation; and
- new and more efficient terminals in outer areas close to national and international trade routes.

16.6 New York Sustainable Streets Plan

New York scarcely needs an introduction! While the Melbourne case study might be seen as primarily driven by a top-down strategic approach, which will blend with local self-determination, the Sustainable Streets Plan (SSP) is more bottom-up, set within the strategic framework of Mayor Michael Bloomberg's PlaNYC 2030.¹ That strategic framework is about the city economy, environment (including climate change) and quality of life, with the transportation network being seen as a critical opportunity to advance these goals and the SSP elaborating on this from the bottom up.

The SSP is the city's transport policy. It contains five main programme areas:

- **Making the nation's safest big city even safer**, with a goal of halving fatalities from 2007 to 2030. The high transit mode share is an important contributor to this result.
- **Improving travel in a thriving city**, a mobility programme to increase the role performed by transit (particularly bus), high-occupancy vehicles more generally, cycling, ferries and walking. Parking management initiatives are an integral part of this programme area.
- **World-class streets for a world-class city**, a programme area with a strong urban design and public realm focus intended to enable streets to perform better a wide range of functions for their communities, not just operate as travel thoroughfares.
- **Delivering outstanding roadways, bridges and ferries**, where the emphasis is on ensuring and maintaining high standards of infrastructure condition.
- **Greening the department from top to bottom**, an environmental programme to lead by example, which includes a target of cutting greenhouse gas emissions from municipal buildings and operations by 30 per cent by 2017.

The Plan also includes an internal programme focus on providing a twenty-first-century transportation department. Some 164 actions are identified to deliver on the Plan, but the Plan does not indicate future funding levels that will be dedicated to these actions. A Sustainable Streets Index helps to assess progress on delivery.

New York's density makes high rates of walking and mass transit use possible and helps explain why the carbon footprint of a resident of New York City is less than one-third that of an average American. In planning for future population growth, the city recognises that its subway system is hitting capacity

constraints and that additional bus capacity, both bus rapid transit and more conventional bus with some BRT advantages (e.g. priority lanes), is the most effective way to add capacity in the near future.

Making space for bus priority treatments highlights the issue of competing demands for street space. Increased bicycling and walking are also priorities within the treatment of street spaces. Street design for safe cycling, and to give cyclists a greater sense of belonging, is an important objective. Reducing auto use is a corollary of the efforts to increase use of low-impact modes. The New York solution to the problem of competing uses has started with a comprehensive review of the use of streets and other public spaces in the city.

In revising the roles and uses of streets, the SSP takes a strategic perspective and argues:

The best cities in the world today approach streets as vital public places that foster social and economic activity, in addition to their more traditional role as corridors for travel. In New York, with some of the densest development in the world, the streets are literally our front yards . . . [The Plan therefore seeks] . . . more varied and lively streetscapes to make our streets great destinations. (NYSSP 2008, p. 20)

Urban design and transportation planning skills are then brought to bear to deliver against this perspective. In functional terms, this results in (for example) a focus on some streets as primarily walking streets, others as bus routes and others as freight routes. 'People-friendly grand boulevards' are another important focus, a theme that is coincidentally common with the Melbourne intent, reflecting the urban realm and urban design emphasis behind both.

An important ingredient of the New York streets programme area is the aim for provision of a public plaza within a ten-minute walk for every community. Underused road space provides much of the raw material, and local engagement is a key part of the process of developing such plazas, in areas where they are lacking or in short supply.

In essence, the New York approach starts at the street level and works upwards to the achievement of community economic, environmental and social goals, whereas the Melbourne approach is doing the reverse. Both are valid approaches to integrated land use/transport policy and planning. In contrast, an urban transport plan for a city that stays within a conventional transport framework would not 'stray' into this broader integrated domain and, in consequence, is missing opportunities to maximise potential public value for the

communities being served. Taking the broader integrated perspective throws up ideas and opportunities that would never emerge from more narrowly conceived perspectives. For example, the idea of grand boulevards depends on crossing functional boundaries to see a range of connected outcome possibilities and their broad design drivers. The changes already taking place in New York are ample evidence of the success of that city's approach.

16.7 Sustainable funding

Implementation is a critical element for any metropolitan strategy or plan, and funding is a key component of implementation. New funding streams will be required in most cities if transport systems and services are to be improved. As discussed in Chapter 12, these funding streams are likely to include: user charges (including tolls, which could be extended to help fund level crossing abolition); value capture measures, which extend user pays charging to beneficiary pays mechanisms (e.g. special rates, a metropolitan liveability levy); asset sales, with the revenue hypothecated to land use transport; and further borrowings over time. All levels of government will usually need to be partners in the funding process, which may be assisted by the use of an intergovernmental agreement.

Arguments of efficiency and fairness support a greater reliance on user pays and beneficiary pays pricing systems, provided distributional considerations are recognised and tackled. User pays systems have the allied benefit of reducing the size of the apparent investment backlog (by encouraging behaviour change), provided equity concerns are handled. In the long term, user pays pricing is perhaps the most important policy change that is needed in most transport infrastructure funding. It is also the most difficult to implement, because of a lack of political support.

Metropolitan land use/transport strategies or plans should specify how various funding sources will be combined to fund the transport initiatives (capital and operating) required in the plan, in sufficient detail to provide comfort that implementation over the first ten years or so will be achievable as planned. This implies that an infrastructure plan should accompany a land use plan and transport plan.

16.8 Regulation for the environment and safety

In addition to spatially focused policy directions and the general funding opportunities, improved land use/transport integration also requires a suite of more generic policies and programmes to improve transport outcomes.

Some of these were summarised at high level in Table 16.1. They include (for example) road safety treatments to reduce risks to travellers (including pedestrians and cyclists), similar programmes for public transport users (e.g. safe walking routes to bus stops, and safety at rail stations), cyclists and walkers, parking controls, traffic-calming programmes, emissions control regulations to manage externalities that are not tackled through pricing reform, behaviour change programmes to support modal change and other behavioural responses, working hour regulations to improve heavy vehicle safety, and suchlike.

16.9 Conclusions

The economic, social and environmental outcomes from cities benefit from land use and transport policy makers and planners taking an integrated land use/transport perspective on policy and programme directions. The engagement of closely related sectors, such as education, health and welfare, is also beneficial (although not elaborated above), for reasons such as the influence these sectors have on the location of population-serving activities and, hence, on travel requirements. This engagement also provides opportunities for improving the alignment of local employment and housing as a city grows and changes (e.g. by opening up schools for wide community uses).

Setting a future vision for a city is the starting point in taking an integrated approach. A values- and outcomes-based approach was illustrated for Melbourne. The New York example illustrates an approach that has high-level goals but works from the street level upwards towards goal achievement. In both cases buy-in from multiple stakeholders is critical for success. Strategic land use considerations should lead the process of land use/transport integration, although the particular strategic directions that are desirable in any particular city are likely to emerge from a continuing learning engagement between land use, transport and communities. Aligning policy directions across levels of government is likely to enhance prospects of successful implementation, because of the wide span of interests that are affected by the success, or otherwise, of cities.

In the current uncertain international financial environment, funding has become perhaps the biggest single obstacle to successful land use/transport integration. Developing plans is probably the easiest part. Finding the political courage to raise the revenue to implement them is another matter. The authors believe that the failure to take tough decisions on funding is likely to result in a slow and insidious move backwards in terms of high-level economic, social and environmental outcomes for many cities.

NOTE

- 1 The website for details of this plan is <http://www.nyc.gov/html/plany2030/html/home/home.shtml>.



REFERENCES

- Benfield, K. (2010), 'The definitive study of how land use affects travel behaviour', *Switchboard*, 4 June, available at: http://switchboard.nrdc.org/blogs/kbenfield/massive_study_confirms_that_de.html (accessed 5 June 2012).
- Bento, A.M., M.L. Cropper, A.M. Mobarak and K. Vinha (2005), 'The effects of urban spatial structure on travel demand in the United States', *Review of Economics and Statistics*, **87** (3), 466–78.
- Borck, R. (2007), 'Consumption and social life in cities: evidence from Germany', *Urban Studies*, **44** (11), 2015–2121.
- Brookings Institution (2008), 'A bridge to somewhere: rethinking American transportation policy for the 21st century', Brookings Institution Metropolitan Policy Program.
- Buehler, R., J. Pucher and U. Kunert (2009), 'Making transportation sustainable', Metropolitan Policy Program at Brookings Institution, April.
- Cervero, R. (2001), 'Efficient urbanization: economic performance and the shape of the metropolis', *Urban Studies*, **38** (10), 1651–71.
- EC (2011), *White Paper on Transport: Roadmap to a Single European Transport Area – Towards a More Competitive and Resource-Efficient Transport System*, Luxembourg: European Commission, Directorate-General for Mobility and Transport.
- Economist* (2012), 'Liveability ranking: Australian gold', *Economist*, 14 August, available at: www.economist.com/blogs/gulliver/2012/08/liveability-ranking (accessed 21 March 2013).
- Ewing, R. and R. Cervero (2010), 'Travel and the built environment: a meta-analysis', *Journal of the American Planning Association*, first published on 11 May (iFirst).
- McKibbin, M. (2011), 'The influence of the built environment on mode choice: evidence from the journey to work in Sydney', *Australian Transport Research Forum 2011 Proceedings*, Adelaide, 28–30 September.
- NMTC (2011), *National Transport Plan 2010–2019*, English version, Oslo: Norwegian Ministry of Transport and Communications, available at: http://www.regjeringen.no/upload/SD/Vedlegg/NTP/Binder1ntp_engNY.pdf (accessed 5 May 2012).
- NYSSP (2008), *Sustainable Streets: Strategic Plan for the New York City Department of Transportation: 2008 and Beyond*, New York: New York City Department of Transportation.
- Stanley, J. and S. Barrett (2010), *Moving People: Solutions for a Growing Australia*, Canberra: Australasian Railways Association, Bus Industry Confederation and International Public Transport Association.
- Stanley, J. and A. Smith (2012), 'Workshop 3A: governance, contracting, ownership and competition issues in public transport: looking up not down', *Research in Transportation Economics*, **39** (1), 167–74.
- Sydec (2007), 'Long term transit expansion prospects', Commission Briefing Paper 4M-05, prepared for National Surface Transportation Policy and Revenue Study Commission, 16 May.

Index

- access restrictions 300
- accessibility 14, 119, 121, 183–7, 224, 235, 326
 - destination 318, 320
 - local 249
- accessibility elasticity 218
- accidents 154–62, 176–7, 204, 225, 295
- acid rain 137
- active transport 204
 - modes 165, 265
- administrative feasibility 224
- administrative simplicity 224
- advanced land use transport interaction models 213
- affect 124
- agglomeration 206–19
- agglomeration assessments 210
- agglomeration benefits 25, 211, 214, 221, 249, 310
- agglomeration economies 116, 206, 221, 235, 244, 250, 310, 317, 325
 - locational 320, 322
- agglomeration effects 184, 208
- agglomeration elasticities 211, 216
- agricultural wastes 201
- air pollution 29, 52, 62–3, 129, 136, 139, 141, 151, 180, 225, 244, 265, 303, 310, 314
 - costs 151
 - monetary evaluation 75
 - policies to reduce 140
- air quality 136–40, 203
- air rights 238
 - development 236
- airshed model 139
- alcohol
 - and safety 158
- allocation
 - efficient 28
- alternative fuels 146, 199, 200, 201–203
- alternatives
 - discovery of 38
- appraisal 48, 53–4
- area-based
 - congestion pricing schemes 233
- Arrow's impossibility theorem 32
- asset sales 240, 330
- average vehicle ridership 285
- axle mass loading 297
- beneficiary pays 224, 240, 330
 - approach 224
- benefit/cost ratio 64–5, 250–51
- benefits 64, 83
 - external 300
 - marginal private 69
- bicycle lanes 273–4
- bicycles 12, 273–4, 281
 - secure storage 281
- bicycling 141, 204, 274, 280–82, 322, 325, 329
- bi-fuel vehicles see dual-fuel vehicles
- bio-diesel 201
- bio-fuels 201
 - second-generation 201
- Blaise Pascal 10
- bottleneck 176, 178
- bottom-up 328
 - approach 287
- bounded rationality approach 38
- bridges
 - freight vehicle access to 297
- bus 10–12, 15–6, 41, 123–7, 172, 184, 188–9, 243, 245, 249–52, 254, 259
 - fixed-route 254
 - on-demand 255
- bus lanes 275
- bus priority 274, 329
 - lanes 190
- bus rapid transit 16, 73, 245, 255, 322, 329
- business case 61
- business models
 - social enterprise 128
- business user benefits
 - value of 212
- busways 245

- ul style="list-style-type: none; padding-left: 0;">
- butanol 201
- bypass lane
 - ramp metering 272
- CAFE regulations 203
- capabilities 24, 33, 120
- capability 124
- capacity 101, 102
 - increases 185
- car dependence 125
- car sharing 204
- carbon monoxide 136
- carcinogenic 138
- carpool
 - lanes 272
 - parking spaces 280
- carpool-matching programmes, 279
- cars 12
- catalytic converter 136
- catastrophic risk 84
- chain of responsibility legislation 302
- change moment 288
- charging system
 - road user 298
- city
 - size of 13
- clearways 270
- club goods 23
- coercive strategies 276
- cold
 - soak 138
 - start 138
 - start mode 168
- communities
 - inclusive 314
 - liveable 314
- community
 - buses 255
 - development approach, 287
 - severance 129
 - transport, 243
- compensating variation 67
- competitive
 - economies 314
 - tendering 259
- complementary benefits 262
- compressed work weeks 278
- computable general equilibrium 213
- conflict resolution 41
- congestion 1–2, 18, 23, 29, 40, 57, 69, 72, 92, 99, 101, 103, 105, 111–14, 170, 171, 175–6, 225, 230, 244, 248–50, 262, 265, 295–6, 310, 314
 - benefits 262
 - costs of 92
 - feedback component 227
 - incident-related 176
 - mitigation 190, 250
 - pricing 72, 187, 189–90, 232–4
 - pricing schemes, 233
 - recurrent 176
 - reduction 188, 204
 - spreading 188
 - tolerance for/of 187, 192
- consensus-seeking 44
- consumers' surplus 67, 217, 250
- consumption 124
- cordon-based
 - congestion pricing schemes 233
- cost
 - attribution model 231
 - marginal social 72
 - per destination 184
 - per kilometre of travel 184
 - recovery 82, 222, 225, 234, 250–51
 - savings public transport 258
- cost–benefit analysis 27, 32, 60, 65, 67, 81–2, 89, 214
- cost-effectiveness 27
- cost-effectiveness analysis 60, 76
- costs 64, 83
 - air pollution 151
 - attributable 82
 - average fixed 81
 - average private 72
 - average total 82
 - average variable 82
 - avoidable 82
 - congestion 111, 114, 175, 188, 190, 226, 230, 232, 252, 262, 295–6, 316
 - external 82
 - fixed 81
 - generalised 184
 - GHG emission 152
 - joint 82
 - marginal 82
 - marginal congestion 113
 - marginal external 82
 - marginal social 69, 72–3, 82, 222, 224, 227–8, 232–5, 298, 300

- mental health 127
 - noise 152
 - road accident 155
 - road damage 231
 - social 117
 - total 82
 - variable 82
 - welfare 127
- damaged vehicles
- removal of 177
- deadweight loss 26, 72
- death 154
- decibels 148
- A-scale 148
- decision making 44
- dedicated turn lanes 267
- delay 95
- demand 101
- cross-price elasticities of 68
 - derived 53, 107, 207, 311
 - income elasticity of 68
 - price elasticity of 68
- demand curve 68
- demand drivers 261
- density 318
- jam 96
 - traffic 96
- design
- street network 318
- design standards 40
- destination
- choice 108
 - accessibility 318, 320
- destinations
- location of 183
- developer contributions. *see* development impact fees
- development impact fees 237
- discount rate 82
- distance headway 95
- distance-based
- congestion pricing schemes 233
 - charging 230
- distributional considerations 224
- distributional impact
- of increasing fuel taxes 228
- distributional weights 65, 70
- diversity
- of land uses 318
- driver behaviour 157
- aggressive 159
- driving
- impairment from alcohol 159
- drugs
- and safety 158
- dual-fuel vehicles 201
- EA 135
- economic assessment 81
- economic competitiveness 294, 316
- economic evaluation 60–61, 63, 73, 172.
- economic performance
- of a city 206
- economic rent 71
- economically efficient resource allocation 222
- economics
- welfare 30
- economies of scale 208
- effective density 310
- EIS 135
- elasticities
- impact 318
- Embarcadero Freeway 133
- emissions evaluation processes 304
- employment densities 206
- employment density
- effective 210
 - relative effective 210
- encroachment 299
- energy crisis 197
- energy insecurity 244
- energy security 195, 314
- enforcement 161
- personal safety 163
- environment
- walking 321
- environmental
- assessment 134–5
 - compensation 89
 - impact assessment 63
 - impact statement 135
 - impacts 135
 - laws 132
 - legislation 134
 - significance 134
 - sustainability 294
- environmental performance
- vehicle 297

- ethanol 201
- Euro VI standards
 - for heavy-duty vehicles 303
- evaluation 5–6, 27–8, 31, 37–8, 42, 45, 48, 51, 53–4, 55–7, 64, 67, 75–7, 80–81, 84–6, 88–90, 117, 151, 172, 213, 215, 289, 304, 312
- economic 60–63
- external benefit/benefits 207, 247
- external impacts 121
- externalities 3–4, 22, 24–5, 82, 127, 132, 151, 154, 173, 227–9, 232, 236, 314
 - consumption 25
 - environmental 132, 151, 232
 - production 25
- externality-reducing tax 227
- financial deficit 250
- fleet fuel efficiency 191
- flexible work hours 276
- flow
 - maximum 100
 - traffic 96
- FONSI 135
- food
 - intake 165
 - production 201
- fossil fuel dependence 225
- fossil fuels 144, 195
- four-step modelling process 108
- free riders 288
- freight 18, 53, 152, 154, 161, 222, 232, 291–4, 299–302, 305–307, 317–8, 324, 325, 327
- freight movement task 293
- freight network
 - sustainable 318
- fuel
 - cells 202
 - economy standards 203
 - production 201
 - security 203
 - taxes 191, 223, 227, 314
- fundamental diagram of traffic flow 98
- funding
 - sustainable 318
- future vision 331
- general equilibrium benefit 217
- generalised cost 72, 102, 107, 122, 184, 190, 218, 247
- generalised cost elasticity 218
- generated traffic 218
- goal setting 61, 116
- goals 36, 38, 42, 45, 46, 49, 62, 331
 - social 116
- goal-setting approach 37
- government
 - assets sales 240, 330
 - borrowings 239
 - business enterprise 257
- GPS technologies 225
- green wave 268
- greenhouse gas emissions 29, 53, 111, 125, 136, 142–4, 146, 195, 203, 225, 234, 244, 314
- greenhouse gases 142
- growth
 - road traffic 292
- growth rate
 - freight 291
- guaranteed ride home programmes 280
- habit 111
- harmonised charging regimes 231
- health 314
- hearing damage 147
- high-level social goals 49, 50, 53
- high-occupancy-vehicle lanes 272
- high-speed rail/trains 12, 217
- horizontal integration 57
- hot soak 138
- hot start 138
- hybrid engine 202
- hybrid vehicles 142
- hydrocarbons 136
- hypothecation 222
- hypothetical compensation test 61
- impact pathway 75
- impacts
 - construction 136
- implementation 46, 330
- income
 - marginal utility of 62
- incrementalist approach 38
- Individual barriers 121
- information
 - diffusion 289
 - lack of 28
- injury 154

- input volume(s) 176, 178
- Institutional
 - barriers 121
 - restrictions 127
- integrated approach 57
- integrated land use/transport 125, 216–7, 309, 318, 322–30
- intelligent transport systems 177
- intergenerational equity 49, 79
- internal combustion engine 11
- internal scale economies 208
- intersection
 - control 268
- intersections 266
- jobs/work balance 320
- jobs/housing balance 321
- joint development 236, 238
- journey times
 - unreliability of 180
- kiss-and-ride 283
- KonSULT
 - project 55
 - tools 309
- land use 192
- land use and transport 219
- land use model 109
- land use/accessibility dimension 125
- land use/transport integration 240, 307, 310, 315, 322–30
 - see also* integrated land use/transport
- land value capture 235
- lane widths 271
- lanes
 - restripe 271
- length
 - vehicle 297
- level of service 103–105
- life
 - value 172
- link-based
 - congestion pricing schemes 233
- liveability 207–208, 316, 323–4, 330
- localisation economies 208
- locally preferred alternative 135
- locus of control 124
- low capital 266
- maglev 254
 - magnetic levitation vehicle. *see* maglev
- many-to-many demand patterns 255
- many-to-one transport function 253
- marginal social cost pricing 222, 224, 234, 281, 300
 - of road freight 232
- market
 - failure 73
 - failures 22, 29
- markets
 - limited extent of 27
- mass transit 247
- mechanical integrity 159
- memorandum of understanding 135
- merit goods 21, 24
- metropolitan improvement levy 237
- minimum service levels 126
- mixed scanning approach 45
- MOBILE 139
- mobile phones
 - and driving 159
- mobility 2, 14, 24, 47, 53, 116–30, 183–6, 192
 - increasing 317
 - preferences 187
 - realised 123
- mode choice 108
- model
 - normative 37
- models
 - decision making 35
 - governance 130
- monitoring 51, 54, 312
- monopolies 26
- monopoly 71
 - natural 26
- monorail 254
- MOVES 139
- multi-criteria analysis 60, 76, 90
- National Environmental Policy Act 134, 136
 - need 50, 120, 125–6, 128–30, 244, 246–7, 257, 311, 318
- negative affect 119
- negotiated
 - exactions 238
 - performance-based contracts 259
- neighbourhood disruption 151, 180
- NEPA. *See* National Environmental Policy Act

- network assignment 108
- niche markets 255
- no-build alternative 135
- noise 147, 180, 225, 265, 300, 310
 - aircraft 150
 - properties 148
 - reduction 149
 - reference levels 148
 - road 148
- noise levels
 - ambient 149
 - mitigating 150
- noise standards 149
- non-coercive strategies 286
- non-excludable 22
- non-rivalrous 22
- notice of intent 134
- obesity 154, 163
 - trends 164
- objectives 36–9, 42, 45–6, 48–51, 79, 128–9, 221–2, 227, 259, 294, 315, 317
- oil
 - consumption 198
 - deposits 196
 - reserves 197
 - supply 198
- oil extraction
 - economics of 198
- omnibus 10
- operating mode 138
- operating regimes
 - harmonised 305
- optimal fuel taxation 228
- optimal road user charges 225
- optimism bias 88, 239
- option benefits 237
- organisational process approach 40
- outcome indicators 53
- output
 - capacity 176, 178
 - flow rate 102
- overweight 163
- oxides of nitrogen 136
- oxides of sulphur 136
- ozone 137
- Pareto improvement 32, 61
- Pareto optimum 21, 69
- park-and-ride 283
- parking 282
 - removal 269
- parking charges
 - for commuters 282
- parking tax 282
- particulate matter 136
- passenger car equivalent 103, 296
- passenger-loading capacities 252
- peak oil 197–9, 301
 - non-existence theory 199
 - pessimistic theory 199, 200
 - plateau theory 199, 200
- peak spreading 188, 279
- pedestrian fatalities 160
- PER 135
- performance
 - transport system 183
- performance benchmarking 258
- personal safety
 - perception 162
 - policy directions 162
- PESTEL 46
- physical quantification 75
- Pigovian tax 227
- place-based approaches 128
- planning process 37
- pluralistic 44
- policy cycle 6, 46, 48, 51, 57, 60
- political bargaining approach 41
- pollutant exposure 167
 - mitigation 169
- pollutants 195
 - exposure to 154
- positive affect 119
- potential Pareto improvement 224
- power sharing 42
- PPM 20–8 133
- preference(s)
 - individual 24, 30, 54, 61–3, 65, 56, 77–8, 83–5, 90, 117
 - public 78
- price–volume curve 101
- pricing measures 187
- primary social goods 24, 33
- proactive, 39
- problem-simplifying 44
- procedural goals 50
- producers' surplus 70

- production function 209
- productivity 207
- programming and budgeting 41, 45
- property damage 154
- property rights
 - allocation of 182
- PROPOLIS 51
- proven oil reserves 196
- public environment report 135
- public good 22
- public
 - goods 21
 - involvement 134
 - participation 54, 57, 77, 90
- public transport 16–8, 29, 95, 117, 122, 125–7, 129–30, 140, 141, 162–3, 165–7, 179, 188–90, 200, 204, 213–14, 218–9, 223, 228, 234–5, 237, 246–7, 251–3, 255–8, 260–63, 265, 280–88, 315–8, 322, 325
- contracts 258
- fares 284
- future role of 261
- generalised cost of 190
- information 281
- minimum service levels 125
- mode shares 261
- public–private partnerships 239
- pure time preference 84
- qualitative measures 76
- rail 10–12, 15–6, 25–7, 42, 80, 84–5, 137, 150–51, 156, 169, 180–81, 213–8, 235–9, 243, 246, 248, 250–60, 279, 291–3, 296, 299–302, 305–306, 311, 317, 321–2, 324–7
 - freight vehicle access to 299
 - heavy 253
 - light 253
- ramp metering 272
- random breath testing 159
- rational actor approach 36, 45
- rationale 50
- reactive 39
- recurring congestion 178
 - causes 178
 - effects of 180
- regulation 127
- regulations
 - harmonisation 304
- relative equity weights 124
- resource-allocative 44
- resources
 - community mobility 128
- response times 177
- restrictions 127
- revealed preferences 73
- ride sharing 279
- risk 86
 - assessment 88
 - aversion 81
- risk factors
 - of social exclusion 118
- risk management approach 145
- risk management procedures 87
- risks
 - fuel 301
- road accidents
 - causes of 157
- road damage 225
- road environment 160
- road pricing 50, 55, 117, 190, 222–5, 227, 234–5, 240
- road rage 154, 170
- road user charge(s) 187, 283
- roads
 - freight vehicle access to 297
- roundabouts 268
- route choice 108
- safety 129, 244, 314
 - from accidents 154
 - personal 154, 161
- safety and health 294
- safety devices
 - in modern cars 159
- satisficing approach 37
- satisficing model 38
- secure bicycle storage 281
- self-sustaining
 - change 289
- sense of community 120, 123
- sensitivity testing 88
- service elasticity 261
- service levels
 - public transport 130
- showers and clothes lockers at work 281
- silos 127
- Smarter Travel 286

- social agglomeration
 - benefits 217
 - economies, 218
- social capital 119–20, 123–4, 127–9
- social enterprise 128
 - mobility-oriented 129
- social exclusion 24, 117, 119, 151, 180, 244, 249, 310, 314
 - risk of 123
- social goal 130
- social impacts 19
- social inclusion 127, 294
 - barriers to 117
 - benefits 237, 249
- social marketing approach 286
- social opportunity cost 85
- social time preference rate 83
- social transit 247
- social transit role 249
- social value 249, 300
- social welfare function(s) 29–34, 63
- solar 202
- sound walls 150
- space–time diagram 93
- spatial integration 51
- special assessments 236
- speed 97, 138
 - free-flow 97
 - space mean 97
 - time mean 97
- speed limits
 - appropriate 157
 - arbitrary 158
 - non-uniform 158
 - unrealistic 158
- speed–flow 112
- speeding 157
- speeds
 - 85th percentile 158
- stakeholders 47
- standard operating procedures 40
- standards
 - performance-based 298
- stated preferences 73
- station spacings 253
- strategic land use plans 300
- strategic plan
 - national freight 307
- strategic transport planning 43
- streetcar 253
- subsidising public transport fares 281
- supply 101
- supply curve 69
- supply–demand equilibrium 109
- sustainability 60–61, 63, 79, 89, 142, 146, 195, 207, 294, 306, 309, 317
- sustainable freight distribution
 - definition 294
- sustainable
 - funding 240
 - pricing 240
- systems analytic approach 36
- tax increment financing 236
- telecommuting 277
- Thredbo Conference Series 5, 257–9
- tidal flow lanes 270
- time
 - value of 74, 112, 215
- time headway 95
- TOD see transit-oriented development
- toll finance 315
- tolling technologies 225
- tolls 283, 330
- top-down 287, 328
- traffic
 - accidents 155
 - congestion 310
 - delays 180
 - flow 265
 - lanes 269
 - lights 268
 - noise 129
- traffic signals
 - area-wide control 269
 - bus phases 275
 - bus pre-emption 275
 - fixed-time 268
 - traffic-actuated 268
- tragedy of the commons 23, 182
- trains 12
- tram 253, 325
- transit-oriented development 238, 300, 321, 322
- transport
 - and land use 121
 - disadvantage 120, 121
 - system management 141

- transport systems
 - effective 314
- travel demand
 - reducing 317
- travel demand management 141
- travel forecasting
 - purpose 107
- TravelSmart 286
- travel-time budget 16
- trip distribution 108
- trip generation 108
- trip purpose 108
- trusted others 288
- trusting partnership 259
- UK WebTAG 211
- uncertainty 86, 88
- uncertainty-avoiding 45
- underground railway(s) 10, 12
- urban primacy 206
- urban settlements
 - efficient 207, 311
- urban sprawl 189, 203–204, 311, 327
- urbanisation 179
 - economies 209
 - efficient 206
- user
 - benefits 247
 - charges 330
 - pays 224, 240, 330
- utilitarianism 61
- utility 62, 107, 124
 - diminishing marginal 83
- valuation
 - monetary 73
 - non-monetary 75
- value
 - capture, 240, 330
 - existence 80
 - intrinsic 79
 - judgement, 30, 32, 33, 61, 63, 83, 117
 - non-use 80
 - option 81
 - total economic 81
 - user 80
- values 36
- variable message signs 177
- vehicle
 - solar-powered 202
- vehicle emission
 - reducing 317
- vehicle emission standards 139, 140
- vehicle excise duties 223
- vehicle miles travelled charging system 224
- vehicle utilisation 317
- vehicles
 - low-emission 191
 - zero-emission 191
- vehicular emissions 180
- vertical integration 51
- vibration 180
- volatile organic compounds 136
- voluntary travel behaviour change 141, 286–90
- voting 31
 - paradox 31
- walking 12, 141, 204, 317, 320–21, 322, 325, 328–9
- warm stabilised 138
- weight and exercise 164
- welfare 62
- welfare losses 64
- welfare-maximising revenue 227
- welfare-optimising framework 227
- well-being 118, 120, 127, 130
 - psychological 123
- wheel 9
- wider economic benefits 207, 212, 214
- width
 - vehicle 297
- willingness to accept 61, 67
- willingness to pay 61, 67, 76
- win–win policies 153, 167
- win–win strategies 151

Introduction to Transport Policy

This comprehensive and accessible textbook introduces the basic concepts of transport policy and decision-making to students of transport policy, transport planning, urban transport, transport evaluation and public policy.

It presents the foundations and rationale of transport policy, incorporating a review of the policy formulation process and models of decision-making appropriate to public sector policy-makers.

Topics covered include:

- The basics of transport planning and traffic theory deemed necessary to understand policy implications of issues including congestion, safety and parking.
- Potential solutions to problems such as road user charges, travel demand management, voluntary travel behaviour change, transport system management and public transport investment.
- Prescriptions for technological change.
- Discussion of the need for an integrated land transport policy along with a case study to illustrate how this might be developed for a typical metropolitan area.

Peter Stopher and **John Stanley** are in the Institute of Transport and Logistics Studies at the University of Sydney, Australia.

EDWARD ELGAR
A FAMILY BUSINESS IN INTERNATIONAL PUBLISHING

The Lypiatts, 15 Lansdown Road, Cheltenham, Glos GL50 2JA, UK
Tel: + 44 (0) 1242 226934 Fax: + 44 (0) 1242 262111 Email: info@e-elgar.co.uk
William Pratt House, 9 Dewey Court, Northampton, MA 01060, USA
Tel: +1 413 584 5551 Fax: +1 413 584 9933 Email: elgarinfo@e-elgar.com
www.e-elgar.com www.elgaronline.com

KQ-987-115



ISBN 978-1-78195-244-3



9 781781 952443