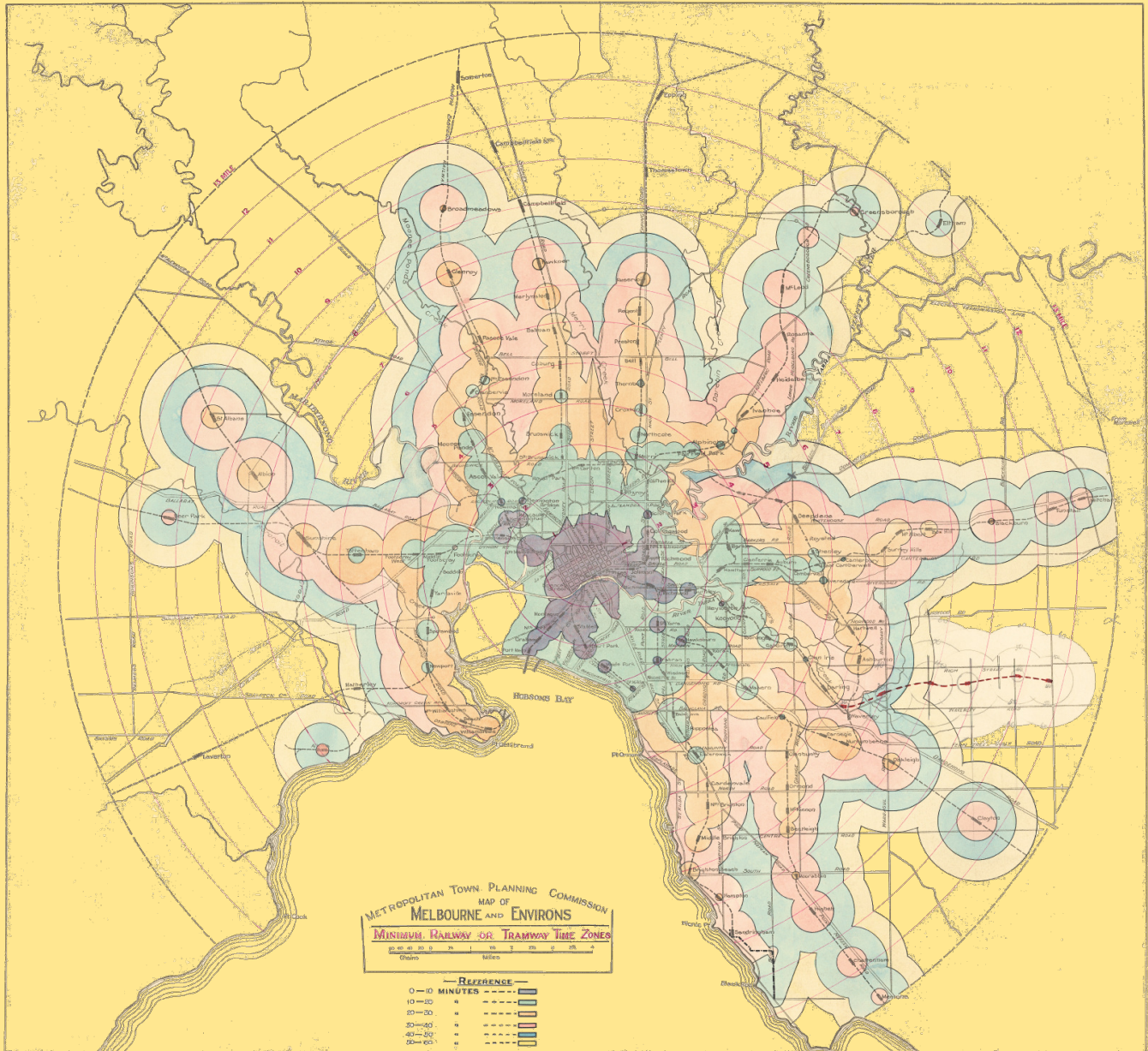


Elements of Access

Transport Planning for Engineers • Transport Engineering for Planners



David M. Levinson
Wes Marshall • Kay Axhausen

DAVID M LEVINSON,
WES MARSHALL, AND
KAY AXHAUSEN

ELEMENTS OF ACCESS

NETWORK DESIGN LAB

Copyright © 2017 David M Levinson,
Wes Marshall, and
Kay Axhausen

PUBLISHED BY NETWORK DESIGN LAB

<https://tufte-latex.github.io/tufte-latex/>.

Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at <http://www.apache.org/licenses/LICENSE-2.0>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

First printing, December 2017

The cover is Metropolitan Town Planning Commission: Map of Melbourne and Environs. Map restoration by Cameron Booth www.cambooth.net. Original from State Library of Victoria.

Contents

I	Introduction	21
1	Elemental Accessibility	23
1.1	Isochrone	26
1.2	Rings of Opportunity	27
1.3	Metropolitan Average Accessibility	29
II	The People	31
2	People	33
2.1	Stages, Trips, Journeys, and Tours	37
2.2	The Daily Schedule	41
2.3	Coordination	42
2.4	Diurnal Curve	44
2.5	Travel Time	47
2.6	Travel Time Distribution	49
2.7	Social Interactions	50
2.8	Activity Space	52
2.9	Space-time Prism	53
2.10	Choice	55
2.11	Principle of Least Effort	57
2.12	Capability	59
2.13	Observation Paradox	61
2.14	Capacity is Relative	62
2.15	Time Perception	63
2.16	Time, Space, & Happiness	65
2.17	Risk Compensation	67
III	The Places	69
3	The Transect	71
3.1	Residential Density	73
3.2	Urban Population Densities	76
3.3	Pedestrian City	77

3.4	Neighborhood Unit	80
3.5	Bicycle City	82
3.6	Bicycle Networks	84
3.7	Transit City	86
3.8	Walkshed	91
3.9	Automobile City	93
4	Markets and Networks	95
4.1	Serendipity and Interaction	98
4.2	The Value of Interaction	100
4.3	Firm-Firm Interactions	102
4.4	Labor Markets and Labor Networks	104
4.5	Wasteful Commute	106
4.6	Job/Worker Balance	109
4.7	Spatial Mismatch	111
IV	The Plexus	113
5	Queueing	115
5.1	Deterministic Queues	118
5.2	Stochastic Queues	120
5.3	Platooning	122
5.4	Incidents	124
5.5	Just-in-time	125
6	Traffic	127
6.1	Flow	129
6.2	Flow Maps	131
6.3	Flux	132
6.4	Traffic Density	134
6.5	Level of Service	135
6.6	Speed	136
6.7	Shockwaves	138
6.8	Ramp Metering	140
6.9	Highway Capacity	141
6.10	High-Occupancy	143
6.11	Snow Business	145
6.12	Macroscopic Fundamental Diagram	147
6.13	Metropolitan Fundamental Diagram	148
7	Streets and Highways	149
7.1	Highways	151
7.2	Boulevards	153
7.3	Street Furniture	156

7.4	Signs, Signals, and Markings	158
7.5	Junctions	160
7.6	Conflicts	161
7.7	Conflict Points	163
7.8	Roundabouts	165
7.9	Complete Streets	169
7.10	Dedicated Spaces	171
7.11	Shared Space	172
7.12	Spontaneous Priority	174
7.13	Directionality	176
7.14	Lanes	180
7.15	Vertical Separations	183
7.16	Parking Capacity	185
8	Modalities	187
8.1	Mode Shares	190
8.2	First and Last Mile	192
8.3	Park-and-Ride	194
8.4	Line-haul	196
8.5	Timetables	197
8.6	Bus Bunching	198
8.7	Fares	200
8.8	Transit Capacity	202
8.9	Modal Magnitudes	204
9	Routing	205
9.1	Conservation	207
9.2	Equilibrium	209
9.3	Reliability	211
9.4	Price of Anarchy	213
9.5	The Braess Paradox	215
9.6	Rationing	218
9.7	Pricing	220
10	Network Topology	223
10.1	Graph	227
10.2	Hierarchy	229
10.3	Degree	231
10.4	Betweenness	233
10.5	Clustering	235
10.6	Meshedness	237
10.7	Treeness	239
10.8	Resilience	240
10.9	Circuitry	242

11	Geometries	245
11.1	Grid	247
11.2	Block Sizes	249
11.3	Hex	251
11.4	Ring-Radial	253
V	The Production	255
12	Supply and Demand	257
12.1	Induced Demand	260
12.2	Induced Supply & Value Capture	262
12.3	Cost Perception	264
12.4	Externalities	266
12.5	Lifecycle Costing	268
12.6	Affordability	269
13	Synergies	271
13.1	Economies of Scale	272
13.2	Containerization	273
13.3	Economies of Scope	275
13.4	Network Economies	277
13.5	Intertechnology Effects	279
13.6	Economies of Agglomeration	280
13.7	Economies of Amenity	282
VI	The Progress	285
14	Lifecycle Dynamics	287
14.1	Technology Substitutes for Proximity	289
14.2	Conurbation	291
14.3	Megaregions	294
14.4	Path Dependence	296
14.5	Urban Scaffolding	299
14.6	Modularity	301
14.7	Network Origami	303
14.8	Volatility Begets Stability	305
15	Our Autonomous Future	307

List of Figures

1	Creative Commons License	15
1.1	Access across America by automobile	23
1.2	Transit accessibility in Minneapolis - St. Paul region	25
1.3	Melbourne isochrone	26
1.4	Rings of opportunity	27
1.5	Accessibility over time	29
2.1	Shibuya	33
2.2	Travel demand model	36
2.3	Trips	37
2.4	Trips with tours	40
2.5	A day in the life	41
2.6	Minnesota State Fair	42
2.7	Diurnal curves on the London Underground	44
2.8	Travel time budget	47
2.9	Travel time distribution	49
2.10	Distribution of social interactions	50
2.11	Activity space	52
2.12	Space-time prism	53
2.13	Minnesota Valley National Wildlife Refuge	55
2.14	Alternative routes for travelers	57
2.15	Disability parking space	59
2.16	Multimodal Beijing	61
2.17	Tokyo subway pushers	62
2.18	Traffic simulation	63
2.19	Happiness	65
2.20	Happiness	66
2.21	Helmeted bicyclists	67
3.1	Transect	71
3.2	Humboldt's transect	72
3.3	Accessibility, mobility, and density	73
3.4	Density declines with distance	76
3.5	Area inside Roman walls	77

3.6	Town Plan of Radburn	80
3.7	Map of Houten	82
3.8	Protected bike lane in Vancouver	84
3.9	Low-stress island	85
3.10	Streetcar	86
3.11	Walkshed	91
3.12	Route of Metro Transit 87 Bus	92
3.13	Map of Southwest Corridor LRT, Minnesota	93
4.1	Farmer's market at Marylebone, London	95
4.2	Pedestrian street in Singapore	98
4.3	Happy family statue	100
4.4	Supplier - firm - customer flowchart	102
4.5	Montgomery County Planning Department headquarters	104
4.6	Bicyclist in Cergy-Pontoise	106
4.7	Ratio of job and labor accessibilities	109
4.8	Labor accessibility by transit in Minneapolis - St. Paul region	111
5.1	Queueing on Franklin Avenue, Minneapolis	115
5.2	Deterministic queues	118
5.3	Stochastic queue	120
5.4	Signal schedule	122
5.5	Traffic incident	124
5.6	Replica of early UPS truck	125
6.1	Microscopic fundamental diagram	127
6.2	I-94 Minneapolis	129
6.3	Flow maps	131
6.4	Delft woonerf	132
6.5	Red Queen	134
6.6	Freeway level of service	135
6.7	Charlotte Motor Speedway	136
6.8	Shockwaves without bottlenecks	138
6.9	Shockwave	139
6.10	Ramp meter	140
6.11	I-94 Mississippi River Bridge	141
6.12	Single occupant vehicle driver	143
6.13	Snow business	145
6.14	Macroscopic fundamental diagram	147
6.15	Metropolitan fundamental diagram	148
7.1	Street in New York City	149

7.2	'Stroad' in Detroit	150
7.3	Bobby Jones Expressway in Augusta	151
7.4	Highways	152
7.5	Octavia Boulevard in San Francisco	153
7.6	Bike boulevard in Sydney	154
7.7	Sidewalk and boulevard, Minneapolis	155
7.8	Parklet in Long Beach	156
7.9	Street furniture	157
7.10	Signs	158
7.11	Protected bike intersection	160
7.12	Bank Junction, London	160
7.13	Conflicts	161
7.14	Conflict points	163
7.15	Additional illustration of conflict points	164
7.16	Rotary intersection in Bourne	165
7.17	Roundabout in Vancouver	166
7.18	Neighborhood traffic circle in Berkeley	167
7.19	Complete street in Sydney	169
7.20	TX-366 in Dallas	171
7.21	Denver Pedestrian Mall	171
7.22	Boston bicyclist	171
7.23	Conventional design v. shared space	172
7.24	Spontaneous priority	174
7.25	One-way street in Evanston	176
7.26	Comparison of intersection conflict points	178
7.27	One-way to two-way conversion in Denver	179
7.28	One-way collision in Denver	179
7.29	Narrow lanes in Cambridge	180
7.30	Acorn Street in Boston	182
7.31	City of Tomorrow	183
7.32	Skyway in Atlanta	184
7.33	Tunnel in Boulder	184
7.34	Parking prices in Brunswick Centre, London	185
8.1	Brisbane Tram	187
8.2	Mode typology	189
8.3	Commute mode share in Minneapolis - St. Paul	190
8.4	Map of ratio of transit to auto accessibility	191
8.5	eTuk vehicles, Denver	192
8.6	Park-and-ride lot	194
8.7	Line-haul cost breakdown	196
8.8	Timetable	197
8.9	Bus queue	198

8.10	Fare map of Boston	200
8.11	Metro Transit Green Line train	202
8.12	I-94 traffic counts	204
9.1	Sydney by air	205
9.2	Parking enforcement	207
9.3	Two route equilibrium	209
9.4	I-35W Mississippi River Bridge	211
9.5	Price of anarchy	213
9.6	Macy's Broadway, New York	215
9.7	Braess' paradox	216
9.8	Toll plaza, Golden Gate Bridge	218
9.9	Marginal cost pricing	220
10.1	Representations of Metropolis	223
10.2	Cartoon of Metropolis	227
10.3	Hierarchy	229
10.4	Degree representations of Metropolis	231
10.5	Betweenness	233
10.6	Clustering	235
10.7	Meshedness	237
10.8	Downed tree	238
10.9	Treeness in the Twin Cities	239
10.10	Resilience	240
10.11	Circuitry	242
11.1	Boston transit network	245
11.2	A section of the Manhattan grid	247
11.3	City street networks	249
11.4	More street networks	250
11.5	Hexagonal networks	251
11.6	Glinks chessboard	252
11.7	Street network in Moscow.	253
12.1	Consumer and producer surplus	257
12.2	Induced demand	260
12.3	Value capture flowchart	262
12.4	Toll Booth in Guilin, China	264
12.5	Smog in Beijing, China	266
12.6	Cashflow	268
12.7	Waiting for the bus	269
12.8	Household income spent on transport	270
13.1	World Wildlife Zoo train	272

13.2	Containerization	273
13.3	Train station in Tokyo	275
13.4	Route map, Eastern Airlines	277
13.5	Highway Patrol Service	279
13.6	Skyline in Hong Kong	280
13.7	Chicken restaurants in Shanghai	282
14.1	Ramp meter deployment curve	287
14.2	St. Anthony Falls in Minneapolis	289
14.3	Hastings	291
14.4	Commuting networks	294
14.5	Shortest path in Singapore	296
14.6	Train station entrance in Arnhem	299
14.7	Bricks in the Road	301
14.8	Network origami	303
14.9	Komatsu bulldozer	305
15.1	Our autonomous future	307

List of Tables

3.1	Residential densities	74
3.2	Road coverage statistics	94
4.1	Spatial interactions and networks	97
4.2	Strategies for addressing spatial mismatch	112
6.1	Highway capacity	142
6.2	Elasticity of demand and crashes with respect to snow	146
8.1	Transit capacity	203
9.1	Price of anarchy in the Minneapolis - St. Paul region	214
13.1	Different types of synergies	271
13.2	Sources of agglomeration economies	281
14.1	County seats in Minneapolis - Saint Paul region . .	293

Copyright

©David M. Levinson, Wes Marshall, Kay Axhausen

Licensed Under: Creative Commons Attribution-NonCommercial 4.0 International You are free to:

- **SHARE** — copy and redistribute the material in any medium or format
- **ADAPT** — remix, transform, and build upon the material

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

- **ATTRIBUTION** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **NONCOMMERCIAL** — You may not use the material for commercial purposes.
- **NO ADDITIONAL RESTRICTIONS** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation. No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

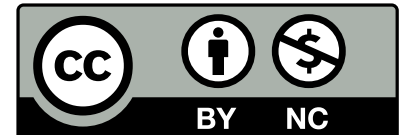


Figure 1: Creative Commons License <https://creativecommons.org/licenses/by-nc/4.0/>

Acknowledgments

We gratefully acknowledge the careful review of an earlier version of the text by: Renato Arbex, Russ Degnan, Alexa Delbosc, Damilola George, Eric Goldwyn, Taina Haapamäki, Bahman Lahoorpoor, Sarah Jo Peterson, Andrew Saffrey, Kylie Segaert, Hibai Unzueta, and Steve Wilson. While we have generally incorporated their input, we remain responsible for all opinions and errors.

We appreciate figures created for this book by Geoff Boeing and Cameron Booth.

Foreword: Nothing in Cities Makes Sense Except in the Light of Accessibility.

“Nothing in Biology Makes Sense Except in the Light of Evolution.” – Theodore Dobzhansky.¹

¹ (Dobzhansky 1973).

Cities, like organisms, take on unique forms that are shaped by millions of factors, but most of those factors relate to accessibility. The location of activities in space depends almost entirely on the location of other activities in space, and how easily they can be reached.

So we adapt the famous Dobzhansky quote:

“Nothing in Cities Makes Sense Except in the Light of Accessibility.”

Transport cannot be understood without reference to the location of activities (land use), and vice versa. To understand one requires understanding the other. However, for a variety of historical reasons, transport and land use are quite divorced in practice. Typical transport engineers only touch land use planning courses once at most, and only then if they attend graduate school. Land use planners understand transport the way everyone does, from the perspective of the traveler, not of the system, and are seldom exposed to transport aside from, at best, a lone course in graduate school.

This text aims to bridge the chasm, helping engineers understand the elements of access that are associated not only with traffic, but also with human behavior and activity location, and helping planners understand the technology underlying transport engineering, the processes, equations, and logic that make up the transport half of the accessibility measure. It aims to help both communicate accessibility to the public.

This book unpacks the idea of accessibility, introduced in the first part: [Elemental Access](#),² into its constituent elements. We group

² Part I.

³ Part II.

⁴ Part V.

⁵ (Levinson and Krizek 2007).

⁶ Part III.

⁷ Part IV.

⁸ Part VI.

these into: [The People](#),³ about human behavior; [The Production](#),⁴ about economics; and, adapting the title of another book,⁵ [The Places](#) (the land use)⁶ and [The Plexus](#) (the network),⁷ respectively. The final part of the book: [The Progress](#)⁸ examines the dynamic co-evolution of place and plexus over time.

While for pedagogical reasons, we try to make these distinctions crisp before muddying them up (it is easier to go from clear thinking to a muddle than the other way), we also note there is a lot of fuzziness and interaction between the transport and the land use elements of the equation. We have designed the book as a hypertext as much as a linear narrative, so readers can drift back and forth between the two. While there are not quite an infinite number of paths through the book, there are many. We have allocated sections to chapters and chapters to parts. For the most part, these are straightforward. In some cases, they have the scent of the arbitrary bush. So be it.

There are more topics that could be included but aren't. Concision is at war with completeness, and the more time we spent writing before releasing, the less likely you would be reading it now. Additional topics may be included in future editions.

Most readers will want to read from beginning towards the end, and that is probably the easiest, but we encourage you to jump around. One of our reviewers suggested that you can read it from front to back if you are a planner and from back to front if you are an engineer. This is an interesting thought, but not quite where we are; we want both planners, engineers, and anyone else reading to know everything and not give up on the unfamiliar terrain.

The equations have, for the most part, been confined to margin notes, which are available to read for the more engaged reader, but can be skipped or glanced at by others. The margin notes also contain footnotes, links to references, and hyperlinks to other parts of the text, as well as some illustrative figures. The space is busy but worthy of attention when more detail is wanted.

The layout of the book is inspired by Edward Tufte's books, which like this are information dense. It was laid out in LaTeX by the authors using the templates listed earlier in the Colophon.

– David Levinson, Wes Marshall, and Kay Axhausen,
December 2017

Part I

Introduction

1

Elemental Accessibility

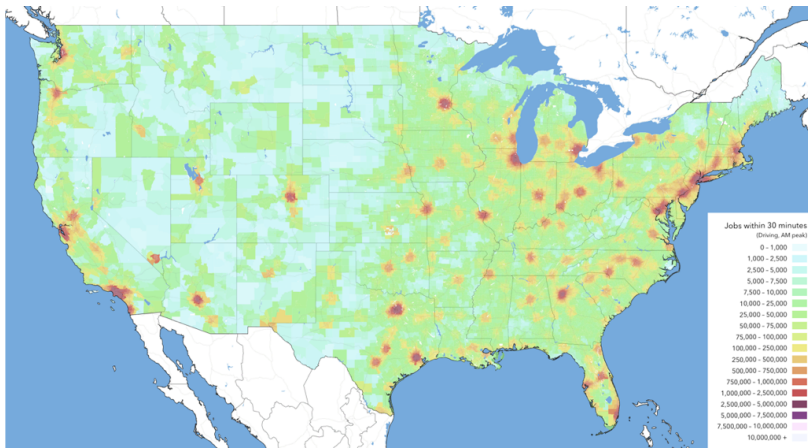


Figure 1.1: Access across America by automobile. Source: Accessibility Observatory.

WHAT MAKES FOR A GOOD TRANSPORT SYSTEM?

Most people answer that question by talking about moving people and goods between point *A* and point *B*. If moving people and goods is the goal, then improving transport typically means increasing performance measures such as road **speed**¹ and **capacity**² – or what we call increasing **mobility**.³

But is that really our goal?

In most situations, the answer is ‘no.’ The underlying goal of transport is instead about getting people to activities and opportunities like work, school, shopping, restaurants, medical care, parks, concerts – and getting goods to market – or what we call ‘accessibility.’

Accessibility, as we mean it, measures the ease of reaching destinations.⁴ We define it more formally as the mathematical product of the number and quality of destinations that can be reached, and the general cost (combining time, money, and other

¹ §6.6.

² §6.9.

³ §6.5.

⁴ The idea is attributed to Hansen (1959).

factors) of reaching them.

- ⁵§1.2. More specifically, a **cumulative opportunities**⁵ accessibility measure evaluates the number of destinations that can be reached in a given time threshold. For instance, Figure 1.1 shows accessibility for the contiguous United States, by car, at 8 am, on Wednesdays in 2015 using GPS traffic speed data and employment data for every Census block, with jobs as the destination and 30 minutes as the threshold. Places where more jobs can be reached in 30 minutes are darker red (or even purple in some large cities); places with few jobs are light blue. In a sense, this is taking the 30-minute **isochrone**,⁶ summing the jobs in the isochrone, and doing this for all eleven million census blocks.
- ⁶§1.1.

Jobs are not the only important destinations, so accessibility can be computed for different places as well as for different modes, on different days, in different years. If computed consistently with a cumulative opportunities measure, accessibility can be compared and can help explain observations that land values and productivity are higher where access is higher, and travel times are lower. Transit mode share is higher where transit accessibility is relatively closer to auto accessibility.⁷ Since one aim of transport is connecting people with places – rather than simply moving quickly – accessibility, which considers both mobility and land development patterns, is far more valuable than many performance metrics traditionally used by state and local departments of transport.

⁷(Owen and Levinson 2015).

The map of Figure 1.2 compares the area accessible by transit in the Minneapolis-St. Paul region.

If we are looking for a place to live, these maps tell us where we might be able to work, shop, and seek entertainment via transit. Transport is also about connecting people with people. Thus, accessibility maps can also tell us about our potential dating pool.

If we are locating a business, maps like these can shed light on the geography from which we might be able to attract employees or customers. This isn't to say that St. Paul would be a poor choice for locating a business compared with Minneapolis. The point is that understanding accessibility can inform our decision-making and help identify the tradeoffs when locating our businesses and our homes.

Accessibility is about more than accessing destinations; it brings us freedom, independence, and opportunity. *Elements of Access* takes the core of these accessibility calculations – some of the fundamental ideas in transport – and examines them in depth. Which destinations? Why certain thresholds? What travel times?

For example, seemingly simple concepts such as travel time have

different meanings at different scales. The traffic engineer measures **speed**⁸ on links, but this is only part of the story. It needs to be coupled with an understanding of human reactions, such as **induced demand**,⁹ that can offset our expected gains. It needs to consider the perspective of the network planner, who considers the **topology**¹⁰ and **directness**¹¹ of the network. There are many other interrelated questions, but we can only find the broader solutions we seek when we first understand the fundamentals this book presents and then ask the right questions.

⁸ §6.6.

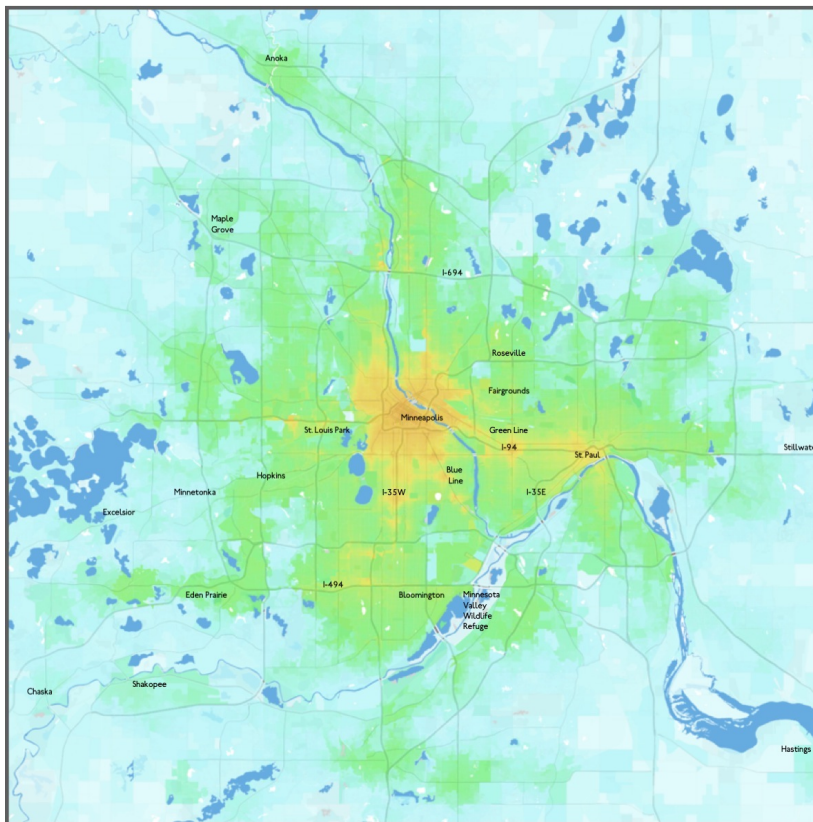
⁹ §12.1.

¹⁰ §10.

¹¹ §10.9.

Minneapolis

Minneapolis-St. Paul-Bloomington, MN-WI



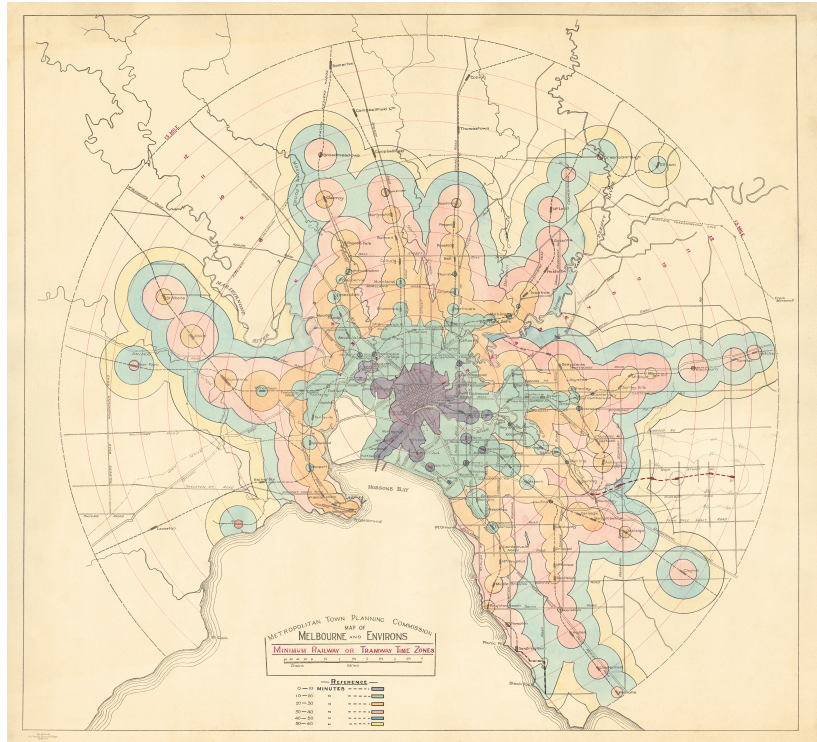
Jobs within 30 minutes by transit, averaged 7 - 9 AM



Figure 1.2: Transit accessibility in Minneapolis - St. Paul region. Source: Accessibility Observatory. Annotated with places mentioned in book.

1.1 Isochrone

Figure 1.3: Metropolitan Town Planning Commission map of Melbourne and environs. Minimum railway or tramway travel time isochrones for journeys to a common central point. Map restoration by Cameron Booth www.cambooth.net. Original from State Library of Victoria.



TRAVEL TIMES INCREASE WITH DISTANCE.

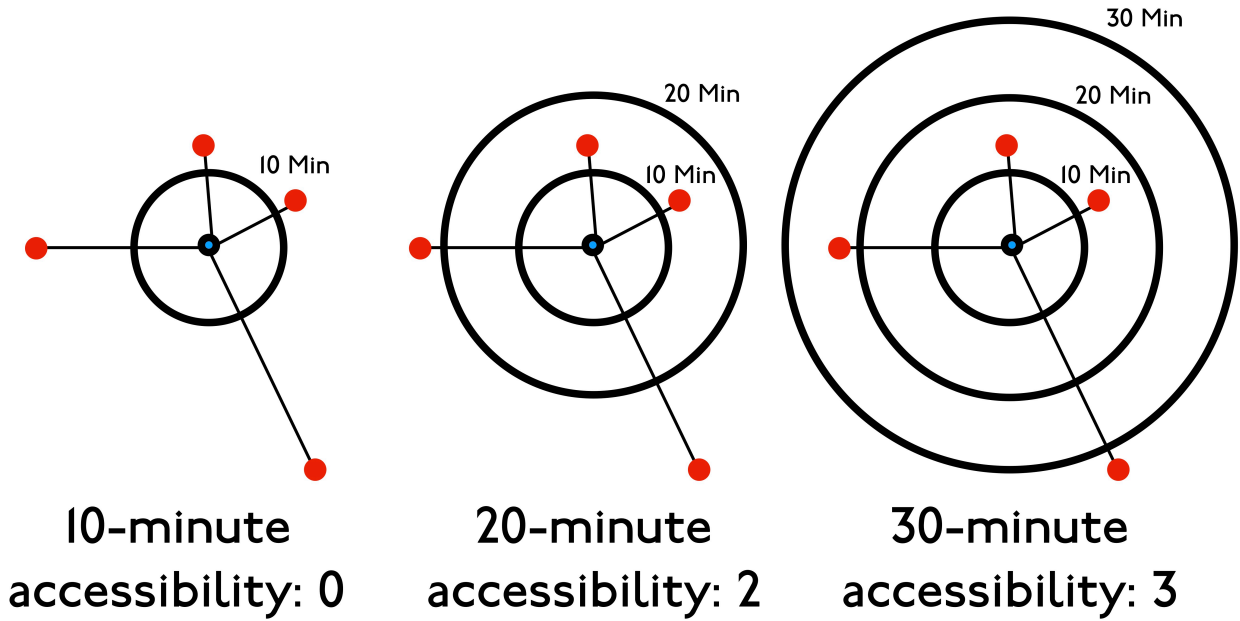
Figure 1.3 for Melbourne, Australia c. 1925, shows how travel times increase for the local public transport network in what is formally called an ‘isochrone’ (‘iso’ meaning ‘same’, ‘chrone’ meaning ‘time’). Areas that are near-in can reach the center quickly. Areas far from the center take longer. Transport networks distort the uniformity of this arrangement by increasing speeds in selected corridors. Even then, stations farther away take longer, and people who live farther from stations take more time than those who are adjacent to stations because of [access costs](#).¹² This map was obviously drawn by hand, and simplifies by assuming people can walk in a [straight line](#)¹³ from their home to the station. If they are walking along a [grid](#)¹⁴ of streets, those circles around the transit stops would be much more diamond like.

¹² §8.2.

¹³ §10.9.

¹⁴ §11.1.

1.2 Rings of Opportunity



THERE ARE NUMEROUS MEASURES OF ACCESSIBILITY, SOME OF WHICH ARE MORE COMPLICATED THAN OTHERS.

The simplest, a binary measure of access, asks: is *A* directly connected to *B*? The answer is either ‘yes’ or ‘no’ (1 or 0).

We can look a bit deeper: can *A* reach *B* on the network (or graph)¹⁵ (even if it has to pass through *C*, *D*, or *E*)? The answer is still ‘yes’ or ‘no’ but now more likely to be ‘yes.’ You can solve the second if you know the first - since if *A* is connected to *C*, and *C* is connected to *D*, and *D* is connected to *E*, and *E* is connected to *B* - then we deduce *A* can reach *B*.

The **cumulative opportunities measure**¹⁶ asks: can *A* reach *B* within a given time threshold? This is slightly more sophisticated but can still be directly measured if we now weight our network graph with the travel cost of each link. Travel cost typically means travel time, but it can be a more general cost that combines time and money and other factors. First we ask: is the time on link *AC*, plus the time on link *CD*, plus the time on link *DE*, plus the time on link *EB*, less than time threshold *T*?

Then we sum it up at each origin for all destinations of interest. The destinations of interest might be jobs, grocery stores, schools, or

Figure 1.4: Rings of opportunity. The figure and a more mathematical discussion of accessibility can be found in (Levinson and Krizek 2017).

A measure of accessibility to employment (for a given origin *i*) can be represented as follows:

$$A_i = \sum_j E_j f(C_{ij}) \quad (1.1)$$

where:
A_i = Accessibility to employment from origin *i*
E_j = Employment at destination *j*
f(C_{ij}) = function of the travel cost (time and money) between *i* and *j* (the higher the cost, the less the weight given to the employment location)

¹⁵ §10.1.

¹⁶ §1.

In cumulative opportunities measures:

$$f(C_{ij}) = 1 \text{ if } C_{ij} < T, 0 \text{ otherwise} \quad (1.2)$$

We can be more sophisticated still and weight the rings differently.

Destinations 10 minutes away are more valuable than equivalent destinations 30 minutes away because people generally dislike travel. The $f(C_{ij})$ can be a distance decay (§3.2) function, like:

$$f(C_{ij}) = e^{-\theta C_{ij}} \quad (1.3)$$

This is typically referred to as a gravity model of accessibility, where theta (θ) is an empirically estimated weight (with a value less than 0) that effectively discounts the value of destinations by their increasing travel cost.

¹⁷ §1.1.

¹⁸ Many jobs do not follow the traditional 9 to 5 schedule, and so may be inaccessible by peak-serving transit.

airports.

This is illustrated in Figure 1.4. In this case, there are three measures of accessibility: one at a 10-minute isochrone,¹⁷ one at 20, and the last at 30. We add up the number of dots in each ring and that is the cumulative opportunity measure for that time threshold for the origin.

Clearly, accessibility varies by location, by time of day,¹⁸ by day of week, by season, by purpose of trip, by type of destinations of interest, by mode of travel, and by individual, among other things. But even basic accessibility metrics like cumulative opportunity provide strong explanations for house values, wages, and travel behavior.

1.3 Metropolitan Average Accessibility

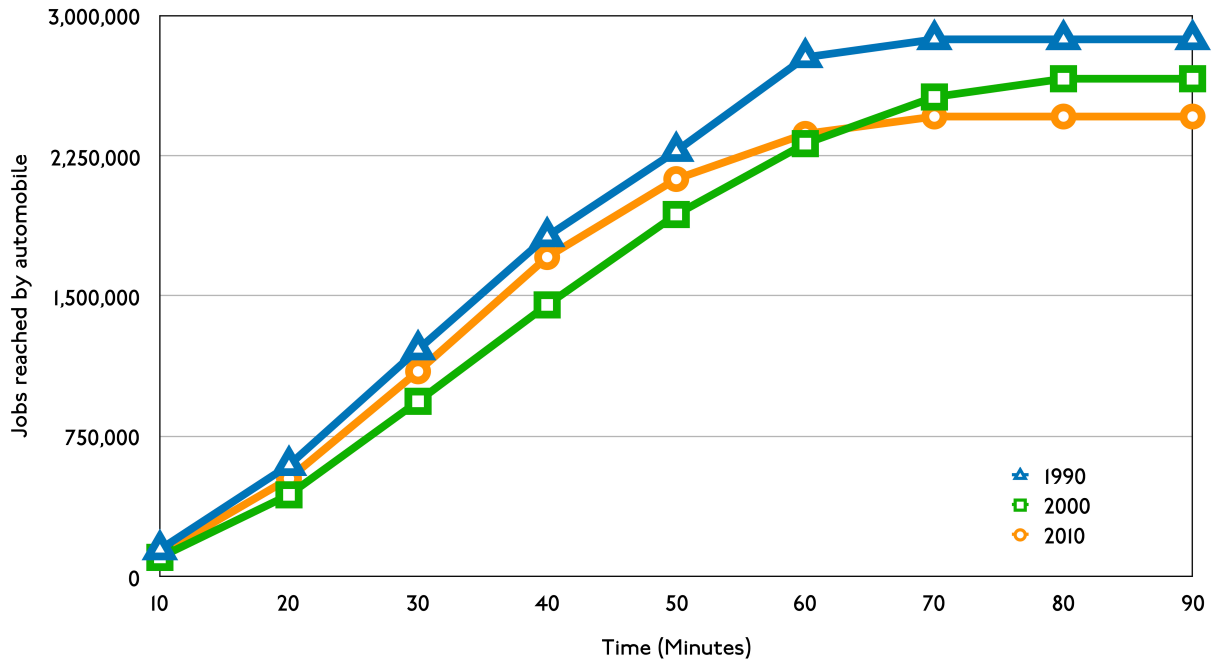


Figure 1.5: Accessibility over time in the US. Source: (Levinson 2013).

OFTEN DETAILED GEOGRAPHIC LAND USE OR NETWORK INFORMATION IS NOT AVAILABLE, ESPECIALLY IF ONE WANTS TO LOOK BACKWARD IN TIME.

Metropolitan Accessibility¹⁹ can be approximated using a variant of the classic formula of the area of a circle: ($A = \pi r^2$).²⁰

Typically when we think of the area of a circle, we think of radius as a distance. But in accessibility, the relevant question is a time radius (minutes), which measures how far you can travel in a given unit of time. So we need to convert this time into a distance, which depends on the speed of the network and its *circuity*²¹ (both of which vary for each origin-destination pair). Accessibility²² is also not a simple area, but rather a number of opportunities (say jobs), so the area is multiplied by the employment *density*²³ in that area.

For metropolitan areas, this macroscopic accessibility measure tends to underestimate accessibility at long time thresholds and overestimate accessibility for short thresholds compared to a more microscopic, geographically accurate analysis. This bias is because average speeds are too high for short trips and too low for long trips, and because this measure ignores job opportunities outside

¹⁹ §1.

²⁰ (Levinson 2012).

²¹ §10.9.

²² §1.

²³ §3.1.

Metropolitan area accessibility can be represented as:

$$a_t = \pi \left[\frac{V_n t}{C_t} \right]^2 \rho_{emp} \quad (1.4)$$

where:

- ρ_{emp} = Urban area employment density (jobs per km^2).
- t = time threshold.
- V_n = Average network velocity in km per hour.
- C_t = Average circuitry.

the finite boundary of the metropolitan area. The measures are comparable at a time threshold just above 30 minutes.

Part II

The People

Accessibility begins with people. If people didn't have places to go, things to do, or people to see, then very little of what we talk about in this book would matter all that much. Since reaching goods, services, and activities clearly matter a great deal to people, then understanding where things are and how transport can be used to access them is important. While many factors impact access to such activities, we first need to understand the people themselves. This includes digging into our daily schedules, figuring out why we make the choices we do, as well as recognizing our perceptions, capabilities, and limitations. This part of this book – The People – does just that.

2

Modeling People



Figure 2.1: Shibuya, Tokyo, Japan. Photo by D. Levinson

‘ALL MODELS ARE WRONG, SOME MODELS ARE USEFUL’ – GEORGE BOX

In the 1950s, the transport community led by Douglass Carroll in Detroit, and then Chicago, taking advantage of a new generation of mainframe computers, developed what is now referred to as the ‘travel demand model’ or ‘urban transportation planning model’ or the ‘four-step model’ of trip generation, trip distribution, mode split, and traffic assignment as shown in Figure 2.2.¹ The aim at the time was to develop forecasts of the behavior of people, and in particular their future traffic – how many trips, where are they going, how many would drive, which routes would they use. These

¹ The history of this can be found in (Boyce and Williams 2015).

forecasts could be used to locate and size freeways being deployed with the upcoming Interstate Highway System. In one sense it was enormously successful, as the model spread from the Midwest of the United States across the globe, and has been used to conduct analyses, inform, and justify projects. The early application of these models gave rise to the idea of [accessibility](#),² following modeler Walter Hansen's 1959 paper.³ It also spurred enormous methodological advances, one of which earned Daniel McFadden a Nobel Prize in economics for his work on developing random utility choice models.⁴

² (§1).

³ (Hansen 1959).

⁴ See travel demand applications in (Domencich and McFadden 1975).

The models turned out to be not terribly accurate. While one can understand the naïveté of early modelers in the 1950s and 1960s (who undoubtedly well understood the limitations), by the 1970s (and certainly by the 2010s), the futility of accurate forecasting should have become apparent to those both within and outside the field. The forecasts are driven by the assumption that behavior in the future, given identical characteristics, will be the same as today. Culture is outside the scope of models, with good reason, but if culture matters, or anything else that is also outside the model's data, there will be misses. Modelers may claim data issues, or poor inputs, and those certainly matter, yet estimation of models across time is never done in practice. There are always reasons – incompatibility of surveys, time, budget, and so on. The excuse for using cross-sectional analysis in the 1950s was there was no time series, only one survey (at most) had ever been done in any metropolitan area. The excuse today is what?

In addition to behavior being static in these models, technology is as well. The use of *stated preference* models to examine what would happen given a new technology attempts to push the boundaries of this, but it fails to say what technologies will actually be around, which will affect demand in ways we just have to admit we cannot accurately foresee. This issue is increasingly important as new modes like shared autonomous vehicles are being considered, and autonomous vehicles (even if unshared) change the character of automobile travel. That a forecasting tool considering 30 years into the future cannot consider the possibility of such change in any reliable way suggests that it is probably not the right tool.

For this reason, these travel demand forecasts, at one time the most sophisticated analyses done by humans with their early use of mainframes, fall into the same trap as much simpler forecasts: under-estimating growth in the early years of a technology's lifecycle and over-estimating in the late stages. Models might be useful for short-term analyses of minor changes, scenario analyses of alternatives,

but most definitely not long-term forecasts. This requires changing evaluation procedures and government regulations. However there are enough problems today that remain unsolved, so that looking for problems 20 to 30 years down the road seems futile.

The models did innovate by systematically breaking down the travel demand question into a set of components. In reality these components are highly interconnected. In the model they are dealt with linearly in sequence, though more advanced models may solve the system iteratively, with feedbacks from later stages to earlier. Similarly, from a narrative perspective, we present the text linearly, but the actual sequence in which these questions are addressed is not nearly so tidy.

We begin first with **trips**.⁵ In *what* activities are travelers engaging daily? The sequence of those activities matter, and may affect not only travel, but the choice of activities themselves. The ability to chain trips together gains efficiencies that allow more activities to be undertaken.

⁵ (§2.1).

While the early versions of the model do not address sequence, more recent models do. The question of temporal distribution, *when* activities occur arises. We begin with the idea of the **schedule**.⁶ The individual schedule has the form it does because of the need for **temporal coordination**⁷ in production. That gives rise to the very real problem of **peaking**,⁸ leading to higher travel times on the road (and lower accessibility by car) in the peaks. However the peaks often see more transit service provided, thus lower waiting times (and higher accessibility) by transit when it operates on its own **right-of-way**.⁹

⁶ (§2.2).

⁷ (§2.3).

⁸ (§2.4).

⁹ (§7.10).

The second step of the classic model, trip distribution or destination choice, examines the spatial distribution of activities, that is, *where* they take place. We begin with the notion of the **travel time budget**,¹⁰ which suggests the amount of time available for travel is limited. There is a **travel time distribution**,¹¹ showing how many trips are of a given duration.

¹⁰ (§2.5).

¹¹ (§2.6).

This distribution, people's willingness to travel, due both to preferences and constraints, explains **social interactions**¹² and the decline of population (and employment) **density**¹³ with **distance**¹⁴

¹² (§2.7).

¹³ (§3.1).

¹⁴ (§3.2).

Spatially that gives rise to a daily (weekly, monthly) **activity space**¹⁵ over which travel behaviors occur. We can combine willingness to travel, travel time budgets, and activity space to get a three-dimensional **spacetime prism**.¹⁶

¹⁵ (§2.8).

¹⁶ (§2.9).

After we decide where we are (or can) go, we can think about *how* to get there. We make a **choice**¹⁷ about mode of travel. This choice is highly constrained though, and depends very much on the pattern of accessibility, which explains the pattern of **mode shares**¹⁸ we see.

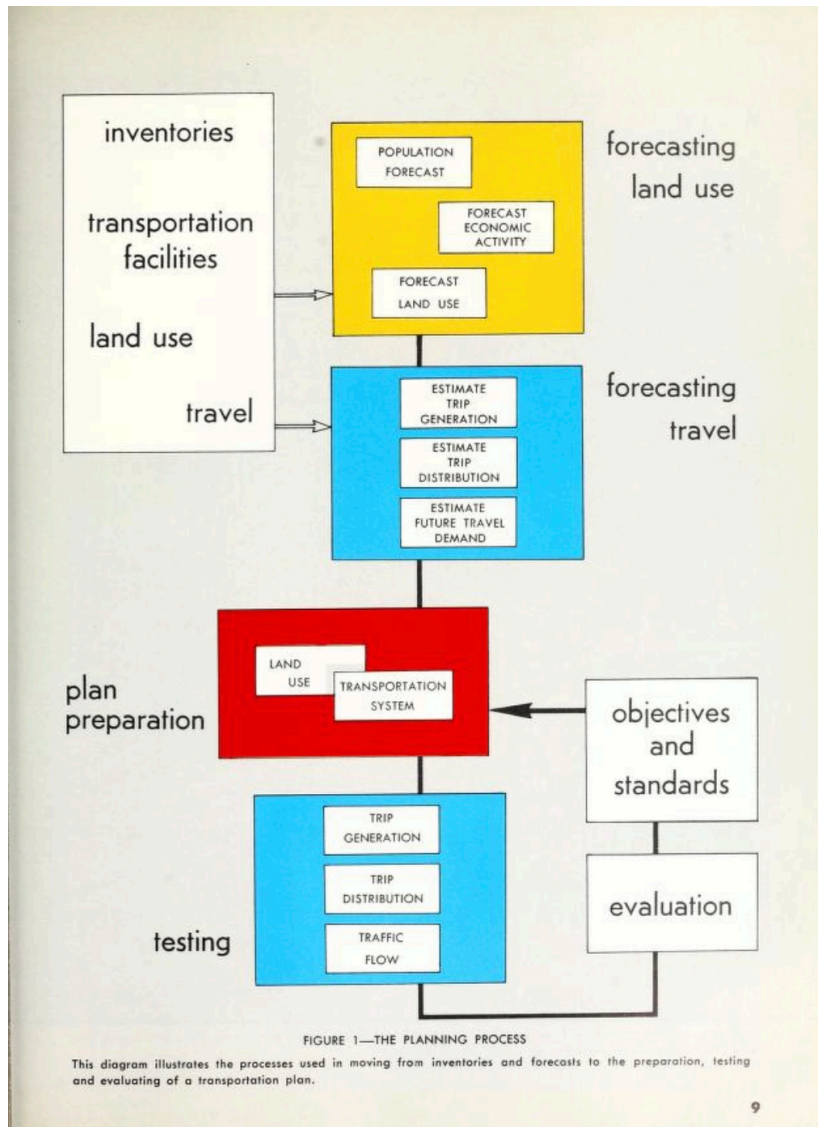
¹⁷ (§2.10).

¹⁸ (§8.1).

¹⁹ (§8). More detail about the individual modal technologies¹⁹ is found later in the book.

Traffic assignment, or as we would now say, route choice, is dealt with in the chapter on Routing.²⁰

Figure 2.2: Travel demand model. Source: (Carroll 1959), p.9.



2.1 Stages, Trips, Journeys, and Tours

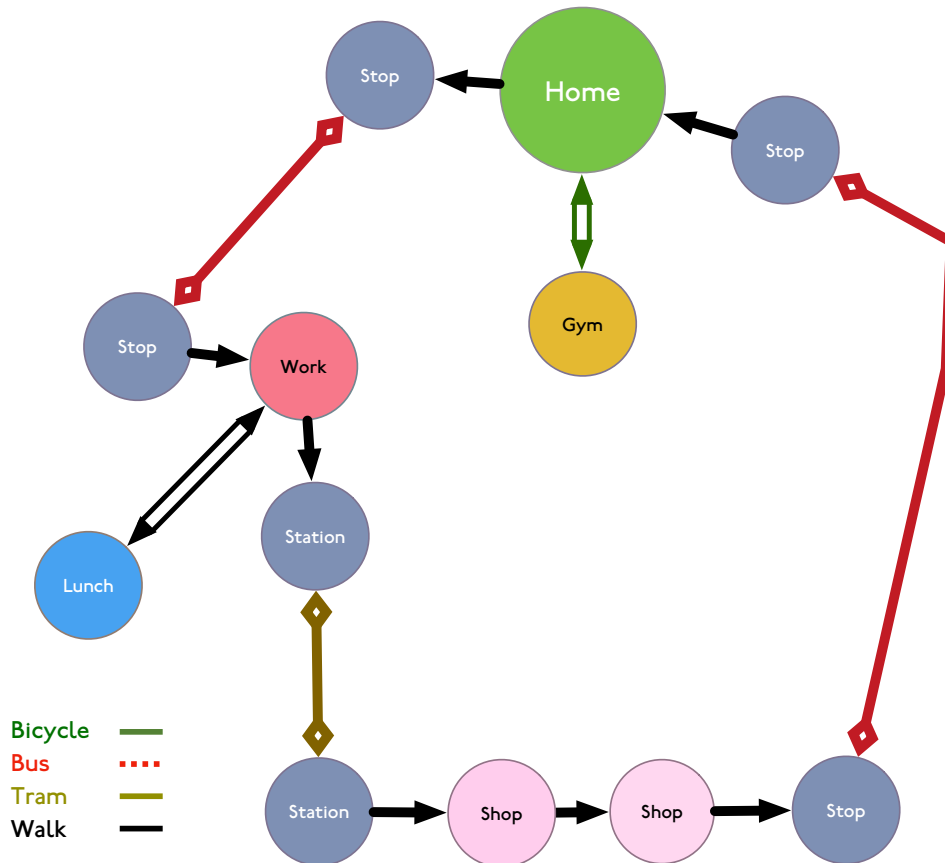


Figure 2.3: Trips.

NO SURPRISE, BUT PROFESSIONAL AND EVERYDAY LANGUAGE OVERLAP IN THEIR VOCABULARY, WHILE NOT BEING IDENTICAL IN THEIR MEANINGS.

In many fields this is not a big problem, as the technical and professional discourse do not overlap and the general public rarely reads or hears the technical discourse, aside from students at the beginning of their training. In transport it is a problem, as the professionals have to address the public: voters, decision makers, or survey respondents. They talk with each other continuously.

As expected, transport planners and engineers have developed a detailed vocabulary to talk about movement; unfortunately, it varies even by mode of transport. *Roget's Thesaurus* gives as synonyms: trip, journey, excursion, cruise, expedition, foray, jaunt, outing, run, swing, tour, travel, trek, errand, hop, junket, peregrination, and

ramble. They clearly have connotations, which makes one more suitable for certain occasions, but they can be used interchangeably for a movement from *A* to *B* and back.

When observing and thinking about movements for a moment, it becomes clear that a 'trip' will have smaller building blocks and might belong to something bigger, say a vacation. The professionals have given names and a structure to these, so that they can measure and talk about them clearly. More progress can be made if the same language is understood by all parties.

The key terms are 'stage,' 'trip,' 'tour,' and 'daily schedule' with their variants in different countries and industries. A 'stage' is the smallest unit, the movement from *A* to *B* with one mode or one vehicle of that mode: walking from home to the bus stop or flying from the first airport to the hub airport are stages. It sometimes is used synonymously with the [first and last mile](#)²¹ problem. The airline industry talks about 'legs' and means the same thing, as do American planners when they talk about 'unlinked trips.' In logistics, the stage with the longest distance is generally called the 'main haul.'²²

A sequence of stages from one activity to the next is a 'trip,' which now requires a definition of a destination activity, as in [Figure 2.3](#). Following the example of time use studies and sociology, the activity is defined as a meaningful interaction with another person or task. In transport, a trip is always one-way, unless identified as a 'round trip.'

A sequence of trips from *A* via various other locations back to *A* form a 'tour,' illustrated in [Figure 2.4](#). The term 'journey' often specifies tours starting and ending at home, though it is often used in a one-way context, for instance the 'journey to work.' One runs into problems if one wants to talk about tours within tours, for example going to lunch and coming back to the workplace. Some parts of the literature refer to these as 'sub-tours.' We also need a word to talk about the movement from home to the primary stop of the tour. You will find the word 'commute' or 'half-tour' to describe just this, even if the commute includes stops such as dropping off the children at school, a quick coffee at Starbucks, and/or time at the gym before arriving at work. Commute is often reserved for just the trip from home to work (and back) though. More generally, a 'chain' refers to any sequence of trips (such as from home to restaurant to movies). The [daily schedule](#)²³ includes all the tours undertaken between getting up and going to bed again.

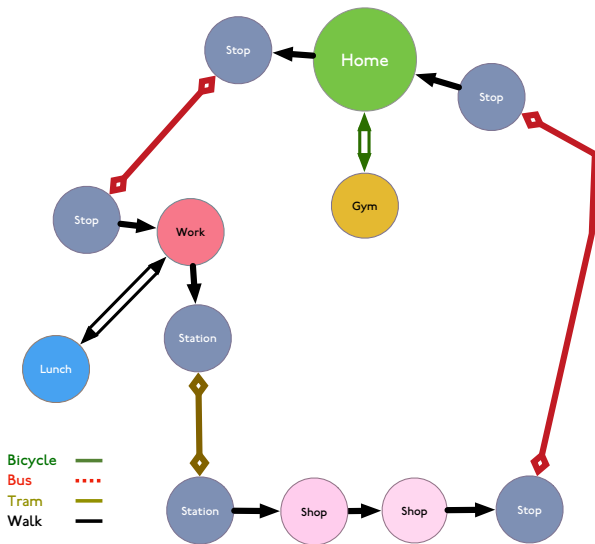
A discussion about [mode choice](#)²⁴ should always refer to the element talked about. Walking (stages) will always be part of the

trip to reach the vehicle(s). Walking will therefore always have the highest mode share among the stages but not of the distance traveled. At higher levels, planners have to decide which mode they allocate to the trip or tour. Normally, they choose the mode of the stage with the longest distance. In this process, the other stages are forgotten and often their distance allocated to the main mode. The chances for confusion are endless, unless this is made clear.

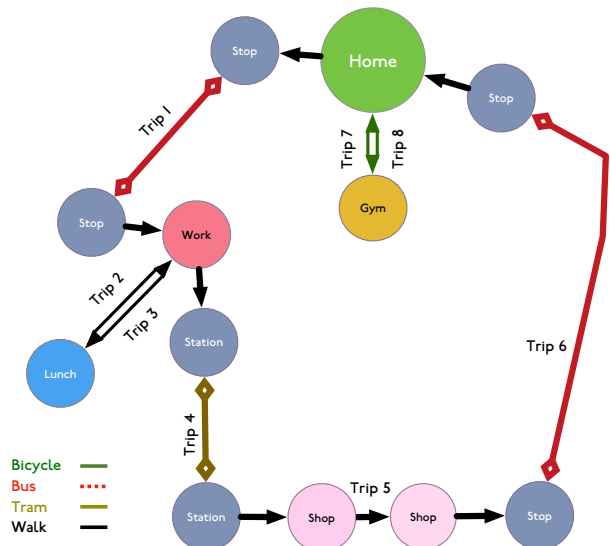
Each trip has two 'trip-ends,' the origin and destination. Confusingly, the Institute of Transportation Engineers in their *Trip Generation Handbook*²⁵ computes trip generation at each trip-end based on land use patterns, so if you simply add up the number of trips generated across all land uses, you wind up with two trips 'generated' for every trip. So be careful. Planners and modelers often say trips are 'produced' at home and 'attracted' to non-home destinations, but that is not very helpful for clear thinking.

Remember: Tours are sequences of trips, which are sequences of stages.

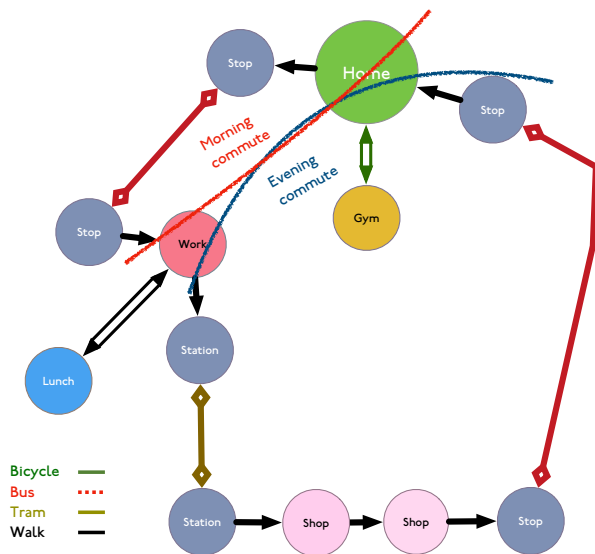
²⁵ (Institute of Transportation Engineers 2004).



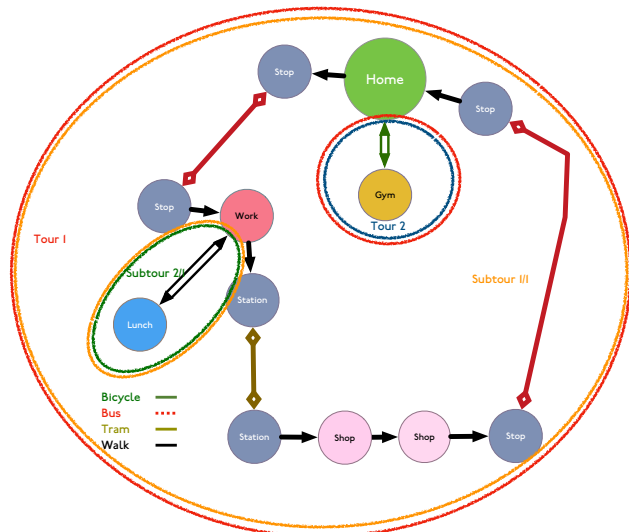
(a) A daily schedule.



(b) There are 8 trips and 6 access/egress trips by walk, so 14 stages. The bicycle trips are assumed to be door-to-door.



(c) From home to work via bus (and two walk stages) comprises the morning commute. From work to shop (via tram) to home (via bus) comprises the evening commute.



(d) There are two tours. Tour 1 contains two subtours.

Figure 2.4: Trips with tours.

2.2 The Daily Schedule

THE DAILY SCHEDULE IS OFTEN DOMINATED BY A MAJOR OUT-OF-HOME ACTIVITY.

We begin with Figure 2.5, the schedule of a day in the life of a typical worker, Paul. Paul awakes from bed (7:00) and performs various domestic and personal maintenance activities. He then travels from home to work by bus (8:00-8:20), arriving a few minutes early for work. He spends 10 minutes in the smoker’s lounge (at least until it’s prohibited). If he avoids falling into a dream, at work, he executes various tasks, meets and socializes with colleagues, takes lunch, and resumes work until the close of business. Paul returns home (5:00-5:30).

Sometimes the home-work-home orbit of *trips*²⁶ has appended other activities: shopping, errands, meal-taking, providing transport for others, etc. For children, the pattern is mimicked in a home-school-home cycle. For the retired, some recreation (avocation) activity may substitute for work (vocation). The weekday is dominated by these three activities: time at home, time at work or school or avocation, and time in travel between the two. This pattern is repeated, though differentiated, daily, for the better part of one’s life.

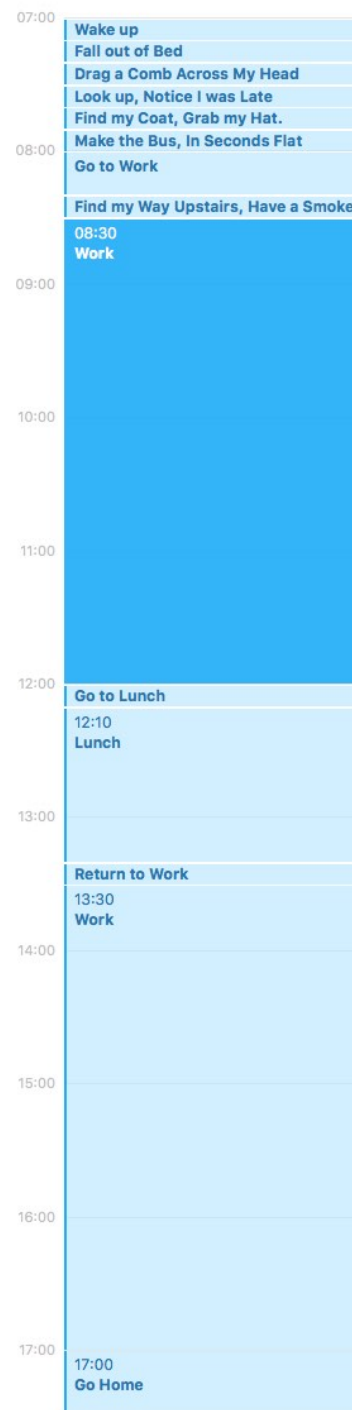
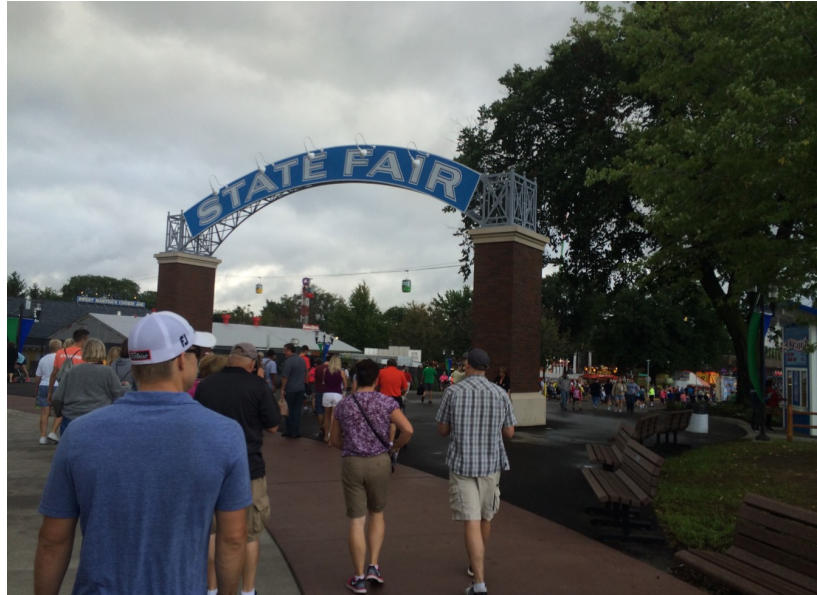


Figure 2.5: A day in the life of a typical worker.

²⁶ §2.1.

2.3 Coordination

Figure 2.6: Minnesota State Fair. Photo by D. Levinson. The sign was lost in the woods for 30 years before being recovered and restored.



WE CAN BE SPATIALLY COORDINATED TO REDUCE OUR SCHEDULING COSTS, OR WE CAN BE TEMPORALLY COORDINATED SO THAT WE HAVE LOWER SPACE COSTS.

The classic multi-purpose room in 1960s era elementary schools, hot-desking, or shared parking between office, stadiums, retail, and churches are examples of temporal coordination to share a scarce resource by using it at *different* times, thereby reducing land and structure costs. Most temporal coordination, though, aims for people to engage in the same task at the *same* time, and thus consume more space. Cities provide both spatial and temporal coordination, putting people close together and having them do the same things at the same time.

Cities work to reduce temporal coordination costs. This is one of the many ways they enhance [economies of agglomeration](#).²⁷ But they do so by increasing spatial coordination costs. Two people cannot occupy the exact same latitude and longitude at the same time without going vertical. This adds to the cost of construction. We do not have freedom to use our land any way we want to; we must share some rights to it because society demands it. This diminishes our freedom of action.

One expects that improved information and communication technologies will reduce the need for in-person interaction, and we

certainly see some of that. But reducing the call of the city does not eliminate it. So long as some physical interaction is required, city-like places will emerge. The need for young men and young women of different genetic lines to somehow interact in person is one such call upon the pattern of the city.

Two-hundred years ago, the city was barely what it is today; two-hundred years from now, the city may differ again. While unlikely, cities may return to being seasonal, like the classic medieval trade fair or the vacation community.²⁸ These once comprised entirely temporary structures, which gradually became permanent. Look at your local fairgrounds for examples of the temporary becoming permanent. Today we construct state fairs with permanent buildings (Figure 2.6). In contrast world's fairs, which do not repeat annually, have temporary structures. While not made of paper maché, the buildings of Chicago's 1893 White City or even New York's 1963 World's Fair are largely gone, while the world's fair is a lot less significant than it once was.

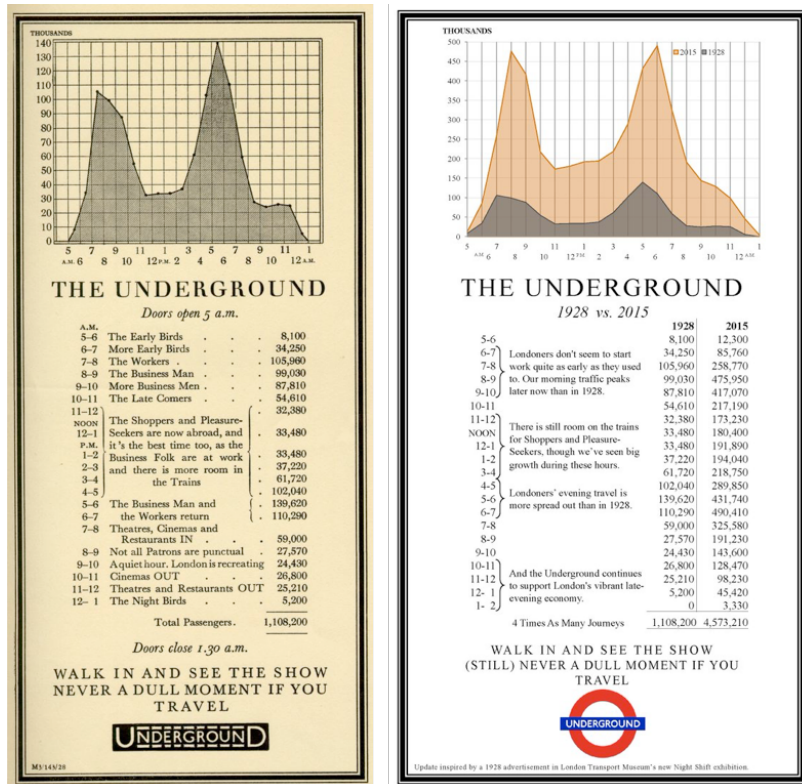
If people were ever to lose their need for daily physical interaction (which is, again, highly unlikely), we would expect a thinning of the urban support system, less reliance on costly permanent infrastructure, and more reliance on the ad-hoc. Humans will still require shelter, and those shelters may still cluster, so long as transport still has costs. Yet, we can imagine a world where advanced technology means we don't need to commute or shop more than weekly. And that could also mean we don't need to live as close together. And with advanced [driverless cars](#),²⁹ even that burden (the need to focus on the task of driving) is lifted, enabling even more spread.

²⁸ Even vacation communities are becoming year round, as telecommunications enables remote work.

²⁹ §15.

2.4 Diurnal Curve

Figure 2.7: Diurnal curves on the London Underground. Source: Transport for London.



WE SOMETIMES THINK OF THE CITY AS A COLLECTION OF PEOPLE AND OBJECTS IN SPACE THAT EXIST FOR THE PURPOSE OF REDUCING THE COSTS OF HUMAN INTERACTION; THE CITY IS ALSO A COLLECTION OF ACTIVITIES IN TIME.

Taking the long view, cities once did not exist (the time before the founding of the city), and eventually may not exist again. The list of abandoned cities is long and, though this may sadden us, will undoubtedly grow longer.

However, the city also operates at shorter timeframes. There is the multi-decade cycle of infrastructure renewal and replacement. There is the multi-year (though somewhat random) cycle of sports team victories. There is the annual cycle of the city operating through the seasons, with winter and spring and summer and fall events. There is the daily cycle of flows of people into and out of the city.

People possess Circadian rhythms; they operate on a 24-hour cycle, and about half that time is daylight. Going to the place where

that activity (work, school, other) occurs follows a pattern: leave home early enough to arrive at the destination at a desired time. Do something there. Leave (after say 8 hours) and return home. There are many complexities.

Figure 2.7 from Transport for London has two peaks: morning and evening. These peaks are the ‘rush hours’ of common complaint, when more people want to use the transport system than capacity is immediately available, leading to congestion. This graph shows both the supply provided by the public transport system (more seats are made available during the peak) and the demand of users. The supply clearly responds to the demands. The afternoon or evening peak is usually higher (and almost always broader) than the morning peak, as we organize more activities after work than before.

Why do we see diurnal patterns of flows? Why are there morning and afternoon peaks, or what we refer to as the ‘rush hours’? The answer is to ensure some set of people (peak commuters) are generally in the same place at the same time. And we do this to reduce inter-personal [coordination](#)³⁰ costs. If we are generally in the same place, we don’t need to pre-arrange meetings, we run into each other in the hallways, I can easily knock on your door, and I see you on the sidewalk. Our temporal coordination costs drop. And even if we do need to pre-arrange, it is relatively simple. An instructor might tell the students in class: ‘I am here because you are here; you are here because I am here.’ In contrast, if we are not generally in the same place, we do need to pre-arrange meetings, and I will not randomly run into you. Our temporal coordination costs rise.

³⁰ §2.3.

In the US, most trips are not commuting trips, even during rush hour. However, work trips with their tight scheduling overload the system at peak times.

There are lots of people for whom the congestion costs of the peak outweigh benefits of organizing work on the ‘standard’ schedule. Many people with shifts in organizations that operate more than 8 hours a day (including medical, police and fire, manufacturing, transport,³¹ retail, some construction, and media) travel in the off-peak. For some, this is necessary (you don’t want to change bus drivers in the middle of the peak); for others’ convenience (why travel at rush hour when it is unnecessary).

³¹ Many bus drivers cannot take the bus to work, since they drive the first or last shift of the day.

In the United States’ Central Time Zone, that peaking pattern is partially dictated by what happens on the East coast. People using Central Time tend to go to work earlier than they otherwise would to ensure a greater overlap in time at work with those back east. Similarly, people involved in international trade may keep odd hours

locally to coordinate with their customers or clients elsewhere in the world. In other parts of the world, schedules similarly adapt to the needs of trade as well as local custom. In some places, work lasts until very late, but there are mid-day breaks.

This temporal coordination imposes the cost of increased loads on the transport system, as people converge and diverge at the same time, requiring either more capacity or causing crowding (and slower speeds). We can (and do) smooth the flows on transport systems, encouraging peak spreading (some of which the market does by itself) through differentiated [prices](#).³²

³² §9.7.

³³ §1.

[Accessibility](#)³³ varies by time of day. When travel by car is slower in the peak hours due to congestion, accessibility drops with it. When travel by transit is faster because of higher frequency of service, accessibility rises. It not only varies by hour, it varies minute-to-minute. The accessibility by transit is much higher a minute before the bus departs than the second after.

Planners typically assume that opportunities are available 24 hours a day. This, of course, is not true. Many jobs expect you to be on-site during certain time periods and arrive at a certain time. School and stores are open during certain hours. For instance, there is no access to eating out when restaurants are closed. We make this assumption because of lack of data of the hours of operation, and may not need to in the future, when we can truly map a 24-hour city.

2.5 Travel Time

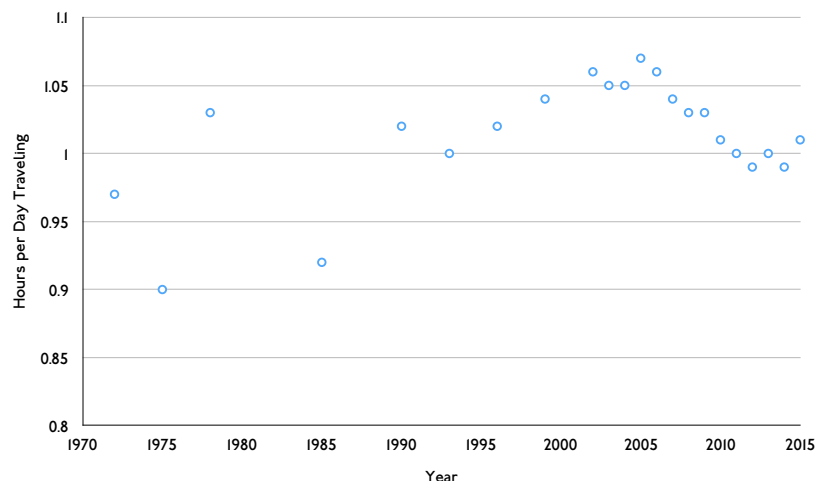


Figure 2.8: Travel time budget.
Source: (National Travel Survey 2014 2015).

ONE OF THE MOST (IN)FAMOUS CLAIMS MADE ABOUT TRAVEL BEHAVIOR IS THAT THE TIME SPENT ON IT IS CONSTANT OVER THE YEARS.³⁴

This claim is generally made for whole populations not individuals, where personal introspection and observation tells us that the time spent changes with age, family responsibilities, as well as new workplaces or homes. It is a claim made for regular daily travel, happening in the usual environment of the traveller. Figure 2.8, from the UK, depicts an example. There are at least three budgets we can consider. The day has 1,440 minutes; this is fixed, and all of the time within the day must be allocated to activities or travel. There is the total amount of time spent traveling daily for all purposes. And there is the total amount of time spent commuting, or travel to and from work.

Yacov Zahavi, and all those in his tradition, base the claim of daily travel time budgets on a striking similarity of the reported numbers for total daily travel time in local, regional, and national travel diary surveys. A figure of 60 minutes was proposed, but this has crept up over the years, as about 100 daily minutes of travel are reported in Switzerland, for example.³⁵

The claim is powerful, when linked to Downs' 'law of peak-hour-expressway congestion' or 'triple convergence,'³⁶ or more formally **induced demand**,³⁷ which observes that travelers will respond to changes in the transport system by changing their behavior until they

³⁴ (Zahavi 1974; Marchetti 1994).

³⁵ (Schafer and Victor 2000). For a review see (Mokhtarian and Chen 2004).

³⁶ (Downs 2005).

³⁷ §12.1.

cannot find a way to improve their situation further. The changes can be caused by travelers improving their daily schedule by leaving or arriving earlier, by them changing to a more attractive mode, by them switching the workplace or the residential location for something better, or by them doing more things outside the home: meeting friends, attending civic meetings, or watching a child's soccer game.

An investment in transport will first generate travel time savings, and thus accessibility, but in the longer term, these time savings are lost as the increased speed leads to increased distance traveled by car. You may say the investment was in vain, and when you look only at time use or motorway speeds, you can essentially make the political argument that all such investment is pointless.³⁸ Or you can look to see whether these long-term changes are indicators that more people can now use the new capacity to do things they want to do ... when, where, and how they want to do them.

Next to political assessment is the empirical question: is all of the change converted into longer travel? If taken seriously, the constant travel time budget implies an elasticity of minus one in terms of travel distance with respect to changes in travel time. Only then can the budget stay constant.³⁹ There is some work on this, but not as much as the size of the claim would justify.⁴⁰ So some of the gains remain in the transport system, mostly as better daily schedules for travelers.

On the one hand, capacity additions can improve conditions for travelers who can take advantage of them, even after considering changes to travel patterns. On the other, rising population increases travel demand (and congestion), and in some places population has increased faster than capacity. These offsetting factors help explain the relative stability of daily travel time budgets.

There is a third hand though. Travelers may have preferences for a certain amount of travel. They may not want to live too near the workplace and desire some spatial separation; this is called a 'positive utility of commute.'⁴¹ 'Rational locators' may also recoil at commutes that over time, become too long. In other words, a commute that was 25 minutes when they moved in has now eroded to 35 minutes due to rising congestion from increased population. Rational locators will periodically readjust their home and/or work location to keep this from getting out of hand, for instance moving to a new location that is again only 25 minutes from work or taking a job nearer home (which may be longer in terms of distance, as suburban routes and urban routes have different speeds).⁴²

³⁸ See (Metz 2008), although he is not quite as radical as the political discussion.

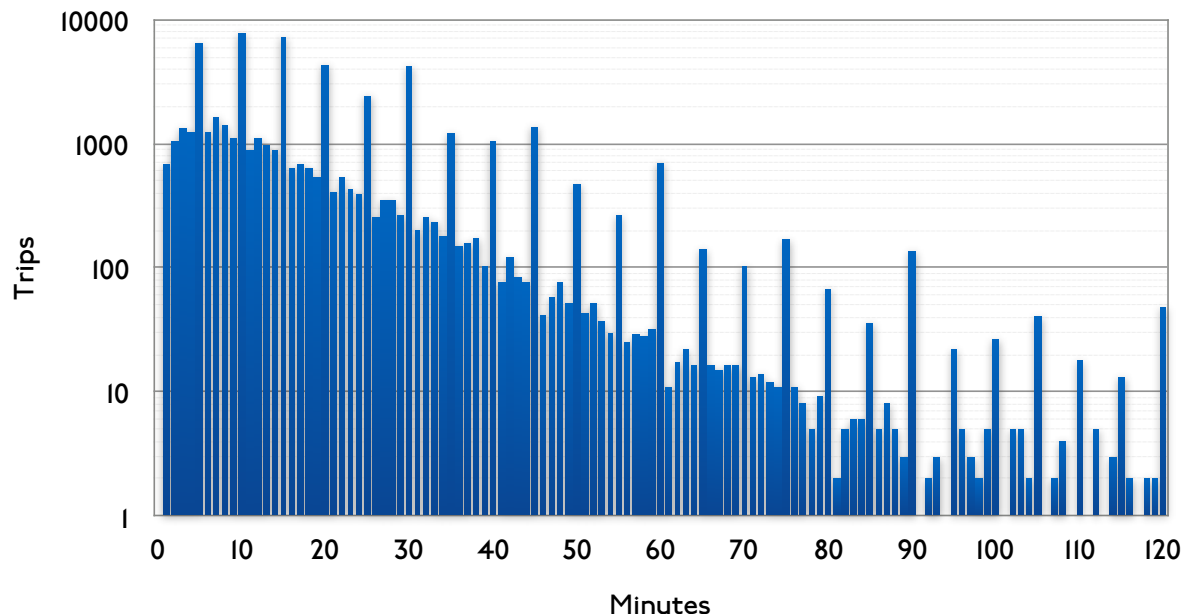
³⁹ Elasticity is the percentage change in an output with respect to percentage change in the input. So an elasticity of -1 means a 1% increase in travel time leads to a 1% decrease in distance.

⁴⁰ (Cervero and Hansen 2002; Weis and Axhausen 2009). All of the existing work indicates that the elasticity is large but not nearly 1.

⁴¹ (Redmond and Mokhtarian 2001).

⁴² (Levinson and Kumar 1994).

2.6 Travel Time Distribution



WHEN WE COMBINE OPPORTUNITY EXPANSION WITH [DISTANCE DECAY](#),⁴³ WE GET THE TRAVEL TIME DISTRIBUTION.

The more opportunities in a location, the more likely it will be chosen. The farther away some place is, the less likely it will be chosen. These are offsetting.

It leads to distributions that look like Figure 2.9, which was drawn from the Travel Behavior Inventory for the Twin Cities for commute trips by automobile.⁴⁴ In the places near home (at the left of the plot), it leads to an increasing willingness to make a longer duration trip. In places farther from the origin (at the right), it leads to decreasing willingness, as the additional opportunities fail to outweigh the longer durations.

Figure 2.9: Travel time distribution from the Twin Cities Travel Behavior Inventory. Log scale on y-axis. Source: [\(Brosnan and Levinson 2015\)](#). Many people round their start time and end time in the survey, and they usually round up, as shown in Figure 2.9. So if a trip were 14 minutes in duration, it would be rounded to 15 minutes. More precisely, since start and end times are reported, and durations computed, if it began at 7:36 and ended at 7:50, it might very well be reported as beginning at 7:30 and ending at 7:45. If it were 22 minutes, it might be rounded to 25 or even 30 minutes. This indicates that self-reported times are significantly biased in travel analysis. Until recently, that was the only data available. But with the advent of GPS devices and cheap sensors tracking traffic across networks, we can get individual travel time and speed data that is both more precise and more accurate.

⁴³ §3.2.

⁴⁴ [\(Brosnan and Levinson 2015\)](#).

2.7 Social Interactions

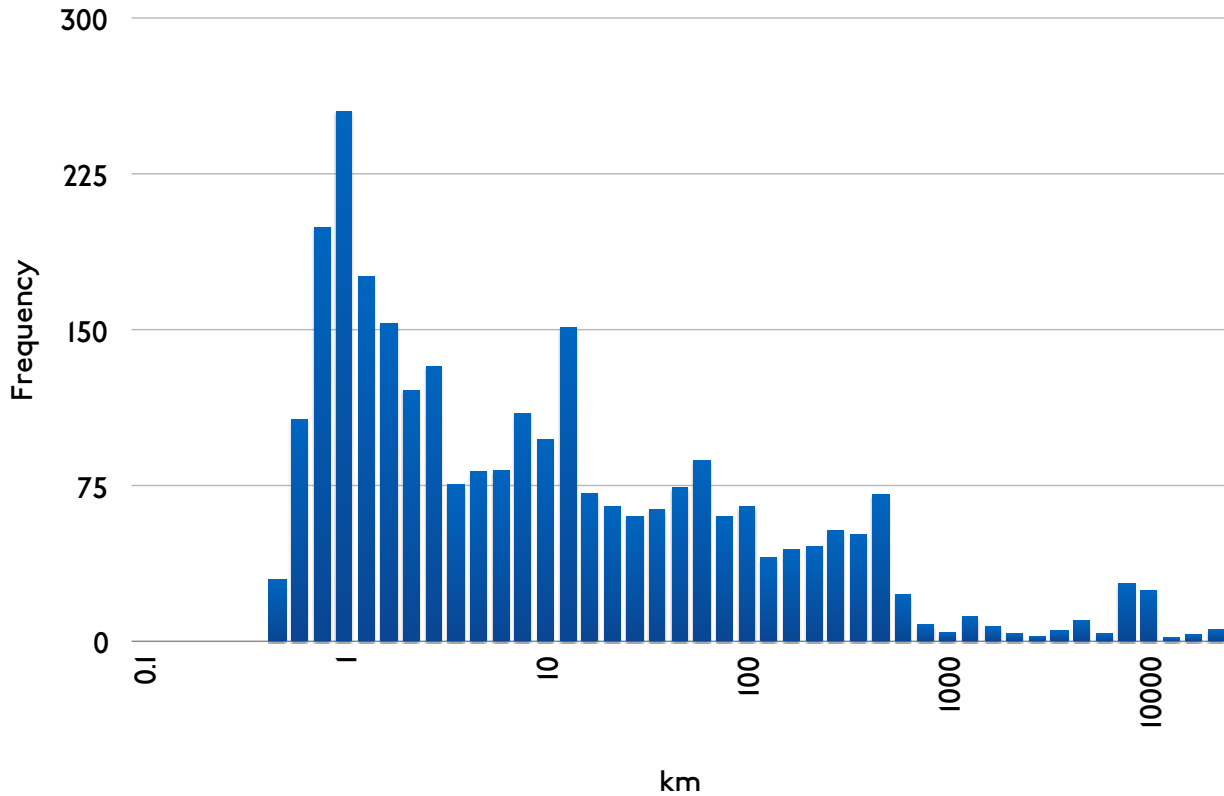


Figure 2.10: Distribution of social interactions. Log scale on x-axis. Source: (Kowald et al. 2013; Axhausen and Kowald 2015).

LEISURE IS THE LARGEST AND FASTEST GROWING SEGMENT OF THE TRAVEL MARKET IN ADVANCED SOCIETIES.

In the industrialized world, leisure travel makes up about 40% of all trips and 40% of all distance traveled. Leisure is a catch-all category in standard travel (diary) surveys: not work, not education, not shopping, not personal business, and not picking up or dropping somebody off. It covers many different activities: from window-shopping to meeting friends for a weekend hike.

Some of these leisure activities are regular, such as attendance at church or going to the gym, but others are irregular or unique: that visit to a friend last seen ten years ago or going on the Hajj. These activities don't have the same constraints as work or school, to which we are committed through a contract or a legal requirement. Some are spontaneous, but others express deeply held

commitments such as a pilgrimage or even a trip to the gym. While we often think of them as ‘discretionary,’ their social nature makes them location-binding for us: 80-90% of these activities involve other people: family, friends, the three other golfers, other players of team, the other 9 worshippers waiting for you to make the minyan (quorum) in the synagogue; never mind the dog expecting her walk.

This overwhelming social nature of leisure implies that the activities are also joint decisions, as one has to account for others when setting dates and locations. In some cases, the choices have become so habitual that the organizer does not think anymore – say for club, civic, or religious events – but for most others, the negotiation is a (large) part of the preparations, which can be seen in the large amount of text message, social network, email, and telephone traffic involved beforehand.

As the effort involved in participating face-to-face in an event, meeting, party, or get-together involves, at minimum, the travel to get to its place, the spatial distribution of friends becomes crucial. The wider our circles, the more travel and associated greenhouse gases we will produce. Yes, the higher the effort, the less likely that we will meet certain persons⁴⁵, but there will be a certain minimum frequency to honor our links: attendance at the wedding of a cousin, being at the funeral of your friend’s wife, or the annual joint hike.

⁴⁵ §3.2.

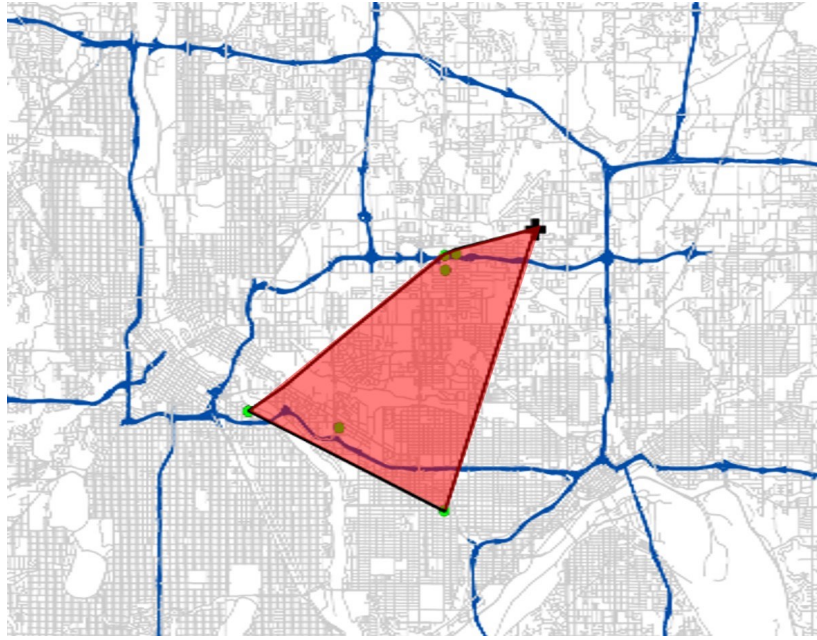
While sociologists have long studied the structure of social networks, they gave little and generally cursory attention to the spatial distribution of such networks. Recent work by joint teams of transport planners and social scientists has shed light on the distances involved (such as Figure 2.10).⁴⁶

⁴⁶ (Kowald et al. 2013; Axhausen and Kowald 2015).

In this typical example from a Swiss study, the bulk of social contacts lives within a 30-minute car ride (supporting the idea of a 30-minute threshold in a typical accessibility analysis), but there is a substantial share living much further away, including some overseas. This distribution should add more long distance contacts, as travel and communication become cheaper with low-cost flying and effectively free video conferencing. So, indeed, the home addresses of our friends and our wish to meet them is one driver in travel and greenhouse gas production.

2.8 Activity Space

Figure 2.11: Activity space. Dots indicate activities, plus indicates home, shaded area is the total convex hull of the activity space which contains all the dots. Source: (Parthasarathi et al. 2015).



THE ACTIVITY SPACE REPRESENTS THE PLACES THAT A PERSON OR HOUSEHOLD ENGAGES OR OCCUPIES IN A GIVEN PERIOD OF TIME, TYPICALLY A DAY.

An illustrative activity space for one traveler is shown in Figure 2.11. The *potential activity space*, on the other hand, represents the maximal area over which the traveler could engage in activities. The activity space is only a small part of all places an individual has knowledge about and could go, which is referred to as the *action space*. The extent of the activity space depends on the individual, their preferences, the opportunities available both within and outside the space, and the character of the network in that space.⁴⁷

⁴⁷ (Parthasarathi et al. 2015).

Areas with high accessibility (where more destinations can be reached in less time) have smaller activity spaces because people can accomplish their daily wants and needs closer to home. Households with more cars, more income, and more workers have larger spaces, as mobility increases the viability of farther away destinations.⁴⁸ Activity space examines the actual travel of an individual or family. *Accessibility*⁴⁹ considers the potential.

⁴⁸ On the other hand, higher incomes may allow a household to live in a more accessible area, which could facilitate a smaller and more efficient activity space.

⁴⁹ §1.

2.9 Space-time Prism

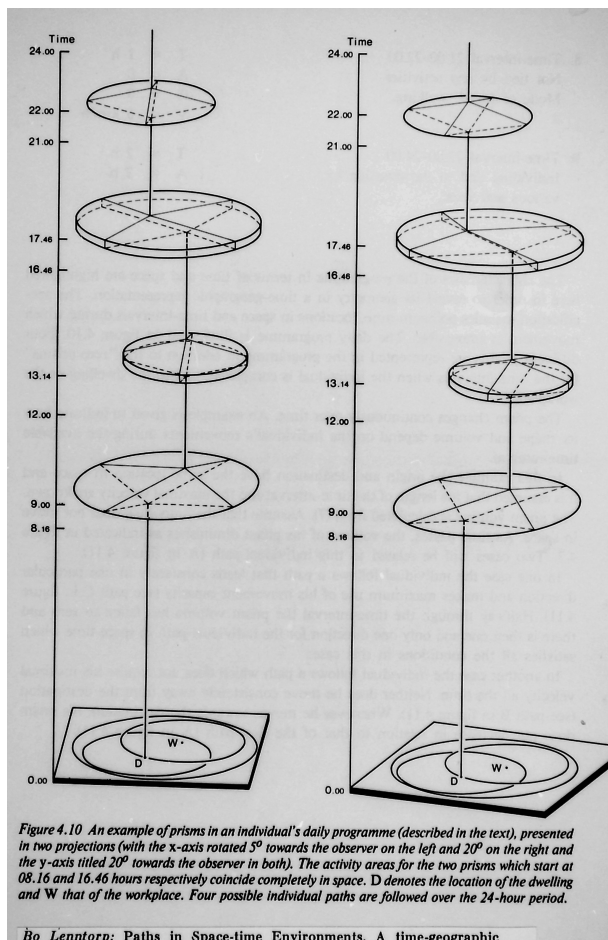


Figure 2.12: Space-time prism.
Source: (Lenntorp 1977).

THE DAILY SCHEDULE IS A COMPLEX OPTIMIZATION PROBLEM.

You might think of your [schedule](#)⁵⁰ as what shows up on your calendar or daily planner. But there are many activities you probably don't typically record: driving to work, going out to lunch, sleeping, and so on. To engineers and mathematicians, the daily schedule is a complex optimization problem filled with objectives and constraints: there are periods of time (windows) when you have to be at certain places in order to be available to others, you have to be able to get to these points with the (mobility) tools you can bring along, and you want to spend certain amounts of time at each point to be able to achieve your goals; all of this within the [24 hours of the day](#)⁵¹ and within your commitments and monetary [budgets](#).⁵²

So how do people solve this daily problem? Often, we start with

⁵⁰ §2.2.

⁵¹ §2.4.

⁵² §2.5.

what happened yesterday (or last week on the same day) as a model. We also have many previous occasions of when we wanted to combine certain activities at the same location. We have building blocks, which reduce the complexity of the problem enormously. In addition, there are constraints that make many combinations of places infeasible within the allotted time.

Social scientists have thought about this problem for a long time. The diagrams in the figure above are from Torsten Hägerstrand, a famous Swedish geographer, who identified and visualized (Figure 2.12) one set of these constraints in the 1970s.⁵³ His insight was to see that some activities in time and space are much more firmly committed to than others. Think of the work schedule of a nurse or teacher, or the need to drop a child off at school. These firm commitments form anchors within the daily schedule, constraining the time available for remaining activities and where they might take place. We are all caught in time and space.

The 'space-time prism' explains why motorized modes are so attractive. First streetcars in the late 1800s, then automobiles in the early 1900s increased the domain that travelers could reach between their committed activities compared with walking. In theory, that increase translates into more satisfaction with the activities undertaken. It also makes clear why pedestrians like very dense environments, as there they can reach many alternatives on foot, while mass transit and cars (especially) slow with congestion.⁵⁴

This reach is especially important, as people generally plan with firm commitment only a part of their day and leave the rest available for spontaneously arising opportunities and ideas. It also makes clear why an overly committed schedule is so stressful, as it robs people of the chance to act spontaneously.⁵⁵ It requires care to avoid disturbing the clockwork of interweaving movements and activities.

As in the figure, the traveler can leave the first commitment at the speed of the mobility at hand: a bicycle, a public transport season ticket, or a car. Generally, the traveler has generally to leave for the second commitment with the same mobility tool and at the same speed. These two funnels define that part of space-time available for an activity or set of activities. You have an *isochrone*⁵⁶ about you at any given time defining where you can go given constraints.

⁵³ (Hägerstrand 1970).

⁵⁴ §6.5.

⁵⁵ (Levinson 2016a).

⁵⁶ §1.1.

2.10 Choice



Figure 2.13: Minnesota Valley National Wildlife Refuge. Photo by D. Levinson.

I came to a fork in the road
 Right or left to what abode
 A chance step along the way
 What would be, who can say.
 So toss a coin, let it fall
 As good a way as any at all.
 – *The Poetry of Jack Sewitch, Volume 2*

LOCATORS CHOOSE WHERE TO LIVE AND WORK. TRAVELERS CHOOSE WHAT TO DO, WHERE TO GO, WHEN TO LEAVE, AND HOW TO GET THERE.

Since the 1970s, analysts have converged on a family of models, called qualitative choice models, to better understand this process.

To express the logit model simply (and there are textbooks on this, so this is a mere surface treatment):

$$P_m = \frac{e^{U_m}}{\sum_{m=1}^M e^{U_m}} \quad (2.1)$$

The probability of choosing alternative (mode) m (P_m) is the ratio of the exponentiated utility of m (U_m) to the sum of all such exponentiated utilities. The utility is typically empirically estimated based on people's actual choice (revealed preferences) or what they say they would do in an experiment (stated preferences). It might be of the form:

$$U_{ij,m} = f(C_{ij,m}, S) \quad (2.2)$$

So the cost of travel between i and j by mode m depends on the Cost ($C_{ij,m}$), which might include different travel time and financial components. It might also depend on socio-economic characteristics (S) of the traveler. Other factors posited to affect the results that can be measured should also be incorporated.

So suppose a model says that the probability of an individual taking transit is 13%, driving is 77%, and walking is 10%. We then use a random number process and draw a number between 0 and 1.

If it falls between 0 and 0.13, the traveler takes transit, if it is between 0.13 and 0.90 the traveler drives and above 0.90, the traveler walks.

A mathematical transformation of the denominator of this equation, called the 'logsum', gives a measure of accessibility (§1) sometimes used by academics or those doing travel demand models. Using a value of time derived from the model, this accessibility can be monetized.

It is also multi-modal. However it is not directly measurable, or comparable, and so is unlike cumulative opportunities (§1.2) measures.

⁵⁷ §8.1.

Unlike typical regression models, the outcome of choice models is a probability of making a choice (like the probability of choosing transit vs. driving vs. walking for a trip to work). This choice depends on the 'utility' of the choice-maker. The probability depends on how much benefit each alternative provides relative to the alternatives. That it is probabilistic rather than deterministic (the choice-maker just picks the best, or highest utility, alternative) is because there is uncertainty along the way: about the measurement of the inputs, about how the choice-maker perceives the inputs, and about how well the analyst measures the choice-maker's preferences.

The most widely used choice model in transport is called the 'logit' model, a particular version of which was formalized by Daniel McFadden in the 1960s and 1970s, and for which he was awarded a Nobel Prize in Economics in 2000 based on work he did to model the mode choice⁵⁷ of travelers in the San Francisco Bay area back in the 1970s.

2.11 Principle of Least Effort

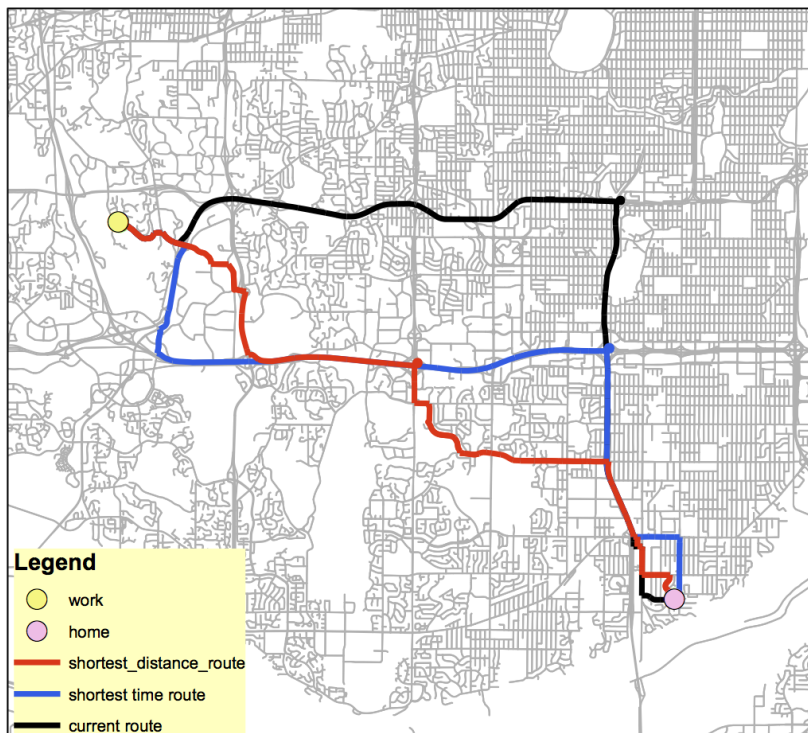


Figure 2.14: Alternative routes for travelers.

THE ‘PRINCIPLE OF LEAST EFFORT’ MAINTAINS THAT PEOPLE TRY TO MINIMIZE THEIR ENERGY WHEN ENGAGING IN AN ACTIVITY.

In concept, the principle of least effort is thus analogous to the principle of least action in physics. For instance, people may use the shortest path between an origin and destination. This assumption is embedded in more or less literal form in most [routing](#)⁵⁸ models.

⁵⁸ §9.

In fact, as in [Figure 2.14](#), we observe that people don’t use the shortest travel time path, for a variety of reasons, among them [time perception](#),⁵⁹ knowledge of the network, computational burden, search costs, making decisions emotionally rather than rationally, and caring about factors beyond travel time like [reliability](#),⁶⁰ the travel time of others, or weighting different elements of travel time (like stops) more than other elements of travel time (like moving at free-flow speed), which again leads to different behaviors.

⁵⁹ §2.15.

⁶⁰ §9.3.

Still, as a general idea, people will tend to choose the nearest satisfactory destination, the mode requiring the least time, cost, and inconvenience, and likewise a route, and certainly don’t do the opposite. Cities exist to maximize [access](#),⁶¹ the number of places

⁶¹ §1.

that can be reached in the least time (cost, effort), subject to individual preferences. By making things convenient, people can do more with less. Less time is spent traveling to reach specialized workplaces, vendors, customers, shops, churches, family and others. While certainly cities grow to the point that congestion becomes a headache, they still produce far more access than the alternative: people spreading themselves out to maximize the space between them.

2.12 Capability



Figure 2.15: Disability parking space. Source: <http://parkinglotstripingdenver.com>.

Based on work with Rania Wasfi.

THE ABILITY TO UNDERTAKE ACTIVITIES DEPENDS ON THE ABILITY TO ACCESS THE ENVIRONMENTS WHERE THOSE ACTIVITIES TAKE PLACE.

People who are limited in their ability to access are termed ‘transport disadvantaged.’ They may be disadvantaged for any number of reasons, including, for instance, physical or mental disability of some kind, inability to speak the common language, inability to drive because of youth or slow reaction times, or lack of resources to possess a private vehicle. ‘Disadvantaged’ thus is broader than a measure of disability that counts defects or impairments within an individual, and instead focuses on barriers people face interacting with the environment. It should be noted that while the most visible disabilities (like being in a wheelchair) attract a lot of attention (both socially and politically), and are allocated [parking spaces](#)⁶² as in [Figure 2.15](#), there are many invisible disabilities as well.

⁶² §7.16.

People with disabilities are a large and growing population whose needs must be considered for designs and plans to be successful. About 1 in 5 Americans has some kind of disability, and 1 in 10 has a severe disability. The number of people aged older than 65 is also steadily increasing. Because the population is aging and the likelihood of having a disability increases with age, the growth in the

number of people with disabilities can be expected to accelerate in the coming decades.

People with disabilities face more challenges interacting with the built environment than those without. Disability is a complex phenomenon involving interaction between features of the person and features of the overall context in which the person lives. The ‘social model of disability’ shifts the concept of disability from counting or categorizing defects or impairment within an individual to a focus on barriers people face within the environment.⁶³ The social model argues that activity limitations are not caused by the impairment, but rather by social institutions. For example a person with a vision impairment that prevents driving a car will likely rely heavily on public transit. If there is no accessible transit, this can limit participation in the community and the ability to live independently, shrinking the [space-time prism](#).⁶⁴ Reliance on other [modes](#)⁶⁵ like paratransit or taxis can become prohibitively expensive. Hence, increasing [access](#)⁶⁶ to services and facilities in the environment is an important aspect of ensuring full participation of people with disabilities in their communities and, in turn, improving their health and wellness.

Studies linking health and the built environment have concluded that the design choices we make in our homes, schools, workplaces, communities, and transport systems impact health.⁶⁷

⁶³ (Drum 2009).

⁶⁴ §2.9.

⁶⁵ §8.

⁶⁶ §1.

⁶⁷ (Dannenberg et al. 2003).

2.13 *Observation Paradox*



Figure 2.16: Multimodal Beijing.
Photo by D. Levinson.

PEOPLE CONSISTENTLY OVERESTIMATE THE **CROWDEDNESS**⁶⁸ OF TRANSPORT FACILITIES LIKE ROADS, BUSES, AND TRAINS.

⁶⁸ §6.4.

There is a logical reason for this overestimation, which we call the ‘observation paradox.’

Imagine there are 2 buses, one carries 49 people, the other carries 1 person. The average number of people on each bus, what the field calls the load factor, is 25 $((49+1)/2)$. There are, however, 49 people who think there are 49 people aboard, and 1 who thinks 1 is aboard. If you compute the perceived load factor: $(49*49 + 1*1)/50$, you get 48. So instead of an actual average load of 25, people perceive an average of 48 people on-board.

The same is true of roads; more people perceive roads to be congested because they themselves are in congestion (or as we might say, because ‘the people are congestion’). It’s even worse since speeds under congested travel conditions are slower, and if you weight travel by the number of minutes experienced (rather than by distance), congestion appears worse.

All of which is to say that anecdotal evidence is not a reliable measure of congestion or transit use, but objective measures will not align with individual experience. This creates policy and political problems and can lead to mis-investment.

2.14 *Capacity is Relative*

Figure 2.17: Tokyo subway pushers. Photo source unknown.



⁶⁹ We consider *anomalies* (§6.1) – such as the fact that drivers don’t necessarily follow the rule-of-thumb when it comes to giving two-second headways – and acknowledge moderate capacity increases in the real-world versus the theoretical. We take into account *HOV and HOT lanes*, (§6.10) in that increasing the number of passengers per car can increase the overall person throughput of a road. We also bear in mind the promise of autonomous cars may even combine shorter headways with increased loading.

⁷⁰ Packed vehicles also have slower load and unload times. This impacts running time and reliability and causes bunching.

⁷¹ §6.5.

⁷² §6.3.

‘CAPACITY’ IN THE TRANSPORT WORLD IS TYPICALLY CONSIDERED A RELATIVELY FIXED ELEMENT.⁶⁹

Capacity, however, is a relative term, much more so than we give it credit for. Figure 8.11 comes from a light rail train in Saint Paul (especially crowded as it was opening day). For those familiar with transit in most of the United States, this train is relatively full.

Figure 2.17 comes from Tokyo, where uniformed attendants routinely cram commuters onto a busy train while future passengers stand and look forward to the same mistreatment.

To those from Japan, the Saint Paul train is essentially empty. To those from the United States, the Japanese train is insanity.

Capacity has less to do with the physical amount of space in the train than what we are willing to tolerate in terms of personal space and comfort.⁷⁰ The same can be said for *Level of Service*⁷¹ (LOS), which is a metric based primarily upon driver’s comfort. LOS A is not better than LOS D from any perspective other than those driving. If every road in your city operates at LOS A, B, C, or even D, then your city is likely lacking economically and in terms of overall vitality. Moreover, few things help other modes become competitive more than LOS values of E or F. And as discussed in the section on *flux*,⁷² bikes are far narrower than cars, SUVs, and trucks. Thus, a seemingly full road might have a lot more capacity – even before Automated Vehicles – than we currently see possible.

2.15 Time Perception

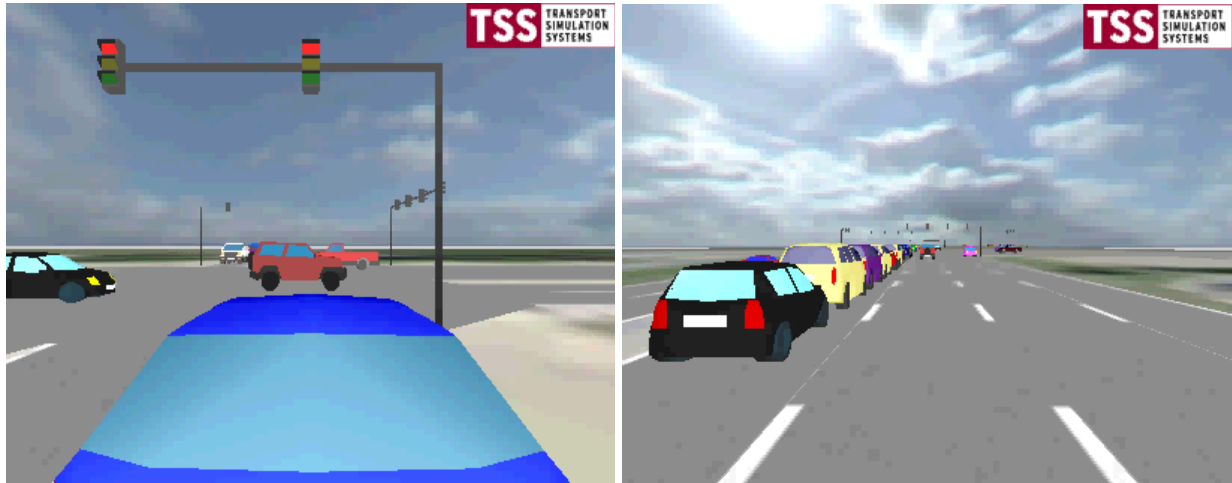


Figure 2.18: Screen shot of a traffic simulation. Source: (Wu et al. 2009).

TIME FLIES WHEN YOU ARE HAVING FUN; IT CRAWLS WHEN YOU ARE NOT.

Time perception and the quality of the experience, it should be no surprise, depends on the environment in which that experience occurs.

Sometimes people think places are farther away than they really are, and other times they think they are closer. Freeways seem to take less time than they really do, local streets longer. This, in part, has to do with task complexity, or the ‘mental transaction costs’ involved in traveling.⁷³

When making many small driving and navigation decisions, like on a signalized route with lots of turns, the driver focuses on the driving task more often. Each time, the driver engages her conscious brain in traveling decisions, and more brain-space is occupied by traveling thoughts.

Other factors include temporal relevance (is the trip important?), temporal expectancies (what does the driver think the travel time will be?), temporal uncertainty (how reliable is the estimate of travel time?), affective elements (what is the emotional state of the traveler?), absorption and attentional deployment (is the driver paying attention to the task at hand?), and arousal (how physically activated is the driver, is she on drugs?).⁷⁴

When driving on an uncongested freeway, many drivers avoid such thoughts. Driving is less salient. Time passes faster.

⁷³ (Parthasarathi et al. 2013).

⁷⁴ (Carrion and Levinson 2012a).

⁷⁵ (Wu et al. 2009).

Vierordt's Law also claims people are more likely to over-estimate short times and under-estimate long times. However, we did not corroborate this with a driving simulator study (illustrated in Figure 2.18) for waiting at a traffic signal.⁷⁵ Perceived and actual waiting time were virtually identical for the first 30 seconds, but for times greater than 30 seconds, actual waiting times were higher than perceived waiting times, up to 120 seconds. At 120 seconds, the trend was for perceived time to overtake actual time, but that was the cut-off for the experiment, so perception findings in this situation require more information. However, the annoyance level at 120 seconds of waiting was much higher than the annoyance of waiting 30 seconds. Further, people hated stops.

⁷⁶ (Levinson et al. 2006).

Of course, with all of this, it depends on how the question is framed, what is asked, and what travelers were expecting. For instance, comparing a computer-administered survey that asked about time preferences (in this case, comparing a mix of stop-and-go traffic with time waiting at a ramp meter) with one in which travelers were in a driving simulator completely flipped preferences for traveling (waiting for free flowing travel versus muddling through congestion).⁷⁶

⁷⁷ (Fan et al. 2016; Lagune-Reutler et al. 2016).

Time perception may be even more important for transit use. Real-time bus or train arrival information helps reduce the anxiety of uncertainty associated with waiting for transit. But more basic experience is important. People overestimate wait times in general. They overestimate it more when there are no amenities (like benches and shelters) and less when there are. When the environment is polluted and near high traffic roads, time is overestimated more. Women in particular overestimate waits in what are perceived as unsafe or insecure surroundings. The presence of trees reduces the travel time estimate.⁷⁷

The relevant time for individuals assessing their own accessibility is the perceived time. This differs from the objectively measured time used by the analyst.

2.16 Time, Space, & Happiness



Figure 2.19: Happiness.
Drawing by B. Levinson
Carpenter.

WE SPEND TIME TO AFFORD MORE SPACE.

We commute further to get more land. But the more time we spend traveling to our remote land, the less time we have to appreciate it.

If we work 8 hours per day, sleep 8 hours per day, the maximum daily commute would be 4 hours each way. In that extreme case, we'd be driving 4 hours for a bed and have no time to appreciate it (ignoring holidays and weekends). Thus, where we lived would only matter for days without work (aside from other family issues).

If we worked at home, we would have 0 minutes of commute with 8 hours to enjoy our home and neighborhood. Where we lived would be very important. This too is complicated by other family members work, school, etc.

People generally [choose under a 30 minute commute](#)⁷⁸ (each way), which theoretically leaves 7 hours a day to do other things. This includes both appreciating the neighborhood environment and the physical structure itself.

⁷⁸ §2.5.

Consider a hypothetical daily time budget, which is largely locationally independent:

- 8 hours sleeping
- 8 hours working

- 1.5 hours traveling (say for a worker, 60 minutes commuting and 30 minutes other travel)
- 1 hour at other out-of-home activities
- 3 hours in front of a screen at home
- 2.5 maximum number of hours to enjoy your location.

So every 1 minute less spent traveling is 1 minute more at the margin to enjoy your location. If an extra minute spent traveling (from 90 to 91 minutes say) reduces time available to enjoy the place (from 150 minutes to 149) minutes, we ask if those 149 minutes at the newer place are 0.67% 'better' than the 150 minutes at the older place. Maybe they are.

Spending 30 minutes more travel (15 minutes each way) reduces time available from 150 to 120 minutes. Now we have to ask if the minutes at the new location spent are 20% 'better.' It is hard to expect to be 20% happier, or 20% more likely that you will be happy, from physical surroundings when so much of your life will be similar.

The data on happiness are complicated. Individuals rarely quantify it, or think about it in these terms. But research finds that people in small towns near big cities are about 10% happier than people in cities.⁷⁹ That does not compare directly with our 20%, since it encompasses the happiness of the whole day, not just the marginal time. But even if commuting is the least pleasant thing people do, a little bit might be worth it.⁸⁰

An elementary school science fair project asked kids 2 questions: how happy they were (based on a Likert scale of 5 happy-to-sad face cartoon pictures) and then what mode of transport they took to school. The results showed that kids that walked or biked were significantly happier than those who were driven or rode the bus, see Figure 2.20.



Figure 2.20: Happiness Likert scale used in an elementary school.

⁷⁹ (Berry and Okulicz-Kozaryn 2011).

⁸⁰ (Crabtree and Index 2010).

2.17 Risk Compensation



Figure 2.21: Risk compensation theory predicts bicycle helmets won't make bicyclists much safer. Photo by Wes Marshall.

REDUCING RISK ENCOURAGES RISK-TAKING.

'Kids should always wear a helmet when they are bicycling.' It is hard to question that statement, and in many places, it is the law (even in a few places without mandatory motorcycle helmet laws). A funny thing happens when riders wear a helmet; they ride faster and more recklessly (and cars drive more closely).⁸¹ Why? The helmet makes the riders feel safer, and the risk of getting hurt (or hurting someone for the driver) subconsciously fades away.⁸² Thus, they behave differently. Changing one's behavior due to a change in perceived safety risk is called risk compensation.⁸³ But who is really safer: a reckless rider with a helmet or a careful rider without one?

Risk compensation is ubiquitous. You see it in sports like American football and hockey where additional protective equipment facilitates bigger hits.⁸⁴ You also see it in racecar driving, a notoriously dangerous sport, so much so that the sport has gone to great lengths to improve safety with better helmets, seat belts, roll cages, fire retardant uniforms, and soft-wall technology.⁸⁵ All of these efforts have reduced the chance of a fatality when a crash occurs. Yet, the research shows that as the casualty rate (per crash) drops, the number of crashes increases. Drivers can push the limits of their racecar to an even greater extent because they feel safer knowing that the risk of death or severe injury is relatively low.

In terms of road safety for the rest of us, the outcomes are not all

⁸¹ (Swaminathan 2007).

⁸² Studies also show that the health benefits of riding without a helmet generally outweigh the safety benefits of riding with a helmet, so if helmets discourage use, they are reducing public health. (De Jong 2012).

⁸³ (Adams 1995).

⁸⁴ (Keating 2001).

⁸⁵ (Potter 2011).

that different. For most of the last fifty years, the conventional approach to improving safety focused on vehicle improvements such as seat belts, air bags, and crumple zones – and road designs that were wide and straighter with increased sight distances and clear zones. This safety paradigm emanated from the US Congressional road safety hearings of the 1960s.⁸⁶ The new mindset focused on engineering measures – such as better vehicle and street designs – that were far easier to influence than the behavior of millions of drivers. While some of these efforts did in fact improve road safety, this was not always the case.⁸⁷

⁸⁶ (Nader 1972; Weingroff 2003).

⁸⁷ (Noland 2003).

Many of the so-called safer road designs, for instance, did not fulfill their promise. If behaviors remained constant, the underlying theory would have been successful. Unfortunately, a driver feeling safe on the road can profoundly impact behavior.⁸⁸ Whether the driver is more likely to speed, divert their attention from the road by talking on the phone or listening to music, or even fall asleep at the wheel, the research suggests that many such road safety improvements actually decreased overall safety.⁸⁹

⁸⁸ (Dumbaugh and Gattis 2005).

⁸⁹ (Vanderbilt 2008).

The problem is risk compensation. It's the same reason you now rely on your vehicle's back-up camera instead looking over your shoulder and using the camera to enhance the information you used to gather manually. For transport engineers and planners, the line between safe and unsafe is not always clear. To have a better chance for clarity, we need to better account for risk compensation – and the impact of the resulting behavior changes – in our designs.

The observant reader will note that risk compensation is the same idea as a **constant travel time budget**⁹⁰ coupled with **induced demand**.^{91,92} Faster speeds (at least in part) are used to increase distance not reduce time. Safer travel is used (at least in part) to increase speed, not reduce risk. An increase in speed has knock on accessibility benefits. People select a driving speed based on their feeling of safety. If they feel safer, they will drive faster, and thus have more accessibility via automobile. Similarly, bicyclists choose routes based on a trade-off of safety and speed. The safer they feel, the faster their trip, and more accessibility via bicycle results.

⁹⁰ §2.5.

⁹¹ §12.1.

⁹² There is one difference with induced demand. Induced demand won't increase total travel time: if the new conditions were slower than the baseline, the trip wouldn't occur. Risk, however, has external effects. Perceived risk for drivers might reduce at the cost of higher pedestrian/cyclist risk, or for people in smaller cars, actual and perceived risk might rise with a reduction in perceived risk of driving a larger vehicle.

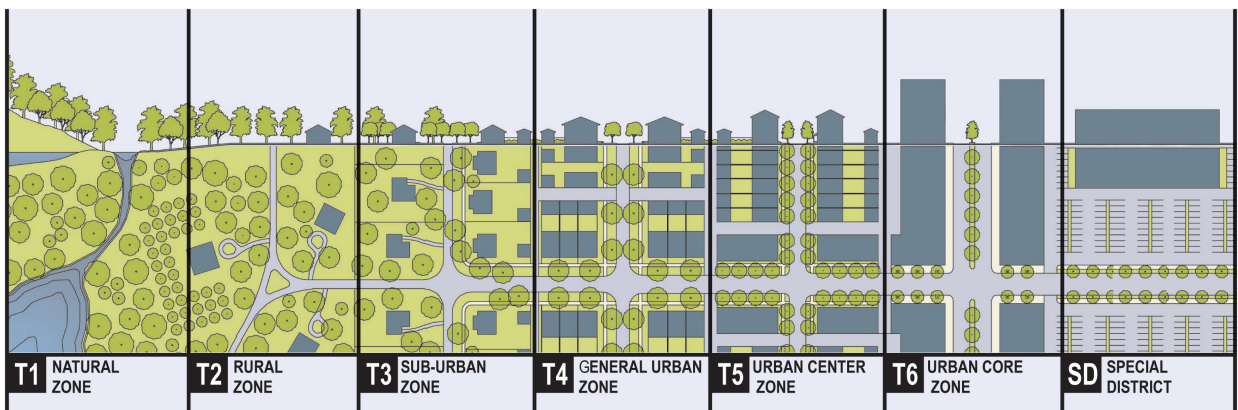
Part III

The Places

Now that we've learned about the people, the next building block of understanding accessibility is about the places we go. Since not everything we want to do nor everyone we want to see is at our fingertips, we need to be able to access the locations where these things are or could be. This means delving into where we live, where we work, and where we play; how we build our regions and cities; and how proximity and land use decisions factor into the transport choices that are viable for people. The Places is where this happens.

3

The Transect



THE TRANSECT IS A SPECTRUM OF CONTEXTS USED FOR TRANSPORT AND LAND USE DESIGN.

At one end of the continuum is the natural/agricultural area and at the other end is the central business district. In between, we find varying urban intensities ranging from rural to suburban to more urban zones.

The current popularity of transect-based design, like Figure 3.1, emerged from Andres Duany's work and the Congress for the New Urbanism. However, similar analytical tools were used as early as 1793 with the Prussian naturalist and explorer, Alexander von Humboldt. Figure 3.2 depicts his transect-based, vertically-magnified look at the tip of South America from the Atlantic to the Pacific oceans. Other naturalists employed similar concepts such as Ian McHarg in his influential 1963 book, *Design With Nature*.

In many instances, the transect is used in conjunction with

Figure 3.1: Transect. Image due to Duany and Plater-Zyberk. Source: https://transect.org/rural_img.html/.



Figure 3.2: The transect concept of Alexander von Humboldt, c. 1793.

¹ §7.1.

² §7.2.

³ §3.1

⁴ §3.2.

⁵ §3.3.

⁶ §3.4.

⁷ §3.5.

⁸ §3.6.

⁹ §3.7.

¹⁰ §3.8.

¹¹ §3.9.

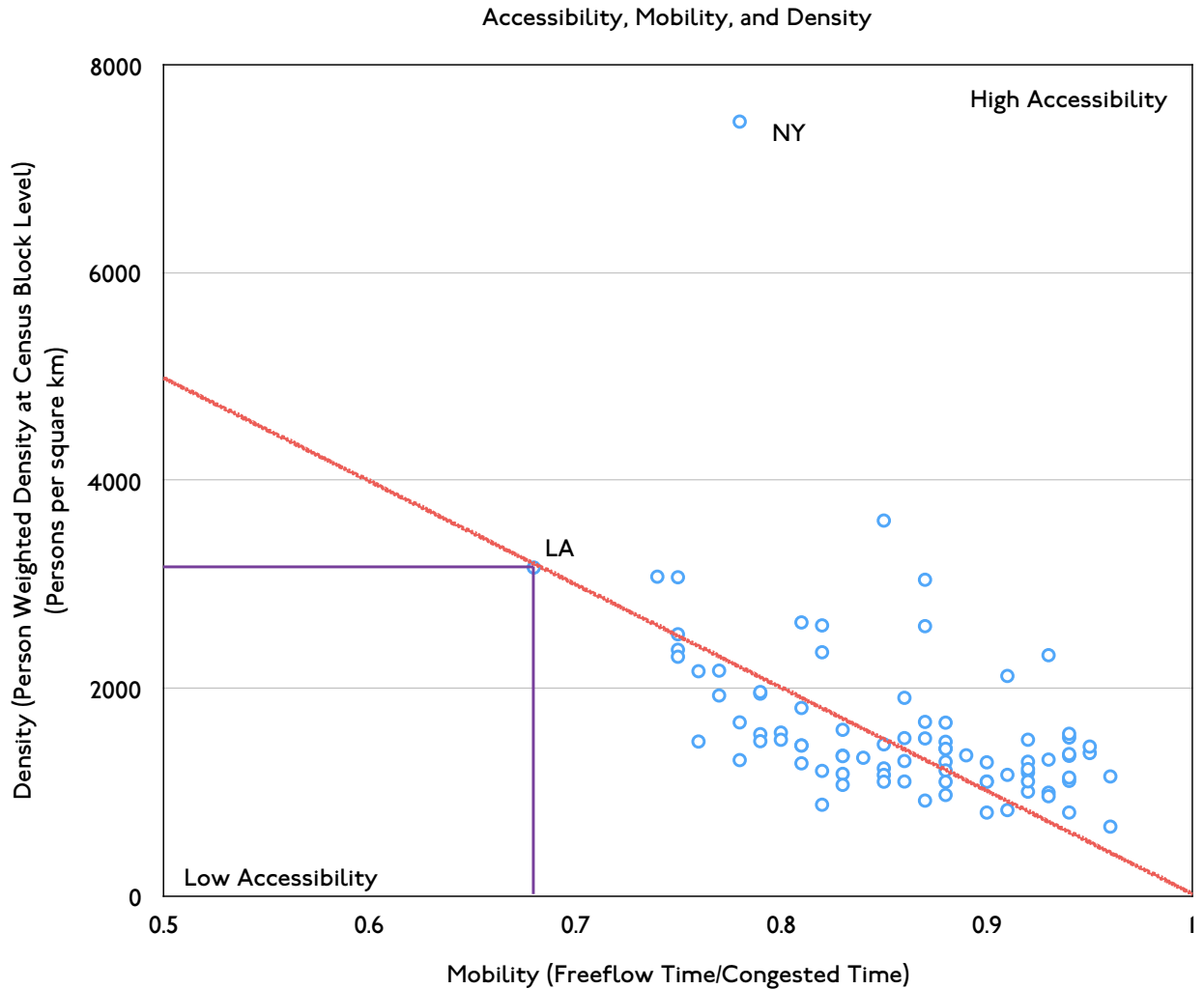
form-based zoning codes. However, the big picture idea is that we should base our designs – in terms of both buildings and transport – on context much more than conventional Euclidean zoning or functional classification. For instance, residential land uses in terms of placement, height, frontage, as well as general character should differ across the rural-to-urban spectrum. When moving towards the right side of the spectrum, housing is more likely to be attached and oriented to the street with smaller setbacks.

Transport design needs to change across this continuum as well. Prioritize walking, bicycling, and increasingly higher-capacity and more frequent transit in the more urban transect zones. However, great design also includes altering the character of a street. One useful example of this concept is US-50 in the Washington DC area. Well beyond the city limits, where the context is rural T2 and suburban T3, US-50 is a [limited-access highway](#).¹ As US-50 moves into the T4 general urban zone near Washington DC, it transitions into a typical arterial road. Once US-50 moves into the city, it becomes New York Avenue. While still large, US-50 has become a much more urban street with sidewalks, medians, on-street parking, and [street trees](#).²

The transect neatly maps to the idea of [residential density](#).³ which is higher in the center of the city and [decays with distance from the center](#).⁴ The center of the city, the old historic core, was often developed around the [pedestrian](#).⁵ Select [neighborhoods](#)⁶ might still be. Larger areas can be designed to be traversed by [bicycle](#)⁷, as is often done in some northern European cities. Safe separated [bicycle networks](#)⁸ can be retrofitted into cities that cannot be dedicated solely to that mode. Cities from the late 19th and early 20th centuries are often lower density than older ones because they were designed as [transit cities](#).⁹ The key factor in such cities is the ability to [walk to transit](#).¹⁰ The lowest density areas are built around the [automobile](#)¹¹ and map to the suburban built environment. The access that is realized by potential travelers in a part of the transect depends on the mode around which it operates.

Context-based thinking needs to be considered with nearly all of our transport and land use decisions and designs. Often, our work even needs to be based on what we hope our places will become as opposed to their current context. Whatever the case, the transect can be a useful tool.

3.1 Residential Density



DENSITY IS A MEASURE OF SOMETHING PER UNIT AREA OR VOLUME.

When looking at cities we might be interested in residential density, measured as population per square kilometer, or in the case of traffic, the number of vehicles per lane-kilometer. How it is measured, and averaged, can affect the result. So when comparing reported densities, it is important to understand how each was measured. Here, we illustrate with population density.

The density of the United States as a whole averages high density areas (like New York City) with low density areas (like Alaska) as

Figure 3.3: Accessibility, mobility, and density. Figure by D. Levinson, Source: (Levinson 2012).

Area	Density (persons per km^2)
50 United States (2015)	35.0
New Jersey (densest state)	467.2
Alaska (least dense state)	0.5
Minnesota (30th)	26.6
Los Angeles Urbanized Area (densest urbanized area)	2,702.5
New York Urbanized Area (5th densest in US)	2,053.6
Minneapolis-St. Paul (MSP) Urbanized Area	1,001.7
MSP, MN-WI Combined Statistical Area (19 counties) (2010)	128.0
MSP MN-WI Metropolitan Statistical Area (16 counties) (2015)	199.0
Metropolitan Council of the Twin Cities Region (7 counties)	390.1
Hennepin County	778.5
City of Minneapolis	2,890
Prospect Park, Minneapolis	2,300
One house in Prospect Park (residents divided by floor area)	16,666.7
Minneapolis census block-based person-weighted density	3,306.7
Minneapolis census block group-based person-weighted density	1,888.9
MSP Metro Area census tract-based population-weighted density (2010)	1,306.8

Table 3.1: Residential densities in different areas. Source: US Census.

shown in Figure 3.3 and Table 3.1. This density is a straight-forward calculation but not exactly how people perceive the world.¹² People tend to live near other people, so we perceive a higher density than a national average suggests. The concept of person-weighted density has been suggested, which instead of summing total population and total area, multiplies the population density in a smaller area by the number of people who are there to experience it, and then sums that across the total area and divides by the total population. This person-weighted density is higher than the unweighted density, and closer to the perceived density, but it still raises questions about which areas to sum up: a parcel, a city block, a block and its neighbors, a block group, a census tract? There is an arbitrariness to this, but the most important part is internal consistency. [Cumulative opportunity accessibility](#)¹³ measures address this in a much more systematic way.

¹² See the [observation paradox](#) (§2.13).

¹³ §1.2.

As shown in Figure 3.3, density is inversely proportional to mobility. Areas with high population density tend to have higher traffic density and more congestion. At the same time, high congestion can lead people to value location efficiency and help spark increased residential density.

Later we discuss [traffic density](#).¹⁴

¹⁴ §6.4.

3.2 Urban Population Densities

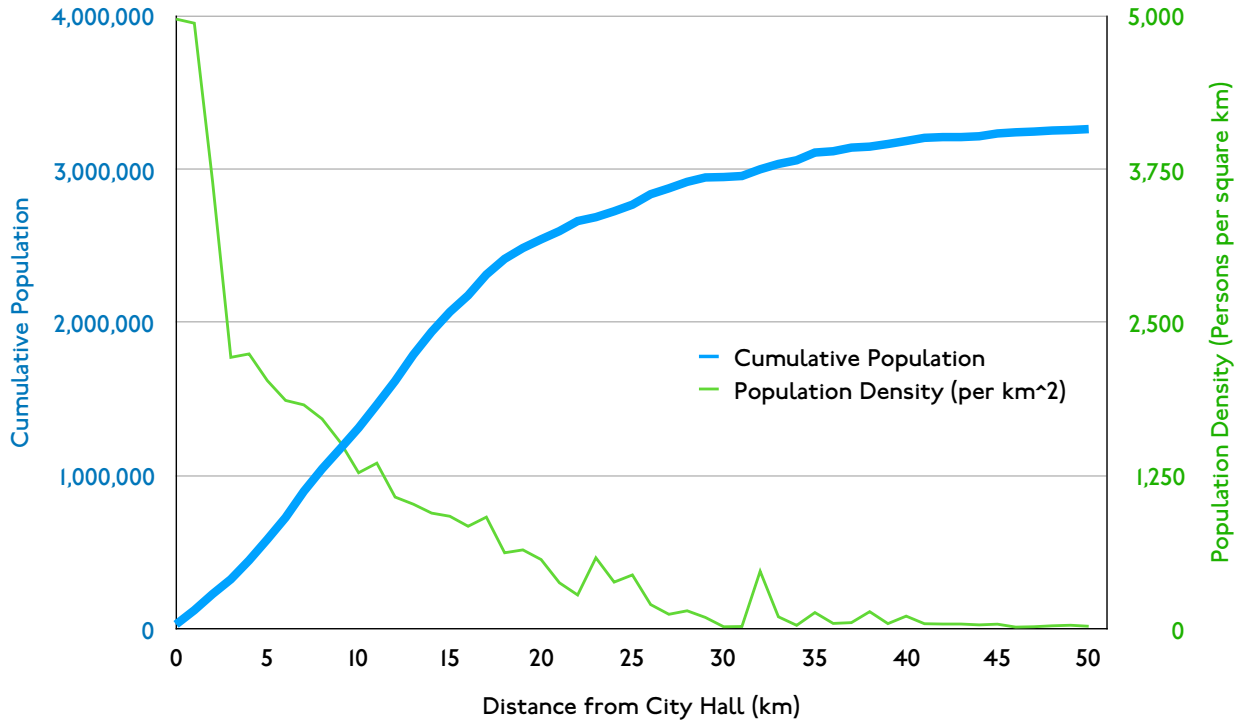


Figure 3.4: Density declines with distance. Source: US Census Data.

Total population (or jobs) increases with distance from the center of a region (as it must). The more area there is, the more people can and do live there, even if they do so at decreasing densities. It appears that for a large distance, each ring has approximately the same population. Julian Marshall observed that “newcomers to urban areas occupy twice the land area per capita of existing residents.” (Marshall 2007). Eventually, the city comes to an end, and rural densities take over, so as density comes to a minimum, the growth in population slows down.

¹⁵ §3.1.

¹⁶ The center here is functional, not geographic.

¹⁷ (Clark 1951).

POPULATION DENSITY DECLINES WITH DISTANCE FROM THE CENTER OF A REGION

An example shown in Figure 3.4, for person-weighted density¹⁵ from the Minneapolis region, with a peak population density of 5000 persons per km^2 , at a distance of 0 km to a density near 23 at a distance of 50 km. The center of a city is more valuable – by definition, the most valuable place to be near becomes the ‘center’¹⁶ – and so more people want to be there, or near there, to minimize their transport costs. Rents are higher there, as is population, and especially, employment density. Land farther out is less expensive, so people can afford more space. This phenomenon has been observed for as long as there have been cities and was quantified by Colin Clark.¹⁷

3.3 Pedestrian City



Figure 3.5: The area enclosed by the Walls of Rome, which ran for a length of about 19 km (12 mi) surrounding an area of 13.7 km^2 (5.3 mi^2). Source: Wikipedia https://en.wikipedia.org/wiki/Aurelian_Walls.

ROME WAS FAMOUSLY BUILT ON SEVEN HILLS.

These hills were not originally one metroplex but rather small villages, populated by different groups, that later melded into a single *conurbation*.¹⁸ For instance, the population of Alba Longa was settled on the Caelian Hill. Some of the Sabine tribe lived in a small village on the Quirinal Hill in the 7th and 8th centuries BCE, and it was incorporated into Rome in the 6th Century during the reign of Servius Tullius. This occurred after the legendary “Rape [or Abduction] of the Sabine Women” when men led by Rome’s founder Romulus sought brides from the nearby Sabine tribe and were refused by their fathers, leading to a conflict mediated by the wives/daughters.¹⁹ The families of the abducted women may have been those who migrated to the Quirinal.

As Rome grew, these villages interacted with each other more and more. One imagines that as populations settled the hills, there was some trading taking place, but most people would work locally in this once agricultural society.

The hill Quirinal is named for a Sabine god of war, Quirinus. The word ‘Quirinus’ comes from ‘quiris’ the Sabine for “spear.” Etymology Online reports “Ancient folk etymology ... traces [the word ‘cry’] to ‘call for the help of the Quirites,’ the Roman constabulary.” http://etymonline.com/index.php?allowed_in_frame=0&search=cry. The Quirites were men of the oaken spear. The word for ‘oak’ is ‘Quercus’ in Latin.

¹⁸ §14.2.

¹⁹ (Beard 2015).

How populous was Ancient Rome? Estimates vary widely.

Storey estimates a peak population of 450,000; on the higher end, Carcopino cites estimates over 1,200,000, (over an area larger than the walled city of Figure 3.5) but suggests that is too high.²⁰ In part, these estimates vary because the area of Ancient Rome varied over time. In part because it's not clear what the Roman Census was collecting data on, whether it was just adult male citizens or the entire human population. It's also not clear how many Roman citizens were residents of the City of Rome, living within the Aurelian Walls.

Let's illustrate with some assumptions:

- ²¹ §2.5.
- One-way travel time budget²¹ ($B = 0.5h$)
 - Walking speed ($S = 5km/h$)
 - Walking network radius ($R_n = S/B = 2.5km$)
- ²² §10.9.
- Network circuitry²² ($C = 1.25$)
 - Walking Euclidean radius ($R_e = R_n/C = 2km$)
 - Walking Euclidean area (potential) ($A_e = \pi R_e^2 = 12.56km^2$)
 - Density ($D = P/A_e$) = unknown
 - Population within travel time budget threshold ($P = DA_e$) = unknown

This area is a walking city if you start from the center. This is close enough to the $13.7 km^2$ enclosed by the walls that we can call Rome a 'pedestrian city.' Someone on the central of the seven hills of Rome, Palatinus, could reasonably 'commute' to the walls and back each day. However, it would take about an hour to cover the diameter of the city.

- ²³ §3.1.
- So the question turns on the assumed population density.²³ As a point of reference, the 2012 population density of Manhattan is $27,227/km^2$, which is enabled by 19th century technologies like elevators and rail transit. Obviously parts of Manhattan are higher and other parts lower. Rome currently has a population density of $2,101/km^2$. At the extreme, the infamous Kowloon Walled City in Hong Kong had some 30,000 people on 2.8 ha ($0.028 km^2$) in the 1960s, giving a density of over 1 million people per km^2 .

If we go with Storey's low-end population estimate of 450,000, this gives a density of 33,000 persons per km^2 , higher than contemporary Manhattan but in line with the density that immigrants on the Lower East Side experienced when Manhattan's population peaked in 1910,

at $39,208/km^2$. And of course, it is higher in places where people live, as there are lots of places where people do other things (work, shop, watch gladiatorial battles, or lounge about and eat peeled grapes). While not impossible, it does imply teeming streets. Estimates of twice that or more seem implausible, but we emphasize that none of us lived there and no one has a time machine (and if we had a time machine, doing an accurate Census of Ancient Rome is not the highest priority).

Also, one can have a pedestrian city that exceeds the one-way walking travel time budget but not a city one interacts with on a daily basis. This is more the equivalent of adjacent and overlapping cities, which likely had multiple cores, perhaps one per hill.

3.4 Neighborhood Unit

Figure 3.6: Town Plan of Radburn, New Jersey. Source: (Regional Plan Association 1929).



TOWN PLAN RADBURN, N.J.



FIG. 15
RADBURN, NEW JERSEY—PLAN OF MODEL COMMUNITY

Courtesy of the City Housing Corporation

[134]

THE TRADITIONAL PRACTICE OF URBAN DESIGN WAS THAT THE HOUSE WOULD SHOW ITS BEST SIDE TO THE STREET AND ENGAGE AND WELCOME THE VISITOR AND PASSERBY.

The rooms inside the building would match this by having the reception rooms facing the street façade and by being recognizable in the fenestration and with suitable detailing of the front. This attitude was maintained throughout the 19th century in spite of the noise of the horses and horse-drawn vehicles, and of the hubbub of pedestrian crowds and the cries of the peddlers.

With increasing wealth and the increasing demonization of the urban environment as dangerous, filthy, and unhealthy, urban middle classes removed themselves to the suburbs but initially maintained the style and expressiveness of their dwelling, emulating the rural villas of their much richer role models. The generally generous lots gave the space to place the house at a distance to the road and from each other, and to produce a park-like environment in the early and more expensive, low density²⁴, hard-to-serve-by-transit²⁵ suburbs.

²⁴ §3.1.

²⁵ §3.9.

The arrival of the car and the much broader urban exodus of the early 20th century changed the equation as streetscapes became less interesting. While many American suburbs maintained a public face through porches and porticos, the reversal of the order began: entry from the back via the garage. Back alleys were now used by the house owners and not as earlier by the servants, coachmen, and delivery boys.

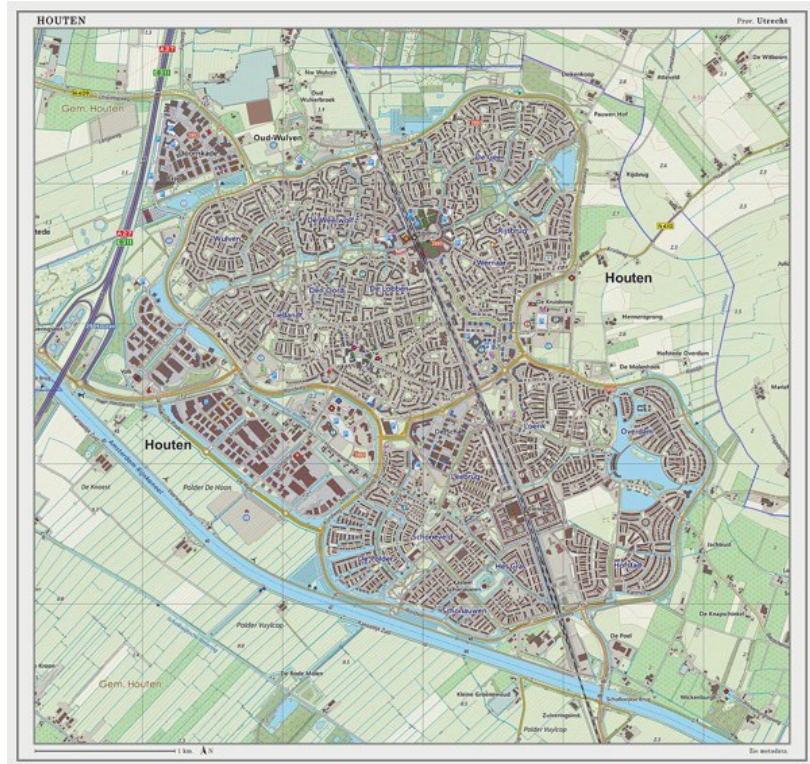
In Europe, both apartment blocks and suburban houses began to reorient themselves towards the garden in the back. The new desire for maximum sunlight invited a flowing transition from the ground floor living room into the garden instead of a first floor reception and dining room with kitchen and pantry on the ground floor together with the servants.

In this context of increasing expectation of quiet and sunlight, the New York designers Clarence Stein and Henry Wright went one step further and reversed the order completely. Their 1928 design for parts of Radburn, (Figure 3.6) New Jersey opened the house to a common park, which flowed into the garden of the house. The front entry would be from a cul-de-sac accessible garage and parking space. The residents, and in particular the children, could walk away from the cars to their destinations, as the different parks were linked by suitable underpasses and overpasses to avoid the streets.

The original design covered only a small area, but the design idea was applied at much larger scales in many post-war suburbs around the world. At this scale, it became clear that the loss of the streetscape was in many cases not balanced by an active communal use of and life in the parks. Many garden gates were locked, the view into and from the park was blocked by hedges and fences.

3.5 *Bicycle City*

Figure 3.7: Map of Houten, Netherlands.



HOUTEN, NETHERLANDS IS BUILT AROUND THE BICYCLE.

Building on an old settlement, Houten was initially planned as a city of 30,000 people. Constructed from the 1960s onward as a reliever for Utrecht, it is connected by a short rail line with two stops in the town. Unlike bikes, cars cannot cross the town but can circumnavigate on a ring road, as shown in Figure 3.7. The industrial and commercial sector is in the southwest of town, with good highway access. Though there is a [balance](#)²⁶ of jobs and workers, most residents work outside the town and most workers commute in, which is not surprising given its rail and highway connection with the rest of the Randstad. The architecture and feel of the place is otherwise very familiar to anyone who has visited a planned US, French, or UK new town from the same era (without the single family homes, most of the buildings are townhouses or apartments).

David toured it on bike one afternoon during the 2014 World Symposium on Transport and Land Use Research with colleagues

and generous local officials. These are his observations:

1. The center of town is the main train station (which was recently rebuilt). The number of tracks were increased, and the station was elevated so it was easier to cross east-west.
2. Under the train station is an enormous bicycle parking facility: Fietstransferium.
3. There are many bike paths through town. Small humps are used to discourage cars, which are prohibited, and motor scooters and mopeds, which are as well, but seem common.
4. The best, most vibrant part of the town is the old town, indicating there is much planners need to learn about recreating places.
5. There are some shared roads, though most prohibit motors officially.
6. The newest part of town is centered on Castellium, inspired by a Roman town.
7. One development is inspired a Norwegian Fjord town.

Some colleagues felt the town too 'sterile,' which is the rap given to new towns, and especially suburbs, everywhere. It is not clear what planning academics are looking for, hypodermic needles on the street? It is of course a suburb of Utrecht, so the core city functions – especially entertainment and culture, will agglomerate there, as cities are where the childless youth seek to find mates. To conduct pop psychology and apply two of the Big Five personality traits, this is a classic case of a trading off Openness to new ideas, which involves exposure to risk, and cities, and Neuroticism, which is fear based, and wants to minimize risk, and seeks more controlled environments (loosely, planned communities or suburbs), which at some level is in part correlated with age and parenthood.

3.6 *Bicycle Networks*

Figure 3.8: Protected bike lane in Vancouver, British Columbia, Canada. Photo by Wes Marshall.



THE CONVENTIONAL LOW-HANGING FRUIT APPROACH TO BUILDING UP A CITY'S BIKE NETWORK WAS TO FIRST LAY DOWN BIKE LANES ON STREETS THAT HAVE THE ROOM AND THEN BY FILLING IN THE MISSING CONNECTIONS WITH SHARED-LINE MARKINGS COMMONLY REFERRED TO AS 'SHARROWS.'

The problem is that establishing a bicycle-accessible city requires more than a network that looks connected on a map.

While some bicyclists feel comfortable riding on almost any road, the vast majority of bicyclists simply don't feel like they can ride right next to relatively fast-moving cars. If a certain trip requires riding on such a street, bicycling is no longer a viable option. So if a city is comprised of bike lanes or cycle tracks as in Figure 3.8 that I do feel comfortable riding that are connected by sharrows on higher-speed streets where I don't feel comfortable riding, I end up confined to a small island of bike-friendliness. That might be fine for recreationally riding around a bit, but it makes it difficult to actually get anywhere and limits bicycling as a utilitarian mode.

Think about a ski mountain that assigns one of four levels of difficulty to each of their slopes: the green circle for beginners; the blue square for intermediates; the black diamond for advanced; and the double black diamond for experts. Now if I'm the type of skier

that would only ride on beginner and intermediate slopes and I take the chair lift to the top of the mountain, I need to be able to get down the mountain using only green and blue slopes. If I start down on an intermediate slope and suddenly reach a point where my only options become black diamonds and double black diamonds, I'm in trouble. Thus, ski resorts make a concerted effort to set all their users up for success.

The same should go for bicycling in cities. The new wave of separated bike facilities are great and help make our city streets – as the former parks commissioner of Bogotá, Gil Peñalosa, would say – 8 to 80 places, where both an 8-year old and an 80-year old can move safely and enjoyably. More often than not, however, we connect these 8 to 80 bicycling streets with double black diamond streets and intersections. These missing links negatively impact people's choice to bicycle in the first place.²⁷

How can we do a better job of connecting our bike networks?

To begin with, we can link our streets and the array of different bike facilities that are now in the bike planning toolbox with who can actually use them in order to get a better understanding of how the bike network functions. For instance, the bicycle level of traffic stress approach, shown in Figure 3.9 – classifies streets based upon the bicycle level of traffic stress (LTS) that they exhibit to the user.²⁸ The methodology uses characteristics such as physical separation, operating space, speed of adjacent traffic, and intersection treatments to assign one of four traffic stress levels to street segments and intersections. By connecting the LTS classification scheme with what has become known as the Portland typology, we can better understand bike network problems.²⁹ The Portland typology categorizes individuals into four basic bicyclist groups: no way, no how; interested but concerned; enthusiastic and confident; and strong/fearless. If I'm an 'interested but concerned' bicyclist who only feels comfortable riding on cycle tracks, bike lanes, and slow-speed streets, asking me to share a lane with fast-moving traffic is like asking an intermediate, blue rectangle skier to head down a steep, icy, mogul-filled run. In the eyes of the rider, the perceived safety risk is the reality. Being forced to ride on streets beyond my comfort zone might just push me back into my car. Just as [reducing perceived risk increases risk-taking behavior](#),³⁰ increasing perceived risk reduces risk-seeking.

So if we want 8 to 80 cities – rather than just a handful of 8 to 80 streets – we need to take a network-level approach to bike accessibility and make sure we don't leave ourselves with disconnected islands of low-stress bike facilities.

²⁷ (Schoner and Levinson 2014).

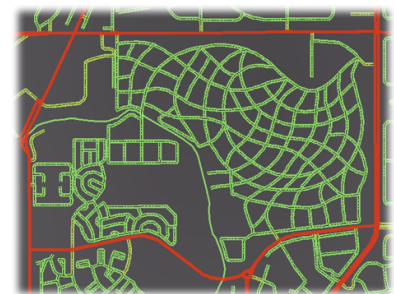


Figure 3.9: Low-stress island created by high-stress roads and intersections in Denver, Colorado. Source: (Bronson and Marshall 2014).

²⁸ (Mekuria et al. 2012).

²⁹ (Dill and McNeil 2013).

³⁰ §2.17.

3.7 Transit City

Figure 3.10: Historic streetcar at Excelsior Works, Excelsior, Minnesota. Photo by D. Levinson.



IN THE UNITED STATES, REGULAR, FREQUENT TRANSIT SERVICE WAS ONCE FEASIBLE FOR NEIGHBORHOODS OF SINGLE FAMILY HOMES.

It is still feasible where economic conditions are favorable. This section conducts some back-of-the-envelope calculations to illustrate the phenomenon.

LAND USE.

³¹ §11.1. Consider the 1-mile (1,600 m) gridded³¹ landscape that is common in the midwest and western United States due to the Northwest Ordinance of 1785. This grid is largely the backbone network of streetcar era land use design.

While there are a variety of ways this grid can be carved up, one common way is to have:

- 10 cross-streets per mile of grid in the long direction (520') (160 m); and
- 20 cross-streets per mile of grid in the short direction (260') (80 m).

This arrangement produces about 200 blocks per square mile (77 per km^2). The size of each 520' x 260' block (center line – center line)

is 135,200 square feet (12,560 m^2).

Considering street and alley space, lots typically have a 40' frontage with 110' depth = 4,400 sq. ft. ($\sim 1/10$ acre) (408 m^2).

This spacing gives 12 houses per block face in the long direction, or 24 houses per block. In this configuration, no houses face the short direction. If there were only housing, this would give 4,800 houses per square mile ($\sim 1,875$ per km^2) after accounting for roads and alleys.

At 2.5 persons per household, typical of the US, this gives us 12,000 persons per square mile (PPSM) ($\sim 4,700$ per km^2) in single family homes at typical built densities.

While some space is devoted to schools, parks, retail, commercial, and industrial activity, among other uses, this should be persuasive that 10,000 PPSM is feasible over large areas without Manhattan-like high density. The City of Minneapolis, for instance according to the 2010 Census, has a density of 7,417 PPSM ($\sim 2,900$ per km^2). At its peak population, it had over 10,000 PPSM (3,900 per km^2).

TRANSIT.

The target density for successful transit is often given as 10,000 PPSM.³²

If we assume that every person originates many short trips (which can be dealt with by walking or biking) and one long trip per day (say going to work, or school, or shopping), the 10,000 PPSM would generate 10,000 transit trips per square mile. So we have 10,000 boardings per square mile. This is roughly streetcar era demand in cities.

If we space transit routes every half-mile (as was typical of streetcars), both east-west and north-south with stops where transit routes crossed and halfway between,³³ the square mile area is served by 21 stops. The four stops at the outer corners are shared with 4 other areas, and the 8 non-corner stops at the perimeter are shared with 2 other areas, while 5 stops are internal to the 1-mile square. This gives us 12 equivalent dedicated stops for the area.

With 10,000 PPSM and 12 stops, each stop serves 833 people per day. If transit vehicles carry 50 people each, that is 17 full transit vehicles per day. Of course transit vehicles do not generally fill up at one transit stop, and over a 17-hour day (assuming no night service), this would be 1 transit vehicle every hour.

If instead we wanted service at 10-minute headways but full vehicles, we would expect each vehicle to fill up 1/6 of its load at each stop (or about 8 passengers per stop). That would be a much higher load factor than generally observed.

³² (Pushkarev and Zupan 1977).

³³ That is 1/4-mile (400 m) spacing between stops.

The maximum walking distance to a transit stop would be 0.35 miles (560 m), and the average walking distance would be 0.175 miles (280 m).³⁴

$$\sqrt{(0.25^2 + 0.25^2)} = 0.35. \quad \text{³⁴ by Pythagoras}$$

MODAL COMPARISON.

So what guarantees people will make 1 transit trip per day? If there were no good alternative, as in the peak streetcar era (Figure 3.10), this is an easy choice.

Today, this depends. The argument for using transit is that in our idealized grid-like city with a grid transit system, the transit system is as direct as every other mode, so there is no lost distance due to **circuity**.³⁵ The only lost time is the schedule delay (which is a maximum on average of 5 minutes, less if people can time their wait to match the transit vehicle), and the time when the vehicle is stopped (and accelerating and decelerating) boarding and alighting passengers, and the transfer time between vehicles. An idealized grid requires, at most, one transfer. Again with a headway, the time between vehicles of 10 minutes gives an average transfer wait of 5 minutes, less if the routes are timed well. Finally with any transit advantages (such as signal timing priority, exclusive lane, or stopping in lane, as opposed to weaving into and out of stops), transit can recover some of the time lost vis-a-vis the automobile.

Where transit is better (faster, cheaper) than alternatives, and frequent enough, people will use it in large numbers. This is observed daily in large cities. Thus, it must be possible to obtain faster, cheaper, and frequent enough service levels. In most places in the US, the transit service and ridership is not there. Let's work through an example.

- 10 minutes: acceleration, stop, and deceleration. For a five mile trip, there will be about 20 stops at 1/4-mile (400 m) stop spacing. If each stop results in 30-seconds lost time (2-3 seconds per boarding plus acceleration/deceleration), that is 10 minutes of time lost there. This will generally result in longer times than an automobile, even with stop signs or red lights every 1/4 mile (400 m);³⁶
- 5 minutes: initial schedule delay, assuming random arrivals;
- 4 minutes: walk access time for the average passenger (walking 0.175 miles (280m));
- 4 minutes: walk egress time to the final destination, though perhaps lower for downtown workers; and

³⁶ The time spent stopped at the stop for a car will be less than for transit, even with transit pre-payment. Unless the auto is stuck behind a transit vehicle and cannot pass, in which case the stopped times are the same.

- 5 minutes: transfer time, on average, if it is effectively uncoordinated.

So now even with our idealized transit system, we have lost something like 10+5+4+4+5 minutes or up to 28 minutes longer than the car for a 5-mile (8 km) trip.

At a value of time of \$15/hour (\$0.25/minute), this is the equivalent of \$7. If the transit fare is \$2, and the cost of gas (at \$5/gallon (\$1.25/liter) and 25 miles per gallon (10 km/liter) is \$1 (not even considering carpooling), net additional out-of-pocket cost for transit is now the equivalent of \$8. Of course, vehicle ownership (\$10-\$20/day) can be avoided, as can parking charges. Also, we are not considering externalities.³⁷

³⁷ §12.4.

THE EXPRESS.

If demand is high enough, we can make transit go faster and have a higher frequency. This is accomplished by having the vehicle stop less often and/or giving it a limited-access right-of-way.

One disadvantage of express routes is a longer access/egress time. Stops can't be spaced as close together if lines are to achieve economies of scale, so stations are on a 1 mile instead of 1/2-mile (800 m) spacing, at best. If that access and/or egress is by transit itself, that imposes additional scheduling time penalties.

We can compensate because now our land use changes take advantage of the express services. At express stations, densities rise. Apartments replace single-family homes.

Express buses and commuter trains often have low frequencies, while modern or modernized subways may have one train every 2 minutes or better.

So if we increase the highest distance to a station for 1-mile (1,600 m) spacing between stations and 1 mile between routes (so every station is a transfer), but increase the frequency to one transit vehicle every two minutes, we get the following:

- 2.5 minutes acceleration, stop, and deceleration (for a five mile trip, there will be about 5 stops at 1-mile stop spacing);
- 1 minute: initial schedule delay, assuming random arrivals;
- 7 minutes: walk access time for the average passenger. The walk access time is twice that of the local transit above, or 0.35 miles (560 m). At 3 mph (5 km/h), this is a walk time of 7.1 minutes on each end, though changes to land use patterns could reduce this;
- 7 minutes: walk egress time (though again, changes to land use patterns may change this); and

- 1 minute: transfer time, on average, even it is effectively uncoordinated.

For a 5-mile trip with transfer, we now lose only 18.5 minutes. This is less than the local transit service above and can be reduced if more people live closer to the station rather than spread out uniformly across the landscape.³⁸

³⁸ If we have higher travel speeds by transit than auto (let's say averaging 45 mph while in motion on exclusive right-of-way instead of 30 on surface streets), for a 5 mile trip the express transit time is 6.67 minutes instead of 10 minutes. But this 3.33 minute savings does not outweigh the lost time due to access and waiting costs. This does not even begin to consider the additional costs of operating express vs. local services, or revenues from the service.

To reduce transport costs with transit-like services, we can arrange cities linearly, thereby eliminating transfers and reducing access costs. This wastes potential accessibility for non-transit modes.

Optimal urban form depends on the technology being optimized. In a city where driving is perceived to cost \$1/trip, and cars save between 18 and 28 minutes per trip compared with transit, it is no wonder the automobile is the dominant mode for long distance trips, even in historically transit advantageous places. Reducing automobile dominance requires changing the perceived (and real) cost of driving for drivers, as there is little that can be done on the transit supply side that will make a significant difference in the absence of that for most markets.

In dense areas, the market takes care of the cost structure by providing expensive parking. In low density areas, there is enough room for everyone's car without charging.

Systematically re-arranging existing cities for transit (or any mode) is putting the cart before the horse. Transport should serve activities, and while transport and land use co-evolve, that co-evolution is slow (over decades) and should be adaptable to alternatives.

3.8 Walkshed

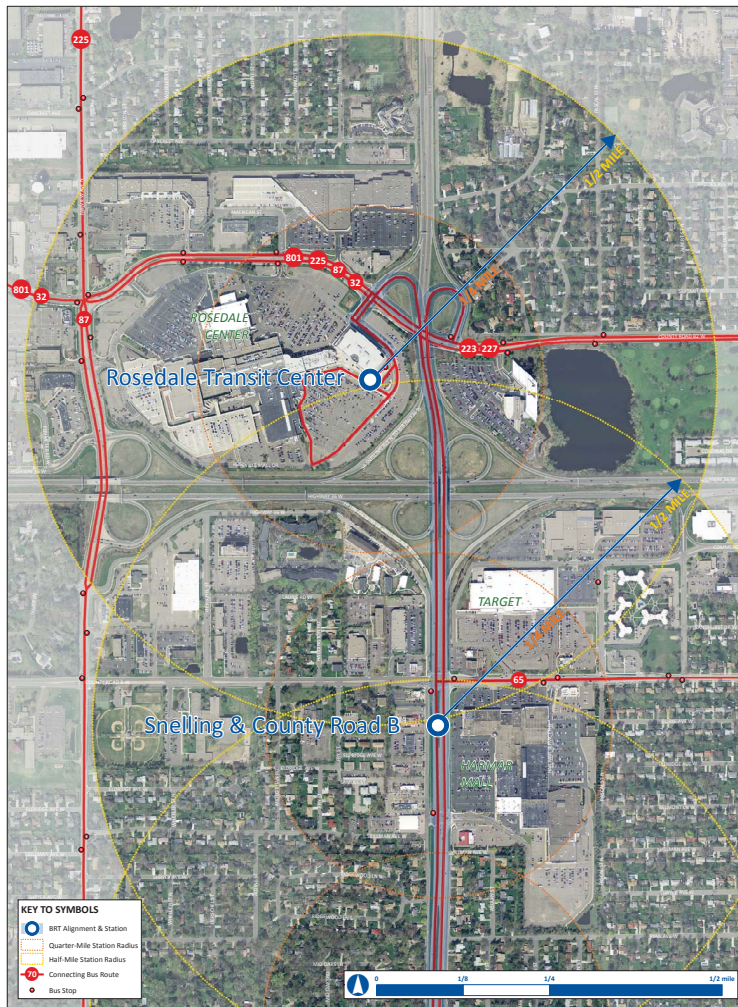


Figure 3.11: Rosedale Mall, conceptual walkshed. Source: Metro Transit.

THE WALKSHED – OR THE DISTANCE PEOPLE ARE EXPECTED TO WALK – SHAPES HOW WE DESIGN TRANSPORT SYSTEMS.

In the United States, the rule of thumb is often that people will walk only 5 minutes³⁹ to get to a transit stop.⁴⁰ In fact, this is too short; a 5-minute walk does not even get you from one end of a large shopping mall to the other, and many people make a full circuit, on two floors, inside the mall, on foot. If there are nice enough environments, planners should expect most people to be able to walk 10 to 20 minutes comfortably with no problem.

³⁹ 5 minutes is about a 400 m or quarter-mile network distance

⁴⁰ The relationship between network distance and air-line (as the crow flies distance) is given by [circuity](#)(§10.9.)

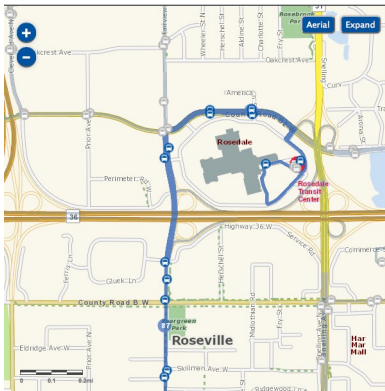


Figure 3.12: The Metro Transit number 87 bus route stops five times at, or within walking distance of, Rosedale Mall. Source: MetroTransit Trip Planner.

⁴¹ §3.7.

⁴² (Walker 2011).

⁴³ (Weinstein Agrawal et al. 2008).

⁴⁴ See time perception (§2.15).

Shopping mall developers do, and they are far more mercenary than the public sector.

Rosedale mall pictured in Figure 3.11 has a Transit Center at the edge of the mall, around which a conceptual walkshed is drawn. Notably, some buses stop at more than one mall entrance, as shown in Figure 3.12.

A longer assumed walkshed has many design advantages. It allows transit providers to increase spacing between stops, which increases running speed, which makes transit more attractive for those already on board. Transit systems trade-off running time for access/egress time (higher access/egress time for lower running time when stops are spaced farther apart).

In dense urban areas, transit stops⁴¹ are generally less than 5 minutes away by foot, so it is unclear what people would do.⁴² The average pedestrian trip to a rail station was 0.47 miles (nearly 800 m).⁴³ A longer distance (10 minutes) is useful (and a slightly better predictor) for the residence end of trips, though shorter distances (5 minutes) at the work-end makes for slightly better predictors.

In short, people will walk longer to transit than we typically give them credit for if they have decent walkable urban routes, high quality and frequent services, and environments that lead people to underestimate the actual time involved⁴⁴ because their minds are not focused on how awful the walk is but about how interesting their surroundings are.

3.9 Automobile City

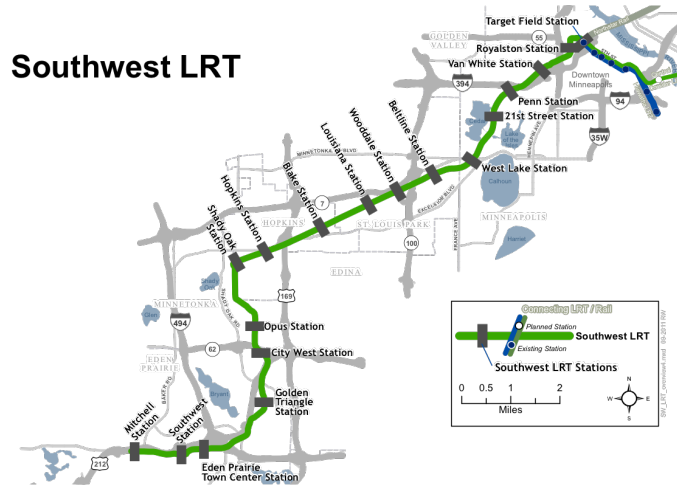


Figure 3.13: Map of Southwest Corridor LRT, (Green Line extension) Minnesota. Source: Metropolitan Council.

THE CAR SHAPES AND RESHAPES URBAN FORM, DEMANDING MORE ROAD SPACE PER PERSON THAN EARLIER TECHNOLOGIES.

Road networks and land uses designed for an earlier technological era were not optimal for the car. The automobile enabled distances to be traversed faster. This meant people could live farther away from their destinations, on larger plots of land, and still satisfy a [travel time budget](#).⁴⁵

How much? In the [Transit City](#),⁴⁶ we saw 1,875 houses per km^2 ; in the 20th century subdivision of the automobile city, this reduces to about half, roughly 1,000 per km^2 , houses on quarter-acre lots, in residential areas. The new road networks also differed. Lower density requires more length of road to serve the same number of people. Imagine it were served by a grid. While the density of the street network (length of network per unit area) might be the same, or even drop, and the traffic on the network might be the same, the extent of the network needs to be twice as long if the density is half as much. There are other possible architectures, that would have fewer roads and lots that are the same width but on average twice as deep, but in practice, we got lots that were the same depth and twice as wide (halving the number of houses fronting the street). Notably, alleys were eliminated and garages placed out front.

The fine-meshed [grid](#)⁴⁷ was also abandoned along the way.

⁴⁵ §2.5.

⁴⁶ §3.7.

⁴⁷ §11.1.

	Persons per <i>km</i> ²	Area <i>km</i> ²	Center line <i>km/km</i> ²	Lane <i>km/km</i> ²	Persons per <i>km</i>	Persons per <i>lane – km</i>	Roads %Area
Minneapolis	2,890	142.20	12.45	29.28	232.19	98.70	10.25%
St. Louis Park	1,642	27.56	9.78	22.73	167.83	72.25	7.95%
Hopkins	1,664	10.57	9.19	20.69	181.10	80.45	7.24%
Minnetonka	713	69.75	7.14	15.87	99.87	44.93	5.55%
Eden Prairie	723	84.05	5.43	12.61	133.28	57.38	4.41%

Table 3.2: Road coverage statistics.

While the macroscopic grid established by the Northwest Ordinance, of roads every mile, typically remains, it is not subdivided into a finer grid, but into a less regular pattern, including more curved roads and cul-de-sacs. This increases network *circuity*.⁴⁸

⁴⁸ §10.9.

The situation in Minnesota is typical. Following the transect of the proposed extension of the Green Line to the Southwest, as shown in Figure 3.13, we move out from the center city of Minneapolis to first (St. Louis Park), second (Hopkins), third (Minnetonka), and fourth ring (Eden Prairie) suburbs, we see the population density tends to drop with distance from the center, but so does road density in terms of center line and lane km. The net is that persons per road km decreases. Hopkins is actually a bit denser than St. Louis Park, but it is also a small town absorbed into the commuting system, illustrating this is not a perfectly smooth process, and Eden Prairie is denser than Minnetonka. In short, there is more pavement per person in outer suburbs than the inner city. Infrastructure is used less intensively.

With fewer persons per lane km, we would expect a lowered *density of traffic*.⁴⁹ And this is true on most roads. But in the end, there are still bottlenecks. Spacing people out upstream of the bottleneck does not reduce traffic at the bottleneck itself. And with a less redundant network, there are fewer reasonable alternative paths, leading to less *reliable*⁵⁰ road conditions.

⁴⁹ §6.4.

⁵⁰ §9.3.

⁵¹ §7.14.

In addition, *lanes*⁵¹ for automobiles are wider in newer suburbs, so not only are there more lane km, there are even more square meters of pavement.

The automobile explodes the spatial requirements of the city, not just in lowered *residential density*,⁵² but in more space for roads. A greater share of the area in the city is devoted to roads, but the suburbs require more road per person.

⁵² §3.1.

4

Markets and Networks



Figure 4.1: Farmer's market at Marylebone, London. Photo by D. Levinson.

"I like being able to fire people who provide services to me." – 2012 US Republican Presidential Nominee Mitt Romney.¹

¹ (Madison and Boxer 1 09).

CITIES, BY CONCENTRATING ACTIVITY, EVOLVE TO FACILITATE HUMAN INTERACTION, OF WHICH MARKET EXCHANGE IS BUT ONE SMALL ELEMENT.

Individuals and firms are not continually changing behavior (taking new jobs, selling their homes, moving their factories, uprooting their lives) at the slightest change in price; any number of other factors preserve relationships over a longer time frame. These processes are fundamentally personal and social – not the abstraction of maximizing automatons. Individuals and firms are

In typically circular fashion, more than one dictionary defines “interact” as “to act on each other.” In other words, to do something onto someone and have them do something back. An interaction can thus be any verb, though we focus things like exchange (both market and non-market, both one-time and continuing) of goods, services, ideas, love, friendship, and just about anything else. Exchange between two parties separated in space is accomplished by transport (of people, goods, energy, money, etc.) and communication (of ideas, bits of information, data, etc.). Exchange can be relatively instantaneous (the one time purchase of a commodity for which a well developed market exists), short term (the negotiation over the purchase price for a good), or long term (an employer-employee relationship, where a flow of money is provided for a flow of work). There has been a long thread in the urban economics literature attempting to describe the tradeoff between location and transport decisions within a quasi-neoclassical framework. But like neoclassical economics and its fictional auctioneer and tâtonnement processes, it fails to provide a realistic mechanism that explains more than macroscopic tendencies.

² (Hanson and Pratt 1995).

³ (Hirschman 1970).

⁴ §4.2.

⁵ (Axelrod 2006).

embedded in complex economic and social networks.

Urban economics concerns itself with three basic issues: land, labor, and linkages. While traditional urban economics treats these relationships in the form of the market transaction, they can be analyzed as centered on the social network (by which we don’t simply mean electronic social networks like FaceBook or LinkedIn, but also the vast majority of human interactions that are not so intermediated).

The market is a place (physical or virtual) where goods and services (which may be stocks or flows) are exchanged in formal transactions. The earliest markets continue with us in the form of farmers’ markets like in Figure 4.1. The social network is the continuing relationships between individuals and between and within organizations. The two relationships mediate one another. In the negative, an exploitive social or intra-firm network relationship can result in one party returning to the market, for instance a worker seeking a new job, a firm seeking a new worker. In the positive, a network can be used to short-circuit market exchanges, as when a firm hires someone who is friend or family to a trusted employee, a common occurrence.² In the terminology of Hirschman, we suggest that networks allow *voice*, markets permit Romney-esque *exit*.³

Interactions can be either repeated or non-repeated. This has important consequences for the strategies that individual participants undertake. Ultimately it influences the *value of the interaction*.⁴ The famous example is the prisoner’s dilemma.⁵ Two accused criminals who are believed to have worked together on the same crime are arrested and held in separate cells. Both prisoners together would be best served if they cooperate with each other and not rat each other out (and thus not cooperate with their captors), but because each does not know the other’s strategy, their best individual strategy is to defect (not cooperate with each other, work with their captors to turn state’s evidence, and help their partner in crime ‘do the time’ in exchange for a reduced sentence for themselves).

If the same game is repeated an indefinite or infinite number of times, cooperation can be rewarded and defection punished. The biblical strategy (‘an eye for an eye, a tooth for a tooth’), dubbed tit-for-tat, is one of the most successful ways to play this game. Cooperation meets with cooperation, a defection with a counter-defection.

Repeated interaction establishes trust (or mistrust). This trust has economic value and may be more important than the simple blind

exchange paradigm of traditional economics. This is one reason why employees and firms may remain together despite small differences in the wage rate between what a wage maximizing employee or cost-minimizing firm could obtain from a switch.

A number of questions that arise in urban economics can be treated in this framework.

- Are labor markets local (submetropolitan) or regional (metropolitan) in scope?
- Are there local technology districts, or can the city be viewed as a whole?
- How do social interactions occur in space?
- How is choice conditioned on and constrained by opportunity?

Subsequent sections consider the elements of interactions between and within households and firms.

The relationships in Table 4.1 connect the two dominant spheres of work and home⁶ and the linkage of resident workers at their home with jobs at their employer through labor markets and the journey to work. The workplace is embedded in social relationships, and social relations are often established at work. Moreover, individuals interact with other social institutions (stores, church, school, state, etc.), and they interact with each other.

⁶ This is what Marx referred to as 'production' and 'reproduction'

	Worker	Employer
Worker	<p>Serendipity and Interaction (§4.1.)</p> <p>Value of Interaction (§4.2.)</p>	<p>Labor markets (§4.4.)</p> <p>Wasteful Commute (§4.5.)</p> <p>Job-Worker Balance (§4.6.)</p> <p>Spatial Mismatch (§4.7.)</p>
Employer		<p>Intra- and Inter-firm Interaction (§4.3.)</p>

Table 4.1: Spatial interactions and networks.

4.1 Serendipity and Interaction

Figure 4.2: Pedestrian street in Singapore. Photo by D. Levinson.



THE PROBABILITY OF INTERACTION BETWEEN ANY TWO PARTIES DEPENDS ON THE PROBABILITY OF INITIAL CONTACT.

This contact can be either planned or unplanned by either or both parties. Unplanned contact is a property of the organization of space and place and increases in people-dense environments like the pedestrian streets of Singapore shown in Figure 4.2. Given rootedness, for instance returning home every night, there is then a physical and practical limit to the amount of space that can be traversed over the course of the day. The [gravity model](#),⁷ which says that the probability of interaction between two places declines with the generalized cost of travel between them, reflects this. Unplanned contact occurs more frequently near the base (home, work, school) than far from it. These contacts then serve as a random seed to future planned contacts with the same individuals. This process of historical [path dependence](#)⁸ shapes future planned contacts so that the distribution of planned contacts resembles that of the unplanned. This occurs even as communications technology annihilates friction of distance, converting it to a fiction of distance.

The desire for any two individuals to interact (and ultimately, firms and other institutions are composed of people, so this applies to them as well) has much to do with affinity.

Members of the same family, individuals who share a genetic

identity, or were raised as if they do, hold the strongest ties of affinity. Similarly, ethnicity involves shared history, a common culture, native language, and religion, particularly important among immigrant groups, provide social cohesion and carries forward trust. Friendship relationships tend to be gendered. Moreover, there remain other sources of affinity – common interests, membership in the same generation, etc.

Finally, interaction is shaped by the traditional economic notions of supply and demand. Desire to interact with a shop to purchase an unavailable good does not make the good available. The amount of interaction is limited by opportunity, including technological feasibility, [space-time constraints](#)⁹, as well as general market conditions. There are also the problems entailed by competition and scarcity. Both the [time in a day](#)¹⁰ and the space on earth are scarce commodities. Available time and opportunities to purchase a market good depend on its availability, the desire of others, and limit opportunities for interaction. Many market ‘goods’ are unique or matching goods. The most obvious is employment. In typical conditions, filling a skilled job prevents others from doing so, since that particular job is unique.

⁹ §2.9.

¹⁰ §2.5.

4.2 *The Value of Interaction*

Figure 4.3: Happy family statue in Singapore. Photo by D. Levinson.



THE PROBABILITY OF INTERACTION DUE TO SPACE, AFFINITY, OPPORTUNITY, AND COMPETITION, ALSO DEPENDS ON THE VALUE OF THE INTERACTION.

Trust and information engendered in repeated interactions on [social networks](#)¹¹ comprise one element of value. There are further values of the interaction: exchange interaction results in a transfer of goods and services for money, social interaction results from – and in – friendship and love, and so on.

[Accessibility](#)¹² measures the quality of a transport network relating how easily places can be reached. A similar notion could be developed for social networks, which would quantify how easily information is exchanged through social means. Relationships can be classified by their directness. Someone with whom you directly communicate is a first order relationship (such as a friend or family (Figure 4.3); someone whom you only indirectly know about is a second order relationship (that is, a friend of a friend); and so on. But not all communications are created equal. Communication to a person who is heavily wired into a social infrastructure may be far more valuable for things like finding a job than communication with a hermit. Knowing lots of people indirectly is still quite valuable, especially for things like a job search, where your colleagues are not the best people to ask, but rather their friends.¹³

¹³ (Granovetter 1973; Tilahun and Levinson 2011; 2013).

The accuracy of information may, as demonstrated in the children's game of Telephone, deteriorate the more ears, brains, and mouths it passes through. This value can in theory be determined empirically, though as a practical matter it may be difficult to measure.

It is important to note that social network relationships need not be symmetric. This occurs in traditional and social media all the time: speeches, radio, television, print, and the internet all differ from the peer-to-peer relationship of a two-person conversation. Someone can post something on the Internet that goes viral. A headhunter matches labor with jobs. When we write a book, you as a reader get information and begin to know us far better than we know you. The author or creator gets value too, accumulating social capital and the like, and is perhaps paid, but the relationship is unequal.

In short, interaction has value and frequency. The value depends on the kind of interaction, whether it is repeated, and a number of other factors. Frequency depends on location in space, affinity, as well as supply and demand. Furthermore, anticipated value shapes frequency, and frequency influences valuation due to the notion of benefits from repeated games. Interaction, which has various properties (including existence, value, kind, frequency, etc.), can in principle be thought of as a multi-dimensional matrix. Access, a weighted measure of existence, value, and frequency for various kinds of interaction, can then be developed as a metric. Information flows are restricted by the number of links on the chain. These flows shape a variety of decisions, not the least of which is job and location choice.

We might think about the following as social accessibility:

Degree1

$$A_{q,1^\circ} = \sum_{p=1}^P t_{qp} \cdot i_{qp}$$

Degree2

$$A_{q,2^\circ} = \sum_{p=1}^P t_{qp} \cdot i_{qp} \cdot (1 + g(A_{p,1^\circ}))$$

DegreeN

$$A_{q,N^\circ} = \sum_{p=1}^P t_{qp} \cdot i_{qp} \cdot \left(1 + \sum_{n=1}^{N-1} g(A_{p,n^\circ}) \right) \quad (4.1)$$

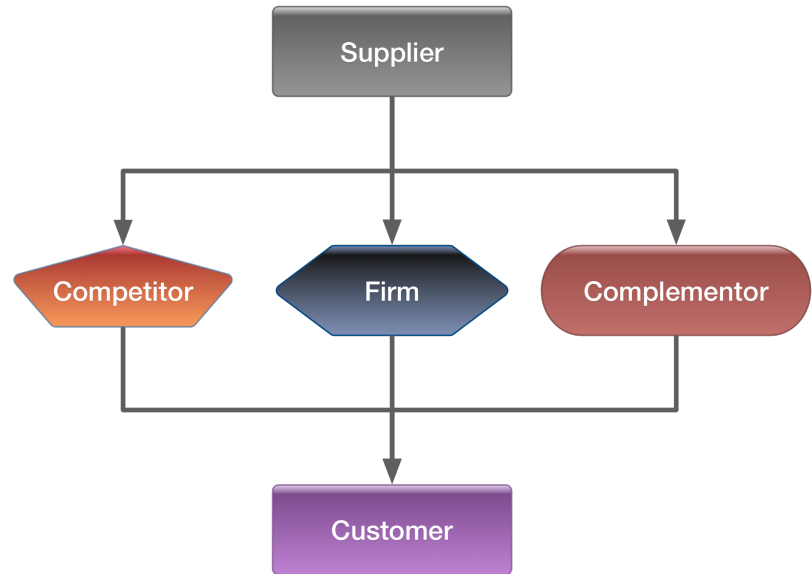
In other words, the first degree social access of a person q , $A_{q,1^\circ}$ is a function of the time spent (t) with other persons ($p = 1$ to P) multiplied by the intensity (i) (information quality) of that relationship. Time spent (t) could be measured over a some period of time, for instance the percent of time in a day, week, month, or time-frame over which t is measured.

The information intensity (i) is clearly qualitative, and lacking other information, could be normalized to 1, or i could be a function of t , or the income or status of person p , etc.

The second order access ($A_{q,2^\circ}$) considers not only the time and intensity but the first order access of the person p being considered, discounted by some indirectness weight or function (g). This is analogous to the gravity or time-weighted cumulative opportunities model. It is a social **distance decay** (§3.2) function. Knowing someone directly may be two times or five times or one-hundred times more valuable than knowing them indirectly.

4.3 Firm-Firm Interactions

Figure 4.4: Supplier - firm - customer flowchart.



ABSTRACTLY, FROM THE VIEWPOINT OF THE FIRM, THE ECONOMY CAN BE VIEWED AS A FLOW OF GOODS AND SERVICES IN ONE DIRECTION AND MONEY FLOWING IN THE OPPOSITE.

The flow of goods and money is illustrated in Figure 4.4. The boxes representing the suppliers and customers are not assumed to be a single level but may involve several transformative steps, hopefully adding value, between raw materials and what is supplied to the firm, or between the processed good and a final consumer good. (The customer, of course, is a supplier to another firm in a full economic web.)

Assuming an underlying validity to this simple model, it can be readily seen why firms want to locate near suppliers (including laborers). Doing so reduces transaction costs of exchange (most easily seen as transport costs but not limited to them). Transaction costs determine which activities are performed within firms and which between, and are therefore central to industrial organization.¹⁴

¹⁴ (Coase 1937; Williamson 1985).

¹⁵ §13.6. With lower transaction costs resulting from [economies of agglomeration](#)¹⁵ due to specialization, an inter-firm division of labor is more viable.

Location choices consider both vertical and horizontal aspects of industrial organization. Location theory demonstrates the vertical

aspects, location with respect to suppliers and to customers for the purposes of minimizing costs, as well as location choice with respect to competitors for serving market areas.¹⁶ In a two-dimensional world, central place theory suggests an optimal hexagonal arrangement.¹⁷ Location choices should also consider the positive aspects of horizontal considerations, location with respect to competitors in an industrial district.¹⁸ Competitors share suppliers and customers, thus the nearby location of competitors reduces transaction costs for co-located suppliers and customers, the gains which can be split some fashion among them.

¹⁶ (Isard 1956).

¹⁷ (Christaller 1966).

¹⁸ (Porter 2011).

The analysis of transactions should not be viewed as being limited to costs however. There are benefits to exchange, not the least of which is the exchange of information and ideas. While one of the supposed advantages of blind market exchange is that it requires much less information, this is not necessarily good. Social networks of various kinds provide information transfer. Network exchange requires communication and thereby enables learning. Not simply using price as a signal to produce more or less, networks enable the use of words to suggest ways to change and improve the quality of goods and to suggest new needs. A blind market exchange cannot easily provide feedback for innovation except randomly when a producer produces something new, something that happens to be in great demand. Networked exchange provides for customer pull in addition to supplier push.

Comparative advantage and economies of scale alone cannot explain specialization, but that this must be complemented technological (absolute) advantages on the part of producer.¹⁹ These absolute advantages are created and renewed through learning. These advantages, as any good monopolist knows, result in quasi-rents. While networks may exist without technological learning; there can be no technological learning without networks. The easiest way to form those networks is co-location. When lots of people co-locate, we call that a city. We measure the amount of co-location using accessibility.

¹⁹ (Storper 1992).

4.4 Labor Markets and Labor Networks

Figure 4.5: The Maryland National Capital Park and Planning Commission - Montgomery County Planning Department headquarters building. Photo by MNCPPC.



WHY DOES AN INDIVIDUAL LIVE WHERE HE DOES? WHY DOES HE WORK WHERE HE DOES? WHY DOES HE TRAVEL AS MUCH AS HE DOES?

When David was a graduate student at the University of California, he dealt mostly with people in and around Berkeley. But because he previously lived in Maryland, he had far more interaction with Maryland than neighboring Arizona, despite the fact that Arizona is closer and more populous. Having grown up and worked in Maryland, he had established **social contacts**²⁰ in that area. This pattern of lumpiness based on history and random events is replicated at multiple scales: international, national, regional, and neighborhood.

²⁰ §4.2.

David's contacts in Maryland helped a graduating classmate at Berkeley obtain a job at the Maryland-National Capital Park and Planning Commission (Figure 4.5), David's former job in fact – something quite improbable if we view labor markets functioning constrained by rigidities and boundaries of a simple **distance decay**²¹ function. While commuting may be limited by a daily **time budget**,²² job search is not, since the possibility of residential relocation always exists.

²¹ §3.2.

²² §2.5.

Traditional models begin this analysis by considering the location of firms as fixed, then consider where people reside relative to them. While in the long term, the location of neither the home nor the firm is fixed (hence the huge suburbanization of jobs over the past three decades), by taking firms as fixed, these researchers hope to identify

'labor markets.'

In contrast with the experience above, some argue that the metropolitan area is too large, and there are sub-metropolitan labor markets. For instance, an analysis of animation studios in Los Angeles "dispel[s] any notion that metropolitan areas invariably constitute the minimum geographical level of local labor market differentiation."²³ Animation studios find work nearby rather than across the entire region. This is not surprising when thinking about central place theory. When looking at the same LA region, others conclude that the area of an individual labor market is large and distances are relatively unimportant.²⁴

²³ (Scott 1988).

²⁴ (Giuliano and Small 1993).

Labor markets need not have delimited boundaries, and need not be classed as large or small. While there are clearly centers of activity, there are less clearly, if at all, definable districts. Instead, we can think of potential jobs for any individual like a magnetic or electric field, while the likelihood of taking any available job diminishes with distance (travel time, [social access](#)²⁵), it has no definitive edge.

²⁵ §4.2.

So, the process of job switching can be briefly outlined. Given a current state of affairs, a specific daily home-work-home orbit, individuals may wish or be required to obtain or change work. On occasion, an opportunity may fall into their lap, as when a friend or headhunter tells them of an opening – all that is required is showing up for an interview and minimal preparation. More often, search is a conscious and directed effort to select from a large but limited number of non-identical options available over a given timespan. This process is semi-continuous, most people are open to the 'right' offer at any time, though what is 'right' is dynamic and fluid. Opportunities appear through both formal and informal means. In this transactional relationship (worker/employer), the search is a two-way process. Depending on the nature of the specific transaction and the various needs of each party at any given time, one side may be more or less active. This process can be analogized to other matching relationships, including home-finding and spouse-finding.

4.5 Wasteful Commute

Figure 4.6: Statue of bicyclist in Cergy-Pontoise, France. Photo by D. Levinson.



IMAGINE WHAT WOULD HAPPEN IF EVERY HOUSEHOLD MOVED FROM THEIR CURRENT HOMES TO A DIFFERENT EXISTING HOME THAT WAS AS NEAR AS POSSIBLE TO WORK, WHILE MAINTAINING THE QUALITY OF THE HOUSE.

The difference between the actual commute time and this imagined minimal commute has been dubbed 'excess' or 'wasteful commuting'. This difference is not small, as the literature shows, implying people care about far more than journey-to-work travel times.²⁶

Commutes are at least somewhat, and perhaps much, longer than the minimum necessary to satisfy housing constraints. There are any number of possible explanations:

1. Information is far from complete. There are workplaces and houses people would prefer, but they simply don't know about

²⁶ (Hamilton and Röell 1982; White 1988; Hamilton 1989; Cropper and Gordon 1991; Small and Song 1992).

them. In large part, that is because **social information networks**²⁷ are central to job and house finding. US evidence shows that, women, in part due to increased domestic responsibilities, seek and find jobs closer to home and children than do men.²⁸ Jobs found through personal contacts have shorter commutes than those who find them through more formal means. This suggests that there are serious information gaps in the formal labor market. This is more often true of lower income/less skilled jobs than higher income jobs, where more formal search procedures will be undertaken on both sides (employer and employee). Networks provide information and serve as screening devices for both the employer and employee. These networks by their very nature respond to and reinforce occupational segregation, a gendered division of labor and illustrate what Hanson and Pratt call “the inseparability of economic and social realms of life” that work is embedded in the social and vice versa.

²⁷ §4.2.

²⁸ (Hanson and Pratt 1995).

2. Commuting has a positive benefit, justifying a non-zero **travel time budget**.²⁹ Several major benefits have been suggested:

²⁹ §2.5.

- (a) Privacy: Time alone in my car (on the train . . .) is valuable, and
- (b) Staking a territory: Travel to work is the projection of the underlying human drive to establish a territory.
- (c) Physical activity: Walk and bike commutes like that in Figure 4.6 provide exercise and improve health and happiness. (Certainly compared with auto commutes).

3. Complications arise due to two-worker households. In a two-worker household, a joint optimization will not reveal itself in the simple cost minimization of minimum commute for each party.

4. Work is one of several activity locations being optimized. This argument suggests that people consider access to shop, school, and other non-work activities in location decisions, activities across the entire **space-time prism**.³⁰ It should be noted that, even for workers, work occupies less than one-fourth of all time; for the population at large, it is nearer one-eighth.³¹

³⁰ §2.9.

³¹ (Handy 1994).

5. **Relocation is not free**.³² While urban economic models oddly ignore transactions costs of moving jobs or homes (implicitly considering them small), experience suggests that they are not. This is odd because urban economists explicitly consider transport costs, which are a very important transaction cost.

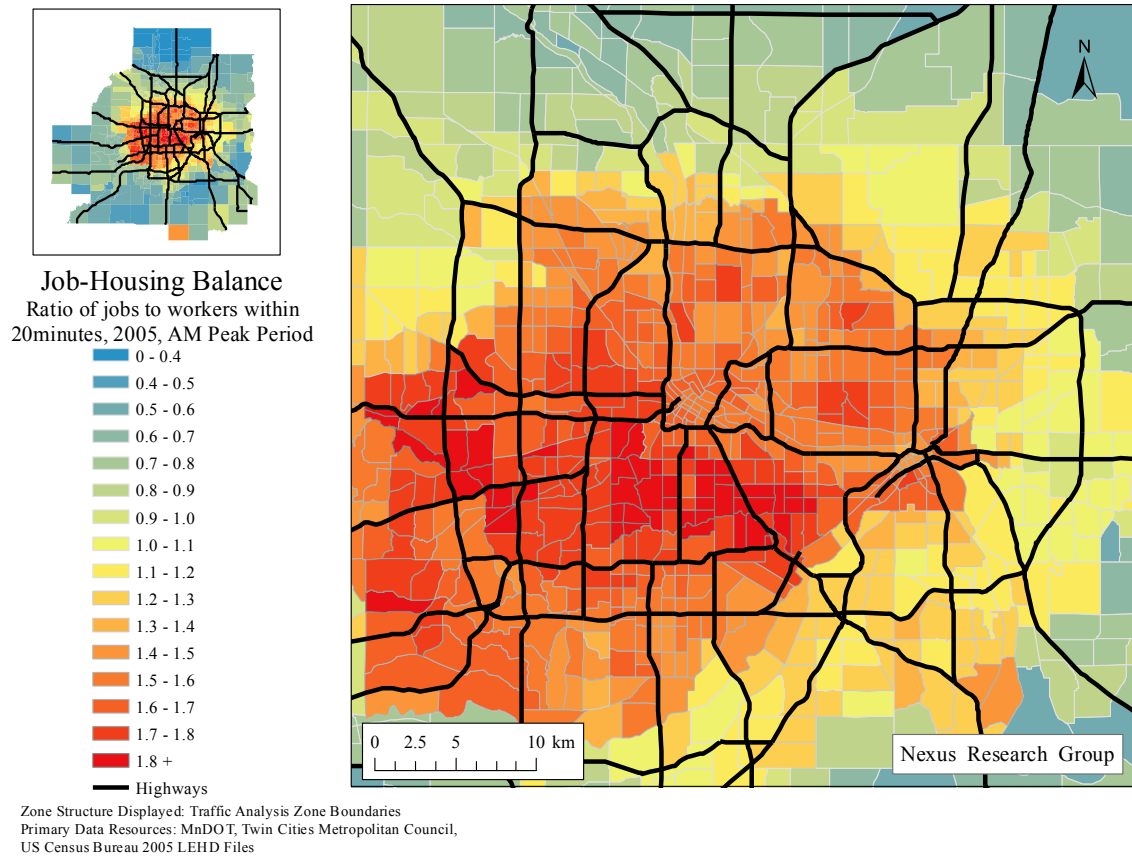
³² §4.4.

Moreover, most houses are not on the market at a given time, so this type of dynamic optimization cannot be easily realized.

- ³³ §4.6. 6. **Job/worker imbalance**³³ (general) - Because of a combination of regulations like zoning and market forces favoring economies of agglomeration, the number of jobs at a local level that can be reached is often much higher or lower than the number of workers who can be reached, requiring longer commutes.
- ³⁴ §4.7. 7. **Spatial mismatch**³⁴ by income (or skill) means that housing for workers in certain groups is available only at a great distance from those jobs which they are qualified. In the US, the related issue is that unskilled work in the suburbs requires workers who cannot live nearby (or that the those working the many service jobs of a CBD cannot afford to live nearby).
8. The value of time and opportunity varies by professional occupation or salary. Higher income jobs are scarcer and should thus draw workers from longer distances. However, there is a certain illogic to this argument, as higher income people have more choice in where they live, and have a higher value of time, and therefore should be more likely to locate or relocate closer to where they work, which is in contradiction to empirical findings. Of course, they probably have higher quality transport, and so may more highly value time in their vehicle as well. They may also place a higher premium on amenities.

All of the above arguments have some elements of truth. All explain why urban economics ought not simply treat utility as being composed of travel time to work and housing value. While those two factors are surely important, they are not all-telling in describing the region. They fail to answer questions about how jobs and houses are found and kept. These dynamics, which depend on both costs of transactions and benefits accruing to relationships, are difficult to capture in a quantitative model, but necessary for a true understanding.

4.6 Job/Worker Balance



HOW MANY JOBS ARE IN YOUR TOWN? HOW MANY WORKERS LIVE THERE? IF THE NUMBERS OF JOBS AND WORKERS IN YOUR TOWN WERE IDENTICAL, IN PRINCIPLE, EVERYONE WHO LIVED IN TOWN COULD WORK IN TOWN.³⁵

The operative word is ‘could;’ in reality, many people don’t actually reside in the same community where they work. There are a variety reasons for this, including:

- The next town over may be mostly resident workers (thus increasing the number of laborers willing to work in your town),
- The next town over may be only jobs (thus competing for your labor),

Figure 4.7: Ratio of accessibility to jobs to accessibility to resident workers. Figure from (Levinson et al. 2017).

³⁵ (Cervero 1989; 1996; Levinson et al. 2017).

- Not all jobs are created equal, skilled jobs require commensurately skilled labor,
- ³⁶ §13.6. • Economies of [agglomeration](#)³⁶ encourage firms to cluster to be more productive, or
- Government regulations, like zoning, preventing a mixture of uses and types of development and densities.

While planned communities often aim for balance, they do not operate in a vacuum.

If we measure it right though, not at the level of the town, but the region, and compare how many jobs can be reached in, say 20 minutes with the number of workers that can reach those jobs, we can see which areas have a surplus of jobs to workers and vice versa. This is approximated by the ratio of job access to worker access, as in Figure 4.7. At the metropolitan area as a whole, this will be about 1.0 (since metro areas are defined as commute regions). Locally, it will not. Areas with more jobs than workers will need to import workers from farther away. Areas with more workers than jobs will export workers, meaning there will be longer commutes than what 'could' be achieved were there balance.

4.7 Spatial Mismatch

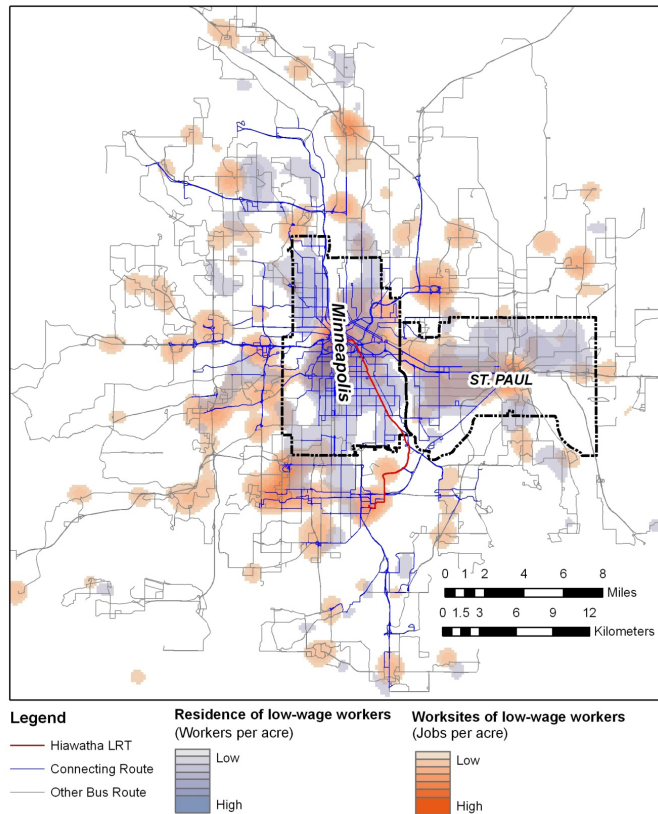


Figure 4.8: Labor accessibility by transit in Minneapolis - St. Paul region. Figure from (Fan et al. 2012).

ARE THE RIGHT JOBS NEAR THE RIGHT WORKERS?

Even if *job/worker balance*³⁷ is achieved, it is no guarantee that commutes will be short. A job is only useful to a worker if she could potentially fill it because she has the right skills. The spatial mismatch hypothesis claims that housing for workers in certain jobs (with certain incomes) is available only at a great distance from those jobs, as illustrated in Figure 4.8.³⁸ In the US context, employers in the suburbs may only be able to find workers for particularly low skill/low wage jobs who can only afford to live in the inner city or in exurbia. In more normal cities outside the US, the wealthy live in the center with the highest accessibility, while incomes fall with distance from the center. In many countries, poor people are priced out of the high-accessibility urban core. Spatial mismatch analysis stratifies job/worker balance by skill, income, or race. Evidence finds spatial mismatch in the US and elsewhere.

With more than five decades of debate about the causes, amount, and solutions, some of the key early, and subsequent review papers include: (Kain 1968; Mooney 1969; Kain 1992; 2004; Raphael 1998; Fan et al. 2012; Tilahun and Fan 2014; Fan 2012; Fan et al. 2014).

³⁷ §4.6.

³⁸ The hypothesis is attributed to (Kain 1968).

Table 4.2: Strategies for addressing spatial mismatch.

Location Strategies	Transport Strategies
Moving jobs nearer workers	Automobility
Moving workers nearer jobs	Specialized transit (employer vans)
Reskilling workers for nearby jobs.	Public transport
Refactoring jobs for existing workers.	Ridesourcing (taxi/Uber/Lyft)

³⁹ §1. While overall (and transit) [accessibility](#)³⁹ tends to be higher in central cities and poor low-skilled workers in the US tend to live in the central city, those areas also have the highest unemployment. In some cases, employers complain that low-skilled jobs in remote suburbs are begging for workers.⁴⁰

⁴⁰ To which the economist retorts: “raise your wages”.

Low-skill workers cannot afford or do not want to move to the suburbs. Employers do not want to relocate to the city. While there are proffered economic development strategies to do both (see Table 4.2), those are longer term and less successful than the more direct strategy of providing transport to connect the workers with the jobs. Constructing a fixed guideway transit line, however, takes years.

Faster ways of connecting workers to jobs include employer shuttles, which would pick up workers near their homes at the beginning of the shift and return them at the end. This still requires an investment of travel time on the part of the worker. One major reason that workers don’t take far away jobs is non-work constraints. What do they do if they need to pick up a child in the middle of the day? This is obviously harder if trapped in the suburbs a long distance from home, without a car.

⁴¹ (O’Regan and Quigley 1998).

⁴² §2.12.

Another strategy that has been relatively successful when tested is providing cars to low income workers.⁴¹ There are still [capability](#)⁴² constraint issues for many people, and providing even more cars may seem counter-intuitive and likely to worsen other problems, but note these are for reverse commute, and often off-peak jobs, so the congestion effects are diminished, even if other social costs remain.

While various forms of ridesourcing (taxi/Uber/Lyft) are expensive on an individual trip basis, and slower than a private car just out the door, if used infrequently and coupled with transit or employer shuttles serving the majority of trips, vouchers for these services might be cost-effective.

Part IV

The Plexus

Understanding plexus is a bit less obvious than people or places but no less important. The Plexus refers to the complex of networks that connects The People to The Places. In this text, we focus mostly on transport networks, but plexus also includes other infrastructure, such as communication and information networks, as well as social networks. Still, transport networks is a broad topic that includes street-level issues as well as the various ways that we might configure or evaluate the street networks that surround them. It also involves the various modes that run on these streets and through these networks as well as some of the fundamentals related to how we measure and understand these modes. Much of this section on The Plexus represents the mobility piece of the accessibility story, but it encompasses much more than what we conventionally call mobility. The Plexus the third building block to truly understanding accessibility.

5

Queueing



Figure 5.1: Queueing on Franklin Avenue, Minneapolis at five-way intersection with East River Road and 27th Avenue. Photo by D. Levinson.

ACCESS DEPENDS ON TRAVEL TIME, AND THE TRAVEL TIME DEPENDS ON THE QUEUEING PROCESS.

In simple terms, we are either traveling unaffected by the movements of others – which engineers call *free-flow* – or we are slowed by others – which we call *queueing* – either directly, due to

Some suggest there are three-phases: free-flow, synchronized flow, and wide-moving jam (or queueing). The existence of synchronized flow, and whether it is a stable phenomenon or just transitory between free-flow and queueing, is debated. Boris Kerner is the leading proponent. (Kerner and Rehborn 1996) Notably his work is published in physics rather than transport journals. Criticism is reported in transport journals. (Daganzo et al. 1999; Treiber et al. 2010)

¹ §2.4.

interactions (such as congestion), or indirectly, because of controls to manage those interactions (like traffic signals).

In free-flow, the traveller or customer can choose her own speed without imposing her choice on other nearby travelers. She can choose her own time and receive immediate service from some ‘server’ (the toll plaza, the empty road). This is directly analogous to the wait at a cash register in the supermarket, the waiter in a restaurant, the ATM, the website for booking a flight, the teller at the bank, the food counter in the cafeteria, and many others.

There are **peaks of demand**¹ that human preferences generate: we want to work during the day and have to work at the same time as our supply chain, customers, and co-workers; we all want to be at the pool during the brilliant summer day; we prefer the fashionably cool restaurant or the coffee shop right next to big office building.

Queues emerge not only when demand exceeds the provided supply but also when the provided supply changes. The most visible example is when there is a physical change in capacity when traveling along a road, such as climbing a hill, seeing two lanes merge into one, or even one lane diverging into two. But the supply can also change dynamically. Think about an historic, staffed toll plaza. Not all the tollbooth lanes are staffed all the time. If supply is roughly in sync with demand, the lines are short. If supply far exceeds demand, there are no lines, but when supply falls short of demand, queues get long. Unfortunately for the customer, no supplier can afford to provide so many counters or so much space in the transport system that travelers always encounter free-flow conditions.

The supplier wants to offer a certain kind of service, offering a particular speed or waiting time until service starts, but without having too much unused space, empty seats in the bus or train, or underemployed staff, which are costs without revenue. Furthermore, providing extra capacity encourages the users to arrive at the peak time, which makes the peaking problem worse over the longer term. Elsewhere we discuss the issue of **induced demand**.²

² §12.1.

To make matters worse in daily practice, the capacity of the serving stations varies randomly around a mean for a variety of reasons: weather, lighting conditions, the experience of drivers, skill level of the person serving, and random crashes in computer networks.

There are two main queueing processes. One is due to an extended period of more arrivals than the server can handle, which is called ‘**deterministic queueing**’³ and occurs due to the peaking of

³ §5.1.

travel. The other is due to random bunching of arrivals or random server times, and that is called 'stochastic queueing',⁴ where stochastic is a fancy word for 'random.' These are discussed in turn. ⁴§5.2.

Applications for queueing include traffic signals, where it is important to platoon⁵ vehicles to try to ensure they arrive when the traffic light is green, and ramp metering⁶ to smooth the flow of traffic entering the freeway. Incidents⁷ tend to cause queues and decrease capacity. ⁵§5.3. ⁶§6.8. ⁷§5.4.

5.1 Deterministic Queues

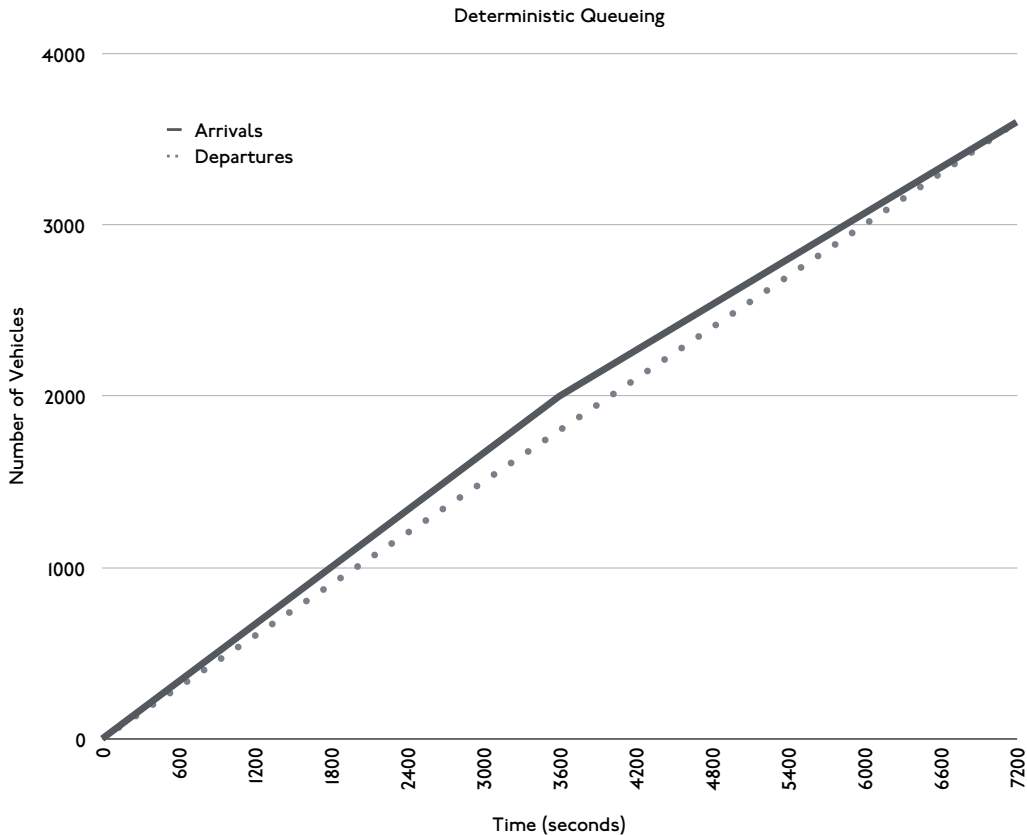


Figure 5.2: Deterministic queues.

DETERMINISTIC QUEUEING OCCURS WHEN TOTAL NUMBER OF ARRIVALS EXCEEDS THE CAPACITY OF THE SYSTEM FOR AN EXTENDED PERIOD OF TIME.

Deterministic is the first of the two types of queueing processes. It can be best understood by an example:

A highway can serve 1,800 vehicles per hour per lane (cars at a two-second following distance). In that hour, 2,000 vehicles arrive. At the end of the first hour, there will be a standing queue 200 cars long.

If in the next hour, only 1,600 cars arrive, the queue will be fully discharged at the end of the second hour (in two hours, 3,600 cars arrived, and 3,600 were served). At the end of the second hour, there will a standing queue 0 vehicles long. If the arrival rate never

dropped below the server rate, the queue would grow forever.

Figure 5.2 plots the cumulative arrival and cumulative departure curves; the delay is the area in between these curves.

The average delay is 200 vehicle-hours divided by 3,600 vehicles times 3,600 sec/hour = 200 seconds per vehicle (3.33 minutes per vehicle). This average is much less than the longest delay (at the end of hour one), which is 400 seconds (6.67 minutes). The first and last vehicles have no delay.

The way it is described here, the arrival and departure curves form a triangle, and the area of a triangle is known to be:

$$A = 0.5(B)(H) \quad (5.1)$$

To make the analysis easier to understand, we break this into two triangles (the first hour and the second hour) (though of course, they both have the same height). The Base (B) for each is 1 hour. The height (H) is the maximum queue, 200 vehicles, found at the end of hour 1, and thus, at the beginning of hour 2. In this symmetrical example, the total delay (D) is given by:

$$D = A = 0.5(B_1)(H_1) + 0.5(B_2)(H_2) \quad (5.2)$$

$$D = 0.5(1)(200) + 0.5(1)(200) = 200 \quad (5.3)$$

Note: This delay is not directly additive to the normal free-flow time, since the vehicles advance in the queue over the distance they would otherwise have spent consuming during their free-flow time. A queue 200 vehicles long is about 1 mile (1.6 km) long. If this were a freeway, it would normally have taken a minute to traverse were there no traffic. So the queued delay of 6.67 minutes is 5.67 minutes more time than would normally have been taken.

5.2 Stochastic Queues

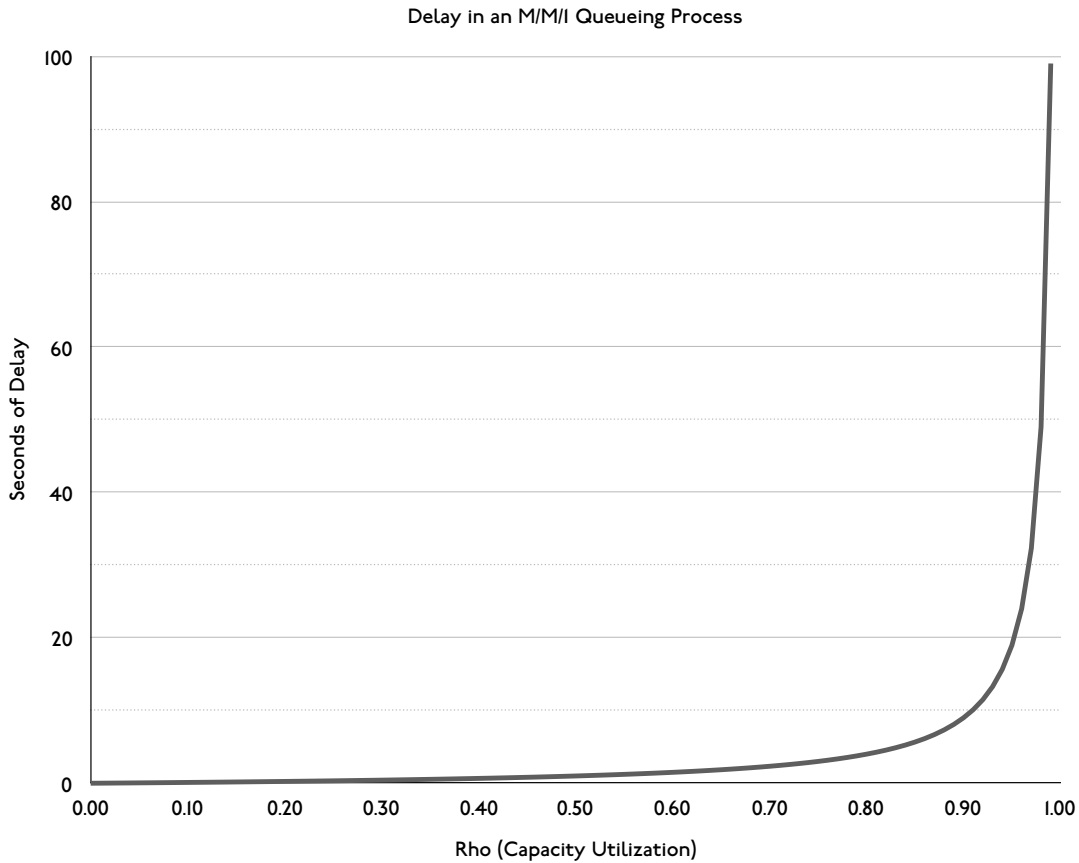


Figure 5.3: Stochastic queues: As capacity utilization approaches 100%, costs rise exponentially.

STOCHASTIC QUEUES OCCUR EVEN THOUGH TOTAL DEMAND IS LESS THAN THE SYSTEM CAPACITY, BECAUSE OF SHORT-TERM FLUCTUATIONS IN DEMAND OR CAPACITY.

⁸§5.1. While **deterministic queues**⁸ occur when demand exceeds server capacity over an extended period, stochastic queues emerge in many transport situations: a group leaving a room having to wait at the door, as only one or two people can physically walk through it at any one moment in time, or a funeral procession occupying both lanes of a road but traveling at a slow speed. A system that works well with a uniform number of arrivals (1 car exactly every 2 seconds) sees queues form when it gets 4 cars in 1 second, even if it gets no cars

for the next 7 seconds and thus has the same overall arrival rate.

Deterministic queues occur when arrivals exceed the capacity of the service. In a sense, *stochastic queueing*⁹ is just the appearance of many short deterministic queues that appear randomly, rather than for an extended period.

Utilization rate (ρ) is the ratio of arrivals (at the back of the line) per unit time (λ) to departures (from the front of the line) (μ) ($\rho = \lambda/\mu$). Crucial for the lived experience, as well as for the planner or manager, is that the waiting time in queues escalates to very high numbers when the degree of saturation reaches 80% to 85% of the available capacity. The figure above shows this explosion for the mathematically simplest-to-capture queue: one server with demand and supply following an exponential distribution of arrival and service times, in the notation of the field, an M/M/1 queue.¹⁰

The right side of this graph dramatically overstates what really happens in traffic since it produces higher delay than a deterministic queue would (given demand cannot exceed supply forever). The left side is more realistic than the zero delay estimated by a deterministic queue when arrivals are less than the server rate. In practice, delay is estimated by models that combine these two features, accounting for both stochastic queueing when supply exceeds demand, and deterministic queueing when demand exceeds supply (for a period of time).

Still, this pattern of delay rising steeply with capacity utilization holds across all systems of queues with any randomness at all.

⁹ §5.2.

¹⁰ The M/M/1 is formatted in Kendall's notation. The "M" stands for Markovian, a random process named for the Russian mathematician, Andrei Markov. The first M indicates the arrivals are Markovian (as opposed to Deterministic), and so described by a Poisson process. The second M indicates the departures are Markovian, that is they have an exponential distribution. The key is that the processes are memoryless, what happens now is independent of what happened before.

The "1" indicates there is a single server (lane).

Expected Waiting Time = $\lambda / (\mu (\mu - \lambda))$.

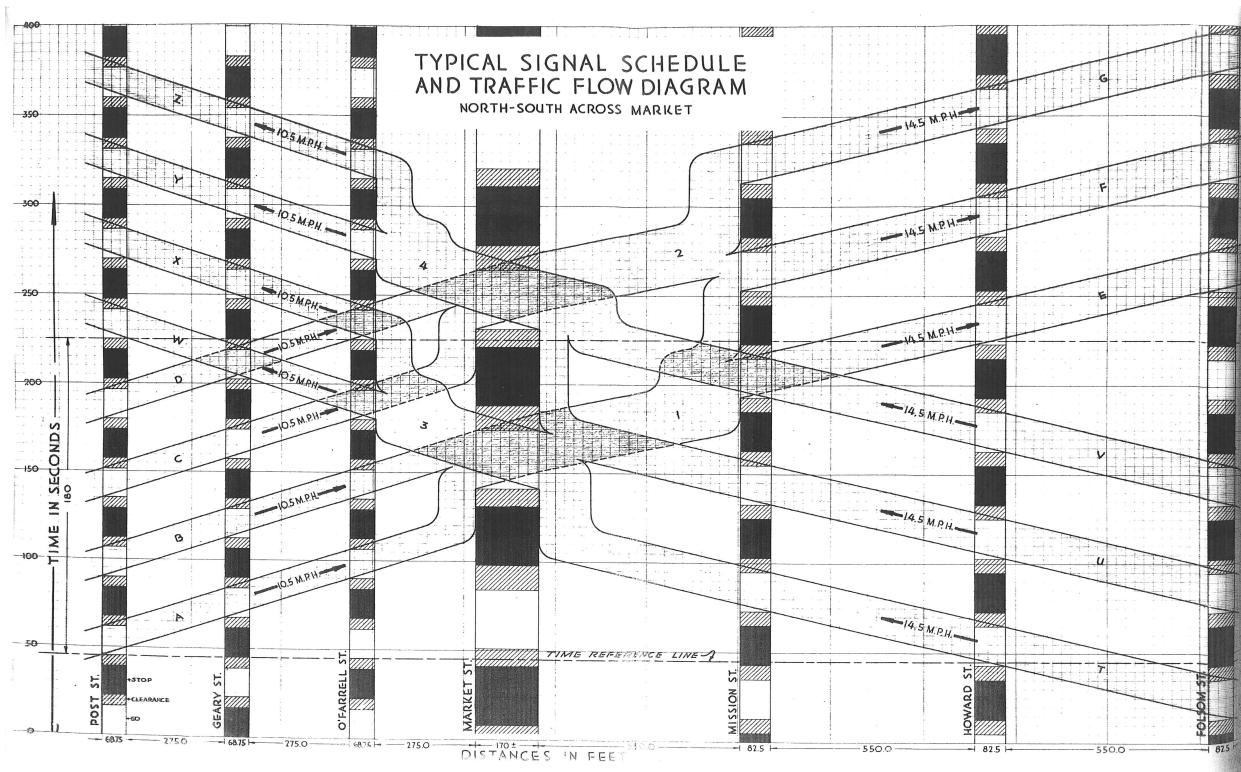
5.3 *Platooning*

Figure 5.4: Typical signal schedule and traffic flow diagram, North-South across Market Street (San Francisco) (1929) From 'Signal Timing Schedule for Traffic Control Plan, June 15, 1929.' Attempted 'green wave': 8.5mph on Market.

WHEN YOU CANNOT GO FASTER THAN THE CAR IN FRONT OF YOU, AND YOU CANNOT PASS THEM BUT WANT TO, YOU MIGHT FEEL FRUSTRATED.

The speed of the driver at the front of this group of cars, which engineers evoking military organizations call a 'platoon,' imposes his or her speed on the others. While this may seem inefficient from your perspective, joining the platoon may be to your advantage.

Traffic lights allocate access to a scarce resource, the [conflict point](#)¹¹ at [intersections](#),¹² When possible, traffic engineers coordinate traffic signals so that if you are driving the recommended speed, you make a series of green lights (this is called a 'green wave').¹³ So if you are in a platoon of vehicles arriving at an intersection, you are more likely to get a green light. This increases capacity utilization and exploits [economies of density](#).¹⁴ Green waves can be established for bicycle or pedestrian facilities as well.

¹¹ §7.7.

¹² §7.5.

¹³ As this is a system optimization, cars in off-peak directions may encounter a red wave.

¹⁴ §13.1.

While signals help coordinate drivers, there are other techniques that can be used as well. Signs sometimes tell drivers what speed the signals are timed for, say 30 miles per hour (48 km/h) so they know that if they speed,¹⁵ they will then get stopped. Electronic variable message signs could be posted to advise drivers what speed to travel to ensure they make a green light.¹⁶

¹⁵ Note: they can also make the green wave if they go a multiple of the speed, like twice the speed or 60 mph (96 km/h). Hopefully this behavior does not occur.

¹⁶ See e.g. Always Green Traffic Control by Nick Musachio. ([Renault 2014](#)).

5.4 Incidents

Figure 5.5: I-10 incident from traffic camera. The image courtesy of TxDOT.



Another estimate: Back of the envelope, people spend 365 hours in a car each year and live 80 years. So $365 \times 80 = 29,200$ hours. (About an hour a day, a rough average for a lifespan, too high for children, too low for active adults).

In the US there are about 40,000 car deaths per year, and assume average life expectancy is 80 years. In 80 years there will be 3.2 million car deaths. If US population is about 320 million, there is about a 1% chance of dying in a car crash over a lifespan (compare with the 1 in 113 dying from official statistics, and this is close). Assume the death occurs at the median age 40, there will be 40 years of life lost (or 350,400 hours). A 1% chance of a 40 year loss gives an expected lost of 3,504 hours.

If everything were linear (which it isn't), for each hour in a car you lose about 0.12 hours (7.2 minutes).

¹⁷ (National Safety Council 2016; Aho 2011).

¹⁸ According to US Federal Highway Administration, "The three main causes of nonrecurring congestion are: incidents ranging from a flat tire to an overturned hazardous material truck (25 percent of congestion), work zones (10 percent of congestion), and weather (15 percent of congestion)." (US Department of Transportation 2017).

¹⁹ §5.

CARS KILL 1.3 MILLION PEOPLE GLOBALLY PER YEAR IN AN ONGOING HUMAN TRAGEDY THAT IS TOO OFTEN ACCEPTED AS COMMONPLACE.

The average American has a 1 in 113 chance of dying in a car crash (and 1 in 672 from a pedestrian incident), and the average driver will file a crash insurance claim once every 18 years. Property-damage crashes are more frequent (10 million per year in the US) than injury crashes (2 million, many of which lead to temporary or permanent disabilities) or fatalities (37,500).¹⁷

For those in crashes, it disrupts your day, if not your life (making you late if not 'late'). It also has a spillover effect, disrupting traffic for many others. About half of congestion is non-recurring, due to incidents of various kinds (of which crashes are the most common), so reducing crashes not only saves lives, it reduces traffic problems.¹⁸ The crash can be thought of as reducing the capacity of the road, and because it was unexpected, not reducing the demand. So as we saw with [queueing](#),¹⁹ the load factor, and thus the expected queue length will increase. Sometimes crashes close roads entirely, for a time, resulting in a queue growing without end, at least until travelers do divert to alternative routes. That of course requires information, which is in short supply for unexpected incidents.

5.5 *Just-in-time*



Figure 5.6: Replica of early UPS truck, Istanbul Turkey. Photo by D. Levinson.

INVENTORY, THE STORAGE OF GOODS BEFORE THEY ARE USED OR SOLD, IS A DEADWEIGHT LOSS.

Storing inventory cuts both ways; while it can increase reliability,²⁰ it can be dangerous to the economics of firms.

²⁰ §9.3.

On the one hand, if you are a manufacturer, holding an inventory of inputs can ensure the production line operates smoothly, and an inventory of outputs ensures customers can get what they want when they need it. On the other hand, holding an inventory of poorly manufactured inputs can result in a lot of rework or the need to cancel production, while holding an inventory of unsold goods can be a huge cost if customer preferences change.

Instead of storing inputs, just-in-time production brings them to the production line immediately before they are used. Rather than a month of inputs, there might only be a days worth, or a few hours. This ensures if there is an upstream manufacturing problem, it can be identified immediately, and the supplier contacted. It reduces costs and improves quality. But just-in-time production requires a just-in-time logistics system. Trucks and trains steadily and reliably bringing goods from suppliers to customers, with a minimum of delay. This can occur with proximity, suppliers located adjacent to their customers, or a fast and direct network. While proximity works well if there is a single supplier for the customer, it does not

serve broad markets well. A factory can be adjacent to one or two customers; it cannot be adjacent to dozens because reliance on [proximity](#)²¹ reduces choice in the market place and makes the system more vulnerable.

²¹ §14.1.

On the other side, getting outputs to final market should also be done as quickly as possible. Holding a six-month inventory of last year's model is not going to work well for goods that turnover frequently. Customers too are used to getting what they want when they want it, through services like UPS in Figure 5.6. While pizza delivery has long been just-in-time, that is now being extended to a wide range of goods that promise not merely overnight delivery, but same day, and for a premium, same hour delivery. When it is faster to order something for delivery than getting it yourself, the nature of sales changes markedly.

In effect, a manufacturing production process is a [deterministic queue](#)²² with an arrival rate just equal to the departure rate. This can be achieved by storing inputs (cars) upstream of the manufacturing process in inventory, or having inputs arrive as needed.

²² §5.1.

The access of commercial producers to their suppliers in a reliable way is thus critical for modern manufacturing, and inventory is an expensive substitute for network reliability.

6

Traffic

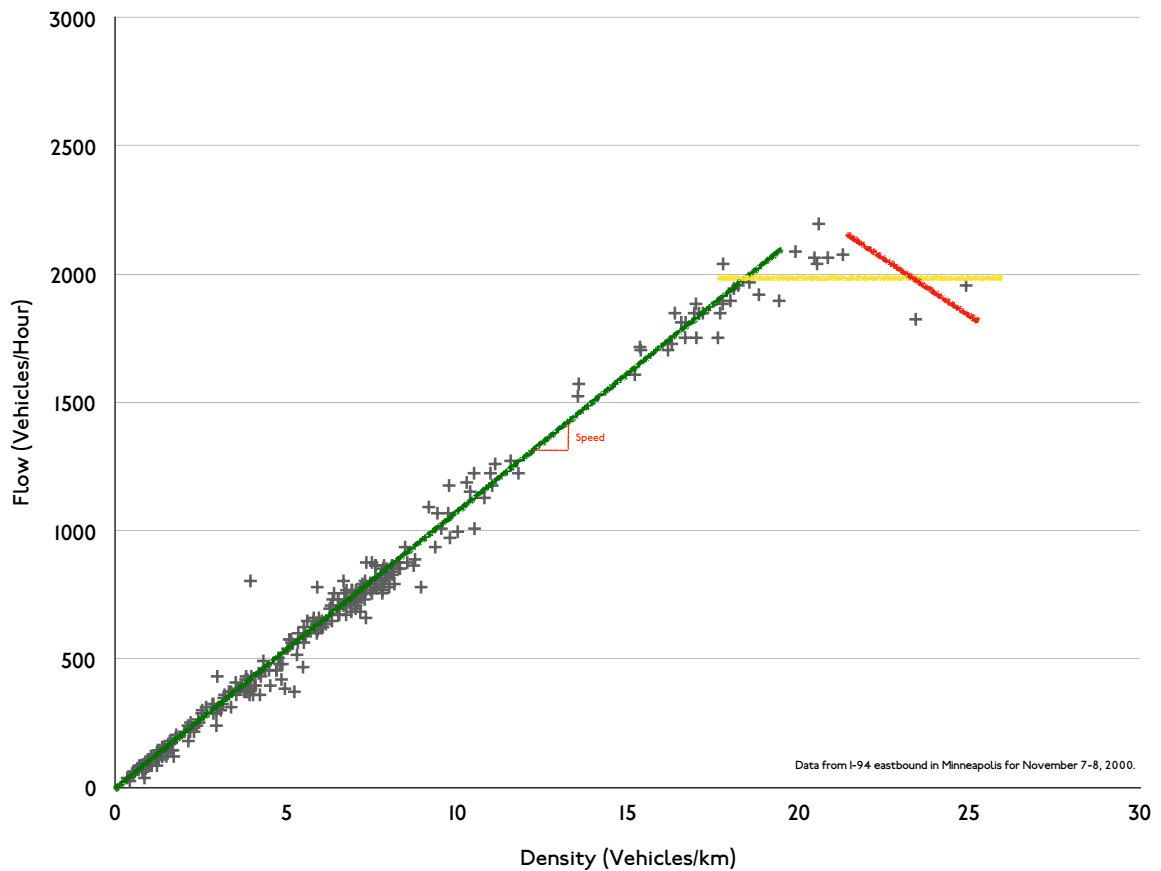


Figure 6.1: Microscopic fundamental diagram of traffic.

THE SPEED OF TRAVEL DEPENDS ON THE NUMBER OF TRAVELERS.

The 'fundamental diagram of traffic', shown in Figure 6.1, relates traffic flow (q) (vehicles per unit time),¹ density (k) (vehicles per unit distance),² and velocity (v) (how fast those vehicles are going in

¹ §6.1.

² §6.4.

³§6.6. terms of distance per hour).³

This can be expressed as an equation:

$$q = kv \quad (6.1)$$

Why does the fundamental diagram have the basic shape it does?

The graph as shown here has two axes, flow on the y-axis and density on the x-axis. Since speed is flow over density, speed is indicated by the slope (rise over run) of the curve from origin to the point of observation. Each point is a five minute average.

When density on the highway is zero, the flow is also zero because there are no vehicles on the highway. As flow increases (following the green line), density must increase if speed is constant. It turns out that we can drive the speed we want, unaffected by other drivers, at lower densities. So if there were only 1 car per kilometer, that car travels at what is called 'free-flow speed,' and if there were 2 cars per kilometer or 5 cars per kilometer, each car is still largely unaffected by the other cars. However as density rises (to empirically 15-20 vehicles per km), cars are eventually slowed by interactions with other vehicles.

Maximum throughput (capacity) is determined by the minimum following headway (here, following the yellow line, it is just about 2,000 vehicles per hour). In good conditions, drivers can travel at free-flow speed up to that point, but then congestion sets in, and as more cars are added, everyone slows down, as shown by the red line. At jam density, flow is back to zero. The example in the figure does not observe jam density, but rather serious congestion for only a few minutes.

The right side of the curve is far more chaotic than the left side of the curve in terms of the relationship of speed and flow for a given density.

We typically describe the fundamental diagram for automobiles, but there are such diagrams for every mode, including pedestrians (think about leaving a stadium after a game), bikes, and transit vehicles (think about [bus bunching](#),⁴ though that is complicated by stopping).

⁴§8.6.

⁵§1.

[Access](#)⁵ is the product of travel times and the location of activities. Travel times depend on the speed of travel on the network and the shape of the network. This chapter explores traffic flow theory, which helps us understand the speed of travel on networks.

6.1 Flow



Figure 6.2: I-94 in Minneapolis.
Photo by D. Levinson.

DRIVERS DON'T FOLLOW THE 'TWO-SECOND RULE.'

Flow (q) is the number of vehicles past a point per unit time. **Capacity**⁶ is the maximum flow. This depends on the road, the vehicles, and the drivers. If we were all racecar drivers, at the same spacing between vehicles, we could get a lot more vehicles past a point over a given period of time. Alternatively, if we drove a lot closer together at the same speed, we could also increase throughput. However, there are reasons we don't do this. We have to consider human reaction times. When taking driver's education, you may have learned the two-second rule: in good weather, follow the driver ahead of you with at least a two-second headway (alternatively, some driver education programs suggest one car length of spacing for every 10 mph). If everyone did this, we could get 1,800 vehicles per hour per lane past a point.

That 1,800 is not random. Recall there are 3,600 seconds in an hour; dividing by 2 seconds per vehicle gives us 1,800 vehicles per hour ($3,600/2 = 1,800$). In fact, limited-access freeways in good weather during peak periods have a much higher throughput, sometimes observed as high as 2,600 vehicles per hour per lane, though more typically between 2,000 and 2,200, indicating that people are following more closely than the two-second rule would suggest. The reason for the two-second rule is that if the driver

A note on terminology and the use of 'flow' versus 'volume.' Traffic professionals sometimes use the term 'volume' when they mean 'flow.' This is needlessly confusing, as the word 'volume' in common use is the measure of a 3-dimensional space, which is more akin to **density**(§6.4) than flow. Moreover, these same professionals use a volume-to-capacity ratio. If you were asked the volume of a cup, it would be the capacity of the cup (though the professionals probably mean to analogize with the volume of fluid in the cup rather than the total capacity of the cup).

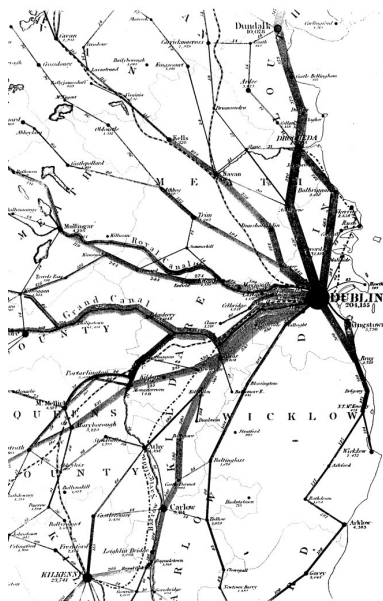
⁶§6.9.

ahead of you slams on the brakes, you have some amount of mental processing to see the brake-lights, tell your foot to move from the accelerator to the brakes, and push hard yourself (also known as perception-reaction time).

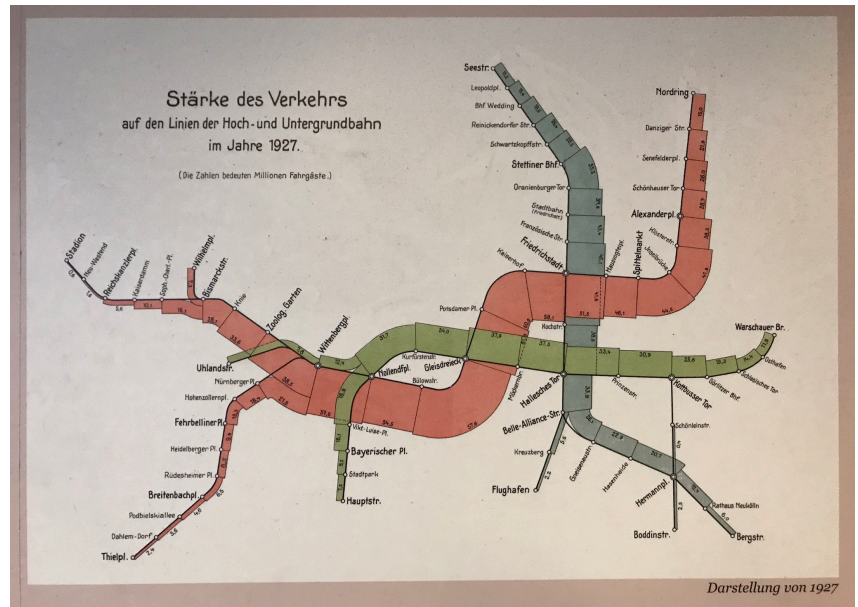
⁷§15. Robots won't need to follow the 'two-second rule.' While human drivers are instructed to follow the car ahead of with a two-second following time, which is more honored in the breach than the observance, we expect that [autonomous](#)⁷ (self-driving) vehicles could follow even more closely and precisely, and thus increase throughput on roads. A world of connected vehicles might be closer still, since information about acceleration and braking could be broadcast as it is happening from downstream cars, rather than requiring sensors, thus allowing connected and autonomous vehicles to safely follow even more closely.

⁸§6.9. Robots also won't need lanes to be as wide, since they will be able to drive with much greater precision. Both closer following and being closer laterally will, at least hypothetically, significantly increase the number of vehicles that can use a given amount of [road space](#)⁸ and may allow us to shrink our roads.

6.2 Flow Maps



(a) Dublin (1837).



(b) Berlin subway (1927).

FLOW MAPS INDICATE THE LEVEL OF TRAFFIC ON PARTICULAR TRANSPORT LINKS OR IN CORRIDORS, using the width of the line to indicate greater flow.

Typically as flows approach a city center, as in Dublin on the left of Figure 6.3 or Berlin on the right, the lines get wider and wider, which reflects greater flow on the corridor.

Figure 6.3: Flow maps of Dublin and Berlin. The Dublin map is credited as the oldest flow map, created in 1837 by Henry D. Harness of the British Army, showing traffic flows (§6.1) between Dublin and the rest of Ireland. Source: Michael Batty.

6.3 Flux

Figure 6.4: Biking on a woonerf in Delft. Photo by D. Levinson.



UNDERSTANDING THE FUNDAMENTAL DIAGRAM CAN HELP EXPLAIN WHY SOME LANES APPEAR UNDERUTILIZED WHILE OTHERS SEEM JAM PACKED.

⁹§6.1. Imagine there were a traffic lane full of cars, say a **flow**⁹ (q) of 500 vehicles per hour, going very slowly, an average speed (v) of say 5 km/h because of congestion, traffic signals, unloading trucks, and the like. The lane will appear full because the **density**¹⁰ is high. ¹⁰§6.4. Density (k) is vehicles per kilometer, and the relationship between flow, density, and speed is given by $q = kv$ (or equivalently $k = q/v$ or $v = q/k$). In this example, the density is 100 vehicles per kilometer, or about 10 vehicles every 100 meters, or 10 meters between vehicles, which is a pretty high density. Not quite jam density (minimum vehicle spacing, maximum vehicles per kilometer, on the order of 150), but close.

Imagine there were a parallel lane for bicycles. They are traveling at 20 km/h. The spacing is one bicycle every 40 meters, or a density of 25 bicycles per kilometer. Yet, the flow is an identical 500 vehicles per hour ($q = kv$). The lane looks one-fourth as full (even less, because bicycles take up much less space than cars) but serves just as many vehicles as the crowded lane.

¹¹§7.14. ‘Flux’ is flow per unit area (and flow is vehicles per unit time). When all **lanes**¹¹ are identical, the difference between flux and flow is

unimportant, but when comparing modes that have different widths, it can matter. Consider our previous example from the [fundamental diagram](#).¹²

¹² §6.

The bike lane is narrower than the car lane (Figure 6.4), so if we were to look at bicycles per square meter, accounting for a car lane of 3 meters (typically 3.65 meters, but narrower in cities) and a bike lane of 1.5 meters, we only need a density of one bicycle every 80 meters to get the same flux as the congested motor vehicle lane. One bicycle every 80 meters is about 1 bicyclist per block at a given time. In contrast, that congested lane of cars has at least 8 vehicles in it for the same length block.

6.4 Traffic Density

Figure 6.5: Red Queen and Alice. Carroll, Lewis (1871) *Through the Looking-Glass And What Alice Found There*. Illustrator John Tenniel.



'I'm sure I'll take you with pleasure!' the Queen said. 'Two pence a week, and jam every other day.'

Alice couldn't help laughing, as she said, 'I don't want you to hire me – and I don't care for jam.'

'It's very good jam,' said the Queen.

'Well, I don't want any to-day, at any rate.'

'You couldn't have it if you did want it,' the Queen said. 'The rule is, jam to-morrow and jam yesterday – but never jam to-day.'

'It must come sometimes to 'jam to-day,'" Alice objected.

'No, it can't,' said the Queen. 'It's jam every other day: to-day isn't any other day, you know.'

'I don't understand you,' said Alice. 'It's dreadfully confusing!'

– Lewis Carroll's (1871) *Through the Looking Glass and What Alice Found There*. (Carroll 1917). Illustrated in Figure 6.5.

DENSITY OF TRAFFIC (k) IS THE MEASURE OF THE VEHICLES PER LENGTH OF ROADWAY.¹³

When you are the only car on the road, the density of traffic is low. When there are many cars on the road, the density is high. The density of traffic indicates the [Level of Service](#)¹⁴, which is a grade that traffic engineers apply to roads, and ranges from A to F. From the driver's point-of-view, A is better than F, just like your report card. However, as we will see in the Level of Service section, it isn't quite that simple.

The maximum density – called 'jam density' – occurs when vehicles line up end to end, and none can move until the car in front moves first. If cars were on average 5 meters long, and literally 'bumper-to-bumper,' there could be 200 cars per kilometer per lane. In practice, cars are longer, and they are only figuratively bumper-to-bumper, so jam density is more like 200 cars per mile per lane or 125 cars per km per lane. But as the saying goes, 'your mileage may vary,' and this result depends on many factors. At jam density, traffic does not flow.

¹³ The letter ' k ' is used because it comes from the German word 'konzentration.'

¹⁴ §6.5.

6.5 Level of Service

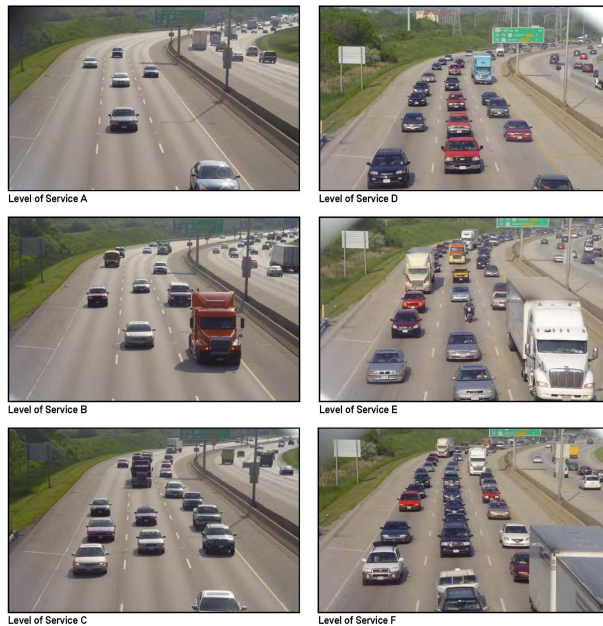


Figure 6.6: Freeway level of service ranging from A to F. Image courtesy of James Morrisey, Wisconsin Department of Transportation.

ROADS ARE OFTEN DESIGNED TO SATISFY A PARTICULAR ‘LEVEL OF SERVICE’ (LOS) STANDARD.

Level of service typically considers *density*,¹⁵ but may look at other measures as well, as shown in Figure 6.6. The *Highway Capacity Manual* defines level of service for different modes and conditions.¹⁶ As a description, LOS is useful. As a normative standard, it is problematic and can create perverse incentives.

¹⁵ §6.4.

¹⁶ (Transportation Research Board 2010).

We said Level of Service A is better ‘from the driver’s point-of-view.’ From society’s point-of-view, however, that might mean we spent too much money and are far too inefficient. Building enough capacity so that roads are level of service A (or even D) during rush hour is costly in terms of dollars and space. The benefits in terms of travel time savings are small. Those resources could have been spent somewhere else. Widening a road to improve its level of service necessarily worsens conditions for crossing pedestrians, who now have more pavement to traverse. As we will see with *induced demand*,¹⁷ a wider road will also attract more traffic.

¹⁷ §12.1.

Some jurisdictions, most notably the state of California, are now moving away from LOS standards, instead looking towards other measures such as vehicle miles traveled (which are harder to measure but much more useful).¹⁸

¹⁸ (San Francisco Planning Department 2016).

6.6 Speed

Figure 6.7: Charlotte Motor Speedway. Photo by D. Levinson.



It turns out space mean and time mean speed are related. Space mean speed (v_s) is also the harmonic mean of speeds passing a point during a period of time, in contrast with the arithmetic mean of the vehicles passing that same point, which is the time mean speed. These numbers are typically similar, yet the time mean speed is never lower than the space mean speed.

$$v_t = v_s + \frac{\sigma_s^2}{v_s} \quad (6.2)$$

Were all vehicles traveling at the same speed, the standard deviation of the speeds (σ_s^2) would be zero, and the two numbers would be identical (Knoop et al. 2009).

¹⁹ §6.

SPEED DEPENDS CRUCIALLY ON MEASUREMENT.

In the discussion of the microscopic [fundamental diagram](#),¹⁹ we talked about speed, and with a minimum of arithmetic, one sees that if $q = kv$, then $v = q/k$. The v here is what is referred to as space mean speed (v_s), so technically, $q = kv_s$. What is space mean speed? Imagine you get in a hot air balloon, look down at all the vehicles on a segment of roadway at one instant, and average the speeds of the vehicles over that roadway, then you will have the space mean speed.

Traffic engineers sometimes refer to another speed, the time mean speed (v_t). Time mean speed refers to the speed of vehicles measured over time at one point in space. If you are standing by the side of the road with a radar gun measuring the speeds of vehicles as they pass and average that number, you will get a time mean speed.

So how and what you measure matters, at least a little bit (and sometimes a lot). Often, it is more convenient to use point measures rather than longer space measures. Detectors built into the roadway, or even cameras at a point, can easily measure flow, occupancy (how long a detector has a vehicle covering it), and with some help, density. Using a measure of time mean speed and flow to estimate density would result in an underestimate (since time mean speed is always higher than space mean speed).

Use of GPS (satellite navigation) devices to measure speed leads

to other types of issues, as first the sampling is non-random, and second, the speed can be measured in various ways, often as the distance traversed divided by time, and so is neither space mean speed nor time mean speed.

6.7 Shockwaves



Figure 6.8: Shockwaves can form without bottlenecks. Source: (Sugiyama et al. 2008).

THE REPERCUSSIONS OF A CAPACITY DROP IN THE PRESENCE OF HEAVY TRAFFIC ARE SUDDEN AND PROPAGATE RAPIDLY UPSTREAM.

You are happily traveling down the road at free-flow speed, but on rounding the corner, you see ahead of you a queue of cars with their brake-lights on. You too will brake, and the queue of cars gets longer. The car behind you will shortly brake, and so on. A shockwave is propagating backwards.

²⁰ §6.1.

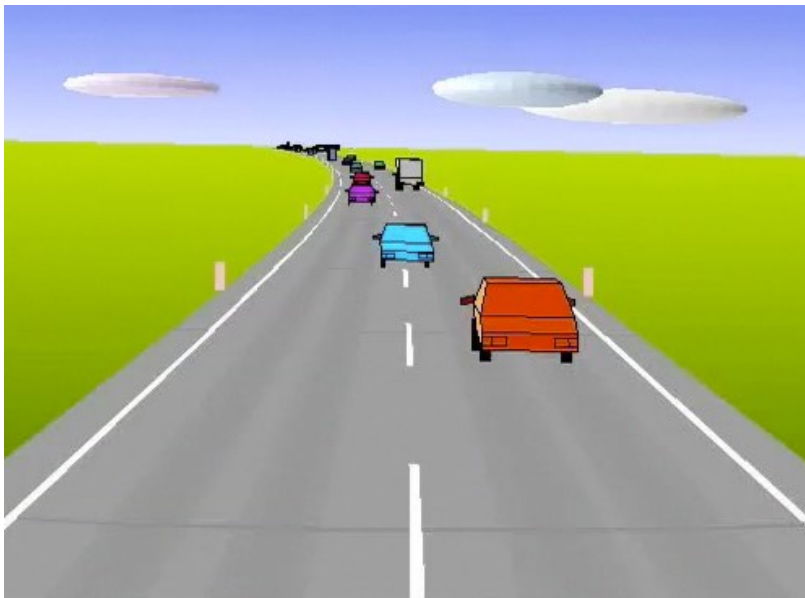
²¹ §6.4.

²² §6.6.

Recall that $q = kv$, that is, flow²⁰ equals density²¹ times speed.²² This is true both before you hit the brakes and after. The flow, density, and speed may change in the two areas (free-flow upstream and congested downstream), but the fundamental relationship between flow, speed, and density is stable. Because you decelerated, your speed drops, and the density of traffic increases because cars

are closer together. It turns out the speed of the shockwave (how quickly it propagates backwards) is the ratio of the difference in flows to the difference in densities.

These shockwaves can move very quickly, especially in heavy traffic, and queues can get very long, especially if the cause of the shockwave doesn't change. The causes are many. It might be a permanent feature of the road (a steep grade, a lane drop, a merge) or temporary (an **incident**²³ of some kind). It may not require a specific cause other than heavy traffic at all, as shown in Figure 6.8. The consequences of unpredicted incidents are far worse since travelers have not built it into their **schedule**²⁴ and lead to **unreliability**.²⁵



Before you hit your brakes, the relationship between flow and density was:

$$q_1 = k_1 v_1 \quad (6.3)$$

After, it was:

$$q_2 = k_2 v_2 \quad (6.4)$$

The speed of the shockwave (v_w), denoted by the brake-lights of vehicles, is moving backwards at a rate of:

$$v_w = \frac{q_2 - q_1}{k_2 - k_1} \quad (6.5)$$

²³ §5.4.

²⁴ §2.2.

²⁵ §9.3.

Figure 6.9: Shockwave. Screenshot from simulation by Martin Treiber. <http://www.mtreiber.de>.

6.8 Ramp Metering

Figure 6.10: Ramp meter, 4th Street at I-35W Northbound, Minneapolis. Photo by D. Levinson.



RAMP METERS ALLOCATE SPACE IN A DOWNSTREAM BOTTLENECK BETWEEN TRAFFIC ALREADY ON A MAJOR ROAD AND TRAFFIC ENTERING IT.

By limiting total inflow into the freeway, traffic managers can ensure that *queues*²⁶ do not form on the freeway, keeping traffic on the left side of the *fundamental diagram*:²⁷ below capacity and at free-flow speed.²⁸ This keeps total throughput near its maximum. By making the flow entering the freeway as smooth as possible, we can ensure that platoons of cars don't enter all at once, causing extra delay at on-ramps, and making merges inefficient.

The ramp meter will generally save travel time for the traffic already on the major road and transfer delay to the vehicles trying to enter at the ramp. This undoubtedly causes queueing on the ramp, where it can be better controlled. It may even lead to spillovers onto arterial roads if the queues get too long and discourage traffic from using the freeway in the first place.²⁹

²⁶ §5.
²⁷ §6.
²⁸ Where the green and yellow lines meet in Figure 6.1.
²⁹ (Levinson and Zhang 2006; Zhang and Levinson 2010).

6.9 Highway Capacity



Figure 6.11: I-94 Mississippi River Bridge. Photo by D. Levinson.

WHAT IS THE CAPACITY OF A HIGHWAY?

Let's make some assumptions:

- 1,800 vehicles per hour per lane (today, 2-second headways) vs. (with automation,³⁰ 1-second headways) 3,600 vehicles per hour per lane.
- 4 (12-ft (3.65m)) lanes today vs. 8 (narrower, 6-ft (1.8m)) lanes with automation
- 4 passengers per vehicle max vs. 1.5 passengers per vehicle today.
- 4 lanes in each direction, 2 directions

³⁰ §15.

We also need to figure out the average length of trip, which is a bit less obvious.

We illustrate this case with I-94 in Minneapolis and St. Paul, which parallels the Green Line of the [transit capacity example](#) later in the book.³¹ Current two-way Average Daily Traffic on I-94 on the peak section (near Riverside Avenue) is about 164,000. The number of entering trips (based on sum of Eastbound entering vehicles at on-ramps between and including Hennepin Avenue (Minneapolis) and Dale Street (St. Paul) is 161,000 (A distance of about 14 km or 9

³¹ §8.8.

Table 6.1: Highway capacity

	A	B	C
Hours	24	24	10
Vehicles per Hour per Lane	3600	833	750
Vehicles per Day	86400	20000	7500
Person Capacity per Vehicle	4	1.5	1.5
Numbered Exits	9	9	9
Lanes	8	4	4
Average Trip Length (in Exits)	1	3	3
Directions	2	2	2
Daily Person Capacity	49,766,400	720,000	270,000

miles)). The average number of Eastbound through trips in the Lowry Tunnel is about 87,000.

Average length is vehicle km traveled divided by number of trips. If we have an average flow of 80,000 vehicles in each direction for 14 km, this is 1,120,000 vehicle km traveled in each direction for this long section. The number of entering trips from Lowry Tunnel to Dale Street (eastbound) (inclusive) is about 240,000. This implies that the average vehicle uses I-94 between the cities uses the facility for 5 km (3 miles). Obviously, this is an approximation, but it is probably not too far off.³² Table 6.1 shows some capacities. This assumes 4 lanes in each direction throughout, which is also not strictly true.

In a world of automated vehicles (Scenario A), if everyone made short (1 exit) trips, in fully loaded (4 persons) per car, fully utilized over 24 hours per day, I-94 could carry about 50 million people per day over this stretch. In contrast, today (Scenario B) it carries about 720,000 people.³³

If we were to constrain it further, so it only operated 10 hours per day (recognizing people travel only during certain hours), it would carry fewer people (Scenario C). This is just a thought experiment to get some magnitudes. But clearly, we have a lot of potential capacity in the years ahead as automated vehicle technology becomes mainstream, if we manage our roadspace and our vehicles more carefully. This argues against capacity expansion.

³² There are 16 entrances over 9 interchanges over this span. Some have parts a, b, and c, and in the sequence (Exit 231 to Exit 240, inclusive). Note Exit 232 is missing. So we will go with 9 exits, or 1.6 km (1 mile) between exits, and an average trip of 3 exits).

³³ If we take the capacity of the parallel Green Line LRT as 2 lanes and I-94 traffic per lane, I-94 produces more person trips per lane than the Green Line.

6.10 High-Occupancy



Figure 6.12: This single occupant vehicle driver is ineligible to use the high-occupancy vehicle lanes unless she pays a toll. Photo by D. Levinson.

INCREASING AUTO OCCUPANCY REDUCES TRAFFIC FLOW FOR THE SAME NUMBER OF PEOPLE, AND THUS INCREASES SPEED (AND ACCESS), AT THE COST OF MORE COORDINATION BETWEEN PASSENGERS (DECREASING ACCESS).

In the early 1970s, high-occupancy vehicle (HOV) lanes were proposed and deployed in the United States to increase the person-moving throughput of roads. The idea was that by restricting the use of lanes to only cars carrying 2 or more (HOV 2+) (or 3 or more (HOV 3+), or 4 or more (HOV 4+)) passengers (or buses only), this lane would have fewer vehicles but more people, and be faster than congested general purpose (GP) lanes. While the idea is sound, if too few people use the lane, it moves fewer people than general purpose lanes. If the GP lanes are uncongested, the HOV lane doesn't save time and is operating on the left side of the [fundamental diagram](#).³⁴ It turns out most of the lanes are underutilized.³⁵ It also turns out that the lanes, instead of inducing much carpooling, typically served 'fampools,' family members (spouses, parent and child) who were commuting in the same direction anyway.

The idea came about to allow non-HOVs³⁶ to use the underutilized lanes in exchange for paying a toll. This would increase throughput on the lane, and as long as the toll was

In very congested corridors, 'Casual Carpooling' or 'Slugging,' emerges, in which a single-occupant vehicle would take strangers on as passengers in order to use the HOV lanes. This occurs most notably on the Shirley Highway in Virginia. Whether those travelers would otherwise take transit or drive is an important question.

³⁴ §6.

³⁵ See: (Dahlgren 1998; 2002).

³⁶ That is, SOVs in an HOV-2 case, or SOVs and two-person carpools in an HOV-3 case.

managed properly, keep the lane from getting congested and slowing down. Because different travelers have different values of time, it would be worth it to some people to pay the toll but not others. Because individual travelers have different values of time at different times, everyone benefits by the availability of this route that has a guaranteed free-flow travel time, rather than a guaranteed price of 'free.'³⁷

³⁷ HOT lanes are generally credited to Fielding and Klein, who published a Reason Foundation report and later a paper on the idea (Fielding and Klein 1993).

³⁸ §6.4.

³⁹ §6.5.

⁴⁰ (Janson and Levinson 2014).

These high-occupancy/toll (HOT) lanes, sometimes called express or managed lanes, typically charge a varying toll to single-occupant vehicles (SOVs), with the toll increasing during more congested periods. The toll is usually tied to time-of-day or to the **density**³⁸ of vehicles in the HOT lane. The purpose of raising the toll with congestion is to discourage enough demand so as to maintain a high **level of service**³⁹ (LOS) in the HOT lane. However, the HOT lane toll may act as a signal of downstream congestion (in both GP and HOT lanes), causing an increase in demand for the HOT lane.⁴⁰ It is hypothesized this is because the toll, rather than being much of a deterrent, instead acts as a signal of the level of congestion in the GP lanes, and thus a proxy for the amount of time savings.

6.11 Snow Business



Figure 6.13: There's no business like snow business. Photo by D. Levinson.

THE MEDIA SOMETIMES ASKS: 'WHY DO WE BECOME SUCH BAD DRIVERS WHEN IT SNOWS?'

When it snows:

Roads are slippery and require longer braking distances. People recognize that roads are slippery and give increased spacing ('following headway' in the jargon) to the car in front. Instead of following at a two-second headway (remember the [two-second rule](#),⁴¹ which is more honored in the breach than in observance), they may follow at a three-second headway. Since there are 3,600 seconds in an hour, a two-second headway implies 1,800 vehicles per hour (traffic engineers will note of course that capacities per lane on freeways are often greater than this in good conditions, implying a shorter than two-second headway). A three-second headway implies a service flow or capacity of 1,200 vehicles per hour. If the underlying demand (those who want to use the bottleneck at that time) remains unchanged at 1,800 vehicles per hour (say it snowed surprisingly in the middle of the day), then instead of serving 1,800 cars, a bottleneck would serve only 1,200 in an hour. This implies a queue 600 cars long. That is non-trivial.

⁴¹ §6.1.

Roads are slippery. People recognize that roads are slippery and drive slower to reduce braking distances, especially on roads which curve. "Free-flow speed is affected by pavement conditions, visibility, and wind

⁴² (Kyte et al. 2001).

speeds.”⁴²

Roads are slippery. People insufficiently recognize that roads are slippery and instead of giving increased spacing, they crash into the vehicle in front of them. This temporarily reduces capacity to zero as the drivers sort out the situation. A “significant increase was observed [in] winter snow event injury and non-injury crash rates (crashes per million vehicle kilometers) ... compared with equivalent winter non-snow event injury and non-injury crash rates.”⁴³

⁴³ (Khattak and Knapp 2001).

Snow does in fact reduce demand. Some people, choose not to go out when it snows. For instance, research on Minnesota travel patterns statewide finds elasticities shown in Table 6.2. If it snows, there is a 5.9% reduction in demand and 63.9% increase in crashes in the 3am to 9am time period. The reduction in demand seems to be less than the reduction in capacity, so queueing increases on roads at or near capacity in the absence of snow.

So we don’t become bad drivers when it snows. We are bad drivers, we just reveal it when the environment changes to the unexpected.

⁴⁴ §15.

This presents one more argument for [robot cars](#).⁴⁴ They can’t overcome the physics of ice or braking distance, or even eliminate congestion, but they can, in principle, better assess road conditions and be less likely to crash.

Table 6.2: Elasticities of demand and crashes with respect to snow in Minnesota. Source: (Huang and Levinson 2010).

Time	Demand	Crashes
3am–9am	-0.059	0.639
9am–3pm	-0.092	0.926
3pm–9pm	-0.115	0.752
9pm–12am	-0.091	0.814
all day	-0.079	

6.12 Macroscopic Fundamental Diagram

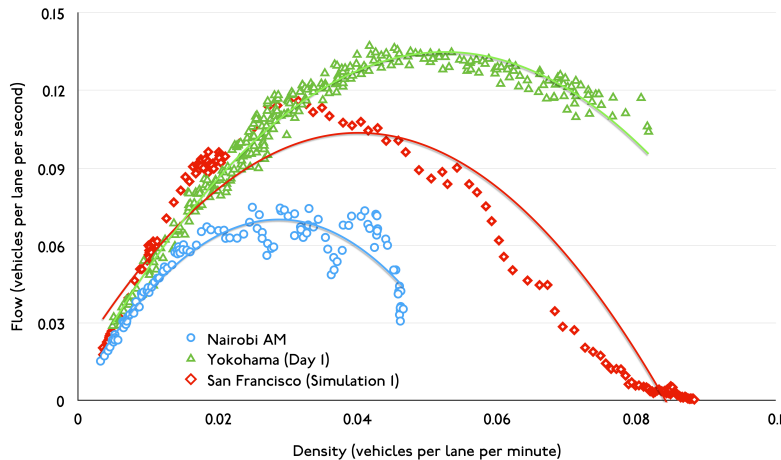


Figure 6.14: Macroscopic fundamental diagram. Source: (Gonzales et al. 2011; Geroliminis et al. 2007; Geroliminis and Daganzo 2008).

THE MACROSCOPIC FUNDAMENTAL DIAGRAM EXAMINES THE RELATIONSHIP BETWEEN SPEED, FLOW, AND DENSITY RELATIONSHIP FOR A NETWORK.

Recall the microscopic [fundamental diagram](#)⁴⁵ which considers the relationship between speed, flow, and density for a single link. We can extend that logic to a larger area.

The number of vehicles entering (or exiting) the network in a given period is the analog to flow, the number of vehicles traveling on the network at that time is the analog to density. While historically this was difficult to measure, with new data sources such as GPS probes, this has become feasible.

It turns out there is a relatively regular (low-scatter) relationship over areas (like a city center).⁴⁶ For more heterogenous networks, for instance the freeway system, the relationship is less stable.⁴⁷ The stability of the relationship suggests control strategies like district metering (analogous in intent to [ramp metering](#)⁴⁸), reducing the flow entering a saturated region to ensure speeds and exiting flows are maintained.⁴⁹ While this means spillovers to nearby neighborhoods or routes, it might be an overall improvement.

⁴⁵ §6.

⁴⁶ (Geroliminis and Daganzo 2008).

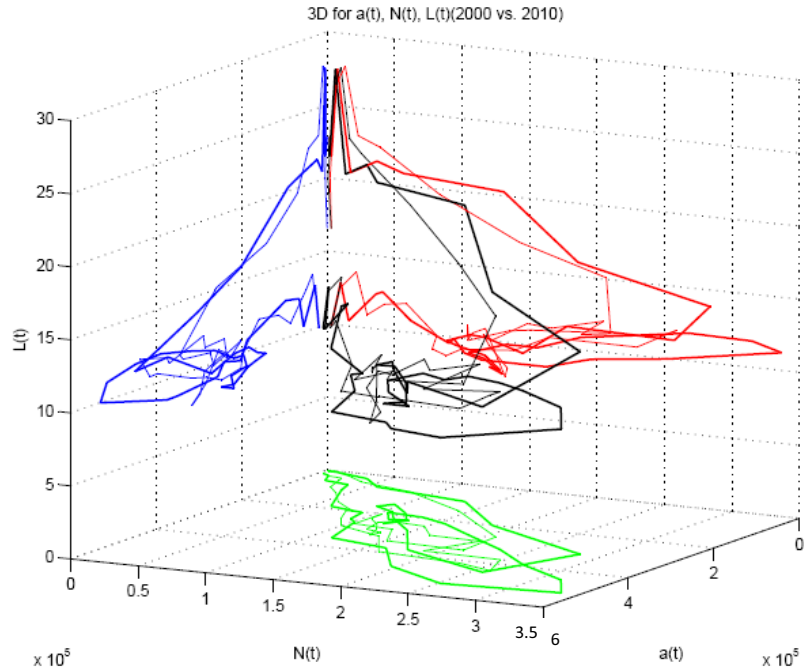
⁴⁷ (Geroliminis and Sun 2011).

⁴⁸ §6.8.

⁴⁹ (Geroliminis et al. 2013).

6.13 Metropolitan Fundamental Diagram

Figure 6.15: Metropolitan fundamental diagram. Source: (Wang and Levinson 2016). The blue lines compare trip length ($L(t)$) and arriving traffic ($a(t)$), the red lines compare length and number of vehicles on the road ($N(t)$), the green lines compare $N(t)$ and $a(t)$, and the black lines are a 3-D projection of all three variables. The thicker lines show the relationship in 2010, and the thin lines show that relationship in 2000 for the Minneapolis-St. Paul region.



THE FUNDAMENTAL DIAGRAM IDEA CAN BE EXTENDED TO THE METROPOLIS AS A WHOLE

⁵⁰ §6.
⁵¹ §6.12.

Recall the [fundamental diagram](#)⁵⁰ relates speed flow and density on a link, and the [macroscopic fundamental diagram](#)⁵¹ extends that to an area.

The Metropolitan Fundamental Diagram, shown in Figure 6.15 gives a more complex three-dimensional relationship. For an individual link, or a small area, trip length can be ignored. For the metropolitan area, trip lengths vary significantly. As shown in the opening figure, these three-dimensional relationships between system flow, density, and trip length appear stable over time for the same area.

7

Streets and Highways



Figure 7.1: Street in New York City. Photo by Wes Marshall.

WHAT'S THE DIFFERENCE BETWEEN A ROAD AND A STREET?¹

While there are general guidelines surrounding the differences, they are hardly followed universally.

First off, roads and streets are generic terms for thoroughfares that facilitate the movement of people and goods. Historically, and in theory, but not always in reality, roads tend to be located in more rural areas and streets in more urban. Both typically intend low-speed vehicle mobility (~25 to 35 mph (40 - 60 km/h)) as well as access to adjacent land uses. Streets are more likely to have sidewalks, on-street parking, and [street furnishings](#).²

What's the difference between a street and an avenue?

Avenues are similar to streets but typically wider with more travel lanes, a greater mix of land uses, and even more street trees. Consider the north-south running avenues in Manhattan against the east-west running streets. Based on the [Commissioners' Plan](#)³ of

¹ Etymologically 'Road' is an active term, deriving from a verb 'to ride,' and 'Street' is a more passive term deriving from a building material ('Strata' in Latin), and these still correspond to their link and place roles - roads being more about connection and streets being more about activity.

² §7.3.

³ §11.1.

1811, most avenues in Manhattan have a 100' right-of-way (although some such as Madison Avenue are smaller and others such as Park Avenue are larger) while most streets have a 60' right-of-way, although a subset of the east-west streets are larger, as in Figure 7.1.

⁴§7.2. **Boulevards**⁴ tend to be even larger than avenues.

Figure 7.2: 'Stroad' in downtown Detroit, Michigan. Photo by Wes Marshall.



So while these trends can help us get onto the same page when discussing different types of streets, the sheer number of exceptions makes for a lot of overlap. We also haven't even begun to discuss lanes, drives, alleys, trails, and mews – or **highways**,⁵ freeways, and turnpikes. There are also new terms – such as 'stroad' – that speak to our flawed history of street design. The term 'stroad,' illustrated in Figure 7.2, coined by Chuck Marohn and the Strong Towns group, describes a street/road hybrid that should have been designed more like an urban street but veered off into a design more befitting a rural highway. Chuck calls stroads the "futon of transportation design." Futons are a hybrid between a couch and a bed, and they end up making for either an uncomfortable sitting or sleeping experience. Similarly, stroads move cars too slowly to be efficient and too fast to make walking, bicycling, transit, or the adjacent land uses all that pleasant. His point isn't simply to define these hybrids; he wants to spark change. If we really want to capture the value of the infrastructure with respect to the private sector in our cities and towns, we need more streets and fewer stroads.

⁵§7.1.

7.1 Highways



Figure 7.3: Bobby Jones Expressway in Augusta, Georgia. Photo by Wes Marshall.

TODAY THE WORD “HIGHWAY” EVOKES IMAGES OF INTERSTATES, BUT THE TERM IN ENGINEERING AND LEGAL PARLANCE IS A BIT MORE GENERAL.

Interstates are indeed highways, such as the Bobby Jones Expressway in Figure 7.3, but when we restrict access and add medians and grade separation with crossing roads, engineers tend to use the word freeway (despite the enormous costs of driving and freeway infrastructure), motorway, or expressway instead. You also might hear them called limited-access highways, or if there is a toll, toll-roads or turnpikes. Highways (like high streets in cities) are the main routes, in contrast with byways, which are secondary, often more circuitous paths.

Highways, more generally, can really refer to any public thoroughfare, often but not always, for vehicular travel. Typically, we use the term highway for major roads and arterials that join different cities. Access can be limited, but most highways such as the ones shown below are at-grade with crossing roads and often include driveways to adjacent land uses. This leads to intersections, or *junctions*,⁶ that are typically not present along freeway corridors. Thus, older (particularly pre-Interstate) highways tend to have stop lights and stop signs instead of the interchanges and ramps we build along freeways.

Speed limits on highways vary quite a bit depending on *context*.⁷

⁶ §7.5.

⁷ §3.



(a) Two-lane rural highway in Connecticut. (US44)



(b) Urban multi-lane highway in Detroit (M-3).

Figure 7.4: Highways. Photos by Wes Marshall.

⁸ §7.14. The same goes for the number of lanes⁸ as well as most other street design features. So in effect, the term highway can apply to both a rural two-lane highway as well as urban multilane highway (Figure 7.4). While their designs are completely different, they both fall under the highway umbrella.

7.2 *Boulevards*



Figure 7.5: Southbound half of Octavia Boulevard in San Francisco, California. Photo by Wes Marshall.

THE USE OF THE TERM ‘BOULEVARD’ HAS LESS TO DO WITH CONVENTIONAL FUNCTIONAL CLASSIFICATION AND MORE TO DO WITH CHARACTER.

Boulevards tend to be wide streets (typically well above a 100’ (30 m) right-of-way (ROW)) in urban areas defined by the prevalence of features such as landscaped medians, generous sidewalks, and an abundance of street trees.

While most boulevards are indeed larger thoroughfares that probably fall under what engineers would consider an [arterial](#),⁹ they often don’t function as such. Arterials, by definition, intend to provide a low level of access to nearby land uses and a high level of through mobility. Boulevards sometimes fit this definition – although such roads are probably better known as parkways – but it is not uncommon for boulevards to also serve local accessibility as well as accommodate a variety of other important street uses: walking, bicycling, transit, and lingering via shopping and residential land uses in addition to public green spaces.

Boulevards have a long and distinguished history, dating back to the 1500s when cities such as Amsterdam converted their medieval walls into tree-lined streets.¹⁰ Although Paris did the same during the late 1600s, boulevards weren’t really in vogue until the City

⁹§7.

¹⁰ (Velasco 2002).

Beautiful movement nearly 200 years later. With the likes of Frederick Law Olmsted and others, the US in the late 1800s and early 1900s saw a number of new boulevards and parkways built – as well as what we now call multiway boulevards.

Multiway boulevards differ from standard boulevards in that they provide separated facilities within the same ROW for the through movement of somewhat fast vehicles (~35 mph (55 km/h)) alongside a slower-moving (~10 mph (16 km/h)), pedestrian-friendly environment with on-street parking and sidewalks. This combination of high mobility and high land access on the same facility simply does not fit into the traditional functional classification system promoted by the association of state departments of transportation (The American Association of State Highway and Transportation Officials or AASHTO). In other words, the functional classification system essentially divides street types into high mobility/low access (freeways and arterial), low mobility/high access (local streets), or medium mobility/medium access (collector roads). Thus, engineers fail to provide a category for this sort of mixed-use street that facilitates high levels of both mobility and access. Due to the rise of functional classification combined with concerns over a high number of potential **conflict points**,¹¹ multiway boulevards disappeared from the toolbox of US engineers beginning in the 1930s.¹²

¹¹ §7.7.

¹² (Jacobs et al. 2002).



Figure 7.6: Bike boulevard in Sydney, Australia. Photo by Wes Marshall.

¹³ (Cervero et al. 2009).

This, however, is beginning to change. Figure 7.5 shows the southbound half of Octavia Boulevard, which replaced the Central Freeway in San Francisco. For years prior, advocates pushed to get the Central Freeway torn down, but engineers were extremely wary of the potential for increased traffic congestion. The 1989 World Series earthquake damaged the freeway enough to force closure, and what happened in terms of traffic congestion? Nothing. The existing traffic using the Central Freeway either disappeared into the grid, changed modes, changed times, or didn't take the trip in the first place. This result provided enough evidence to secure permanent removal and eventually, installation of the first multiway boulevard in the US in more than eighty years. Now, Octavia Boulevard carries over 45,000 cars per day while also functioning as a place in itself where people work, shop, play, and live all side by side.¹³ Moreover, concerns over conflict points have yet to be realized in terms of worse road safety outcomes.

Another use of the term boulevard that is beginning to become more common is the bike boulevard. Bike boulevards, however, look and function quite a bit differently than the boulevards we've been talking about thus far. The idea behind a bike boulevard is to

turn a continuous residential street corridor into an extremely bike-friendly route. This usually includes trying to reduce both the volume and speeds of vehicular traffic through the use of diverters and traffic calming. This usually allows for a greater range of bicyclist types to traverse the corridor with a much lower level of traffic stress than most other types of bike facilities (lowering the perceived cost of travel, and thus increasing the perceived access). Empirical data from installations usually finds that bike boulevards do not negatively impact the adjacent streets in terms of car traffic or speeds either. Figure 7.6 depicts a bike boulevard in Sydney, Australia.

A final use of the term ‘boulevard’ is as the protective grassy strip (or ‘planting strip’ or ‘tree lawn’) between the street and the sidewalk (Figure 7.7), where street trees are often planted. This usage, most common in the midwestern United States, harkens back to the earlier etymology of the word.¹⁴ This area is referred to as a ‘verge’ in Australia, New Zealand and England (where the sidewalk is termed the ‘pavement’ or ‘sidepath.’)

While boulevards now come in various shapes and sizes, the underlying refrain is a general shift in mindset with regard to the underlying purpose of a street away from moving cars and accessing land. Well-designed streets are places in themselves, and boulevards tend to serve as great examples of this aim.



Figure 7.7: A sidewalk in Minneapolis, with a boulevard to the right. A family of wild turkeys are the pedestrian population in this scene. Photo by David Levinson.

¹⁴The word ‘Boulevard’ arises from the Dutch bulwark, or wall of a fortification, and more specifically came into French as the flat top of that fortification. *Online Etymology Dictionary* <http://www.etymonline.com/index.php?term=boulevard>.

7.3 Street Furniture

Figure 7.8: Parklet in Long Beach, California. Source: Chris McCahill, SSTI.



ONE OF THE MOST UNDER-APPRECIATED ASPECTS OF STREET DESIGN IS WELL THOUGHT OUT STREET FURNITURE.

The term “street furniture” collectively refers to all the amenities for public use in the right-of-way such as benches, bus stop shelters, wayfinding signs, drinking fountains, trash receptacles, mailboxes, police and fire department call boxes, and bicycle racks. Phone booths were once a ubiquitous fixture, and perhaps something akin to the Internet kiosks in New York City might eventually take their place. Street furniture not only provides for basic utilitarian uses – such as the ability to sit on a bench – but it adds to the visual interest and aesthetics of our street spaces. The presence of street furniture also clues drivers to the fact that they should expect pedestrians in this area.

With respect to street design, the area on the edge of the sidewalks between the vehicular portion of the street and where pedestrians walk is called the furnishing zone. This is the space where we typically locate street furniture as well as street trees, plantings, landscaped strips (see [boulevards](#)¹⁵), signs, fire hydrants, utility poles, and even cafes. In terms of functionality, the furnishings zone also serves as a buffer between cars and pedestrians. If designed poorly, then the sidewalk becomes an obstacle course, as in Figure 7.9(b).

¹⁵ §7.2.



(a) Street furniture



(b) Street furnishings obstacle course.

Street furniture is especially important to include along retail streets, at transit stops, as well as near important buildings and restaurants. However, they should be considered anywhere where we expect – or want – pedestrian activity.

Many cities have also started to transition on-street parking spaces into “parklets” that provide for an even larger pedestrian realm and more opportunity for street furniture. Parklets – such as the one in Long Beach, CA in Figure 7.8 – are particularly useful when the sidewalk space is limited and the surrounding land uses are complementary. According to city staff, the adjacent shops in this Long Beach example experienced an increase in sales following the parklet installation and had to hire additional employees.

Figure 7.9: Street furniture.
Photos by Wes Marshall.

7.4 Signs, Signals, and Markings

Figure 7.10: Signs.



Adapted from: wikibook *Fundamentals of Transportation/Traffic Control Devices*. Written by author.

SIGNS, SIGNALS, AND MARKINGS ARE CONSISTENT ACROSS THE LANDSCAPE SO THAT DRIVING SKILLS CAN BE TRANSFERRED SPATIALLY BETWEEN PLACES, MAKING ROADS MORE USEFUL AND, IT IS HOPED, DRIVERS SAFER.

Road signs in the United States are standardized in the *Manual of Uniform Traffic Control Devices*, (MUTCD) other countries have similar documents. The history of the evolution of this document is given in a series of papers by Hawkins.¹⁶

Like the networks themselves, signs *evolved*¹⁷ from local practice, cities, and states copying neighbors, inventing what they needed, and then later standardizing (first for rural and urban areas separately, and then jointly) after the value of coordination became apparent when automobile travelers crossed jurisdictional boundaries. From the first center line in Michigan in 1911 and stop sign in 1915, a 1923 recommendation established the basis of the shapes used for road signs today.

This system was improved over time. In 1924, the Minnesota Department of Highways published its *Manual of Markers and Signs* with the same shapes, but the white background was made yellow. In 1924, the American Association of State Highway Officials (AASHO, later AASHTO) adopted the MVASHD plan (with black on yellow); however, red and green on signs were rejected due to lack of visibility at night.

The objective of AASHO in these early years was first to inventory all of the sign characteristics that had been locally deployed, and then to standardized various aspects: shape, word, color, symbol, and uniformity of erection and application. Even as late as 1930, the third National Conference on Street and Highway Safety published a *Manual on Street Traffic Signs, Signals, and Markings* that had either white or black paint for concrete, and white or yellow paint for bituminous (asphalt). A red border and legend on yellow was suggested for stop signs.

Separately, standards were being developed for cities. Traffic signals are largely an urban phenomenon. While the date of the first traffic signal is contested, the electric traffic signal appeared in

¹⁶ (Hawkins 1992a;c;b; 1994).

¹⁷ §14.7.

The Mississippi Valley Association of State Highway Departments (MVASHD) proposed the following: shapes that were most complex to make like the circle and the octagon would indicate most danger and be used least often. Their recommendations for shape are standard in the US, and some globally:

- Round: Warns of railroad crossing;
- Octagon: Signifies a stop;
- Diamond: Indicates ordinary condition of danger requiring precaution at all times;
- Square: Indicates intermittent danger conditions requiring little more than ordinary care;
- Rectangular: Indicates regulatory or directional information; and
- Cut-Out: Identifies highway routes.

Cleveland in 1914, and the first three-color traffic signal in 1920.

Finally, in 1932, a Joint Committee on Uniform Traffic Control Devices met to rectify and combine the separate AASHO and NCSHS manuals for rural and urban traffic into a complete manual. Main initial points were color codes, signs at night, and reduced sign sizes in urban areas. Visibility research was undertaken, sponsored by the Bureau of Public Roads. Minor changes continued after this date, though a modern driver would certainly understand the road at this point. For instance in the 1954 MUTCD, the stop sign changed from black on yellow to white on red; yield sign were introduced as triangle (black on yellow), emulating European standards.

In 1924, the National Conference on Street and Highway Safety (NCSHS) developed urban sign recommendations. Their recommendations for color are recognizable:

- Stop: Red;
- Proceed: Green;
- Caution: Yellow;
- Cross-Roads: Purple or other distinctive color; and
- Centerlines: White.

7.5 Junctions

Figure 7.11: Protected bike intersection. Source: Alta Planning.



WHEN ONE ROAD MEETS ANOTHER, IT'S A 'JUNCTION'.

Although we often use 'intersection' as a synonym, junctions more generally refer to places where people change modes, routes, or directions. Junctions can also be [grade-separated interchanges](#),¹⁸ [signalized intersections](#), [roundabouts](#),¹⁹ protected-bike intersections such as the one in Salt Lake City shown in Figure 7.11, or one of any number of various types.

¹⁸ §7.10.

¹⁹ §7.8.



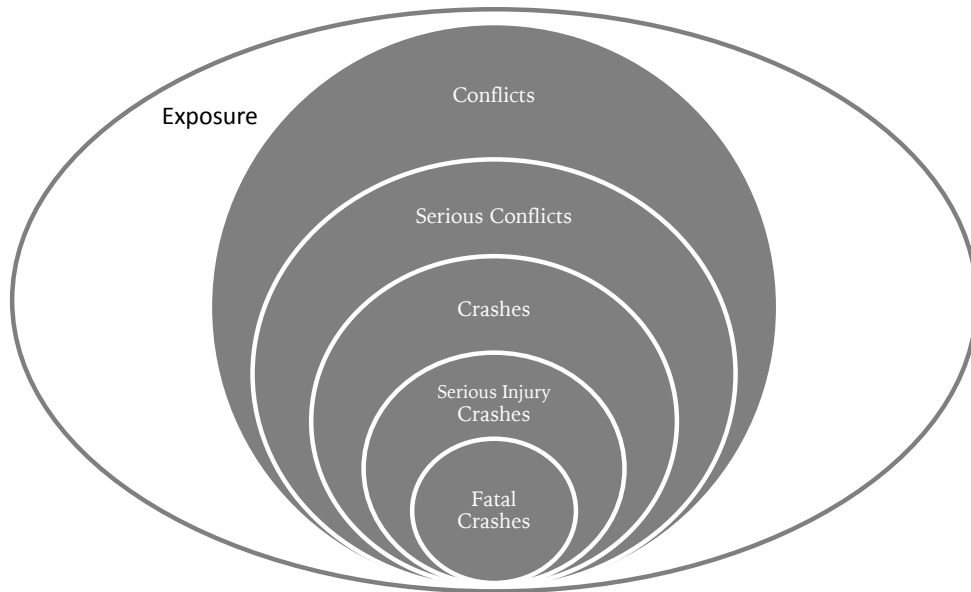
Figure 7.12: Bank Junction, London, c. 1850. Source: VictorianLondon.org.

Historically, places developed around junctions where roads or trade routes intersected. Nine different streets converge at Bank Junction in London, and you can see in Figure 7.12 just what this meant in mid-1800s England. The area not only claims home to quite a few notable sites – such as the Bank of England headquarters and Mansion House, home of London's mayor – and one of the busiest Underground stations in the city.

This sort of development also happened where modes, instead of roads, intersected. New York City evolved, at least in part, due to its harbor with access to the Atlantic Ocean, the Hudson River, the Erie Canal, and land routes such as the Boston Post Road. Chicago developed because it linked major railroads with inland rivers and canals.

So while all intersections are junctions, not all junctions are intersections.

7.6 Conflicts



CONFLICTS REFER TO TRAFFIC SITUATIONS WHERE ROAD USERS WOULD CRASH IF ONE OR BOTH DIDN'T BRAKE, WEAVE, OR MAKE ANOTHER EVASIVE MANOEUVRE.²⁰

For various reasons – such as the relative infrequency of crashes, the under-reporting of crashes, and the chance of incorrect or incomplete crash records – engineers have been using traffic conflicts as a proxy for road safety for more than five decades, starting in 1967 with two General Motors researchers. Perkins and Harris developed what became known as the Traffic Conflicts Technique (TCT) in 1967 in order to show that GM cars were safer than the cars of other manufacturers.²¹

Traffic safety researchers were quick to adopt and refine the techniques, as evidenced by the hundreds of published papers on the subject, under the supposition, as shown in Figure 7.13, that conflicts were associated with crashes and that conflicts could even be used to predict future crashes. Conflicts were categorized by type as well as by severity. They went on to eventually include modes other than the automobile. Yet, it took nearly twenty years for researchers to significantly associate conflicts with actual crash

Figure 7.13: Connecting exposure to conflicts to crashes by type. Image by Wes Marshall.

²⁰ In traffic engineering terms, 'conflicts' don't quite overlap with what most of us would consider near-hit crashes. It's always a bit funny to see these called near-miss crashes because if you nearly missed something, you hit it.

²¹ (Perkins and Harris 1968).

²² (Migletz et al. 1985).

²³ (Svensson 1992).

outcomes.²² Today, one of the most widely researched and used conflict approaches is the Swedish Conflict Technique.²³

Still, there have always been issues with conflict-based techniques in safety research. The definition of a conflict, or even a serious conflict, has long been up for debate. Even when conflicts are well defined and observers well trained, there are inconsistencies in assessing conflicts.²⁴

²⁴ (Chin and Quek 1997).

²⁵ (Hauer and Garder 1986).

The bigger issue has to do with how much conflicts matter in terms of actual road safety outcomes.²⁵ Today, we have empirical examples of shared spaces and multiway boulevards that have intentionally – and successfully – maximized the number of conflicts as a means of increasing safety, applying the theory of [risk compensation](#)²⁶ in reverse. The trick seems to be understanding context when assessing value and meaning with respect to conflicts.

²⁶ §2.17.

Our meaning of conflicts also differs from what we mean by [conflict points](#),²⁷ which will be discussed in the next section.

²⁷ §7.7.

7.7 Conflict Points

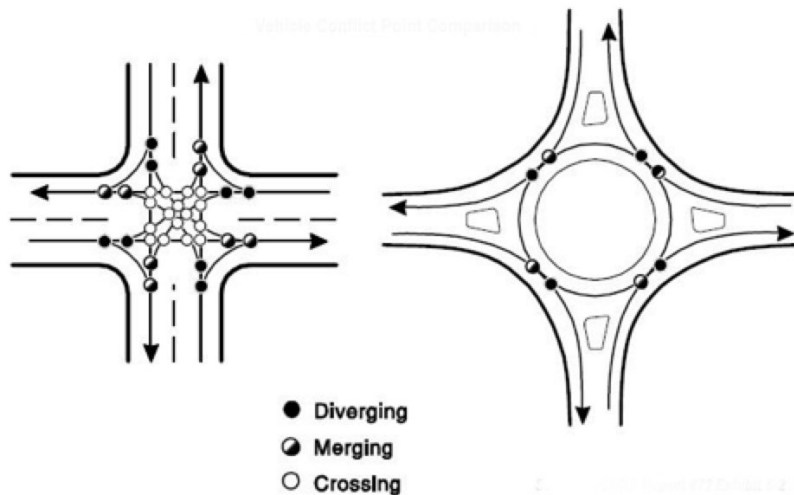


Figure 7.14: Conflict points.

CONFLICT POINTS TELL WHERE WITHIN AN INTERSECTION OR ALONG A STREET ROAD USERS WOULD HAVE CONFLICTS WITH OTHER ROAD USERS.²⁸

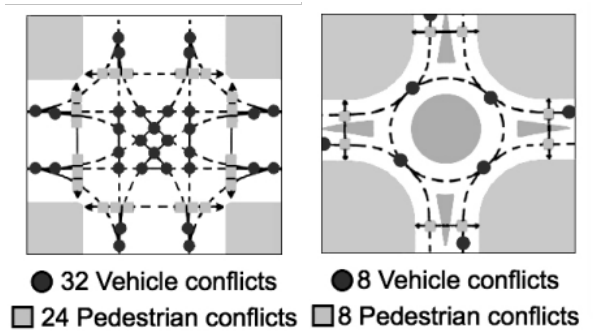
This means mapping and counting the locations, within an intersection for instance, where road users would be crossing, merging, or diverging with other road users.

The theory behind conflict point analysis is that we should minimize them in order to increase safety. For instance, the conflict point diagrams in Figure 7.14 compare the number and type of conflict points for a 4-way intersection against a roundabout.²⁹ The 4-way intersection has 16 crossing, 8 merging, and 8 diverging conflict points; the roundabout has 4 merging and 4 diverging conflict points. In terms of safety, the biggest benefit for roundabouts seems to come in the fact that we managed to design out the crossing-type conflicts, which can be the most dangerous. And since roundabouts have proven to be generally safer than conventional 4-way intersections, we have a winner in terms of conflict point analysis.

One problem is that most every conflict point diagram you see doesn't include modes other than cars. Figure 7.15 adds pedestrian conflicts to the mix. Another problem is that we are using conflict points as a proxy for actual road safety outcomes, which doesn't always work out. For instance the multiway boulevard, which produces on the order of 64 conflict points, often does so with a

²⁸ Unlike conflicts, conflict points don't have much to do with actual interactions between road users in the engineering vernacular.

²⁹ §7.8.



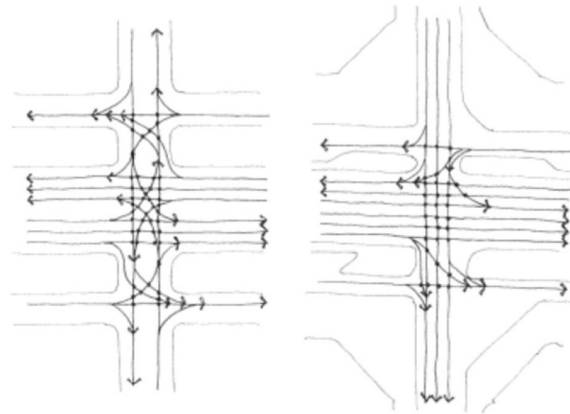
(a) Four-way intersection vs. roundabout. Source: Alex Inhen, nextSTL.

Figure 7.15: Additional illustration of conflict points.

³⁰ (Jacobs et al. 2002).

³¹ Driver's education for senior citizens often makes the same point, routing to avoid left turns (right turns in left-drive nations) for the issue of safety.

³² (BBC 2009; Black 2016; Nottingham City Transport 2017).



(b) Multiway boulevard. Source: (Jacobs et al. 2002)

better safety record than similar-sized, conventionally-designed streets.³⁰

We also consider potential conflict points in terms of street design, particularly with driveways and left-turning vehicles. This leads to what we call access management, which is essentially a set of strategies intended to reduce conflict points and, in turn, conflicts within the built environment. Similarly, the shipping company UPS attempts to minimize the number of left-turns that their drivers need to make.³¹ In fact, they contend that more than 90% of the turns their drivers make are right turns. While the UPS Right Turn Policy was initially envisioned in 2004 as an environmental effort to reduce gasoline consumption and emissions – with which it has been successful, saving an estimated 30 million miles driven and 3 millions gallons of gas annually – it also improves safety. This is the same reason many bus routes in the UK and other left-hand drive countries use anti-clockwise loops at suburban terminals and in city centers.³²

7.8 Roundabouts



Figure 7.16: Rotary intersection in Bourne, Massachusetts, 1960. Source: National Archives catalog.

FOR MOST AMERICANS, ROUNDABOUTS PROBABLY FALL SOMEWHERE ON THE LOVE-HATE SPECTRUM BETWEEN EXTREME DISLIKE AND HATE.

One reason for such an unenthusiastic assessment can likely be traced to some common misconceptions about what we are actually talking about when we talk about roundabouts.

Having grown up near Boston, Wes spent many Friday evenings on hot summer nights sitting in traffic with everyone else trying to get to Cape Cod. The Cape Cod Canal cuts the area off from the rest of Massachusetts, so the only way across by car meant traversing one of the two bridges. For years, both bridges also had what those from New England called a ‘rotary’ on one side or the other, as shown in Figure 7.16. Traffic would routinely back up for miles at both bridges, making what could have been a 75-minute drive considerably longer. The culprit was often these multilane rotaries. Such intersections proved to be inefficient – both in terms of traffic flow and land consumption – as well as dangerous. You may also remember Clark Griswold getting stuck in a London traffic circle in the 1985 movie *European Vacation*. ‘Hey look kids, there’s Big Ben! Parliament!’ Homer Simpson was in a similar situation in a 2003 episode and nearly killed the Queen of England! Examples like these end up giving all circular intersections a bad rap. Not surprisingly, circular intersections were phased out of most US design toolboxes and the minds of many Americans.

Modern roundabouts, not developed until the 1960s, refer to

Figure 7.17: Roundabout in Vancouver, British Columbia. Photo by Wes Marshall.



something quite different. For one thing, roundabouts are much smaller than the old rotary intersections. Instead of outside diameters exceeding 300' (90 m) or 400' (120 m), modern single-lane roundabouts typically range between 90' (27 m) and 180' (54 m). Another thing is that the cars entering must yield to the cars already in the roundabout (this was usually, but not always, the case with the older traffic circles and rotaries). The main defining characteristic of modern roundabouts, however, has to do with speed deflection. Speed deflection refers to angle at which cars enter the roundabout. With the old rotaries, there was little to no horizontal deflection of through traffic so cars could easily exceed 30 mph (50 km/h). A well-designed modern roundabout typically has enough deflection in the angle of this approach to actively manage vehicle speeds to less than 20 mph (30 km/h). It can also still handle truck traffic with design features such as a traversable apron that skirts the inner circle, which can be seen in Figure 7.17 from Vancouver.

So what does the research tell us about modern roundabouts? In most contexts, they move traffic more efficiently and are safer than conventional intersections. Why would this be the case? In terms of efficiency, there is no waiting for the light to turn green when there is no cross traffic. In fact, single-lane roundabouts have been shown to reduce delays as compared to conventional intersections and effectively manage traffic flows as high as 25,000 cars per day. Less



Figure 7.18: Neighborhood traffic circle in Berkeley, California. Photo by Wes Marshall.

idling also means fewer emissions. In terms of safety, roundabouts eliminate [conflict points](#)³³ and the most dangerous types of conventional intersection crashes; while you may get more sideswipe or rear-end crashes, such crashes are far less likely to be fatal or result in severe injury. Also if the roundabout is designed with adequate deflection, these crashes tend to happen at slower speeds. This reduces crash severity to the tune of 78-82% fewer serious injury or fatality crashes as compared to conventional intersections.³⁴

Roundabouts have further advantages for [intersections](#)³⁵ that are not four-way ([degree 4](#))³⁶ as shown in the opening image, as those result in even more delays with signals. They can help enable topologies such as the [hex](#),³⁷ which are complicated from the traditional traffic engineering perspective.

There are valid concerns about pedestrians and bicyclists in roundabouts, but splitter islands, setback crosswalks, and sidewalks – when combined with slower vehicle speeds – help tremendously. Interestingly, many places allow bicyclists to act as either a vehicle or a pedestrian in roundabouts. Other concerns center more on effectively serving those with impaired vision, which is an issue with most roundabouts, but still better than many other intersection designs due to the lower speeds.

While multilane roundabouts are unnecessarily used in many situations where a one-lane roundabout would work well, multilane

³³ §7.7.

³⁴ (AASHTO 2010).

³⁵ §7.5.

³⁶ §10.3.

³⁷ §11.3.

roundabouts still offer many of the same advantages. Compared to single-lane roundabouts, however, they:

- lose some speed deflection when flows are low;
- introduce a new crash type: sideswipe crashes due to lane changes; and
- make things more difficult for pedestrians and bicyclists.

You can also include neighborhood traffic circles – which are even smaller than most modern roundabouts – in this overall discussion of circular intersections. Figure 7.18 from Berkeley, California combines 4-way stop control with a circular intersection. While not quite a roundabout, it is a good example of using a small traffic circle to help manage speeds and improve safety.

Compared to signalized intersections, roundabouts are generally less expensive (where land is plentiful), more efficient, more environmentally friendly, and perhaps most importantly, safer. Furthermore, you never have to worry about a power outage with a roundabout. While there are legitimate reasons not to use roundabouts in some situations – such as highly unbalanced traffic flows or right-of-way limitations – many get eliminated as an option due to our cultural biases against them. All we are saying is, give roundabouts a chance.

7.9 Complete Streets



Figure 7.19: A complete street in Sydney. Photo by Wes Marshall.

ONE OF THE MOST PROLIFIC ACTIVE (HUMAN-POWERED) TRANSPORT POLICY MOVEMENTS FOCUSES ON “COMPLETING THE STREETS.”³⁸

For instance in the US, Smart Growth America founded the National Complete Streets Coalition in 2005, which has been instrumental in helping pass complete streets legislation – intended to compel street designers to consider all road users – in over 1,000 municipalities/agencies and across 33 US States. The underlying goal of a complete streets policy is to compel planning for all modes in all transport projects, but the more commonly held impression is that completing the street means adding design elements like sidewalks, [bike lanes](#),³⁹ and raised medians to almost every major street.

Given the term ‘complete streets’ itself, this line of thinking is not surprising. So despite the broader intentions, the most visible end products of this movement are typically manifested with street design elements intended, as the coalition suggests, to help “build roads that are safer, more accessible, and easier for everyone.” In addition to improving road safety, the National Complete Streets Coalition highlights other benefits, including: higher rates of active transport, less driving and fewer vehicle miles traveled, as well as

³⁸ (Sears 2014; Mustafa and Birdsall 2014).

³⁹ §3.6.

better public health outcomes.

While the fundamental lessons behind complete streets are important, street-level design elements matter, but they need to be considered in combination with street network design.

The premise behind the complete streets movement in terms of planning for all modes is indeed valuable; however, focusing simply on street-level, like Figure 7.19, without consideration of the network-level impacts will leave us in a position where achieving more sustainable transport behaviors – as well as better road safety and health outcomes – is difficult.

In other words, complete streets in an incomplete city don't work. We need both complete streets and complete cities.

7.10 *Dedicated Spaces*



Figure 7.20: TX-366 limited access freeway in Dallas, Texas. Photo by Wes Marshall.

SOME PLACES DO NOT ALLOW FOR PEDESTRIANS OR BICYCLISTS.

When we also have limited-access, we call these freeways. Such facilities prioritize the through-movement of vehicles. The freeway in Figure 7.20 cuts directly through Dallas, Texas, has a high level of mobility, and a low level of land access. This is thought to provide efficiency and safety with respect to moving high traffic volumes at high speeds.

At the other end of the spectrum, we have pedestrian-only streets that generally exclude vehicles. Figure 7.21 depicts the 16th Street Pedestrian Mall in Denver, which allows for pedestrians and buses, but not cars or bicycles. Professor Norman Garrick of the University of Connecticut considers such designs – as well as *shared spaces*⁴⁰ where cars would also be allowed – to be on ‘context time.’ Under context time, the social aspects of the space governs behavior. Thus, such dedicated spaces are multi-functional, culturally defined, personal, diverse, and unpredictable.

In contrast, freeways such as the one in Dallas are on ‘system time.’ Dedicated spaces on system time are single-purpose, regulated, impersonal, uniform, and predictable. There is a time and place for both system time and context time, just like there is a time and place for dedicated spaces that focus on cars and mobility as well as those that focus on other modes and social needs. Just tell the flip-flop-wearing bicyclist (Figure 7.22) that riding on the Southeast Expressway in Boston smells like an unwise life decision.

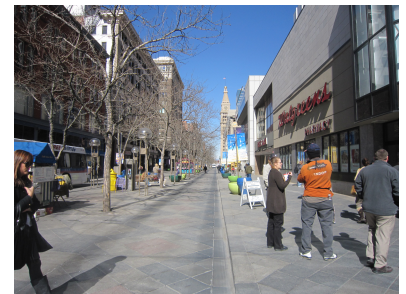


Figure 7.21: 16th Street Pedestrian Mall in Denver, Colorado. Photo by Wes Marshall.



Figure 7.22: Bicyclist on I-93 in Boston, Massachusetts. Source: Boston.com.

⁴⁰ §7.11.

7.11 Shared Space

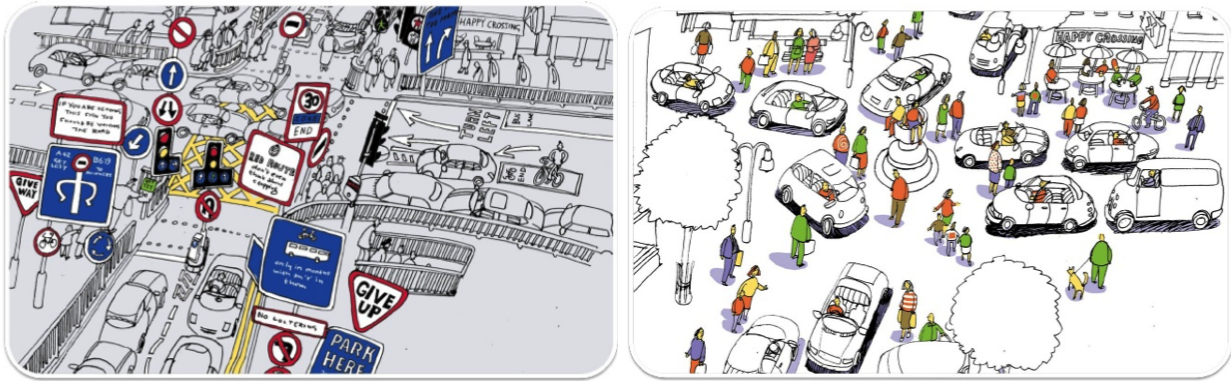


Figure 7.23: Conventional design v. shared space. Cartoons by Ben Hamilton-Baillie and Paul Boston. Used with permission.

ENGINEERS AND PLANNERS TYPICALLY DESIGN TRANSPORT SYSTEMS TO ISOLATE DIFFERENT MODES OF TRAVEL AS MUCH AS POSSIBLE; VEHICLES SHOULD STAY ON THE ROADWAY, BICYCLES IN THE BICYCLE LANES, AND PEDESTRIANS ON THE SIDEWALK.

Some visionary transport engineers and planners seek to do the exact opposite (akin to how we built streets before the automobile) and encourage increasing interactions between different modes by removing horizontal and vertical demarcations, removing all signage,⁴¹ and abolishing the basic rules of the road.

⁴¹ §7.4.

By removing what seems to give us 'order' in the transport system, the theory of shared spaces is that we force road users to react to social cues. In other words, when road users enters what for all intents and purposes is an unregulated situation, they must orient themselves to the situation by observing and building upon the order established by fellow road users as opposed to that instituted by externally-created rules. The thinking is that this creates more awareness, and that perhaps we can achieve even greater 'order' in the transport system.

⁴² (Klein 2006).

To picture this concept, imagine a public ice-skating rink and try explaining to somebody who has never seen one how it works.⁴² If you tell them that dozens of people on sharpened metal blades are moving throughout a confined area at varying speeds, and doing so on a surface made of ice, they'd probably picture total chaos. Such chaos, however, rarely fails to materialize. Rather, the lack of rules and demarcations forces skaters to become aware of their surroundings and fellow users while social cues helps skaters

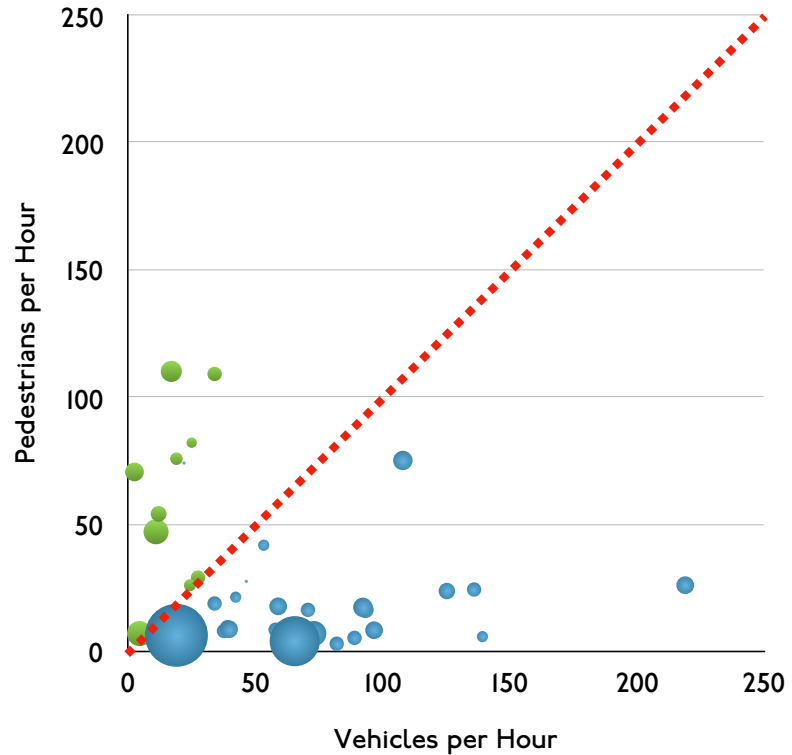
modify their paths and avoid collision with other users of the system. Social scientists term this *spontaneous order*. As with everything, there are good and bad designs.

In the context of a street or intersection, it is exceedingly difficult for traffic engineers to give up such control and cede whose turn it is to cross the street to the road users themselves while hoping for order to spontaneously emerge. It seems like we would be setting ourselves up for madness, but similar to the ice skating rink example, it also seems to work in the transport system.⁴³ Shared space designs, primarily undertaken in European, and Asian cities, as well as Auckland, New Zealand, have somewhat surprisingly been shown to increase both efficiency and road safety over more conventional designs. Whether this design concept takes off in the rest of the world remains to be seen.

⁴³ It should be noted the visually impaired community, especially in the UK have been vocal opponents of shared spaces, which may be harder to navigate when you cannot make eye contact. Also a recent incident on Exhibition Road in London where a taxi injured 11 pedestrians in a shared space suggests the design issues are not fully settled nor understood (Mairs 2017).

7.12 Spontaneous Priority

Figure 7.24: Spontaneous road user prioritization in shared space intersections. Red line = 1:1 ratio of pedestrians to vehicles; green circles = pedestrian-dominated intersections; blue circles = vehicle-dominated intersections; circle size = higher level of modal dominance when conflict arose) Source: (Ferenchak and Marshall. 2017).



⁴⁴ §7.11.

SHARED SPACES⁴⁴ ARE STREETS WHERE ALL SIGNS, TRAFFIC CONTROL DEVICES, STREET MARKINGS, AND SEPARATION OF MODES HAVE BEEN REMOVED.

This way of thinking forces all road users, no matter the mode of transport, to take responsibility for their own actions and negotiate the space via all the other road users by means of eye contact and other social cues.⁴⁵ This is in stark contrast to a conventional street design where modes tend to be separated and movements guided and controlled by traffic signals and the like. In the right context, the result of shared space is not chaos; instead, spontaneous order takes hold, resulting in a space often more efficient and safer than a conventional design.

Shared space is an often misunderstood concept. First things first; the right context is key. Shared spaces would not work everywhere, especially when the focus is mobility and high travel speeds. The surrounding land uses and the way that these buildings

⁴⁵ This section is based on work by Ferenchak and Marshall. (2017).

and activities interact with the street make a big difference. So does the mix of road users. A street dominated by cars would be hard pressed to function like we might imagine a shared space should.

Many people believe living streets and/or woonerfs to be synonymous with shared spaces. However, these street types specifically grant priority in the street space to pedestrians. A true shared space concept does not. Why? Because it doesn't have to. In the right context, this prioritization occurs naturally. [Ferenchak and Marshall. \(2017\)](#) analyzed data from 37 shared space intersections with high levels of interaction between pedestrians and vehicles and assessed which mode acquiesced to which when a conflict arose. When vehicles outnumbered pedestrians, while controlling for other design factors, the pedestrians tended to back off and cede the road space to the cars. However when pedestrians outnumbered cars, this prioritization flipped. Now, the cars were the ones yielding to the pedestrians when a conflict arose. The red line in [Figure 7.24](#) represents the 1:1 ratio of pedestrians to vehicles. What we call the modal dominance index is represented by the size and color of the circles. The green circles signify pedestrian-dominated intersections while the blue circles represent vehicle-dominated intersections. The size of the circle indicates a higher level of dominance over the shared space.

Many shared space designers are tempted to follow the living street or woonerf model and grant pedestrians priority in the street space, to the point where there is a call for what is known as a Pedestrian Priority Shared Space (PPSS). While such designs can be successful and find a multitude of benefits, putting up signs to grant pedestrians priority misses a key point of the shared space concept. A true shared space in the right context doesn't need those signs.

7.13 Directionality

Figure 7.25: One-way street in Evanston, Illinois. Photo by D. Levinson.



WHY DO TRAFFIC ENGINEERS SEEM TO LIKE ONE-WAY STREETS SO MUCH?

The *AASHTO Green Book* points out a handful of efficiency advantages.⁴⁶ By removing the delay caused by left-turning cars, we increase traffic capacity and speed. Fewer intersection conflicts means more efficient signal timings and, in theory, fewer and less severe crashes (by reducing or eliminating head-on crashes). Medians are no longer necessary, so you can often fit in an extra lane of through traffic, which further increases capacity and speed. More mobility with better safety? What's not to love?

Beyond the abundant advantages, AASHTO lists a few disadvantages as well. There is the potential for increased travel distances in cases when you have to travel almost around a whole block to reach your destination. When all lanes begin to back up at traffic lights, emergency vehicles may be blocked. Lastly, one-way streets may confuse visitors. This leads to wrong-way drivers and head-on collisions, as in Figure 7.28.

Given AASHTO's list of pluses and relatively few minuses, it makes sense why so many of our streets send traffic in just one direction. Then again, it's not hard to argue that what AASHTO deems an 'advantage' might be the opposite. If I lived or worked on a one-way street, I'd be pretty hard-pressed to believe that more

cars moving at higher speeds is necessarily a good thing.

Donald Appleyard's early studies found many residential livability advantages on two-way streets, but the one-way street he investigated had far more traffic than the two-way comparisons.⁴⁷ Denver converted a handful of one-way pairs to two-way operation in the early 1990s and found that residents preferred the change. A recent case study out of Louisville looked at a handful of one-way to two-way conversions and found significant increases in pedestrian traffic, property values, and business revenue. These benefits were accompanied by a significant decrease in crime. Charleston, South Carolina⁴⁸ and Lubbock, Texas⁴⁹ also found success in terms of two-way streets helping downtown revitalization.

Such livability benefits are all well and good, but are they worth the increased road safety risks that AASHTO made so clear? The research is beginning to suggest that the safety answer isn't clear cut. Lubbock found no significant change in terms of traffic flows or safety.⁵⁰ Another study from Jerusalem also found no difference in road safety.⁵¹ Despite similar traffic levels on the Louisville conversions, crash rates dropped with the two-way streets.⁵² Moreover, child pedestrian injury rates on one-way streets have been found to be more than double the rates on two-way streets.⁵³

More research is needed on the safety outcomes. However, it is also interesting to ask why the safety benefits of one-way streets would be overblown. First, there are likely to be differences in driver behavior, most notably with faster speeds on one-ways. It is pretty easy to understand see why slower traffic – despite the noted increase in [conflict points](#)⁵⁴ – might help reduce crash severity. Another ITE guide even says the following regarding the safety of one-way streets: 'one-way pairs with good signal progression and high travel speeds seemed to elicit red-light running behavior.'⁵⁵ Another example of [risk compensation](#)⁵⁶ we discussed earlier.

Figure 7.26 comes from the *ITE Traffic Engineering Handbook*⁵⁷. It makes the case for better safety on one-ways by depicting the number of conflict points at an intersection for a two-way street as 32 and for a one-way street as only 5. This is a stark difference that could theoretically result in better safety. Beyond the fact that conflict points are not often well correlated with actual safety outcomes,⁵⁸ the bigger issue is that they are comparing apples and oranges. This diagram compares an intersection where all four-legs have two-lanes (one in each direction) to an intersection where all four-legs have one lane. In reality, the one-way streets would have at least two lanes, if not three as in the image from Denver below or

⁴⁷ (Appleyard and Lintell 1972).

⁴⁸ (Baco 2009).

⁴⁹ (Hart 1998).

⁵⁰ (Hart 1998)

⁵¹ (Hoeherman et al. 1990).

⁵² (Riggs and Gilderbloom 2016).

⁵³ (Wazana et al. 2000).

⁵⁴ §7.7.

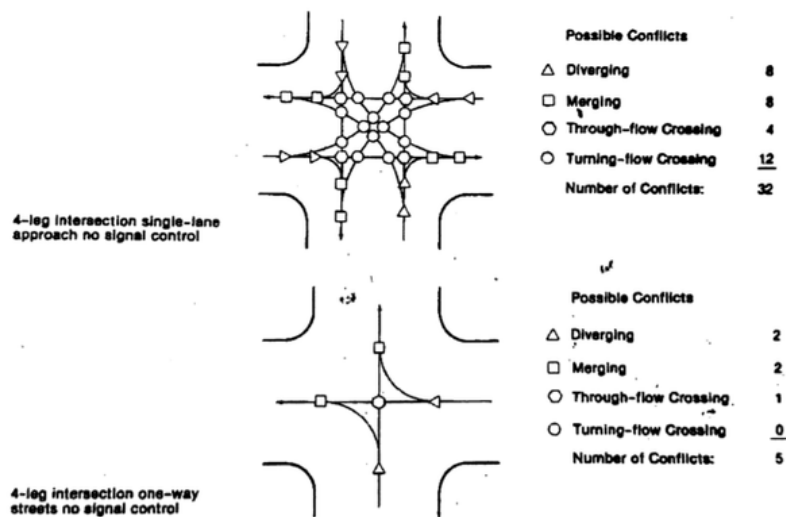
⁵⁵ (Tindale et al. 2004).

⁵⁶ §2.17.

⁵⁷ (Institute of Transportation Engineers 1999).

⁵⁸ (Jacobs et al. 2002).

Figure 7.26: Comparison of intersection conflict points (Institute of Transportation Engineers 1999)



in cases where the median is removed. One-way streets with multiple lanes is a fairer comparison that would substantially increase the number of potential conflict points and deem the comparison in the image below as relatively meaningless. Moreover with regard to conflicts, AASHTO even suggests converting from two-way to one-way operation in situations where an urban street has too many pedestrian-vehicle conflicts. The reduction in pedestrian-vehicle conflicts is supposedly derived from a simpler set of intersection movements.⁵⁹ The real reason for the reduction of pedestrian-vehicle conflicts might be even simpler: fewer pedestrians wanting to cross the street in the first place.

⁵⁹ (AASHTO 2011).

So after all that, the only definitive advantage left for one-way streets is increased traffic capacity. Notably, a capacity advantage of one-way streets can reduce barriers to the implementation of protected cycle lanes or bus-only lanes - and cycle/bus contraflows can mitigate the distance travelled implications. However, increased capacity is also up for debate. Taking into account the decreased accessibility⁶⁰ of circuitous routes,⁶¹ drivers make significantly more turning movements and travel greater distances given the same origins and destinations in a network dominated by one-way traffic patterns.⁶² A network of two-way streets actually has a greater trip-serving capacity - particularly for trips less than 5 miles - as compared to a network of one-way streets.⁶³ When also prohibiting left-turns in the two-way network, this capacity advantage of the two-way network included longer trips as well.

⁶⁰ §1.

⁶¹ §10.9.

⁶² (Walker et al. 2000).

⁶³ (Gayah and Daganzo 2012).

Not only do one-way streets often hinder accessibility and



(a) Before (one-way)



(b) After (two-way)

livability, the traffic engineering benefits don't necessarily hold. While one-way streets are still needed when relatively narrow cross-sections prevent two-way traffic, and may be inoffensive on two-lane roads as in Figure 7.25, in most other urban contexts, it is hard to imagine why so many cities continue to preserve wide one-way streets. Some cities are changing their ways. The before-and-after images of Figure 7.27 are from Larimer Street in Denver where a one mile (1.6 km) stretch was recently converted from a one-way into a two-way. Instead of three high-speed lanes heading toward downtown, there is now one lane in each direction with accompanying bike lanes. With noticeably slower traffic and more active transport use along this corridor, it makes sense why there so many new businesses seem to be popping up, especially when compared to the parallel streets that remain one-way traffic. It might be time for cities to find a new direction – and more research is needed – but it seems like this new direction will run both ways.

Figure 7.27: One-way to two-way conversion in Denver, Colorado. Source: Google Streetview.



Figure 7.28: A head-on collision being cleaned up on a one-way street (where one of the drivers clearly got confused and went the wrong way). Photo by W. Marshall.

7.14 Lanes

Figure 7.29: Example of 9' and 10' lanes in Cambridge, Massachusetts. Source: Adapted from Google Street View.



THE PORTION OF THE STREET CROSS-SECTION TYPICALLY DEDICATED TO MOVING TRAFFIC IS OFTEN SUBDIVIDED INTO LANES.

Each individual lane corresponds to a single channel of traffic – which could include cars, trucks, transit, and/or bikes – and is separated from other lanes via lane markings. In general, in each lane, vehicles follow each other in tandem rather than riding side-by-side, although sometimes skinnier vehicles like bikes and motorcycles split lanes designed for wider vehicles like automobiles.

When designing a street, especially a collector or arterial, an engineer typically needs to make decisions regarding how many lanes to provide and how wide those lanes should be. The number of lanes typically comes down to considerations such as **functional classification**,⁶⁴ **flow**,⁶⁵ **capacity**,⁶⁶ and **level of service**.⁶⁷

⁶⁴ §10.2.

⁶⁵ §6.1.

⁶⁶ §6.9.

⁶⁷ §6.5.

⁶⁸ §3.

However, deciding upon an appropriate number of lanes should also consider **context**,⁶⁸ surrounding land uses, bicyclists, transit, pedestrians – especially in terms of crossing distances – and safety. The number of lanes on the major streets in a city were consistently and significantly associated with travel behavior – both in terms of mode choice and vehicle miles traveled; cities with fewer lanes on their major roads have considerably higher rates of active transport and less vehicle travel per capita.⁶⁹ Cities with fewer lanes on the

⁶⁹ (Marshall and Garrick 2012; 2010a).

major roads also have significantly fewer crashes across all severity levels, including severe injuries and fatalities, holding all other variables constant.⁷⁰ Other studies found similar results in terms of more lanes being associated with more crashes.⁷¹

Transport agencies have also started to recognize the efficiency in focusing on moving people rather than cars. In other words, it is difficult to maximize flux⁷² within our limited-street space with single-occupancy vehicles when transit and bicycles can move a greater number of people with less lane consumption.

In concurrence with determining the number of lanes, engineers also need to establish appropriate lane widths. Many US engineers consider 12' (3.65 m) lane widths to be the 'standard.' The AASHTO *Green Book*⁷³ does not explicitly state this but does say that 12' lanes are 'desirable' in both rural and urban areas. AASHTO also asserts that wider lanes help in terms of safety, capacity, overall driving comfort, as well as with maintenance costs. They do, however, allow for some wiggle room by saying that lane widths can range from 9' (2.75m) to 12'. 11' (3.35m) lanes are acceptable in urban areas with pedestrian crossings or ROW restrictions (which should include almost every urban area). In rural and residential areas, 10' (3.05m) lanes are acceptable on low-speed facilities and 9' on low-volume roads. AASHTO also says that we can use 10' or 11' lane widths on the inner lanes of multilane urban streets when we want to make the outer lanes even wider (12' or 13' (3.95m)) for bicycle use (although this doesn't exactly make for a great bike facility).

So while 12' lanes have become the de facto standard, primarily based on safety concerns, the research doesn't seem to concur with this assertion. Overall, the existing literature suggests that generally wider streets result in higher vehicle speeds⁷⁴ and negative safety implications.⁷⁵ Controlling for posted speed limit, lane widths are the most significant factor impacting speeds.⁷⁶

In terms of safety outcomes, several studies find worse safety outcomes associated with wider lanes.⁷⁷ For instance, Noland gathered data from all 50 states over a 14-year time period to study the safety impacts of what we normally label as 'improvements' to our streets. Instead of improving safety, they actually seem to have harmed it; in fact, Noland attributes more than 900 additional fatalities and 60,000 additional injuries to agencies having increased their lane widths. Wider lanes also mean longer pedestrian crossing distance, and a negative impact on pedestrian safety.⁷⁸

In reality, streets with narrow lanes are still quite common in our older cities, and they continue to function just fine. Figure 7.29 shows a cross-section of a street in Cambridge, Massachusetts with lanes –

⁷⁰ (Marshall and Garrick 2011).

⁷¹ (Vitaliano and Held 1991; Milton and Mannering 1998; Noland 2003).

⁷² §6.3.

⁷³ (AASHTO 2011).

⁷⁴ (Hansen et al. 2007; Martens et al. 1997).

⁷⁵ (Dumbaugh 2006; Ewing and Dumbaugh 2009; Ivan et al. 2009).

⁷⁶ (Fitzpatrick et al. 2000).

⁷⁷ (Harwood 1990; Noland 2003; Potts et al. 2007)

⁷⁸ (Gärder 2004)

Figure 7.30: Acorn Street in Boston, Massachusetts. Photo by Wes Marshall.



that often facilitate truck traffic – of just 9' and 10'. Figure 7.30 shows us Acorn Street in the Beacon Hill neighborhood of Boston with a curb-to-curb cross-section of just 6'. The standard at the time this was built required streets to allow two cows to pass each other at the same time.

So what do the more informed design manuals suggest in terms of lane widths? NACTO (National Association of City Transportation Officials) recommends 10' lanes on any street with a target speed less than or equal to 40 mph (65 km/h).⁷⁹ If a street has high bus or truck traffic, NACTO suggest changing one lane from 10' to 11'.

⁷⁹ (National Association of City Transportation Officials 2013).

The analogy of lane width in the rail sector is gauge, the spacing between rails on railroad tracks, which determines the widths of trains, tunnels, and so on. This however must be more standardized, as the train cannot deviate from the track.

7.15 Vertical Separations



Figure 7.31: ‘This is the City of Tomorrow.’ Source: Life Magazine, July 5, 1937.

By 1960, STOP-AND-GO TRAFFIC WILL BE A THING OF THE PAST.

Before he designed the General Motors Futurama exhibit at the 1939 World’s Fair in New York, Norman Bel Geddes⁸⁰ designed a similar “utopia” for a Super-Shell gasoline advertisement in Life Magazine, shown in Figure 7.31. Bel Geddes proposed three strategies; quoting from the ad:

1. Sidewalks will be elevated – you’ll walk and shop ABOVE Main Street, actually cross over it;
2. Local traffic will use the FULL width of the streets below – no sidewalks, no parked cars. Loading and unloading will be done INSIDE the buildings; and
3. High-speed, long-distance traffic will have its own elevated, one-way lanes, no stop lights or intersections.

The underlying premise of this Bel Geddes quote is that our city transport systems will essentially become three dimensional – or have what we call vertical separation. By separating modes and eliminating as many intersections as possible, we could theoretically increase capacity⁸¹ and efficiency.

We tried this, at least to some extent, in cities such as Atlanta with pedestrian bridges (commonly known as the gerbil tubes,⁸² as

⁸⁰ Norman Bel Geddes was the father of actress Barbara Bel Geddes, most famous for playing Miss Ellie in the TV serial *Dallas*.

⁸¹ §6.9.

⁸² Enclosed pedestrian bridges are more pejoratively called ‘honky tubes.’

⁸³ §12.1.

⁸⁴ Some cities, such as Dallas, have contemplated removing their tunneled pedestrian networks.



Figure 7.32: Pedestrian skyway in Atlanta, Georgia. Photo by Wes Marshall.



Figure 7.33: Pedestrian/bicyclist underpass in Boulder, Colorado. Photo by Wes Marshall.

⁸⁵ (Murphy et al. 2017).

⁸⁶ (Marshall 2017).

shown in Figure 7.32, due to their resemblance to Habitrails) that help people traverse multiple blocks without ever setting foot on the street. Due to land use patterns and issues such as **induced demand**,⁸³ vertical separation didn't solve Atlanta's congestion problems, nor provide much downtown street life.⁸⁴

However, vertical separation is beneficial in other situations such as Minneapolis' skyway network or Montreal's Indoor City on frigid days or the over- and under-passes along a freeway. While the Big Dig tunnels in Boston may have had a few cost overruns and may not have solved congestion, they helped reclaim more than 27 acres (11 ha) of land for public use, which is unheard of for a mature major city. We can't underestimate what underground subways have done for cities around the world, or even cities such as Boulder, Colorado – shown in Figure 7.33 – that have taken to building short underpasses so that pedestrians and bicyclists can cross their major roads unimpeded.

From a safety perspective, removing pedestrians from conflicts with cars must logically reduce the likelihood those pedestrians are in collisions. However, the safety in numbers effect suggests the remaining pedestrians will be less safe as a result.⁸⁵

Elon Musk – founder of PayPal, Tesla, and SpaceX – recently announced that he has grown tired of the Los Angeles traffic congestion and has literally started digging under his SpaceX campus.⁸⁶

Musk says:

You have tall buildings, they're all 3D, and then everyone wants to go into the building and leave the building at a same time. On a 2D road network, that obviously doesn't work, so you have to go 3D either up or down. And I think probably down.

His initial goal is to improve our tunneling technology with a better boring machine, and then to eliminate traffic congestion. While he might be falling into the same-old Super-Shell/Futurama trap with respect to induced demand, he plans on taking things at least a few levels deeper with his Boring Company:

If you think of tunnels going 10, 20, 30 layers deep (or more), it is obvious that going 3D down will encompass the needs of any city's transport of arbitrary size.

Our history has shown us that vertical separation is useful in many situations but is also not appropriate for every context. We'll have to wait and see on this vision of extreme vertical separation.

7.16 Parking Capacity



Figure 7.34: Why pay more to park on the road. Parking sign from Brunswick Centre, London. Photo by D. Levinson.

CARS ARE AT REST NEARLY 23 HOURS PER DAY.

Parking interplays with access⁸⁷ in several ways. First, for car travelers, the distance between parking and the final destination can be a major component of total travel time, and so difficult parking reduces access by car. Second, the cost of parking plays into a full cost approach to access combining time and money cost. Third, space devoted to parking cannot be devoted to other activities, and thus reduces the effective density of activities.

How much parking is there in a city such as Minneapolis? This is not a question for which there is a well-sourced answer.⁸⁸

The Target Center, the downtown arena home to the Minnesota Timberwolves NBA basketball team, says that “there are nearly 25,000 parking spaces in 38 parking lots and ramps throughout downtown.” The Minneapolis Municipal Parking System “has 17 parking ramps and 7 lots. These Ramps and Lots encompass over 20,000 parking spaces.”⁸⁹

In the City there are 7,000 metered spaces, mostly in commercial districts, including downtown and elsewhere in the city.

Outside of downtown requires more estimating.

On-street unmetered parking? The City has 1,670 km (1,100 miles) of streets. Most are residential and have on-street parking. We can assume about 120 spaces per km. If there were no “no parking restrictions,” this gives 220,000 on-street spaces (the vast majority of which are unmetered).

Off-street private parking? There are 155,155 households. If each

⁸⁷ §1.

⁸⁸ Sources:

- Target Center: Plan Your Parking https://www.targetcenter.com/plan_your_visit/parking
- City of Minneapolis Parking Meters: <http://www.minneapolismn.gov/parking/meters/>
- City of Minneapolis Municipal Parking Ramps <http://www.minneapolismn.gov/parking/>
- Length of city streets <http://www.minneapolismn.gov/publicworks/streetsweeping/>

⁸⁹ Subtracting this from the first estimate suggests only 5,000 parking spaces are private.

one has 1 off-street space (some have 2 or 3, some have 0), that would be 155,155 off-street spaces in residential areas.

This doesn't even count off-street parking at businesses, schools, stores, etc. Roughly every car has to have a space at home, work, and shop. In short, there is a lot of parking.

8

Modalities



Figure 8.1: Brisbane Tram.
Photo by D. Levinson.

WHAT CHARACTERISTICS DESCRIBE AND DIFFERENTIATE MODES?

Every mode must differ from every other mode on at least one dimension (otherwise they would be the same mode). This is analogous to the idea of speciation in biology. Figure 8.2 is a first cut at this for surface passenger transport. It distinguishes primarily on the non-mechanical (non-propulsion) characteristics of the service first. Multiple modes that are otherwise obviously distinct: gondolas and subways are much the same from a transport service perspective, but one is underground and uses a train and the other is suspended by a moving cable. Their capacities may also differ. This taxonomy differentiates things that are qualitatively different

rather than quantitatively different.

So the first cut is about time:

- Is a reservation required or not (that is does it need some advance planning)?

The next columns are about time and space:

- Is the service scheduled or dynamic?
- ¹ §8.5. • Are the routes **fixed**¹ or dynamic?
- If the route is fixed, are stops fixed (does the vehicle stop at every stop, or only when called, like a bus)?
- Otherwise, if the routes are dynamic, things get a bit more ad-hoc, as the key question changes.

This table considers the physical network,

- Does the mode serve you door-to-door in part or in full?
- Is the ride shared with other parties going from different origins or to different destinations?
- Is the mode human powered?

Some traditional distinctions (access mode vs. primary mode) are not distinguished here; rather, that would be thought of as at least two stages of a **trip**,² one where you walk or drive to some place (with the purpose of changing modes), and second where you take some form of transit.

² §2.1.

The idea here is suggestive. By looking at the elements that define modes, and adding others not shown here (propulsion technology, payment mechanism, and so on), innovation can occur.

The rest of this chapter considers modes. First we look at how accessibility explains **choice of mode**.³ We next consider the elements of trips, which are especially pertinent for more complex transit trips: The **first mile/last mile**⁴ problem, the storage of cars in **park and ride lots**,⁵ and the **linehaul**⁶ segment. Scheduled modes use **timetables**⁷ to coordinate the system. Sometimes, for very understandable reasons, unreliable systems fall out of coordination and off-schedule, and bus or train **bunching**⁸ occurs.

³ §8.1.

⁴ §8.2.

⁵ §8.3.

⁶ §8.4.

⁷ §8.5.

⁸ §8.6.

While travel time affects access, a full-cost accessibility would also consider **fares**,⁹ which are often inefficient. We turn to ask about the capacity of **transit systems**¹⁰, and then **compare**¹¹ transit use with highway use for a selected corridor.

⁹ §8.7.

¹⁰ §8.8.

¹¹ §8.9.

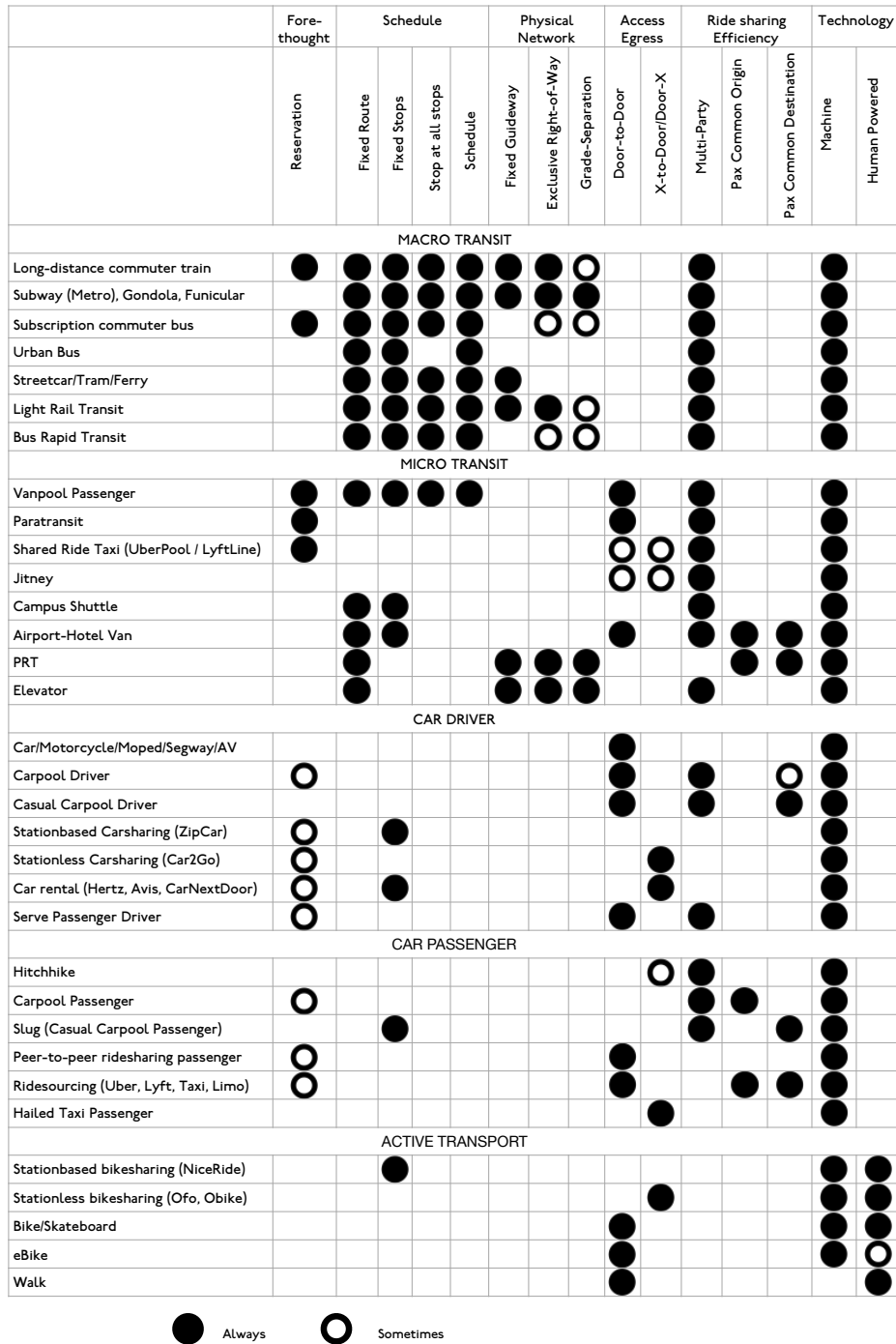


Figure 8.2: A typology of modes.

8.1 Mode Shares

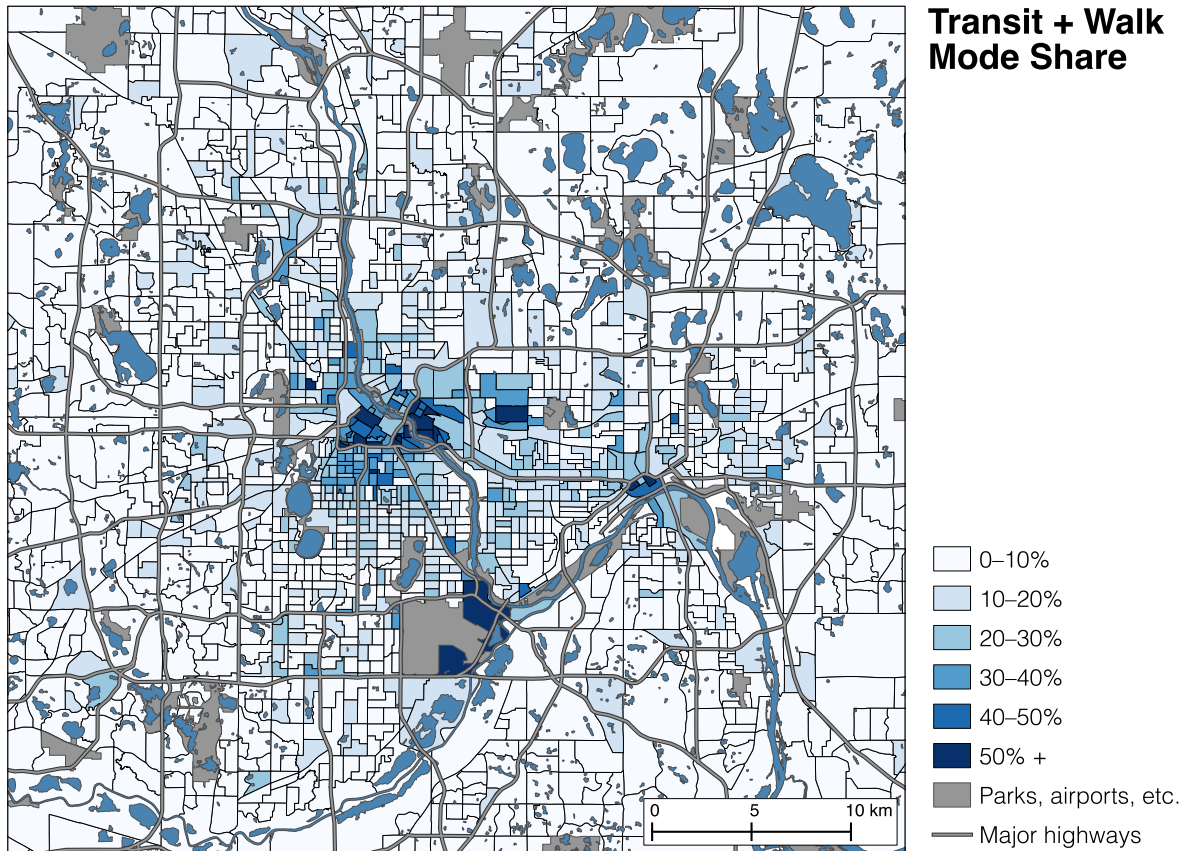


Figure 8.3: Commute mode share of transit + walking in Minneapolis - St. Paul region. Source: ACS 2007–2011 estimates. Drawn by A. Owen.

¹² §2.10.

THE PROBABILITY THAT AN INDIVIDUAL CHOOSES A PARTICULAR MODE DEPENDS ON THE ACCESSIBILITY OF THAT MODE AND OTHERS.

In formal practice, *mode choice*¹² can perhaps be best assessed using a discrete choice model. The difficulty is that utility (the output of choice models) is not observable and can only be estimated in models. And results from models are inherently much farther from a confirmable reality than measured and easily replicable results from direct observation (or databases constructed from direct observations).

Alternatively, we can look at this issue at a simpler level. It is much easier to predict aggregate than individual behavior. It turns out that aggregate mode shares can be explained very well with

observable measures of accessibility.¹³ As shown in Figure 8.4, the areas with the best transit service (where the number of jobs that can be reached by transit is nearest the number that can be reached by auto) are in the center of the city. It also turns out those are the areas with the highest transit plus walk mode share, as shown in Figure 8.3. This is no accident, access is an indicator of the relative usefulness of the alternative modes. Places where transit and walking can reach a higher share of jobs are more likely to get users than where car is the only effective option.

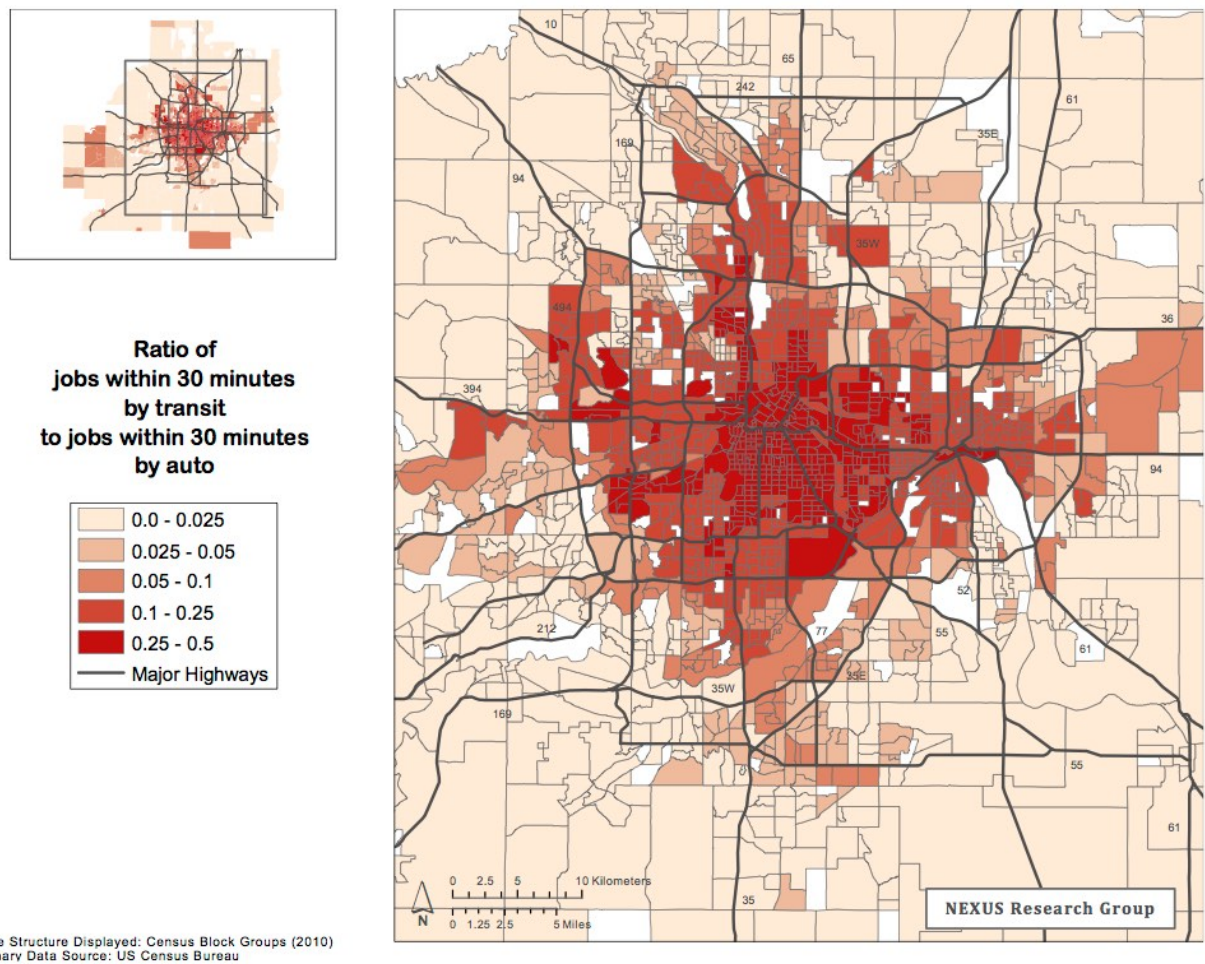


Figure 8.4: Map of ratio of transit to auto accessibility. Source: (Owen and Levinson 2015).

8.2 First and Last Mile

Figure 8.5: eTuk vehicles outside Union Station, Denver, Colorado. Source: City and County of Denver.



ONE KEY ASPECT TO ACCESSIBILITY THAT IS BEGINNING TO GET THE ATTENTION IT DESERVES IS THE FIRST AND LAST-MILE ISSUE.

For instance with transit, we might have reliable and frequent service from the neighborhood we live in to downtown where we work. However, if the transit lets you off a mile or so from your office and we fail to make that last-mile [walkable](#)¹⁴ or provide other viable travel options, the transit service isn't nearly as useful – no matter how reliable and frequent we make it.

¹⁴ §3.3.

Accordingly, our mindset about providing usable transit service needs to be about a lot more than the transit itself. It is not uncommon for the debate to focus on issues such as train versus

bus versus streetcar when it is just as important, if not more, to make sure that the transit is safely and easily accessed by a variety of modes.

In terms of impact, the research suggests that last-mile (near your destination) tends to be more important than the first-mile (near your home) when people are deciding whether or not to use transit.¹⁵ Getting off the train at Union Station in Denver, Colorado, for instance, the last-mile options are extensive with a good walking environment, pedestrian bridges that traverse the nearby highway to the neighborhoods to the north, a bikeshare station that you can't help but walk past, a free bus that runs the length of the pedestrian mall to nearly the other side of downtown, an off-road bicycling and walking trail that can bring you even further across downtown and to the neighborhoods to the south, as well as taxis, pedi-cabs, ridesourcing options such as Uber and Lyft, carsharing vehicles, and even electric tuk-tuk vehicles (Figure 8.5).

¹⁵ (Barnes 2001).

However, the first-mile connections can be especially important in creating a transit system that is useful for those other than commuters. For instance, good transit-oriented developments can go a long way towards solving the first-mile issue and supporting the broader goals of making a transit system that can be used by all ages for all types of trips. Unfortunately, the default in most cities tends towards extensive [park-and-ride](#)¹⁶ facilities at the expense of location efficiency and mixed use. If the intent of a transit organization is to get as much transit ridership as quickly as possible and attract the unfortunately-labeled 'choice' rider, such a focus makes sense. While it can be exceedingly difficult to play the long game and wait for the development of land uses supportive of transit, the end result can make a drastic difference in scope and extent of the first-mile problem.

¹⁶ §8.3.

And if people still want to take an auto to transit, ride-hailing services like Uber and Lyft have been more than happy to oblige, as are more traditional taxis. In fact, Lyft reports that transit stops are their number one category of destination, even more than staples such as airports or restaurants. Lyft is also being subsidized to help with the first-mile issue in cities such as Centennial, Colorado and via the Livermore Amador Valley Transit Authority (LAVTA) in California. This helps solve the first-mile issue without the opportunity cost of dedicating what can be extremely valuable land to subsidizing [parked cars](#).¹⁷

¹⁷ §7.16.

8.3 Park-and-Ride

Figure 8.6: Park-and-ride lot at Louisiana Avenue Transit Center. Source: Metro Transit.



MOST PEOPLE DON'T FIND PARK-AND-RIDE LOTS AESTHETICALLY APPEALING.

Many people don't want them in their neighborhood. Yet transit agencies, as in Figure 8.6, find them economically attractive as a way of solving the [first-mile](#)¹⁸ problem in the [automobile city](#).¹⁹ With the broad availability of [parking](#)²⁰ in the modern world, it's not clear why park-and-ride lots are required, but those spaces may not be where they are needed to conveniently serve a transit stop. Let's do some math.

In one acre (0.4 ha), there are 43,560 square feet (4047 m²). It takes about 300 square feet (28 m²) to store a parked car (including lanes, etc.).²¹ This suggests you can store 145 parked cars per acre (360/ha).

If every one of those parked cars carried 1 person, that means 145 transit boardings from that station in the morning (and 145 boardings elsewhere in the evening, assuming symmetry), generating 290 daily transit [trips](#).²²

In contrast, let's say we had zero park-and-ride spaces. Even if only workers lived there, and they had 100% transit mode share for work trips and another 2 non-work trips per day by transit, that is 145 people per acre. That is the [residential density](#)²³ equivalent of 92,800 people per square mile (35,380 per km²). Those are Manhattan like densities (actually higher).

Low, or even medium, density around the station will not enable

¹⁸ §8.2.

¹⁹ §3.9.

²⁰ §7.16.

²¹ Note, other studies have found this as high as 513 square feet (48 m²) in practice, when including plantings and access lanes (Marshall et al. 2008).

²² §2.1.

²³ §3.1.

as many transit users as the park-and-ride lot. Now that doesn't mean it is cost-effective to build a park-and-ride lot, which depends on the value of land, on maintenance costs, whether park-and-ride spaces are given away for free or can be charged for, and levels of demand. It certainly doesn't mean it is cost-effective to construct a parking structure. It simply suggests that more transit riders might be generated from an acre of parking than a typical acre of transit-oriented development.

However this number should be discounted somewhat. It turns out that people who live within the nominal transit walkshed may drive to available park-and-ride facilities, in part this is due to the poor pedestrian environment created by the park-and-ride lot in the first place.²⁴ The phenomenon of 'park-and-hide' also takes place in semi-urban environments without park-and-ride facilities, transit users park on-street in adjacent neighborhoods.

²⁴ (Truong and Marshall 2014).

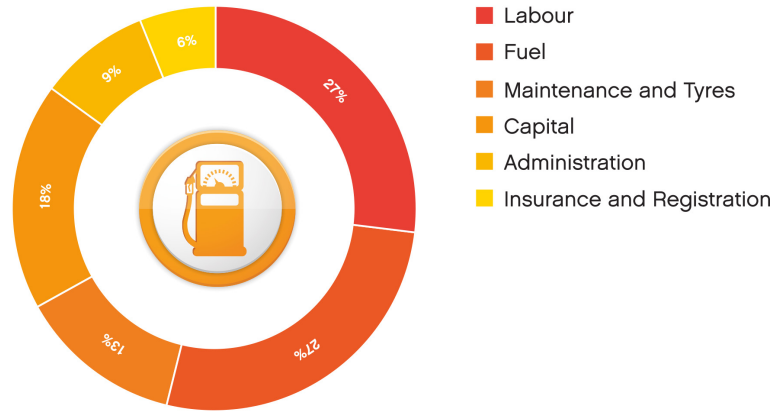
Maintenance costs are surprisingly high: a 288-stall lot generated \$17,000 in plowing and \$16,000 in lighting expense per year.²⁵

²⁵ (Black 2008).

8.4 Line-haul

Figure 8.7:
cost breakdown.
(Beaurepaire 2017).

Line-haul
Source:



THE LINE-HAUL COMPLEMENTS THE FIRST AND LAST MILE.

²⁶ §8.2. While the **first mile**²⁶ concerns itself with getting from the origin to the main (higher capacity and faster) part of the network, and the last mile from that backbone to the destination, the line-haul is the travel along that backbone. The concept is important, distinguishing between the access and egress times and costs at the ends of **trips**,²⁷ and the main part of the trip.

²⁷ §2.1.

Line-haul also refers to the movement of freight, typically by trucks, between cities or ports. The line-haul cost, enumerated in Figure 8.7, is the money and time spent for a certain line-haul while the line-haul rate is the cost per mile or per kilometer. These costs include such line items as fuel, labor, time, administration, tolls, insurance, registration, maintenance, and depreciation; however, some of these costs depend on what's being shipped and its weight.

Line-haul even refers to the transporting of people, particularly between important transit terminals, train stations, and airports. For instance, we might want to compare the line-haul travel times and costs for the various modal options (like a train, bus, taxi, ridesourcing) in getting from a typical Union Station to the airport and back. Sometimes the line-haul takes less time than access and egress, for instance with air travel for shorter distance trips. While not generally called out, the idea applies to automobile travel as well, as travelers use local streets for the first and last mile(s), and highways for the middle of the trip.

8.5 Timetables

Figure 8.8: A typical timetable from 19th Century railroad services. In this case: York, Scarborough, Pickering, and Whitby.

SCHEDULED TRANSPORT SERVICES (SUCH AS BUSES, TRAINS, AND AIRPLANES) FOLLOW TIMETABLES, WHICH INDICATE WHEN A VEHICLE WILL ARRIVE OR DEPART AT A GIVEN STOP OR STATION.

While timetables, like that in Figure 8.8, are designed to serve the daily patterns of demand, they are constrained by many other things (such as there are only so many vehicles available), vehicles may need to be repositioned (dead-heading or the red-eye) for future demands, there are only so many drivers (pilots) and crew available, there are only so many gates available, and so on. So the distribution of actual services is *flatter*²⁸ than consumers would demand in an unconstrained environment. This mismatch of service availability and desired departure or arrival times both creates crowding and is one of the factors that pushes people towards unscheduled, on-demand modes of transport.

²⁸ §2.4.

8.6 Bus Bunching

Figure 8.9: Long bus queue in Putney due to Tour de France. Photo by D. Levinson.



WHEN THEY REACH THE END OF THE LINE, THERE MIGHT BE TWO, OR THREE, OR MORE BUSES ONE RIGHT AFTER ANOTHER, AND THEN NO BUSES FOR 20, 30, 40, OR 50 MINUTES.

²⁹ §8.5. Yet the buses were dispatched on a [schedule](#)²⁹ from the start of the line every ten minutes.

What happened?

Along the way, buses bunched, as illustrated in Figure 8.9. Somewhere near the middle of the route, a passenger waiting for the bus may wait 20 minutes, despite the 10-minute posted headway, and then see two buses arrive at once.

Why does this occur?

³⁰ §5.2. The answer is [stochastic arrivals](#).³⁰ People arrive randomly at bus stops, sometimes many people arrive at once. When Bus *A* gets to the stop, those many people take longer to board than fewer people would. The bus departs later than scheduled. At the next bus stop, more people have accumulated since the actual time between buses was longer than scheduled. Boarding takes longer still. The bus falls farther behind schedule.

Meanwhile, the following Bus *B* gets ahead of schedule, since fewer people are at each stop than expected, since people board the first bus that arrives at the stop after they do. If Bus *A* was supposed to arrive at 8:10 and Bus *B* at 8:20, but Bus *A* was 5

minutes late, Bus *A* then must board 15 minutes worth of arriving passengers and Bus *B* only boards 5 minutes worth of arriving passengers. Eventually Bus *B* catches up with Bus *A* on a long route. The buses are bunched.

There are no perfect solutions to this problem; it is a natural feature of this kind of dynamic stochastic system (and is related to the 'hitchhiker's paradox',³¹ which explains the long delays associated with uniform arrivals and random servers) and occurs at the best run transit organizations. Strategies for minimizing bus bunching include:

- Making buses adhere to static schedules (which works with low frequency and uncongested systems but breaks down as demand rises).
- Holding buses at control points to restore the desired headway (delaying Bus *B*).
- Inserting buses (Bus *C*) midway along the route at control points if the gap between buses becomes too long.³²
- Allowing buses to overtake (Bus *B* passes Bus *A*).
- Skipping stops.
- Allowing passengers on Bus *A* to alight but taking on no (or a limited number of) new passengers.
- Reducing bus delays with traffic signal priority and better designed stops that require payment before boarding.
- Reduce dwell times through all door boarding.
- Replacing buses with rail-based services with higher capacities and longer headway.

Creative people may think of others. Making the system work well is critical to attracting riders. If headways are functionally 20 minutes rather than 10 minutes, [accessibility](#)³³ is lost.

³¹ The setting of the paradox is the following:

- Cars pass a point of a road following a Poisson process.
- Cars arrive on average every 10 min.
- A hitchhiker arrives.

How long should the hitchhiker expect to wait until the next car?

It turns out the mean waiting time is 10 minutes, not 5 minutes. This is because the process is memoryless. As an economist would say 'sunk costs are sunk.' Just because cars arrive on average every 10 minutes is irrelevant, as the cars are independent.

³² Bus insertion sounds like it requires having extra buses and drivers lying around waiting to be deployed. This can be planned for and built into the schedule so there are short versions of long routes. If bunching is known to be a problem on the long route, at some point along the route, there would be another 10-minute headway short route providing the service to the same destinations that dispatches dynamically to preserve headways on the peak direction. In the reverse direction, it is given a different identity, so people know it is the short route.

³³ §1.

Well, did he ever return? No, he never returned and his fate is still unknown.

He may ride forever 'neath the streets of Boston. He's the man who never returned.

Charlie handed in his dime at the Kendall Square Station and he changed for Jamaica Plain.

When he got there the conductor told him, "One more nickel."

Charlie couldn't get off of that train.

Some transit operations only require you pay when you board, and have a flat fare, others require that you "tap in" and "tap out", that is pay when you exit, at least loosely proportional to how far you traveled. The advantage of the distance based fares is their fairness, costs increase with distance traveled, so shouldn't fares? The disadvantage is the increased collection costs (some system needs to track passengers by boarding and alighting location), and the delays imposed on exiting as passengers [queue](#)³⁴ up at exit turnstiles. The same issue arises with road [pricing](#):³⁵ some systems charge based on distance traveled with entry and exit tollbooths, others have payment gantries along the way, and some just charge at the on-ramp. Practical considerations, trading off collection costs for increased payment efficiency determine the system design. However, once it is in place, it is very hard to change; people very much dislike new points of payment, more so than just increased fares.

³⁴ §5.

³⁵ §9.7.

8.8 Transit Capacity

Figure 8.11: A relatively full Metro Transit Green Line train on opening day, June 2014. Photo by D. Levinson.



TAKING ADVANTAGE OF EXISTING CAPACITY IS FAR MORE COST-EFFECTIVE THAN BUILDING NEW CAPACITY.

What is the existing capacity of transit, how much is used? Capacity depends on assumptions and human behavior. The table below works through some scenarios based on assumptions from Minneapolis.

1. How many hours per day is the line operating? (H)
2. What is the frequency within that time period? (T/H)
3. How many cars per train are there? (C/T)
4. What is the capacity per car (cars are rated at 230 passengers (P_{max}), but this includes standees)? (P_{max}/C)
5. How long is the line in number of stations? (S)
6. How long (how many stations) is the average trip (excluding boarding station)? (s)
7. How many directions are you considering? (D)

This measures capacity in terms of daily boardings. (Daily distance traveled is another, perhaps better, measure of capacity

utilization, but this is not tracked since there is no tap-off process at stations in Minneapolis - Saint Paul.)

This is illustrated through an example of the Green Line, a light rail transit line connecting Minneapolis and Saint Paul along University Avenue. Table 8.1 shows some surprisingly high numbers, up to 7 million (under the admittedly silly unconstrained scenario (A) where people only ride the train for 1 stop before alighting, trains run for 24 hours a day, and people are standing at near crush capacity), with more plausible numbers in the 255k territory, assuming everyone gets a seat, but the train can run at 5-minute headways (C).³⁶ The numbers are in contrast very low with the admittedly silly over-constrained scenario (E), where everyone rides the whole line from end-to-end.

The main point is that, even with a September 2017 average weekday ridership of 48,859, there is a lot of capacity on the Green Line yet to go, even if the train only runs 18 hours a day.³⁷

This analysis demonstrates that transport capacity is not the constraint in land development along the Green Line corridor.³⁸

Certainly **load balancing**³⁹ is an issue, much of the capacity is ‘off-peak,’ but that is what **pricing**⁴⁰ is for. Higher loads would increase wear and tear on the cars, and add costs, but hopefully the added revenues would more than compensate.

Given there is a lot of developable land in existing corridors, one (rhetorically) asks, why are new corridors being subsidized for development?

To calculate the capacity, we use the following equation:

$$Q_{max} = H \frac{T}{H} \frac{C}{T} \frac{P_{max}}{C} sSD \quad (8.1)$$

³⁶ Here we are limited by capacity in one section (downtown Minneapolis), which does run at 5-minute headways, but splits the capacity between the Green and Blue lines, which share tracks.

³⁷ The constraint is not the track, but the fleet. 10-minute headway is all today’s fleet can support, to increase headway the operator would need to increase speed greatly or add vehicles.

³⁸ One could similarly demonstrate the underutilization in the north-south direction on buses, and in all directions on roads.

³⁹ §2.4.

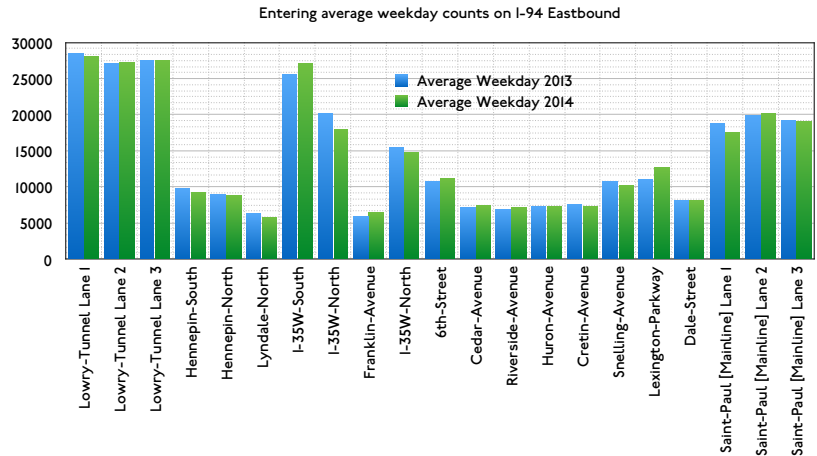
⁴⁰ §9.7.

	A	B	C	D	E
Hours	24	24	24	18	18
Trains per hour	12	12	6	6	6
Runs per day	288	288	144	108	108
Cars per train	3	3	3	3	3
Capacity per car	230	115	115	50	50
Stations	19	19	19	19	19
Trip length (stns)	1	7	7	7	18
Directions	2	2	2	2	2
Daily capacity	7,153,920	510,994	255,497	83,314	32,400

Table 8.1: Transit capacity

8.9 Modal Magnitudes

Figure 8.12: I-94 traffic counts. Source: Minnesota Traffic Observatory and MnDOT via DataExtract program. Analysis is by author.



Car culture remains dominant in Minnesota. As we can see comparing the observations in the [highway capacity section](#)⁴¹ with the [transit capacity example](#),⁴² the number of people using I-94 on a given weekday is about 20 times larger than the number of people using the Green Line.

⁴¹ §6.9.

⁴² §8.8.

A \$1 billion transit investment (the Green Line) is rounding error for the change in traffic count on the parallel highway (comparing entering vehicles between Lowry and Dale for October 2014 (244,103) and October 2013 (244,712) – average weekday traffic). [Induced demand](#)⁴³ may explain part of this, but these two facilities also serve very different markets.

⁴³ §12.1.

Transit investments like the Green Line LRT serve transit users; highway investments like I-94 serve highway users. They are, at this point in history, at this location, barely substitutes. Congestion reduction should not be a selling point for transit investments, just as reducing crowding on trains or buses is not a valid selling point of highways.

9

Routing



Figure 9.1: Sydney by air. Photo by D. Levinson.

The process of which routes individual travelers use, and how many use each route, affects the travel time between location, and thus the accessibility. We started with the [principle of least effort](#)¹ wherein individual travelers seek the easiest route. This is not quite the shortest travel time path, and is far from perfect, but describes a general tendency. A minority of travelers actually minimize their travel times.

¹ §2.11.

Given travelers are traveling, a very important point is the inherent accounting identify of what goes in must come out, or [conservation of flow](#).² With the principle of least effort and conservation of flow, we find that traffic approaches what might be considered a [user equilibrium](#),³ no traveler can do significantly

² §9.1.

³ §9.2.

better given other travelers are also trying to do their best.

⁴ §9.3. These processes tend to be solved for average days, but most days are not average days, there is a high degree of variability from minute to minute and day to day because the network lacks [reliability](#).⁴ For many years the Sydney Harbour Bridge (Figure 9.1) carried all of the automobile and train traffic from the north in the Sydney CBD; from 1992, the Sydney Harbour Tunnel just west of the Opera House provided an additional route, increasing capacity and reliability of the system.

⁵ §9.4. We further note that individuals doing what is best for them is not inherently best for society as a whole, so we can measure a [price of anarchy](#)⁵ as the ratio of the user equilibrium and system optimal travel times. This indicates the inefficiency of the system by letting travelers choose their own routes. The access to the Sydney Harbour Bridge includes an elevated highway in front of Circular Quay, so at least some travelers can switch between the tunnel and bridge routes. While this adds reliability if one of them sees a capacity reduction, it also may add to inefficiency. It is not clear that this is actually a [Braess Paradox](#),⁶ when travel times actually increase with the presence of an additional link (the Cahill Expressway link in this case). While this paradox is unusual, it is also intriguing and sheds light on the complexity of networks.

⁷ §9.6. Managing traffic on networks can be done in many ways. Traditionally traffic is allowed to use roads on a first-come, first-serve basis, but it could be [rationed](#)⁷ so that the right to travel by car on roads is restricted by some other mechanism. The most efficient way to allocate scarce road space involves [road pricing](#),⁸ but this has been politically difficult to achieve.

9.1 Conservation



Figure 9.2: Parking enforcement in London reduces the number of vehicles on the street and violates conservation. Photo by D. Levinson.

OVER A PERIOD OF TIME, THE NUMBER OF CARS ENTERING AN INTERSECTION MUST BE THE SAME AS THE NUMBER EXITING IT PLUS THE NUMBER REMAINING IN THE INTERSECTION.

The number of pedestrians entering a link must be the same as the number exiting it (excluding those with a destination on that link). In many obvious ways, there are conservation laws necessary for the short term analysis of traffic. This accounting identity for flow is critical for understanding the amount of delay that results from traffic (with notable unusual exceptions like in Figure 9.2, where an abandoned and possibly stolen Porsche in London was lifted onto a tow truck, combining two vehicles into one).

There are perhaps other conservation principles. The [travel time budget](#)⁹ posits that people are, on average, conserving travel time. If one trip gets too long, other trips get cut, or people adapt, to keep travel times within bounds. But this is not a strict accounting identity (unlike the hours in a day); it is more of a tendency or preference.

What becomes dangerous is the misapplication of the principle of conservation of flow.

Imagine there are two bridges (Bridge 1 and Bridge 2) across a river, both on the cusp of congestion. Bridge 1 is closed for construction. Bridge 2 sees an increase in traffic, but not all of the traffic from Bridge 1 now crosses Bridge 2. The traffic was not

⁹§2.5.

conserved. Certainly some make the same trip at the same time. Others make the same trip at different times. Some switched modes. Some of those travelers chose different destinations. Others forewent that trip altogether.

¹⁰ §12.1.

Just as the notion of **induced demand**¹⁰ says that when road networks are expanded with new or wider links they see more traffic, when they are contracted they see ‘reduced demand.’ There is a lot of evidence at this point. Studies of the collapse of the I-35W Mississippi River Bridge in 2007 in Minneapolis showed that about one-third of traffic simply ‘disappeared’ after the collapse, and no longer crossed the river daily. The remaining two-thirds used different bridges.¹¹ San Francisco found similar results with the Embarcadero Freeway after the 1989 earthquake, which gave them the wherewithal to eventually remove the freeway.¹²

¹¹ (Zhu et al. 2010; 2017).

¹² See also discussion of freeway removal and **Octavia Boulevard** in §7.2.

¹³ §2.1.

Trip generation is another area where conservation principles are misapplied. Traffic engineers regularly estimate the number of **trips**¹³ entering and exiting various types of land uses. For instance, the number of vehicles coming out of a cemetery assume that it is proportional to the size of the cemetery in acres (generating a rate like 5 trips per acre per day).¹⁴ Yet, logic suggests that building a new cemetery does not increase the number of people who will die.¹⁵ Instead, a new cemetery will attract more of what is a fixed amount of business, and existing cemeteries less. There is for practical purposes a conservation of deceased people, which limits the number of trips to visit deceased people, which suggests that trips are not strictly speaking solely a function of the size of the cemetery. Still, these rates are built into legal code. Creating a facility does not automatically create demand for the facility depending on facility size. Sure, some types of activities may induce overall demand (a library may increase travel to borrow books), but others will not. After construction, a new office building does not generally increase the number of employed, it just moves them around, and thus may increase employment locally.

¹⁴ For an example of trip generation embedded in code, see this example from San Diego. <https://www.sandiego.gov/sites/default/files/legacy/planning/documents/pdf/trans/tripmanual.pdf>.

¹⁵ Aside from the very small number of fatal crashes due to people going to visit the cemetery itself, and objectively, they would have died eventually anyway.

9.2 Equilibrium

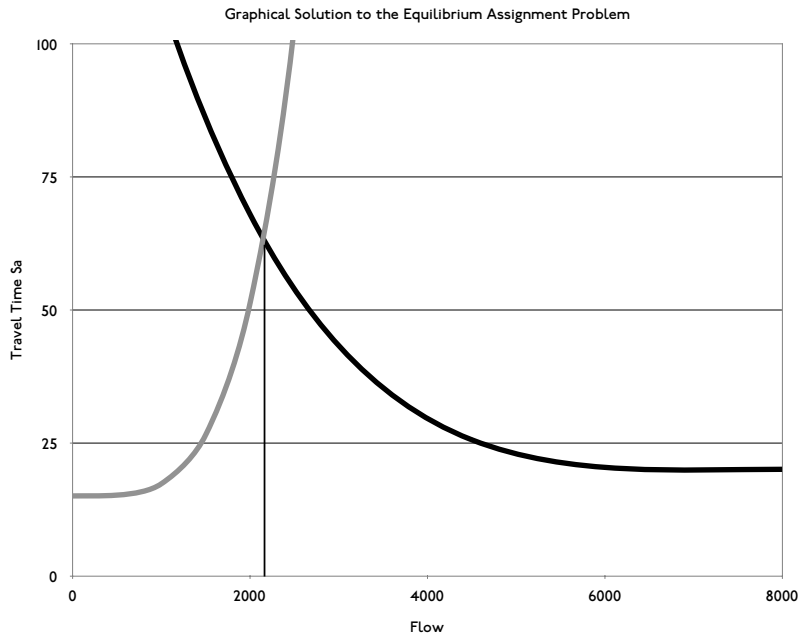


Figure 9.3: Two route equilibrium. Results of equilibrium assignment on two route network. Drawn by Author, Based on results from (Eash et al. 1979).

User Equilibrium: “The journey times in all routes actually used are equal and less than those which would be experienced by a single vehicle on any unused route,”

System Optimality: “At equilibrium the average journey time is minimum” – John Glen Wardrop.¹⁶

¹⁶ (Wardrop 1952).

THERE ARE SEVERAL TYPES OF NETWORK EQUILIBRIUM.

In the classic economic sense, equilibrium occurs when supply equals demand. This implies a transport network that just serves the transport demand placed on it at the prevailing cost, and where the cost of expanding the network by some amount just exceeds the economic value the network owner could obtain from that expansion.

There is also an equilibrium given the existence of the network such that the travel time assumed by travelers when making trips, and used in choosing the number of trips to make, when to make them, where to go, and what mode to use, equals the travel time resulting from the route choice of that demand pattern.

Finally, there is the *user equilibrium*. Given a demand pattern, *rational* road users behave such that their travel costs are minimized. Provided that trip-makers are *omniscient* in perceiving the travel costs on all routes and are able and willing to select the path with

the least cost, a *deterministic user equilibrium* (DUE) will be achieved when all used paths are least-cost paths and all unused paths have cost greater or equal to the used ones. In this situation, no road user can make his travel cost less by unilaterally changing routes, subject to others doing likewise.¹⁷

¹⁷ This is called a Wardropian User Equilibrium (Wardrop 1952). It is equivalent to a Nash Equilibrium in Game Theory, though developed independently (Nash et al. 1950; Nash 1951).

This is illustrated in Figure 9.3. For two routes, the times on the routes are equalized at around 60 units of time, where approximately 2,000 travelers take route *a*, and the remainder chose route *b*.

If the way road users perceive travel cost is not identical (as in DUE), but rather accommodates uncertainty, a more general view of equilibrium, that is, *stochastic user equilibrium* (SUE), can be achieved. At SUE, a road user might select routes probabilistically, accounting for the actual travel time of that route and an uncertainty term describing random perception errors. In reality, such *time perceptions*¹⁸ are not simply random but depend on travel conditions and are likely to be biased.

¹⁸ §2.15.

Implicit above is the notion that drivers act on the information immediately. In other words, they start the trip or switch routes at the same time they receive the information. Unfortunately, this assumption is an unrealistic description of observed reality, where there is usually a time lag between the time the information is collected and the point the driver starts the trip or switches routes. There is also the assumption that the travel time is fixed for the duration of the trip, when in fact travel time on links varies over time. Methods like dynamic traffic assignment (DTA) help address these issues.¹⁹

¹⁹ (Jayakrishnan et al. 1994).

9.3 Reliability



Figure 9.4: I-35W Mississippi River Bridge under construction. Photo by D. Levinson.

THE TIME IT TAKES TO TRAVEL A PARTICULAR ROUTE MAY BE LESS IMPORTANT THAN HOW RELIABLY THE DRIVER CAN PREDICT THE DURATION OF THE COMMUTE.

If drivers can ensure reaching their destinations in a time-certain manner, they may be willing to drive on routes that take somewhat longer rather than risking the use of routes that can be traveled at faster speeds on average, but that entail greater risks of arriving late. This is not a mere theoretical issue, as that situation reasonably describes the differences between signalized arterials on a [grid](#)²⁰ street network (on which travel is slow, but reliable) and a freeway (on which travel is fast, but subject to ‘catastrophic’²¹ failures that may cause all traffic to come to a halt and provide no opportunities for the driver to exit the roadway).

²⁰ §11.1.

²¹ §10.8.

Travel time reliability can be measured in many different ways. One common way is to look at the standard deviation of travel time. Another compares the 95th and 50th percentile travel times.

Rather than considering the value of travel time reliability, conventional planning models assume that drivers select the shortest travel time path. With GPS data about people’s actual routes and the actual travel times on networks, we now have a lot of evidence that people don’t actually use the shortest path.²²

²² (Zhu and Levinson 2015).

It turns out the *reliability ratio*, the ratio of the value of reliability

to the value of travel time, is on the order of 1 – 1 minute of standard deviation is about as costly as 1 minute of travel time. People see real value to improving travel time time reliability.²³

²³ (Carrion and Levinson 2012b).

Reliability is but one of many factors people consider when choosing routes. In addition to **perceived travel time**²⁴ and reliability, we might add tolls, aesthetics, number of stops, familiarity, the availability of services (gas stations, coffee, McMuffins), type of road (some people hate freeways, others hate traffic lights), **circuitry**,²⁵ and perceived safety, among many others.

²⁴ §2.15.

²⁵ §10.9.

Trying to improve reliability is one of the main justifications for any number of traffic management programs, including ramp metering, highway helpers, and **high occupancy/toll**²⁶ lanes.

²⁶ §6.10.

9.4 Price of Anarchy

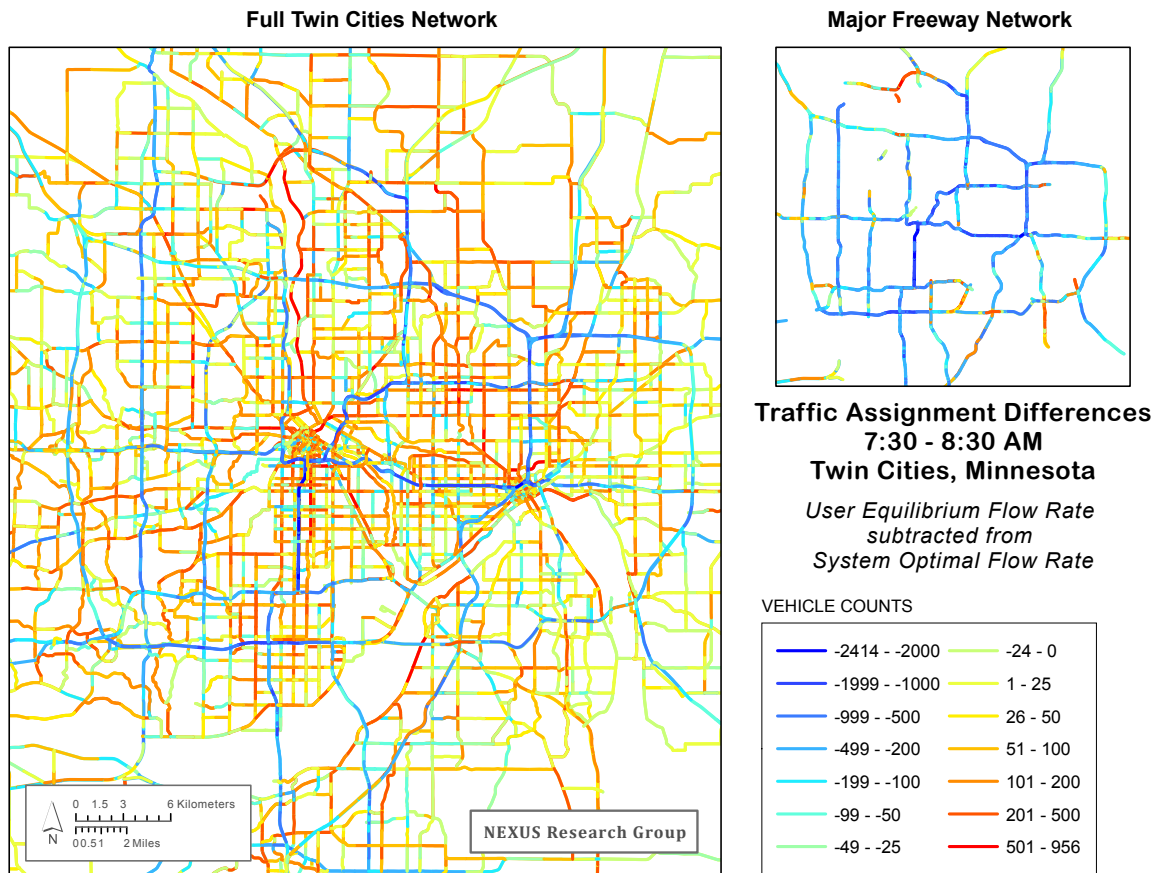


Figure 9.5: The price of anarchy in the Twin Cities is fairly low.

HOW INEFFICIENT IS IT TO LET EVERYONE DECIDE THEIR OWN ROUTES?

The ratio of the travel time that results from each user trying to minimize their own time, subject to everyone else doing the same (the *user equilibrium*²⁷ travel time) and the travel that results from systematically allocating routes to drivers to minimize the total travel time on the network (the *system optimal*²⁸ travel time) is the *Price of Anarchy*. This is a measure of how much inefficiency results from individuals choosing their own routes. One could imagine softly encouraging travelers to take routes for the benefit of others, through for instance, exhortation or traveler information. One could imagine doing so more rigorously through congestion *prices*²⁹ that were set

²⁷ §9.2.

²⁸ §9.2.

²⁹ §9.7.

Table 9.1: Price of anarchy in the Minneapolis - St. Paul region.

	User Equilibrium	System Optimal
Vehicle km	9,326,447	9,373,185
Vehicle Hours	150,964	148,389
Speed (km/h)	61.8	63.2
Trips	632,232	
Time Savings	1.7%	

at a level to ensure that the two travel costs (time + money) were equivalent in the user equilibrium and system optimal frameworks.

However, it turns out that the Price of Anarchy is relatively small on real networks. When there is no congestion, the price of anarchy is 1. When the network is supersaturated, the Price of Anarchy also approaches 1. When traffic is congested, some travelers, we might call them ‘traffic entrepreneurs,’ seek alternative routes. This tends to help the network reach [equilibrium](#).³⁰ The Price of Anarchy is somewhat higher in practice during moderate congestion.³¹

While the Price of Anarchy is small, the flow differences are more significant. As shown in [Figure 9.5](#), in the Twin Cities, application of a travel demand model suggests there is more traffic on freeways in a user equilibrium routing and less on arterials than is the case of a system optimal routing. We can attribute this to various factors, among them, most people don’t know alternative routes very well, as well as habit.³²

³⁰ §9.2.

³¹ (O’Hare et al. 2016). For a more theoretical treatment see: (Roughgarden 2005).

³² Research presented at American Physical Society “The Physics of Traffic Congestion and Road Pricing in Transportation Planning”. Analysis by Shanjiang Zhu and David Levinson (March 16, 2010)

9.5 The Braess Paradox



Figure 9.6: Macy's Broadway, New York. Photo by D. Levinson.

ADDING CAPACITY CAN SOMETIMES INCREASE TRAVEL TIME.

Elsewhere we identify [induced demand](#),³³ which suggests that adding capacity might not reduce congestion because new travelers are attracted to the route. Even if there were no more travelers, adding capacity to a transport network does not guarantee that individual travelers will enjoy shorter travel times. In some cases, the counter-intuitive, paradoxical result that adding capacity leads to an average rise of travel times arises.

This paradox was introduced by Dietrich Braess in 1968. It shows that adding one link to a simple four link network may cause longer travel times for every traveler if all travelers choose to minimize their own travel times.³⁴ In this case, each traveler's decision to act selfishly may achieve a user equilibrium that makes everyone worse off, and thus increase total travel cost.

Results from two widely cited cases present counterintuitive consequences of either expanding the network (Stuttgart) or removing links from the network (New York City).³⁵

In both instances, the Braess Paradox has been argued to explain the unexpected results. Still, research in this field is largely conceptual and usually based on small networks with simplified link performance functions. Ever since this phenomenon was first described in the literature, it has been widely studied due to its

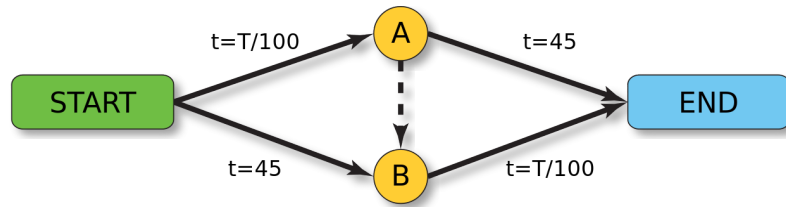
This section was based on work written by D. Levinson and Shanjiang Zhu.

³³ §12.1.

³⁴ The paper was originally written in German (Braess 1968), and later introduced to English-speaking community (Murchland 1970) and finally translated directly (Braess et al. 2005).

³⁵ Stuttgart case: (Knödel 1969); New York case: (Kolata 1990).

Figure 9.7: Illustration of the Braess paradox from Wikipedia, Adding link A-B increases travel time for each traveler. [https://en.wikipedia.org/wiki/Braess'_paradox](https://en.wikipedia.org/wiki/Braess%27_paradox).



significance for network design.

On a more general network, Steinberg and Zangwill conclude that the “Braess paradox is about as likely to occur as not occur” with random rather than planned additions.³⁶

³⁶ (Steinberg and Zangwill 1983).

Clearly, the occurrence of the Braess Paradox depends on link congestion function parameters and the demand for travel.³⁷

³⁷ (Pas and Principio 1997).

A 2008 study further explored this concept and identified links that might trigger the Braess Paradox on sketch networks of Boston, New York, and London.³⁸

³⁸ (Youn et al. 2008).

Although this research was based on maps of real networks, it still assumed aggregate link performance functions (which map traffic flow onto travel time) and unique origin-destination pairs. A series of independent and repeated route choice decisions of participants when facing a Braess Paradox type network in two laboratory experiments and concluded that the paradox was likely.³⁹ Our own classroom experience suggests that engineering students who are exposed to the concept of the Braess paradox are much more amenable to the idea that adding capacity does not necessarily reduce travel times.⁴⁰

³⁹ (Rapoport et al. 2009).

⁴⁰ (Di et al. 2015).

The relative lack of field evidence suggests that the Braess Paradox is primarily a theoretical curiosity and is too extreme to be a widespread real-world phenomenon due to complexity in travel behavior and network conditions.

To date, we have found no convincing studies that empirically demonstrate the Braess Paradox on real large-scale networks. This may be due to:

1. The difficulties in accurately measuring network flow and travel time;
2. Individual valuations on the costs of travel used in selecting routes;
3. Confounding factors contributing to long-term changes in travel demand and pattern;
4. The lack of a clearly defined impact zone isolated from the rest of

the network;

5. The relative rarity of such paradoxes; and
6. The political difficulty of an empirical, large-scale, real-world trial.

However, while ‘absence of evidence is not evidence of absence,’ it is suggestive. After searching for Bigfoot for decades, in a world where everyone has a camera in their pocket, surely the expectation of finding Bigfoot decreases.

Still, the conversion of Broadway in Manhattan, one section of which is shown in Figure 9.6, from a through route is a good candidate for the general proposition that removing a link can increase speeds and the efficiency of the network (in addition to creating great pedestrian spaces), but whether this case is, strictly speaking, a Braess Paradox, as opposed to some other ‘paradox,’ perhaps in this case related to the complexity of [traffic signal coordination](#)⁴¹ when the grid is interrupted by a diagonal route or even ‘reduced demand,’ is unclear. Researchers have also captured new paradoxes under different network conditions.⁴²

⁴¹ §5.3.

⁴² (Fisk 1979; Dafermos and Nagurney 1984; Cohen and Kelly 1990).

9.6 Rationing

Figure 9.8: Toll plaza, Golden Gate Bridge, California. Photo by D. Levinson.



SCARCE RESOURCES ARE RATIONED FOR SAFETY, EFFICIENCY, AND EQUITY.

In many cases, travelers or vehicles are controlled by external devices for safety or equity reasons: air traffic control allocates take-off and landing slots at airports; traffic signals⁴³ allocate the space at intersections to different traffic streams in turn; ramp meters⁴⁴ limit traffic from entering freeways; parking meters⁴⁵ ration the use of scarce curbside roadscape. In each case, the travelers will have to wait for their turn,⁴⁶ but because they arrive at their convenience, randomly from the point of view from the control system, they may experience more waiting time than strictly necessary.

Intelligent Transportation Systems (ITS) are about introducing information technologies into surface transport. The most important information technology is price. Price is not itself a good, but rather a mechanism that provides information about the value at which people will exchange one good or service for another. The price tells you that you will have to give so many dollars in exchange for a widget, the right to ride the bus, or to travel across the bridge.

The first problem in surface transport in advanced economies (where the network and vehicles exist and are widespread) is who gets to use which piece of infrastructure at what time (the problem

of allocation). The second is the problem of paying for the maintenance of existing facilities (the problem of funding). The structural feature at the core of these problems is the lack of an apparent price that is sensitive to time of day, location, and costs.

When travelers drive an untolled road in the United States, they still have a relatively small **personal cost**:⁴⁷ their time and the monetary costs of operating an automobile, including gas taxes. But those prices contain very little information and do not represent the actual costs they impose on the system (that is, the marginal cost of one additional car trip). The cost of fuel does not reflect the cost of traveling during the peak (except to the extent that fuel consumption is higher in stop-and-go traffic), or the cost of traveling on costly or critical facilities. The price travelers face is not real-time or real-space but rather an abstracted expectation of average costs (assuming drivers pay their full costs, which they don't off the freeway, or even on the freeway when you account for **externalities**).⁴⁸

⁴⁷ §12.3.

⁴⁸ §12.4.

Often the most efficient way of rationing a scarce resource is charging for it. This typically reduces demand compared with no price, and if set appropriately, can help ensure **supply matches demand**.⁴⁹

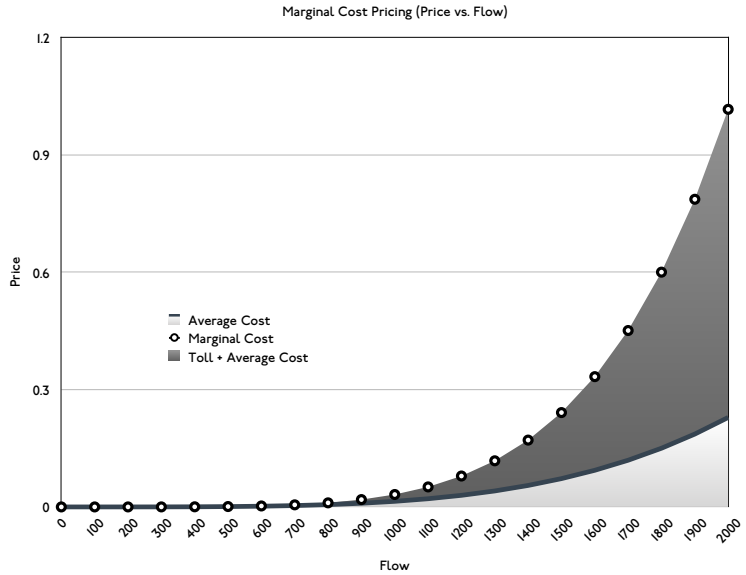
⁴⁹ §9.2.

Early tolls required travelers to stop, as illustrated in Figure 9.8. With modern technology, **prices**⁵⁰ are collected while vehicles are in-motion, and can vary in time and place, to reflect the real costs of travel, just as other goods have prices that vary with demand. When demand is up for gasoline or houses, the prices rise. When supply rises, prices fall. When demand falls or supply rises, the price falls with it. The price represents the matching of consumer's willingness to pay (to the extent the supplier has monopoly powers) with supplier's willingness to accept (assuming competition in the marketplace). This can simultaneously solve both the problem of allocation and reduce, if not eliminate, congestion and the problem of funding. Thus, the problem is less about our technical ability to reduce congestion and more about our lack of political will.

⁵⁰ §9.7.

9.7 Pricing

Figure 9.9: Marginal cost pricing.



PRICING BOTH RAISES REVENUE AND ALLOCATES DEMAND.

Prices can vary spatially, some routes are more expensive than others, or temporally, some times are more expensive than others. It turns out that temporal differences in prices are far more important because the *price of anarchy*⁵¹ is so low.

⁵¹ §9.4.

The right price from a theoretical economic perspective is the marginal cost of travel, or the difference between the *external*⁵² cost (the cost imposed on everyone else) and the *private*⁵³ (or internal) cost (the cost travelers bear themselves). This is defined as the change in total cost with respect to change in flow.⁵⁴ This assumes a more macroscopic perspective than queueing analysis and requires understanding the static relationship between aggregate flow and average travel time (a link performance function), as shown by the light shaded area in Figure 9.9.

⁵² §12.4.

⁵³ §12.3.

⁵⁴ §6.1.

Using more dynamic *deterministic queueing*⁵⁵ ideas, we can see the exact costs one additional traveler imposes costs on everyone else. Imagine a bottleneck that ‘serves’ 1 car every two seconds, but demand is 1 car every second (for 1 minute), and then cars stop arriving. After 2 seconds, 2 cars arrive and 1 car has been served. After 60 seconds, 60 cars have arrived and 30 cars have been served. After two minutes, 60 cars arrived and 60 have been served.

⁵⁵ §5.1.

Wouldn't it be better if cars arrived right when they could be served, rather than queueing needlessly? Prices can incentivize travelers to do that.⁵⁶

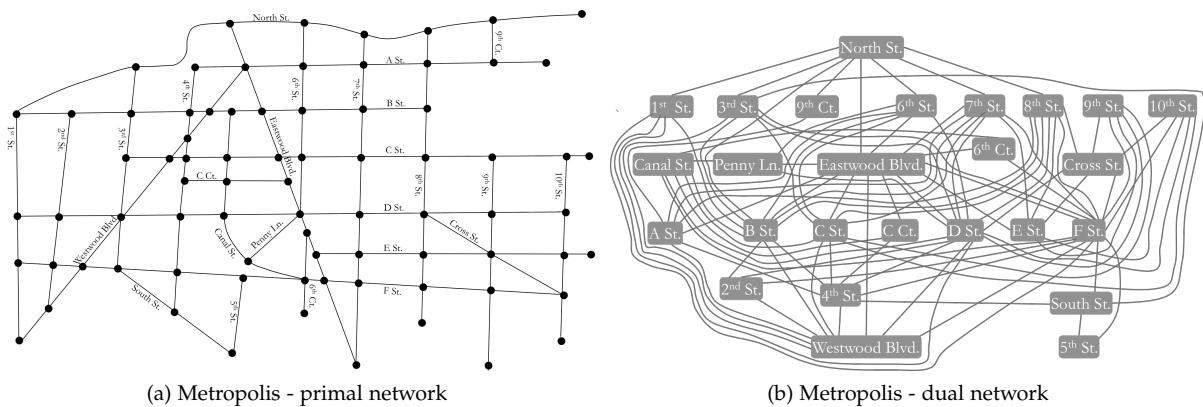
The marginal (or incremental) cost of each car depends on which position the car arrived in. The first car in the queue delays all following cars. The last car in the queue delays no one. Since there are 60 cars in our example, and each is delayed 2 seconds by the presence of a car ahead of it, the first car imposes a total of 120 seconds of delay in external costs. The second car imposes 118 seconds of delay on other following vehicles, and so on until the next to last car imposes 2 seconds of delay upon the last car.

Imagine a value of travel time savings of \$0.25/minute (\$15/h). This doesn't necessarily mean the toll for the first car is 2 minutes * \$0.25/minute = \$0.50. We also need to consider schedule delay, that is, early and late penalties. If everyone wants to arrive at the end of minute 1, but not everyone can, they have already spread themselves out some. Prices just do that a bit more. Thus the highest toll might be at second 58 or second 60, which would guarantee arrival at minute 1, and it would be lower as we get farther from that time, till it gets to zero (or a baseline flat toll) at second 0 or second 120, where people are maximally early or maximally late. The shape of this tolling triangle depends on preferences for early and late arrival. Usually each additional minute of late arrival at work is considered more costly than being a minute early or a spending a minute en route.

Coordinating to the level of the exact second may be challenging, but imagine each second is a minute, and each car is 60 cars, and this occurs over a two-hour rush hour period instead of two-minutes. Spreading the traffic out over time can eliminate all delay and make society as a whole better off. It has the potential to make no one individual worse off, but this depends on the value of time of different travelers, desired arrival times, and their demand patterns, and what is done with the revenue. With the right prices, and the right information, everyone arrives on-time without delay.

⁵⁶ See (Levinson and Rafferty 2004; Levinson 2005).

Network Topology



NETWORKS PLAY A ROLE IN NEARLY ALL FACETS OF OUR DAILY LIVES, PARTICULARLY WHEN IT COMES TO TRANSPORT.

Even within the transport realm lays a relatively broad range of different network types such as air networks, freight networks, bus networks, and train networks (not to mention the accompanying power and communications networks). We also have the ubiquitous street network, which not only defines how you get around a city, but provides the form upon which our cities are built and experienced. Cities are just [social](#)¹ networks embedded in space.

Around the world good street networks come in many different [geometrical configurations](#)² ranging from the medieval patterns of cities like Prague and Florence, to the organic networks of Boston and London, and the planned grids of Washington, DC and Savannah, Georgia. But how do researchers begin to understand and quantify the differences in such networks?

The primary scientific field involved with the study of shapes and networks is called topology. Based in mathematics, topology is a subfield of geometry that allows one to transform a network via stretching or bending. Under a topological view, a network that has

Figure 10.1: Representations of Metropolis. Primal street network with intersections as nodes and segments as links. Dual street network with segments as nodes and intersections as links.

¹ §4.1.

² §11.

been stretched like a clock in a Salvador Dali painting would be congruent with the original, unstretched network. This would not be the case in conventional Euclidean geometry where differences in size or angle cannot be ignored. The transport sector typically models networks as a **graph**³ of nodes and links.⁴ The node (or vertex) is the fundamental building block of the model; links (or edges) are not independent entities but rather are represented as connections between two nodes. Connectivity – and the overall structure of the network that emerges from that connectivity – is what topology is all about. In other words, topology cares less about the properties of the objects themselves and more about how they come together.

³ §10.1.

⁴ (Levinson and Krizek 2007).

For instance if we look at the topology of a light rail network, the stations would typically be considered the nodes and the rail lines would be the links. In this case, the stations are the actors in the network, and the rail lines represent the relationships between the actors.⁵ Those relationships – and more specifically, those connections – embody what is important. Taking a similar approach with a street network, we might identify the intersections as the nodes and the street segments as the links, as shown in the network above based on an early version of Metropolis.⁶ For most street networks, however, the street segments are just as important as the intersections, if not more so. The ‘space syntax’ approach takes the opposite (or ‘dual’) approach with street networks: the nodes represent the streets, and the lines between the nodes (that is the links or edges) are present when two streets are connected at an intersection, as shown using the same Metropolis network (Figure 10.1).⁷

⁵ (Neal 2012).

⁶ (Fleischer 1941).

⁷ (Jiang 2007).

Initial theories related to topology trace back to 1736 with Leonhard Euler and his paper on the Seven Bridges of Königsberg.⁸ Graph theory based topological measures first debuted in the late 1940s.⁹

⁸ (Euler 1953).

⁹ (Bavelas 1948).

The topological approach to measuring street networks, for instance, is primarily based upon the idea that some streets are more important because they are more accessible, or in the topological vernacular, more central.¹⁰ We note that some streets are more important because they are wider, or they are wider because they are more important. This is considered in the **hierarchy of roads**.¹¹

¹⁰ (Porta et al. 2006).

¹¹ §10.2.

Related to connectivity, centrality is another important topological consideration. A typical Union Station, so called because it was a combined train station for different private railroads, is a highly central and important node because it acts as a

hub for connecting several different rail lines. Some common topological measures of centrality include [degree](#)¹² and [betweenness](#),¹³ which we will discuss in more detail elsewhere.

There are also some peculiarities worth remembering when it comes to topology.

When thinking about the ‘size’ of a network, our first inclination might be measures that provide length or area. In topological terms, however, size refers to the number of nodes in a network.

Other relevant size-related measures include: ‘geodesic distance’ (the fewest number of links between two nodes); ‘diameter’ (the highest geodesic distance in a network); and ‘characteristic path length’ (the average geodesic distance of a network).

Density is another tricky term in the topological vernacular. Earlier sections defined [traffic density](#)¹⁴ and [population density](#).¹⁵ When talking about the density of a city, we usually seek out measures such as population density, intersection density, or land use intensity. In most cases, these metrics are calculated per unit area. In topology, however, ‘density’ refers to the density of connections. In other words, the density of a network can be calculated by dividing the number of links by the number of possible links. Topologically, the fully-gridded street networks of Salt Lake City, Utah and Portland, Oregon (as shown in [Figures 11.3](#) and [11.4](#), respectively) are essentially the same. With respect to transport and urbanism, however, there remain drastic functionality differences between the 200’ (~60m) Portland blocks and the 660’ (~200 m) Salt Lake City [blocks](#).¹⁶

As illustrated with the Portland/Salt Lake City example, one limitation of topology is that it ignores scale. However, this can also be an advantage. For instance, Denver might be much closer to Springfield, Illinois than Washington, DC as the crow flies, but a combination of several inexpensive options for direct flights to DC and relatively few direct flight options for Springfield mean that DC is essentially closer in terms of air network connectivity. Topology captures such distinctions by focusing on connectedness rather than length.

While topological analyses such as the above are ‘scale-independent,’ we also need to be careful about use of this term because ‘scale-free networks’ are not equivalent to ‘scale-independent analyses.’

In topological thinking, scale-free networks are highly centralized. More specifically, if we plot the number of connections for each node, the resulting distribution for what is known in topology as a scale-free network would resemble a ‘power law distribution’ with some

¹² §10.3.

¹³ §10.4.

Topological measures were initially developed in papers analyzing social networks ([Freeman 1978](#)) and the political landscape. ([Krackhardt 1990](#)) Since then, topological analyses have been widely adopted in attempting to uncover patterns in biology, ([Jeong et al. 2000](#)) ecology, ([Montoya and Solé 2002](#)) linguistics, ([Ferrer i Cancho and Solé 2001](#)) and transport ([Jiang and Claramunt 2004](#)) and ([Carvalho and Penn 2004](#)). Topology represents an effort to uncover structure and pattern in these often complex networks ([Buhl et al. 2006](#)).

¹⁴ §6.4.

¹⁵ §3.1.

¹⁶ §11.2.

nodes having many connections but most having very few. A hub-and-spoke light-rail system, for instance, tends to exhibit scale-free network qualities with relatively few stations connecting many lines. The nodes in a random network, on the other hand, tend to have approximately the same number of connections. For instance when we define the intersections of a street network as the nodes and the segments as the links (the primal graph), the results tends towards a random network, where nodes have a similar number of entering and exiting links (degree). If we flip the definition again, so that the streets are the nodes and the intersections the links (the dual graph), we trend back towards a scale-free network, where a few nodes (streets) connect many links (have many junctions), but most nodes connect few links.¹⁷

¹⁷ (Jiang and Claramunt 2004; Jiang 2007).

¹⁸ §10.8.

One reason to look at connectivity in these terms has to do with the critical issues of **resilience and vulnerability**.¹⁸ In general, robustness is associated with connectivity. When we have good connectivity, removing one node or link does not make much of a difference in terms of overall network performance. In contrast, scale-free networks are more susceptible to strategic attacks, failures, or catastrophes. However, as shown in a recent paper about urban street network topology during a Zombie apocalypse, good connectivity could actually be a double-edged sword.¹⁹

¹⁹ (Ball et al. 2013).

Access depends on network speed but also network distance. Distance is a product of the network's topological structure discussed in this chapter, and its geometric configuration and morphology, as well as how the network interactions are managed, treated in following chapters.

10.1 Graph



Figure 10.2: Cartoon representation of Metropolis. Source: (Fleischer 1941).

‘The map is not the territory.’ – Alfred Korzybski

THE GRAPH IS THE STYLIZED REPRESENTATION OF A TRANSPORT NETWORK FOR USE IN NETWORK ANALYSIS. IT COMPRISES NODES AND LINKS.

The cartoon of Metropolis (Figure 10.2) compared with graph representations (Figure 10.1) illustrates the point.

A node (or vertex) typically is a junction or intersection between two roads, or any place that traffic can enter the network. Each node has a location in space, typically denoted by latitude and longitude (or planar coordinates). There may also be a z-coordinate to identify elevation. Typically, each node has a unique identification number. In addition, nodes may have other attributes. On an idealized graph, nodes are points and have no size. In practice, junctions do have some physical size, so it may be desirable to measure that size.

A link (or edge) is defined by two nodes and is **directional**.²⁰ So the link from node i to node j differs from the link from node j to node i . Links have many attributes, including length, free-flow speed, capacity, number of lanes or width, functional classification, and some ‘link performance function’ relating travel time to free-flow speed, flow, and capacity.

A more sophisticated analysis may consider the shape of links (the

²⁰ §7.13.

simple definition above implies that links are straight lines between two points).

A turn can be defined by a sequence of three nodes (at-from-to), and may be useful as the cost of making a left or right turn may differ from a through movement at a junction.

CENTERS ARE EDGES, LINKS ARE SPACES

In graphs, nodes are aspatial representations of the intersection of links, which themselves are aspatial representations of the structure of network. However, real nodes, like centers (places) and junctions represented as spaceless points on the graph, take physical space. As such, they provide a spatial separation between areas that adjoin them. In addition to nodes standing in for places in their own right, centers serve as edges (in the sense of 'boundaries,' not 'links') of adjoining neighborhoods.²¹

²¹ §3.4.

Similarly, links themselves are not one-dimensional objects, but at least two-dimensional, as we discuss in the chapter on streets.²²

²² §7.

10.2 Hierarchy

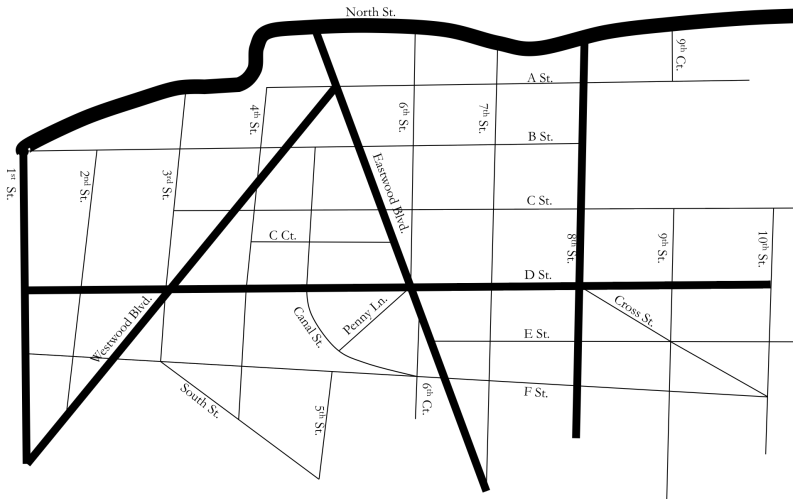


Figure 10.3: Hierarchical street network for Metropolis.

NOT ALL LINKS (OR NODES) ARE CREATED EQUAL, PARTICULARLY WHEN IT COMES TO TRANSPORT NETWORKS.

In binary networks, the focus is on whether or not a connection between two nodes exists. However, when we know about the presence of a link as well as the strength of that link, it is called a ‘valued’ or ‘weighted’ network. For instance when traveling from *A* to *B* in a street network, there is usually discontinuity in street type. In other words, one might move from a local street to a collector road to an arterial road and then back to a collector before reaching their destination. While engineers know this sort of differentiation as functional classification, it is also referred to as the ‘hierarchy of roads.’

Hierarchy, which is embedded in many natural and societal systems such as biologic cells and the Internet, is a common transport complexity that requires a more complicated topological analysis.²³ Typical topological measures such as *degree*²⁴ or *betweenness*²⁵ can be useful in helping understand network hierarchy, particularly in tree-like networks; however, such measures would fail to properly distinguish between streets in a gridded street network. In the Figure 10.3 version of Metropolis’ street network, the major streets are represented by thicker lines and easily discerned, even in a gridded network. Using the basic set of topological metrics, we would have no idea that 8th Street is functionally different from 7th Street or F Street from D Street.

²³ (Tomko et al. 2008).

²⁴ §10.3.

²⁵ §10.4.

These metrics fail to consider other attributes such as urban design, number of lanes, ‘active’ transport infrastructure supporting walking and biking, adjacent land uses, and speed. Topology alone would not necessarily be able to distinguish such streets.

²⁶ (Xie and Levinson 2008).

Working with hierarchical networks often involves dividing networks in multiple layers or tiers. Measurements of heterogeneity have also become common proxies for characterizing hierarchy. To identify heterogeneity among street segments, researchers have used entropy measures as well as discontinuity measures.²⁶ Discontinuity, for example, does not necessarily denote a disconnected network; rather, the reference is to the discontinuity in moving from one street type to another. If we sum the number of times a traveler goes from one type of street to another while traveling along a shortest path route, we find the trip discontinuity. Dividing that number by the length of the trip gives us the relative discontinuity. Other simplistic hierarchy measures calculate the relative percentage of a particular type of road. For instance, we might divide the number or length of arterials by the total number or length of roads to find the relative percent arterials.

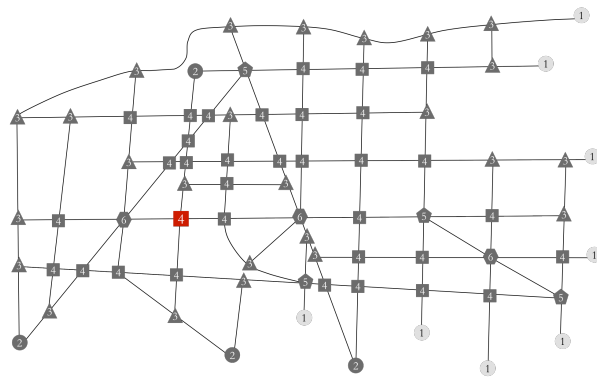
²⁷ (Bern and Marshall 2013).

²⁸ (Jiang 2009).

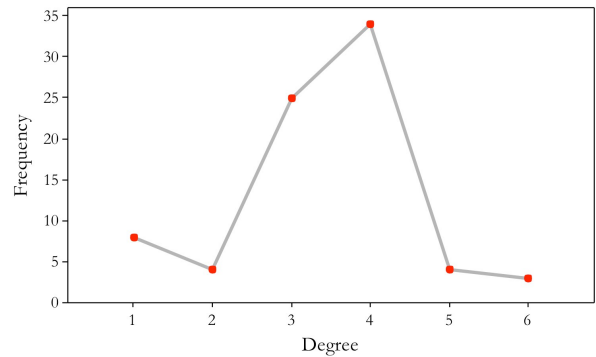
Interestingly, it is not uncommon for large-scale transport models to delete the local streets on the lower end of the hierarchical spectrum for the sake of computational efficiency. Yet, removing such streets creates a bias against more connected networks because less connected networks typically need to be supported by major streets with more capacity than would be needed in more connected networks.²⁷ Some topological researchers – where the focus should be on understanding the full network – unfortunately reach the same conclusion: ‘urban streets demonstrate a hierarchical structure in the sense that a majority is trivial, while a minority is vital.’²⁸ If we only care about vehicle traffic flow, such statements may be true. However, our previous street network research confirms that understanding the full network holds the key to pushing toward improved safety, increased active transport, and better environmental and health outcomes.²⁹

²⁹ (Marshall and Garrick 2009; 2010a;b; 2012).

10.3 Degree



(a) Metropolis - degree values



(b) Metropolis - degree distribution

Figure 10.4: Degree representations of Metropolis.

HOW CONNECTED AND HOW INFLUENTIAL IS A NODE WITHIN THE OVERALL NETWORK?

Centrality measures help gauge the overall importance of a node. One of the simplest measures of centrality is ‘degree,’ which measures the number of connections between a node and all other nodes. For instance if we are considering a street network with intersections as nodes, a nodal degree of 4 would indicate a typical 4-way intersection. Figure 10.4(a) renders the Metropolis street network with a degree value shown at each intersection and a 4-way intersection highlighted in red. When we focus on what is happening at one particular node, it is called the ‘ego network’ (in that we are looking at the network from the perspective of a single node while ignoring all nodes not directly connected, which can be deemed a bit narcissistic). The entire network can be called the ‘complete,’ ‘whole,’ or ‘global network.’ So if we want an overall degree measure, we can calculate average degree, which is the average number of connections for all the nodes within the overall network. When the average degree exceeds 1, every node has at least one connection, on average. When the average degree approaches $\log(n)$, where n equals the number of nodes in the network, every node starts to become accessible from every other node.³⁰ For the Metropolis network, there are 78 nodes with an average degree of 3.4.

Analyzing degree measures for a complete network also entails generating a ‘degree distribution,’ which literally equates to the plotting the frequency of each degree for all the nodes shown in Figure 10.4(b) for the Metropolis street network. The idea is to try to capture the relative differences in connectivity between the nodes in

³⁰ (Neal 2012).

order to gain a sense of network structure. For instance, every node in a homogenous network would have the exact same number of connections and not much of a distribution. A more centralized network might have one node with a high degree value and all other nodes with low degree values.

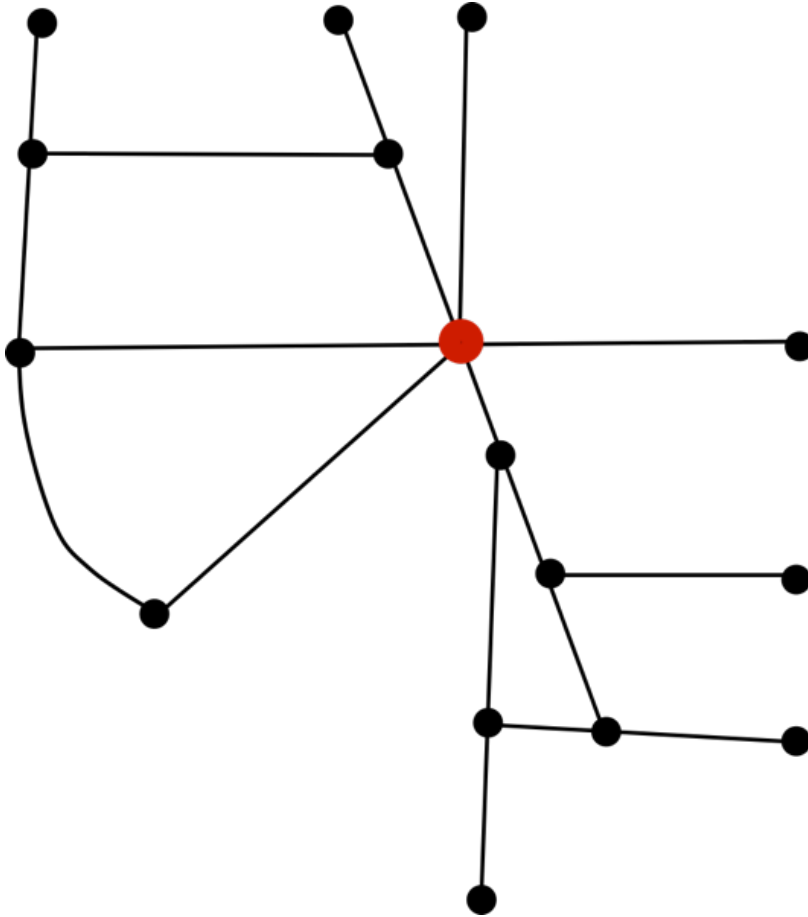
10.4 *Betweenness*

Figure 10.5: Betweenness. Red node indicating highest betweenness value for this network.

'BETWEENNESS' MEASURES CAPTURE RELATIVE FLOW BY QUANTIFYING THE NUMBER OF TIMES A NODE OR LINK IS ON A SHORTEST PATH BETWEEN TWO OTHER NODES.

Degree³¹ is often good for measuring local conditions, but adequately characterizing centrality is a bit more complicated. When trying to figure out centrality in terms of how connected and influential a node or link is, it is useful to get a sense of relative network flow through a particular node or link.

³¹ §10.3.

The first step would be to calculate the shortest path between every origin and every destination. Next, we count the number of times that a particular node or link shows up on a shortest path. The resulting number represents the relative role of a node or link as a connector between clusters of nodes or links. In Figure 10.5, the

intersection highlighted in red must be included in over half of the shortest paths. We call this count ‘betweenness,’ which is essentially an attempt to quantify how necessary a node or link is to get from one side of the network to the other. The Panama Canal, for instance, is a key maritime link connecting the Atlantic and Pacific Oceans. Without it, ships would have to route around Cape Horn at the southernmost tip of Chile or through the Straits of Magellan. For a ship traveling from New York to San Francisco, the Panama Canal – due to its high betweenness value – cuts more than 7,500 miles from the journey. In terms of other transport issues, betweenness usually relates to metrics such as accessibility and traffic congestion.

In addition to revealing relative importance, betweenness also indicates how irreplaceable a node or link may be to a network. In other words, what happens if we remove a certain node or link from the network? Very high betweenness values can indicate a critical connection between various groups of nodes or links. In some cases, this represents a vulnerability where we would want to add redundancies to the network.

In transport networks, if we assume all travelers take the [shortest path](#)³² and treat each traveler as having a unique origin and destination, betweenness is the same as the [flow](#)³³ (number of travelers) on the link. We could call this ‘flow-weighted betweenness.’

³² §2.11.

³³ §6.1.

10.5 Clustering

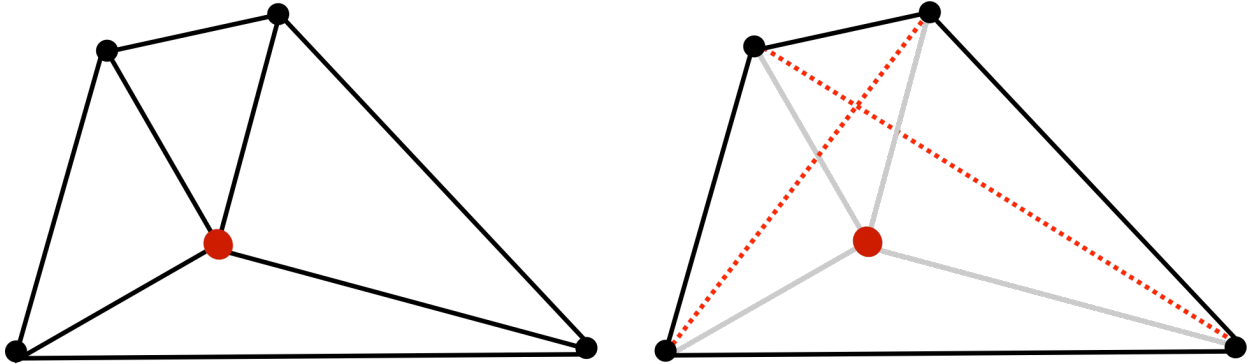


Figure 10.6: Calculating the clustering coefficient.

WHEN WE HAVE NODES OR LINKS WITH HIGH BETWEENNESS VALUES, IT IS OFTEN BECAUSE OUR NETWORK IS SPLIT INTO VARIOUS SUB-GROUPS THAT CAN BE CALLED 'CLUSTERS.'

Clusters tend to have their own unique set of properties, so it is useful to be able to identify clusters quantitatively.

While there are a growing number of clustering algorithms, the basic idea behind them is to capture the degree to which nodes cluster. The 'clustering coefficient,' for instance, represents how likely is it that two connected nodes are part of a larger group of highly connected nodes. It can be calculated by dividing number of actual connections *between*³⁴ the neighbors of a node³⁵ by the number of *possible* connections between these same neighboring nodes. For instance in Figure 10.6, the red node is the node of interest, and it has a *degree*³⁶ of 4. Those 4 neighboring nodes make 4 actual connections *with each other* (shown by the black lines in the figure on the right) but have 6 possible connections (shown by the black lines plus the red dashed lines).³⁷ Thus, the clustering coefficient for the red node is 4 divided by 6 or 0.67.

The value represented by the clustering coefficient ranges from 0 (no clustering) to 1 (complete clustering). If we are interested in the amount of clustering for an entire network, we average the clustering coefficients for all of the nodes. Clustering tends to be higher in real-world networks than in random networks. So when a network becomes more centralized, so that a small percentage of nodes have high connectivity, the overall topology becomes more differentiated and clusters begin to emerge.

Other related terms include 'component' and 'clique.' When a given sub-group of nodes is also highly connected, that is called a

³⁴ §10.4.

³⁵ Neighbors are the nodes directly connected to the node in question

³⁶ §10.3.

³⁷ In mathematical terms, 4 choose 2.

component. When the nodes in a component have few connections to other nodes outside of the component, that is a clique. Understanding clusters, components, and cliques in networks can be useful because they can hold more influence over behavior than overall network structure.³⁸

³⁸ (Neal 2012).

Imagine, for instance, a New Urbanist neighborhood with great street connectivity set into a city with poor overall street connectivity. Analyzing network structure for the overall city might lead us to one conclusion; yet, we could find very different outcomes in the New Urbanist neighborhood. While factors such as land use, street design, and demographics influence transport-related outcomes as well, the concept of clustering holds value for those interested in truly understanding transport networks and accessibility.

10.6 Meshedness

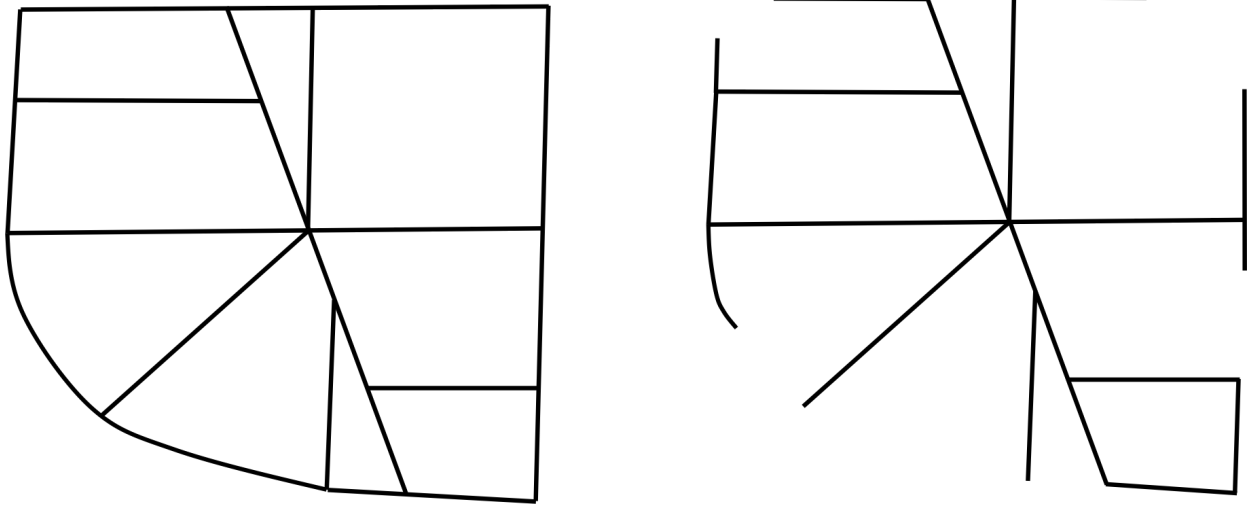


Figure 10.7: Connected and dendritic networks.

URBAN PLANNERS AND ENGINEERS HAVE LONG BEEN INTERESTED IN MEASURING STREET CONNECTIVITY AND TYPICALLY DO SO WITH RELATIVELY SIMPLE MEASURES.

The 'link-to-node ratio' (called β) or the 'beta index' in transport geography), divides the total number of links³⁹ by the total number of nodes.⁴⁰ In Figure 10.7, the connected network has link-to-node ratio of 1.6 while the dendritic network's link-to-node ratio is 1.0 (a link-to-node ratio of 1.4 is typically considered a well-connected street network).

The 'connected node ratio' divides the number of connected nodes (nodes that are not dead ends) by total number of nodes. The networks in Figure 10.7 have a connected node ratio of 1.0 and 0.6, respectively. The underlying intent is to distinguish between well-connected or gridded street networks and dendritic, tree-like networks – as highlighted in the figure – in researching relevant issues such as travel behavior, road safety, VMT, and public health outcomes.

Topology⁴¹ takes a slightly different approach to understanding this issue. The 'meshedness coefficient,' for instance, measures connectivity by looking at the number of cycles in the network with respect to the maximum number of 'cycles' (a cycle is a closed path that begins and ends at the same node with no fewer than three links). A meshedness coefficient of 0 represents full tree structure (no cycles), and 1 represents complete connectivity (every node is

³⁹ Links are defined as road segments between intersections.

⁴⁰ Nodes are defined as intersections including dead ends.

The beta index is formally:

$$\beta = E/V$$

where: E is number of edges or links, and V is the number of vertices or nodes.

For planar graphs:

$$\alpha = \frac{E - V + P}{2V - 5}$$

where P is the number of subgraphs.

For non-planar graphs:

$$\alpha = \frac{E - V}{0.5V(V - 1) - (V - 1)}$$

⁴¹ §10.

⁴² (Buhl et al. 2006).

directly connected to every other node, which is not feasible in a large surface transport network).⁴² In non-planar networks, this measure is also known in transport geography as (α) or the 'alpha index.' The alpha for the connected network above is 0.4 and for the dendritic network, it is just 0.03. For large networks, beta and alpha are highly correlated.

⁴⁴ (Xie and Levinson 2007).

Another useful metric is 'treeness'.⁴³ Instead of counting the number of cycles, treeness divides the length of street segments not within a cycle by the total length of street segments.⁴⁴

Networks with good overall connectivity are called 'integrated' networks. Networks with low connectivity are called 'fractured' networks (although fractured networks can still be comprised of connected components). Again, these measures relate to issues of resilience. When a single node failure can significantly erode network functionality, the system is fragile. Figure 10.8 shows a fallen tree in Lake Oswego, Oregon that cut off more than 50 families from the outside world (or more specifically, the cars of more than 50 households were trapped). If only that network had a little less treeness.

Figure 10.8: Downed tree traps more than 50 households. Image from (Florip 2010).



10.7 Treeness

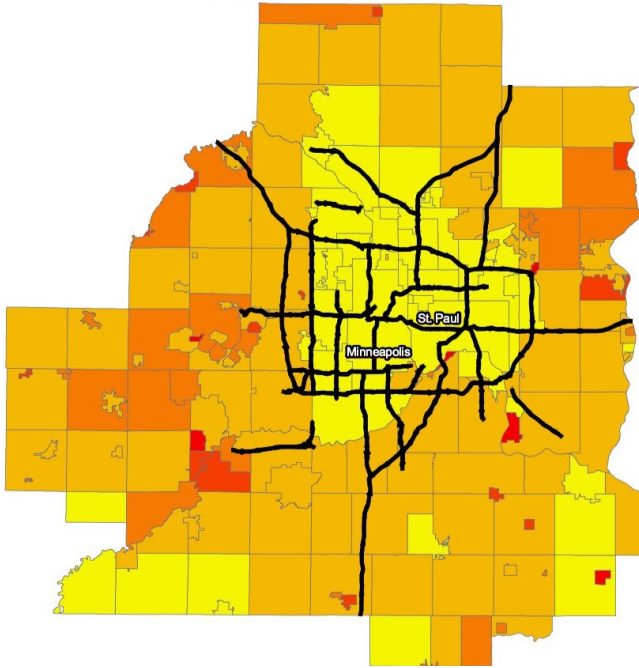


Figure 10.9: Treeness in the Twin Cities region (yellow represents low treeness; red represents high treeness). Not surprisingly, suburban areas have more tree-like networks.

CHRISTOPHER ALEXANDER ASSERTS 'A CITY IS NOT A TREE'.⁴⁵

That is true in some ways. Yet, the suburbs are certainly more tree-like than cities. Our students have measured the 'treeness' of networks, the length of street segments not within a cycle by the total length of street segments.⁴⁶

For instance, we see in Figure 10.9 that treeness is, not surprisingly, higher at the suburban edges of the metropolitan area than in the center, though it declines as we see rural areas, where the sparser network is also more mesh or grid-like.⁴⁷

Many physical infrastructures are better configured as trees, especially if they require a large capital investment (like a wastewater treatment facility). Similarly, stream and river valleys are naturally organized as hierarchies. Transit networks are also often more tree-like or radial than roads, and while may eventually evolve into ring-radial system, don't generally start out that way.⁴⁸ The Boston transit network in Figure 11.1, for instance, looks very tree-like.

⁴⁵ (Alexander 1968).

⁴⁶ (Xie and Levinson 2007).

⁴⁷ (Parthasarathi and Levinson 2012).

⁴⁸ (Roth et al. 2012).

10.8 Resilience

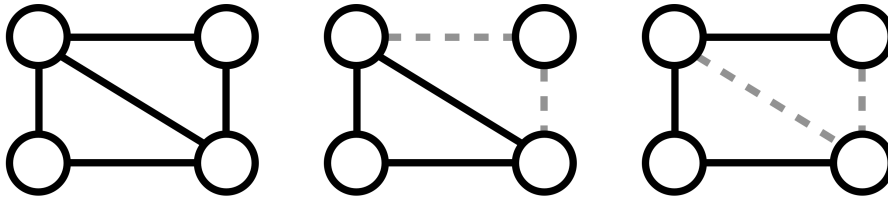


Figure 10.10: Resilience.

WHAT INVESTMENTS SHOULD YOU MAKE TO KEEP YOUR NETWORK ONLINE?

Say you are charged with ensuring that your network keeps operating. It is constantly under threat, not just from terrorists, but also from the long deterioration from lack of maintenance or the sudden onset of mother nature. If you lose connectivity, you will lose [accessibility](#).⁴⁹ Which links are most important to keep operating?

⁴⁹§1.

Graph theory defines ‘resilience’ such that: if graph G has property P , what is the minimum number of edges (E) (links) that need to be removed so that G no longer has P ?

For example, consider the graph on the left of Figure 10.10 and its resilience with respect to connectivity. Removing any one edge leaves a connected graph. It is necessary to remove two edges to produce a graph that is not connected (middle figure). Thus, we could say that this graph has a resilience of 2 with respect to connectivity. Note that this does not mean that removing any two edges will destroy connectivity in this graph. The figure on the right demonstrates the possibility of removing two edges while leaving the graph connected.⁵⁰

Under this definition, a given graph will have different values of resilience with respect to different properties. As a result, the definition is concrete, but flexible, and can be usefully applied to real-world networks where properties are of variable importance from different perspectives.

The example above highlights the difference between random edge removal and targeted edge removal. If edges are removed randomly, a property might survive the removal of many edges. Targeted edge removal implies that the graph is analyzed and edges are chosen to maximize effect. The effect on the network of either type of edge removal depends, in part, on degree distribution.

⁵⁰ This section is adapted from the Wikibook *Transportation Geography and Network Science* en.wikibooks.org/wiki/Transportation_Geography_and_Network_Science/Resilience originally written by the research team. See also (Sudakov and Vu 2008)

Graphs following a power-law distribution (scale-free) tend to be highly resilient to random edge removal because there is a very good chance that the edges removed will connect only low-degree vertices – and therefore the overall graph structure will be affected only slightly. Graphs are much more vulnerable, however, to targeted removal of edges attached to high-degree nodes, especially to the removal of those nodes themselves. In scale-free graphs, these high-degree vertices are critical in connecting subgraphs.⁵¹

⁵¹ (Newman 2003).

A graph with a low resilience with respect to a property can lose that property as a result of only a few edge removals. We can say that the graph is vulnerable with respect to that property.

But this is only half of a complete consideration of vulnerability. The other half has to do with the effect on the network's performance if the property in question has been lost.

In graph theory, resilience is a binary concept: an edge either exists, or it does not; a graph either has a property, or it does not. In real-world transport networks, links have additional properties such as capacity, utilization, demand, and cost.

10.9 Circuity

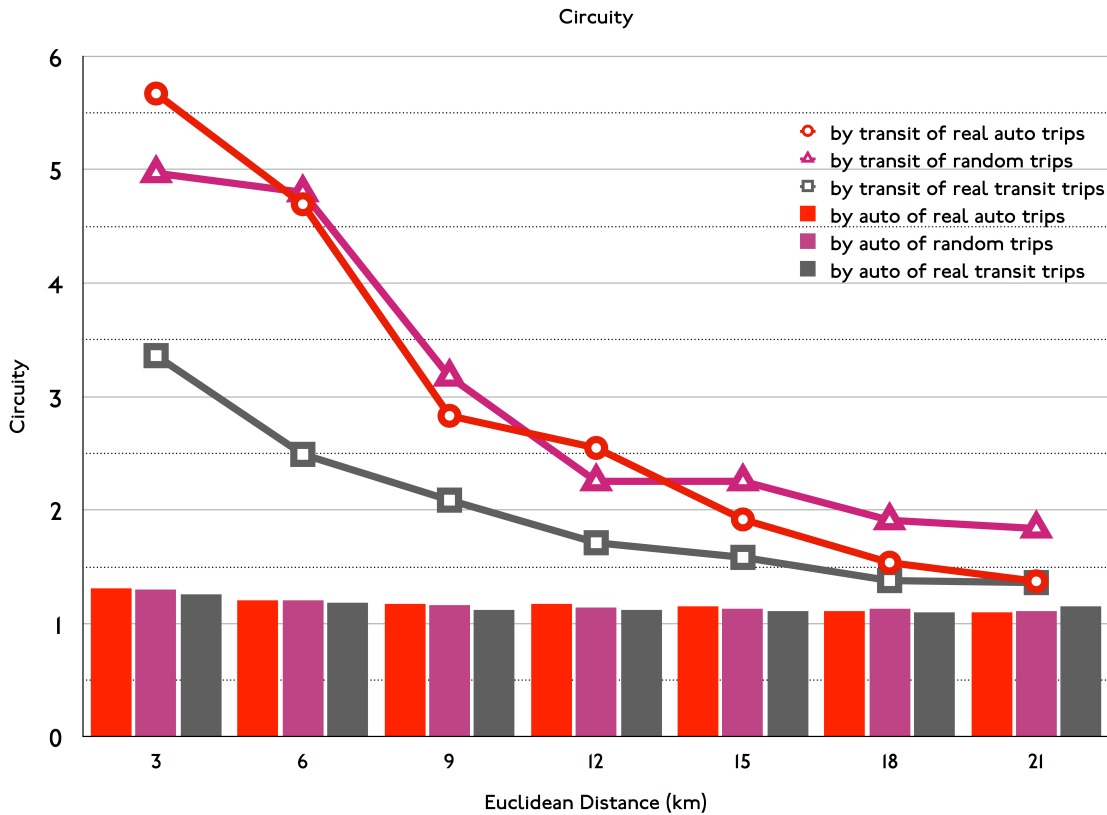


Figure 10.11: Circuity of urban networks. Source: (Huang and Levinson 2010).

‘CIRCUIITY’ IS THE NETWORK DISTANCE BETWEEN TWO POINTS DIVIDED BY THE EUCLIDEAN (OR STRAIGHT-LINE) DISTANCE AND MUST BE ABOVE THE VALUE OF 1.

⁵² Some researchers use ‘directness,’ which is simply the reciprocal of ‘circuity.’

The value of circuity⁵² varies for each origin-destination pair and depends on the physical layout of the network.

$$C_t = \frac{D_{network}}{D_{Euclidean}}$$

where:

- C_t = Average circuity of trips in time threshold t .
- $D_{network}$ = network distance.
- $D_{Euclidean}$ = Euclidean distance.

Grids⁵³ don’t do especially well on this measure. In the worst case, going from the Southwest to the Northeast corner of a square grid, geometry tells us the circuity would be about 1.41 (the square root of two). For a trip from the Southwest to Southeast corner, or the the Southwest to Northwest corner, the circuity would a minimum value of 1. Radial networks reduce the circuity to downtown, but increase it substantially for some suburb-to-suburb trips.

⁵³ §11.1.

It is important to note, however, that people do not just accept poor circuity, they in fact choose their home with respect to their workplace on corridors with less circuity than random. People self-

select homes and jobs that have more direct routes (lower circuitry) to achieve better housing options in the same travel time, as rents are lower **farther from the center**.^{54,55}

Circuitry can help to explain mode choices of commuters.⁵⁶ As shown in Figure 10.11 the circuitry of transit networks is much higher than that of road networks, helping to explain the higher travel times of transit compared with the automobile, and thus the lower mode share.

Transit is generally not operating on an efficient grid. The reasons for this are political; in the US, transit systems choose to expand their spatial coverage at the expense of directness and efficiency.⁵⁷ Transit is especially inefficient for short trips. Transit circuitry exponentially declines as travel time increases, in part because the long trips are more likely served by trains, which are much more difficult to divert than buses for the political expediency of ensuring spatial coverage.

On the highway side, in the US, overall circuitry has increased between 1990 and 2010; random points have not only become farther apart in distance, their shortest network path has become more circuitous, suggesting that the more recently constructed parts of street networks are laid out more circuitously than older parts of the network.⁵⁸

⁵⁴ §3.2.

⁵⁵ (Levinson and El-Geneidy 2009).

⁵⁶ (Huang and Levinson 2010).

⁵⁷ (Taylor 1991).

⁵⁸ (Giacomin and Levinson 2015) though see also (Barrington-Leigh and Millard-Ball 2015).

Geometries



Figure 11.1: Boston's very tree-like transit network.

NETWORKS COME IN VARIOUS SHAPES AND SIZES.

The shapes can be treated [topologically](#),¹ but sometimes we may be interested in them from a less abstract perspective. Consider the angle of intersection of links.

If links are joined at a regular 90° pattern, we get a [rectilinear grid](#),² with a [nodal degree](#)³ of 3 or 4, depending on whether links go through or stop.

If links are joined at a regular 60° pattern, we get a [hexagonal](#)⁴ pattern, and nodes will typically have degree 4 or 6, depending on whether links all join at nodes (or a roundabout), or whether the intersections are offset to reduce traffic conflicts. We can certainly

¹ §10.

² §11.1.

³ §10.3.

⁴ §11.3.

imagine angles other than 60° or 90° (we can go from 0° to 360° in principle), but these are less common. Smaller angles will tend to generate more space devoted to roads (since roads have a minimum width), larger angles will sacrifice connectivity.

⁵ §10.7. We also have radial patterns: **dendritic**⁵ or **ring-radial**,⁶ where a central node may be connected by many radiating routes, like the 10 or 11 of Central Moscow with a low average angle of 36° , but other nodes are connected to only 3 (trees) or 4 (ring) links.

⁶ §11.4.

The ring-radial pattern, and especially the tree-like pattern as shown in 11.1, increases the accessibility of the center at the expense of the edges. The hexagonal pattern lowers transport costs at the expense of using more land for transport, and thus less for activities. The grid has a relatively high transport costs, but conserves land for real estate. Real networks tend to combine these features.

⁷ §11.2. The grid varies across many dimensions, most notably **block size**.⁷

11.1 Grid

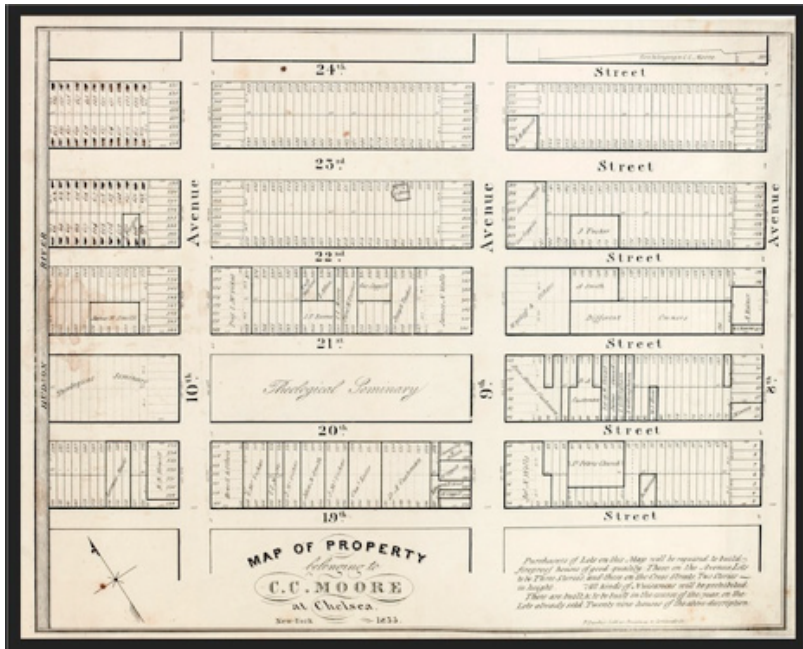


Figure 11.2: A section of the Manhattan grid.

OUR NETWORKS ARE SHAPED BY SQUARING THE CIRCLE.

Just as we have cut the round earth into a grid of latitude and longitude (and knowing that each 'block' of 1 degree latitude by 1 degree longitude gets smaller and smaller as we approach the poles), we similarly cut our cities and rural areas into a finer mesh from that same grid. Much of this arises from the various large scale ordinance surveys that took place in the Americas, Australia, and India. Grids date much earlier, to Miletus and Mohenjo Daro, among many others. Not all grids are aligned with longitude and latitude, some align with local landscape features, but most of the modern ones are. Even grids aligned with longitude and latitude occasionally look mis-aligned, as 1 degree of longitude at the north pole is a lot smaller than 1 degree at the equator. (Where grids of different alignments come together, interesting spaces are created.) Not all grids are squares; most are more like rectangles.

So why should we have 90-degree rectilinear grids? Proponents claim it:

- simplifies construction and makes it easier to maximize the use of space in buildings;

- simplifies real estate by making the life of the surveyor easier;
- simplifies intersection management by reducing conflicts compared to a 6-way intersection; and
- is embedded in existing property rights and impossible to change.

We in the modern world need not be bound to the primitive tools of the early surveyor, the primitive [signal timings](#)⁸ of the 1920s traffic engineer, or the primitive construction techniques of early carpenters. And while for existing development we might be locked into existing property rights, for new developments that doesn't follow.

⁸§5.3.

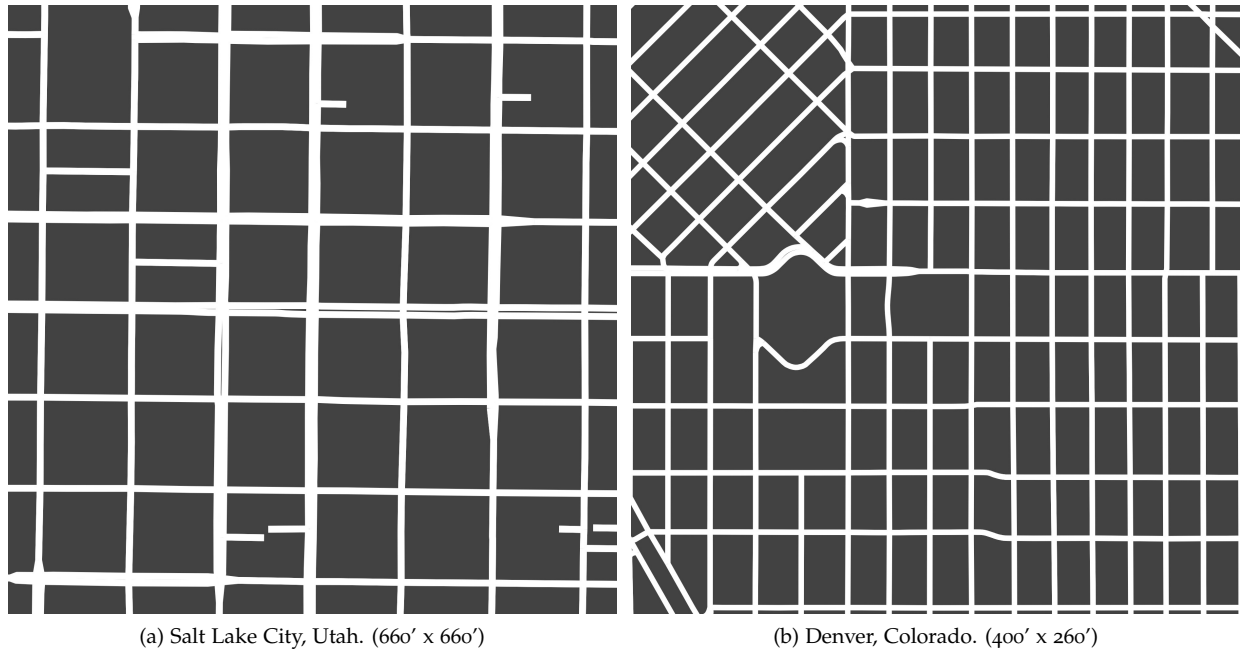
Opponents argue it:

- is among the least efficient way to connect places from a transport perspective;
- reduces opportunities for interesting architecture; and
- wastes developable space by overbuilding roads.

⁹(Ballon 2012).

The surveyors' plan of Manhattan, part of which is drawn in Figure 11.2 was a grid.⁹ But the grid was differentiated. It set aside Avenues, which ran North-South (N-S) and designed to be 100' wide (30 m). Since Manhattan is longer than it is wide, there were many more East-West (E-W) streets, exacerbated because the grid is tighter in the N-S direction than the E-W direction. The standard width was to be 60' (18 m) wide (and the blocks were 200' (60 m) long). But selected streets (4th, 23rd, 34th, 42nd, 57th, 72nd, 79th, 86th, 96th, 106th, 116th, 125th, 135th, 145th and 155th Streets) were wider, set at the same 100' width as the avenues. Some of these (155th, 145th, 125th, 42nd) eventually became the roads that some of bridges and tunnels to the rest of New York would land, though this is not a perfect match. They would also tend to become the sites of stations on the subway system a century later.

11.2 Block Sizes



UNLIKE SIR MIX-A-LOT, URBANISTS DON'T LIKE BIG BLOCKS.

While many cities use *topologically*¹⁰ identical rectangular *grid*¹¹ systems, the scale varies significantly, from Portland at 200' x 200' (60 x 60 m) to Salt Lake City at 660' x 660' (200 x 200 m). Small blocks increase surface area and the amount of building frontage on streets. In the same area that 1 block from Salt Lake has 800 m of street frontage, Portland's 9 city blocks have about 2,160 m of frontage. This is much more interesting to walk around.

Smaller blocks also increase intersection density and connectivity. For every Salt Lake block (Figure 11.3), Portland has 6 additional intersections (Figure 11.4). In practice, Salt Lake bifurcates many of its blocks,¹² with smaller streets or pedestrian cut-throughs, to address some problem of size. The smaller lots enabled by the increase of street frontage are more affordable for home-buyers.

For cars, more intersections means more places for delay, but it also means less delay per intersection, as traffic is spread across more (and skinnier) streets. Dispersion of traffic across a finer meshed grid increases reliability at the cost of average speed.

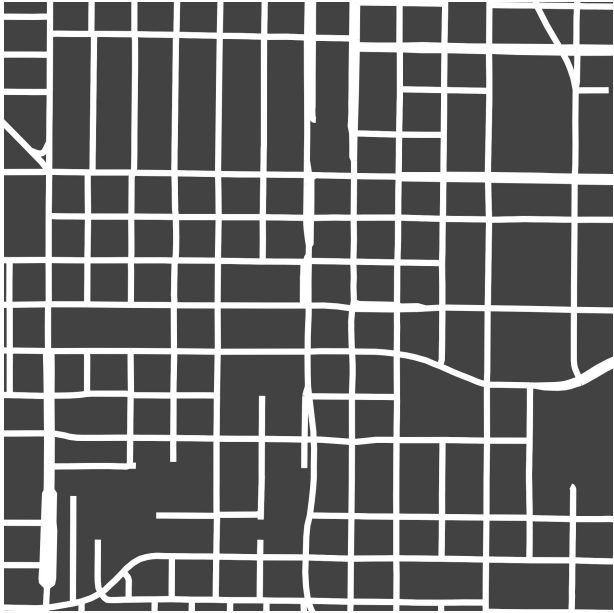
Figure 11.3: City street networks to the same scale. Graphics by Geoff Boeing for this book using Open Street Map data. The methodology is described in (Boeing 2017). Block sizes from (Nelson-Nygaard Consulting Associates, Inc. 2015).

The opening thought is a reference to American rapper Anthony Ray, who uses the stage name Sir Mix-a-lot, and had a massive hit song in 1992 with the body positive "Baby Got Back."

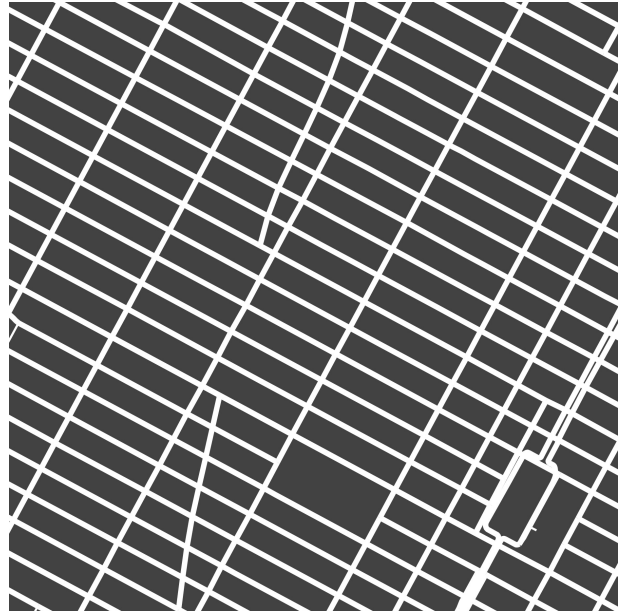
¹⁰ §10.

¹¹ §11.1.

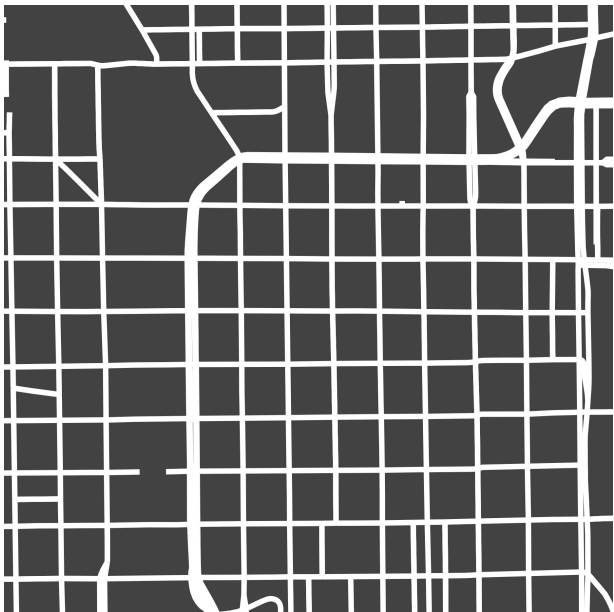
¹² Block bifurcation is more common in residential areas.



(a) Phoenix, Arizona. (300' x 300')



(b) New York City, New York. (922' x 200')



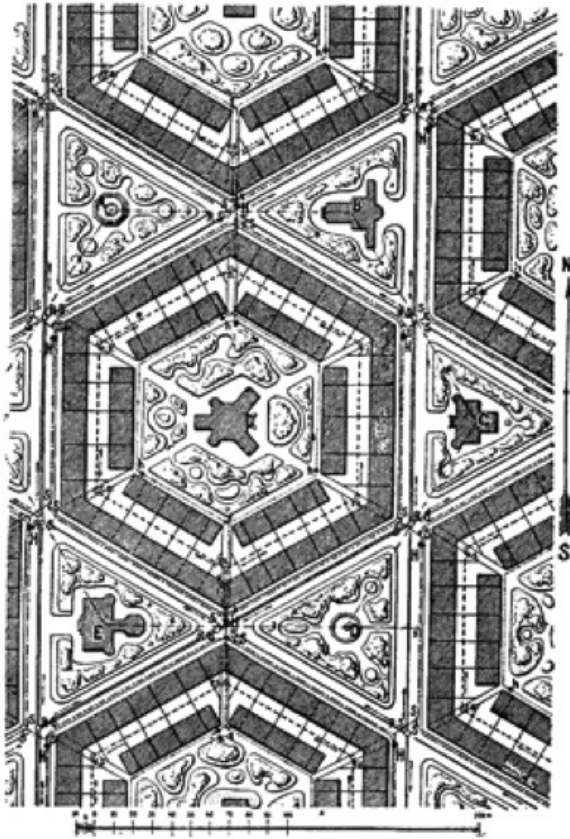
(c) Chicago, Illinois. (300' x 220')



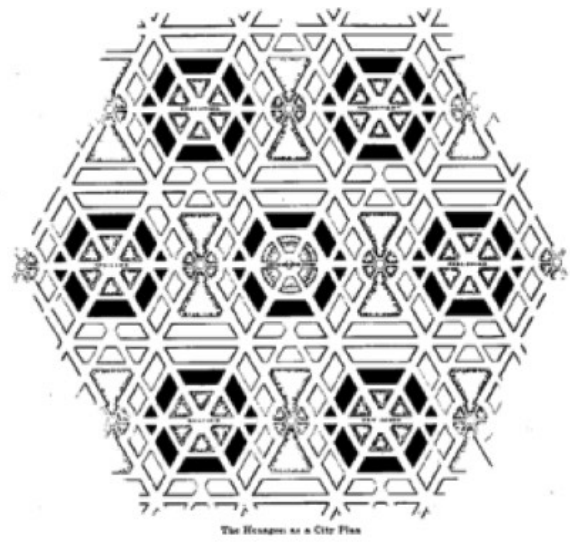
(d) Portland, Oregon. (200' x 200')

Figure 11.4: City street networks. Graphics by Geoff Boeing (Boeing 2017). Block sizes from (Nelson-Nygaard Consulting Associates, Inc. 2015).

11.3 Hex



(a) Network.
(by Mueller)



(b) Network.
(by Lamb)

Figure 11.5: Hexagonal networks.

IN CONTRAST WITH THE GRID, THERE ARE MANY DESIGNS FOR NON-RECTILINEAR STREET NETWORKS.

Most 19th and 20th century designs are simple aesthetic choices, as in Canberra, the planned capital city of Australia, and don't seem to relate to deeper urban organizational issues.¹³

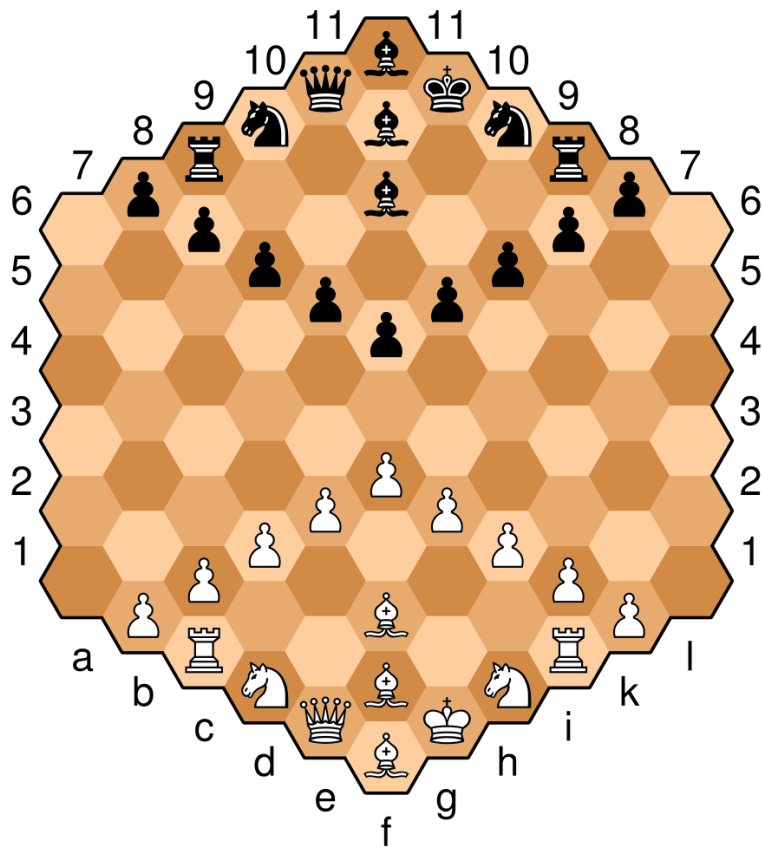
Rudolf Mueller proposed *The City of the Future: Hexagonal Building Concept for a New Division*, shown on the left in Figure 11.5. Mueller's plan offsets the 60-degree streets so that they come together in 4-way rather than 6-way intersections (though they are still at 60-degrees and not bent to make 90-degree intersections). This ensures that the cells in the plan are not bisected by roads and that they are instead hexagonal blocks. This plan loses a lot of areas to ornamental parks in the middle of streets. Charles Lamb's City Plan (right of Figure 11.5) has the streets hexsect the hexagonal cells. In this case, the blocks are really triangles.

¹³ (Ben-Joseph and Gordon 2000).

The hexagon is efficient; it replicates the closest packing of circles. (Take a penny, surround it with pennies so that they are all touching. The central penny touches six others.) Thus following the closest-packing argument, the hexagon as geometrical shape is not sufficient for efficiency; we must also arrange those shapes into an efficient pattern, in this case, something more like the Glinski Chess Board (Figure 11.6).

¹⁴ §11.1. So although we talk about 'grids',¹⁴ as being necessary for connectivity, we can get even more connectivity if we think about a variety of different geometries. No need to be square.

Figure 11.6: The Glinski chessboard: a hexagon-ish shape (technically with 66 sides) comprised of hexagons. Source: Wikipedia



11.4 Ring-Radial



Figure 11.7: Primary street network in Moscow, Russia.
Source: <http://www.flow-n.eu>.

RADIAL NETWORKS MAXIMIZE ACCESSIBILITY OF THE CENTER.

Ring-radial networks nearly maximize that accessibility while providing connections between suburbs. Figure 11.7 highlights the primary street network for Moscow, Russia. Instead of a **grid**¹⁵ or **dendritic**¹⁶ network, Moscow has a ring-radial street network. The circular bullseye routes comprise the rings while the lines converging in the center make up the radial. These circular routes are often called ring roads, perimeters, loops, or beltways (especially when they are highways).

Traveling from anywhere within the region to what would typically be the central business district – and back – is relatively efficient. Traveling from one suburb to another might not be as easy as getting downtown but is made possible by the ring roads.

Ring-radial street networks tend to be less common than ring-radial transit systems. Again, such transit network designs often work well for downtown trips, particularly commute trips. However, as cities and regions continue to grow and become more polycentric, ring-radials can lose much of their efficacy.

¹⁵ §11.1.

¹⁶ §10.7.

Part V

The Production

Accessing destinations almost always involves trade-offs. We may want a certain job, for example, but there is a limit to the amount of time and money we would be willing – or able – to spend to access it. This section on The Production essentially captures the economics of accessibility. When we talk about economics, we often talk about supply and demand, but accessibility economics means talking about issues related to induced demand and induced supply. It also means connections to our previous sections such the cost perceptions of people, the agglomeration effects of places, and the network economies of the plexus. This section on The Production ties these themes together, as they relate to accessibility, under an economic umbrella.

12

Supply and Demand

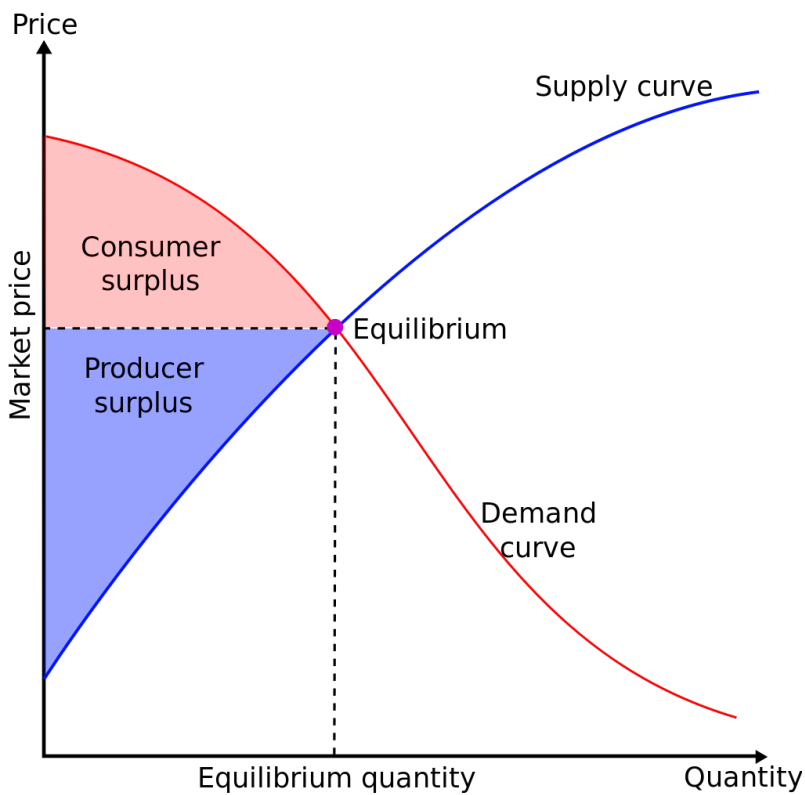


Figure 12.1: Consumer and producer surplus at equilibrium for general supply and demand curves. Source: Wikipedia. Created by User:SilverStar.

THE ECONOMIC BENEFITS OF TRAVEL AND ACTIVITY CAN BE MEASURED THROUGH THE IDEA OF SURPLUS.

Every introductory course in economics covers the topics of supply and demand. As the price falls, the quantity demanded of a good tends to rise. As the price rises, the quantity of a good supplied tends to rise. These two curves intersect at a market

equilibrium.

In microeconomic terms, we can measure total benefits as the sum of consumer and producer surplus. Consumer surplus, shown in Figure 12.1, is the difference between willingness to pay (the demand curve) and the price actually paid. If I would pay \$10 for a trip, but only actually pay \$1, I have \$9 of consumer surplus. The producer surplus (or profit) is the difference between the cost to the producer of providing a service (the supply curve) (say a bus) and the price they can charge. The area between the supply and demand curves in a typical economic graph is thus the total benefit. If we charge more than the equilibrium price, producers might get additional profit, but consumers would lose consumer surplus, and the overall benefit would drop. In a robust, idealized market, many producers and consumers come together and find the equilibrium price, which is where the marginal cost of producing the last unit of a good or service equals the marginal benefit to the consumer of purchasing it, and that point is the equilibrium price. Real markets are less idealized than the graphics of the first chapter of introductory microeconomic textbooks. There may be only one producer, who has a monopoly on production. Public goods may be subsidized and thus garner no profits at all. There are externalities¹ that are not borne by decision-makers.

¹ §12.4.

The relationship between consumer surplus, which is the benefit of actual trips, and accessibility,² which is the benefits owing to potential travel, can perhaps be squared by thinking about real estate markets.

² §1.

In real estate, the price of land depends mostly on its location with respect to other development. When pricing a house, it depends not on the travel time to one particular destination, but all potential destinations. There is a downward sloping curve, willingness to pay for real estate decreases as the number of potential destinations that can be reached declines. However unlike most markets, land is fixed, and the benefits accrue to land holders rather than land purchasers, so the accessibility benefits when assigned to land generates rent, or producer's surplus, rather than consumer's surplus.

The idea of consumer surplus, which is central to economics, was developed by civil engineers in France studying bridge pricing.³

³ (Ekelund Jr and Hébert 1999).

⁴ §12.1.

⁵ §12.2.

The ideas of induced demand⁴ and induced supply,⁵ which are debated in the transport community, are just demand and supply in economic terms. It is the failure to understand this, which was built into simple models of times past, which led to forecasts which understated traffic growth (especially in early years) as it responded to capacity, and failed to consider supply as part of the market,

instead treating it as an exogenous policy variable.

There are also important distinctions in the supply curve (and to a lesser extent demand curve) about who bears the costs. **Externalities**⁶ are costs borne by those outside the **costs**⁷ of the voluntary market transaction we normally consider, and in transport, tend to be significant and the source of many transport problems like congestion, crashes, and pollution. Scarce transport supply is often given away, leading to congestion, but it can be **rationed**⁸ or **priced**⁹ to get a better outcome.

Large infrastructure projects tend to require high initial expenditures for construction as well as an ongoing expenses for operations and maintenance, while revenues come only after the project is opened. Ensuring these costs (and the revenues) balance over time is job of engineering economics and discussed in **life-cycle costing**,¹⁰

The shapes of these cost and benefit curves are also important, and discussed in **synergies**.¹¹ It turns out the demand curve is not always downward sloping, and the cost curve not always upward sloping.

⁶ §12.4.

⁷ §12.3.

⁸ §9.6.

⁹ §9.7.

¹⁰ §12.5.

¹¹ §13.

12.1 Induced Demand

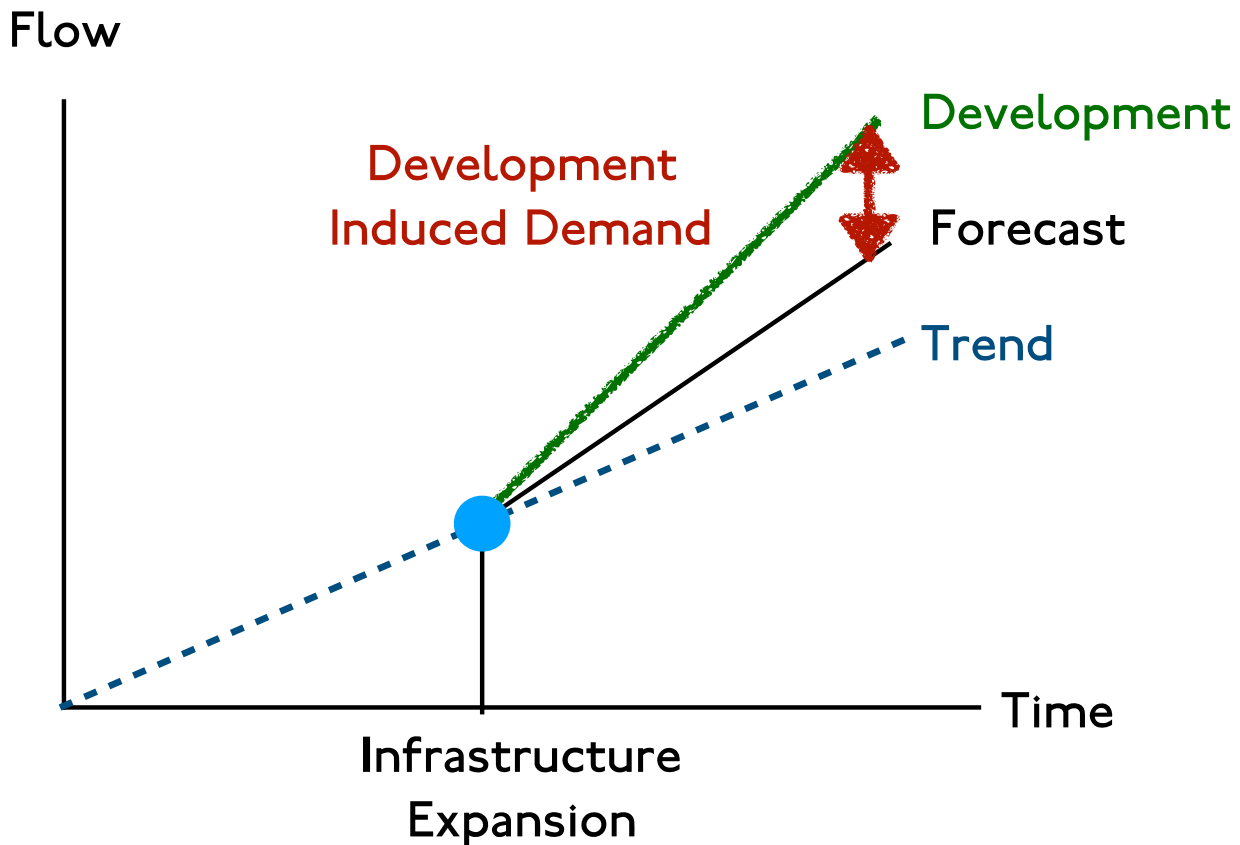


Figure 12.2: Induced demand.

INCREASING CAPACITY SELDOM REDUCES CONGESTION AS MUCH AS EXPECTED.

You already have a congested roadway, and the transport planners predict even more traffic on that road in the near future. What do you do? For most of the last century, the answer was to increase capacity. In the short-term, this seemed to work. Time and time again, over the long-term, the actual amount of traffic after the capacity increase grew far more than expected, as illustrated in Figure 12.2. What seemed like an obvious solution to a congestion problem continued to disappoint. But why?

The reason for these failures lies with the principle of induced demand. Once capacity increases, not only do you get the originally predicted traffic growth, but you also facilitate some often

unanticipated changes in travel behavior. First, existing road users might change the time of day when they travel; instead of leaving at 5 AM to beat traffic, the newly widened road entices them to leave for work with everyone else. Second, those traveling a different route might switch and drive along the newly widened option. Third, those previously using other modes such as transit, walking, bicycling, or even carpooling may now decide to carpool or drive alone instead. Together, these unwanted behavior changes fall under what is termed the theory of triple convergence (also known as the 'Iron Law of Congestion,' and which we discussed earlier in terms of [travel time budgets](#)).^{12,13} This latent demand induces more traffic than originally expected and saps the supposed improvement of the expected benefits.

¹² §2.5.

¹³ (Downs 2005).

Induced demand implies that the [trip generated](#)¹⁴ had a higher [consumer surplus](#)¹⁵ than no trip (or an alternative). So regardless of the congestion, the expansion in capacity had value, the evidence for which is the induced demand.

¹⁴ §2.1.

¹⁵ §12.

The old joke is that adding lanes to cure congestion is like loosening your belt to cure obesity. Empirical results over the last century – due to the principle of induced demand – have borne out that this issue is real and should always be accounted for when considering adding capacity as a solution to congestion.

12.2 Induced Supply & Value Capture

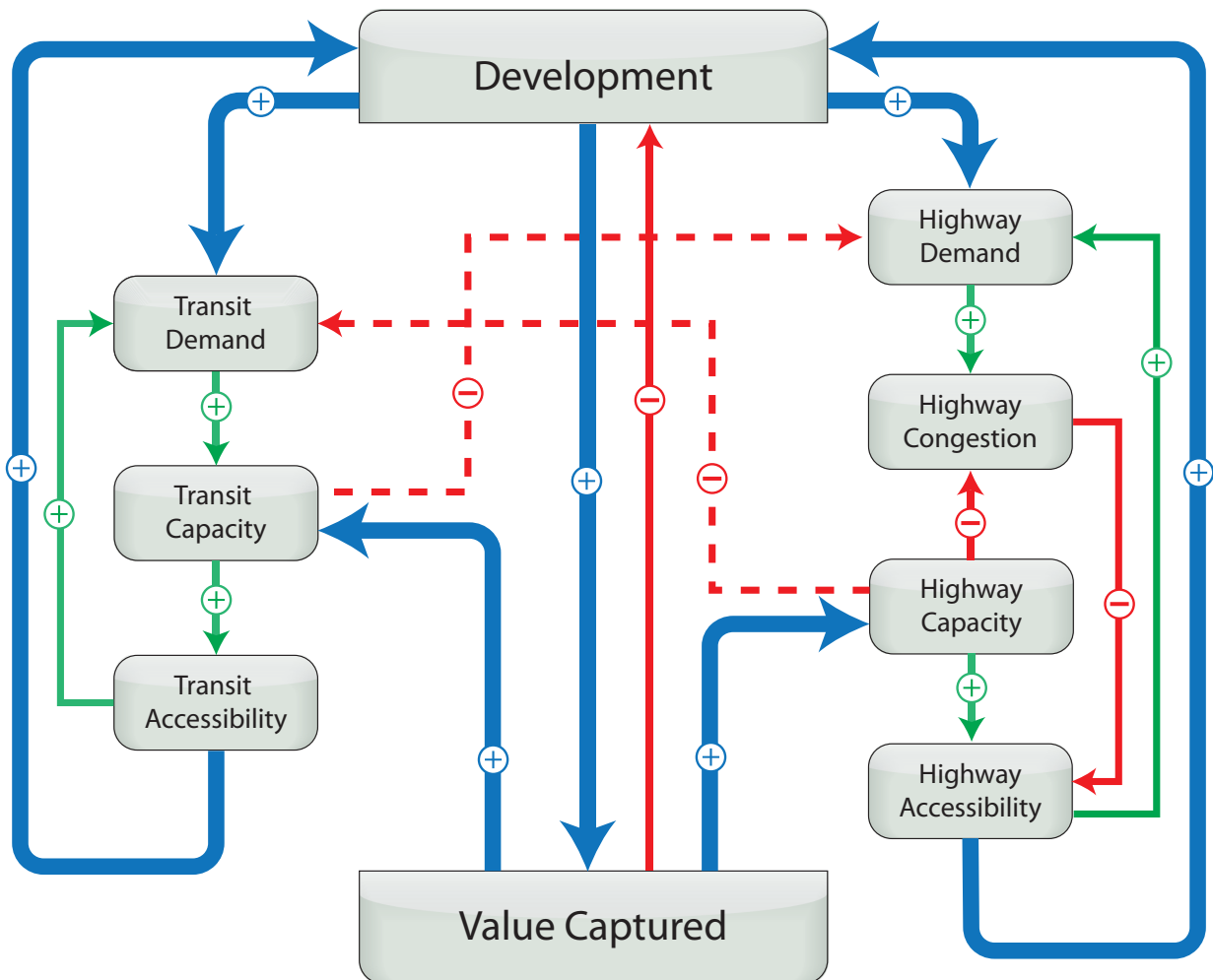


Figure 12.3: Value capture flowchart. Source: (Lari et al. 2009).

THE MIRROR PROBLEM TO INDUCED DEMAND IS INDUCED SUPPLY.

All too often, the response to congestion is to expand roads. This supply was induced by the changing demand patterns. There is a great deal of evidence that supply responds to market conditions in the transport sector, even though in the modern world, it is often mediated by political institutions.

Consider for instance the London Underground.¹⁶ Not only did new underground stations induce¹⁷ new development and

¹⁶ (Levinson 2007; 2008; Levinson et al. 2015).

¹⁷ §12.1.

encourage more people to travel longer distances, more people encouraged rail promoters to build new lines to serve them, the most promising of which were approved by Parliament. It was a positive feedback system of the kind drawn above. In this case, the value was captured by land developers, who benefitted from new stations, new lines, and the accessibility they provided. In some cases, the land developers and the transit line builders were the same people, for instance along the Metropolitan Line in London, where the Metro-Land suburbs were constructed.

This is also not just a transit phenomenon; it applies in principle equally well to the highway side. The same processes continue to occur, though for all modes there are diminishing returns to mature systems, so a route in a city with many uncongested lines is far less valuable than a route in a city with few lines or with crowded lines.¹⁸

¹⁸ (Iacono and Levinson 2016).

To the extent that the benefits from new infrastructure can be captured to pay for new infrastructure, as in Figure 12.3, more infrastructure will be created. This process, dubbed 'value capture,' has many possible mechanisms but in many cases is not implemented in the US context, where the public provides infrastructure, the benefits are captured privately, and then people complain about a lack of infrastructure.

12.3 Cost Perception

Figure 12.4: Toll Booth in Guilin, China. Photo by D. Levinson.



THE COSTS YOU PAY PER TRIP DIFFER FROM THE COSTS YOU INCUR.

If you drive, you might pay for the costs of owning your car, the costs of insurance, the cost of tolls, and the costs of fuel. Generally, the costs of the car and insurance are independent of how much you drive, while the costs of tolls and fuel vary with distance. We say the costs of the car and insurance are ‘fixed,’ as they do not change with the amount of travel, while the cost of tolls and fuel are ‘variable.’ As a result, when thinking about making one more trip, you will consider the variable costs but tend to discount or ignore entirely the fixed costs. This means the ‘marginal cost’ of each additional trip is lower than the ‘average cost’ (the total cost divided by the number of trips), and will bias you to driving more than you otherwise would.

The present system does not reward drivers when they reduce their mileage. It fails to return to the motorist much of cost savings realized from driving less. Changing the cost of driving to a method where drivers pay out-of-pocket for what were ‘fixed costs’ is actually less expensive overall and has the added advantage of meeting public sector goals of reducing [congestion](#)¹⁹ and environmental impacts, and limiting new highway construction and expensive maintenance costs.

¹⁹ §6.5.

To illustrate, let’s assume a vehicle that averages 20,000 km per year. The cost of insurance for that vehicle is \$2,000/year, registration

tags, sales tax and parking are \$1,000/year, and ownership/lease fees are \$4,000/year. Instead of the fixed cost model, convert those costs to variable costs. Thus, the costs of insurance translate to a distance-based charge of \$2,000/20,000 km or \$0.10 per km, sales tax and parking \$1,000/20,000 km to \$0.05 per km. And ownership/lease fees of \$4,000/year or \$4,000/20,000 km translate to \$0.20 per km. Adding these together would add \$0.35 per km to the cost of a trip. This would give a better signal to the traveler of how much each trip cost to them personally, and by raising the out-of-pocket price, presumably reduces the amount of travel.

This does not even consider social costs or [externalities](#),²⁰ like air pollution, carbon emissions, [crashes](#)²¹ (above and beyond what is covered by insurance), and noise, for which cost estimates vary widely but are now clearly underpriced with an effective charge of about \$0/km.

²⁰ §12.4.

²¹ §5.4.

Of course raising the monetary cost of travel reduces the perceived amount of accessibility in terms of the money plus time accessibility. It is now costlier to travel farther. On the other hand, by reducing the demand for travel, speeds should rise.

12.4 Externalities

Figure 12.5: Smog in Beijing, China. Photo by D. Levinson.



MANY COSTS OF TRAVEL ARE NOT BORNE BY THE TRAVELER.

²² §12.3.

Some costs are perceived²² by the traveler, and enter into the decision about what mode to use, what route to take, or where to go. While these may be priced imperfectly, these ‘private’ or ‘internal’ costs are borne by the traveler. Other costs are not. These are termed ‘externalities’ or ‘social’ costs. These include the congestion the traveler causes other people, the pollution they generate, and the noise they make.

²³ §5.

Congestion externalities differ from the congestion costs travelers suffer. The congestion that travelers endure is caused by the queue²³ of cars ahead of them in traffic. The congestion they cause afflicts those behind. We can measure this as the difference in the costs of travel with and without the car in question. This increases travel time for others, which is valued at some value of time.

Pollution comes from a variety of sources, as in Beijing in Figure

12.5. Vehicle air pollution is both breathed in by travelers, and generated from their tailpipe (or the power plant in the case of many electric vehicles). The amount that is generated is costlier than pollution intake because it afflicts not only other travelers, but people anywhere nearby. The cost of this in terms of more particulate matter, CO, volatile organic compounds, SO₂, NO_x, and Pb are measured in terms of the health damage. Looking at how mortality and morbidity vary with pollution levels, and then looking at value of life lets us estimate a price for this due to damages. If the cost of avoiding pollution production is less than the cost of pollution damages, than that should be done.

Carbon emissions (greenhouse gases) have an even wider effect, and a more uncertain cost of damages. The most straightforward way to estimate their price is to estimate the cost to avoid them. This is usually in terms of dollars per ton of emissions.

Noise rarely kills or injures people, but it diminishes the quality of life. This can be monetized by looking at the difference in property values with higher and lower noise costs. Houses near airports and highways are cheaper because of the noise externality. The contribution to the noise externality from each car on each link can be estimated and assigned the price based on the reduces land values nearby.

Safety is also potentially an externality, but much of it is already internalized with insurance costs, which transfers the costs of crashes from the individual to their auto insurance fund.

The exact values of these vary. Congestion clearly varies based on time of day and location, as do the others to a lesser extent. Some estimate that \$1/gallon (\$0.25/l) would cover the external cost of greenhouse gas emissions and air pollution, lower than fuel taxes in many countries, but higher than in the US.²⁴

If travelers considered their social costs, perhaps through higher taxes and tolls, they would travel less. Their effective *accessibility*²⁵ (how many opportunities they can reach on a dollar) would also decline. Today accessibility is heavily subsidized through pollution externalities (borne by society at large, particularly the health care system) and real estate costs in terms of noise (though this is a two-way street, and subsidized accessibility also props up real estate costs). A less subsidized system would help society achieve a more economically efficient amount of accessibility.

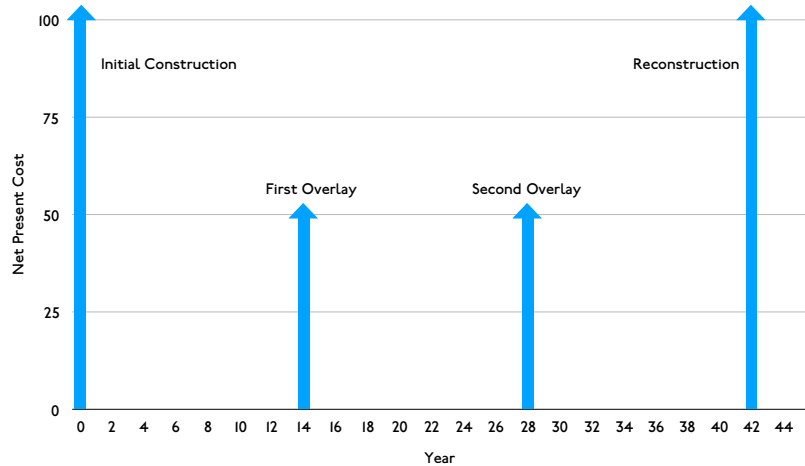
There are a variety of particulates, which vary by size: PM₁₀, PM_{2.5}, and ultrafines and especially the component called black carbon. Small particulates can embed deeper in the lungs and may be more dangerous. Think of your breathing system as a series of sieves of filters. Large particles (say rocks) don't get past your mouth, but smaller particles do. The larger of those are filtered out before they get to your lungs, but very small particles can get very deep into the lungs. Because until recently, the health effects of ultrafines were not measured, they are still not a regulated pollutant in the US. (World Health Organization 2012).

²⁴ (Knittel and Sandler 2011).

²⁵ §1.

12.5 Lifecycle Costing

Figure 12.6: Pavement overlay cashflow diagram.



CASH FLOW DIAGRAMS HELP ENGINEERS AND PLANNERS THINK ABOUT INITIAL AND DOWNSTREAM COSTS AND BENEFITS.

Arrows pointing up indicate expenditures; arrows pointing down indicate revenue (or vice versa, depending on what you are looking at). The arrows have a height indicating the magnitude of money involved and occur at a time along the timeline, the x-axis.

Figure 12.6 shows the cash flow of pavement overlays in a scenario for an existing road. In this example, asphalt roads last 42 years before they need to be completely rebuilt but require pavement overlays in the interim. Overlays are performed during a pavement life cycle to ensure the pavement maintains a certain quality (lack of maintenance would eventually result in road closure).²⁶ For a given road, initial construction occurs in year 0, the first overlay in year 14, the second in year 28, and the road is fully rebuilt in year 42.

²⁶ §10.8.

If you know the discount rate (how much tomorrow's money is worth in today's dollars), you can compare alternative scenarios, for instance using a different material (say portland cement concrete) that has a higher initial (construction) cost than bituminous asphalt but fewer future maintenance costs. It turns out that the answer often depends as much on economic factors, like the interest rate, as well as the cost of materials, as on technological factors like which pavement lasts longer.

12.6 Affordability



Figure 12.7: Waiting for the bus in Globeville, Denver, Colorado. Photo by Wes Marshall.

AFFORDABILITY REFERS TO THE ECONOMIC BURDEN THAT FAMILIES FACE WHEN CONSUMING TRANSPORT SERVICES.

More specifically, it refers to the transport costs that households pay when accessing fundamental destinations such as work, school, healthcare, as well as other necessary goods and services. When the cost of reaching such essential destinations begins to consume a relatively high percentage of a household's income, this leaves less money for housing, food, clothing, and other things.

Housing policies have long been considered under this lens. For instance, according to many policies, the threshold at which the percentage of income being dedicated to housing becomes 'unaffordable' is 30%. Based on this threshold, 55% of U.S. neighborhoods would be considered affordable for the average household.²⁷ However, overall affordability needs to include both housing and transport, and the Center for Neighborhood Technology created their Housing + Transportation Affordability Index under this premise. With respect to transport costs, the literature suggests an affordability threshold of 15%. When including transport costs, only 26% of U.S. neighborhoods are considered affordable for the typical household. There are a few caveats to this kind of analysis. The first is that household sizes vary with location, and are typically larger in the suburbs than in the center city, which affects the mix of spending, as well as absolute amounts. The second is that most of the cost of a house or car for many people satisfies wants not needs. A used car provides most of the same service that a new car does, at a fraction of the cost. Most people prefer new, and many spend money on new

²⁷ (Center for Neighborhood Technology 2010).

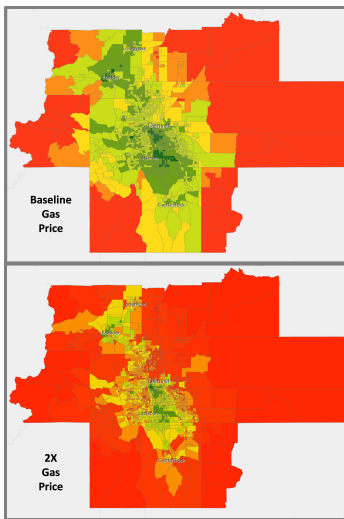


Figure 12.8: Percent of household income spent on transport to work in Denver, Colorado. Source: (Marshall and Henao 2015).

²⁸ (Marshall and Henao 2015).

²⁹ §4.7.

vehicles, but that is of choice not necessity. A third is that high income people could choose to pay a higher share of their income for luxuries (and in practice, pay a lower share), but at low incomes, even 45% for housing and transport might be too much.

In *The Shock Heard 'Round the Suburbs*, we considered the modal options available and modeled what would happen if gas prices doubled.²⁸ Given baseline gas prices, the average household in only 32 out of over 2,000 Census block groups in the Denver region would find transport costs unaffordable. Under a scenario where the gas prices doubled, this number jumped to nearly 500 block groups, as shown in Figure 12.8. In other words, the average household in the exurbs commits what would be considered an unaffordable percentage of household income to commuting in the baseline scenario. While housing is less expensive in those areas, they are also less resilient to shocks to the system, in part because there are few viable modal options available. When gas prices increase, transport costs can skyrocket for such households. On the other hand, the average household in Denver's more accessible neighborhoods might not even notice such gas price increases.

So if a family moves to the suburbs or exurbs – where accessibility is lower – in order to be able to afford the house they want, they should also consider the costs of what might now be a much longer commute. Such a move might also exacerbate the job/worker [spatial mismatch](#).²⁹ Thinking about things in the other direction: one of the best ways to improve transport affordability is to improve accessibility. The problem is that in many cities, there is an undersupply of housing in more accessible locations, and what housing there is often comes at a premium. This can push lower income families to locations where modal options are limited and distances are long, and just because a transport option exists doesn't mean that everyone has the financial wherewithal to use that mode. So when talking about affordability, accessibility should be a critical concern. And when thinking about accessibility, it is also worth considering equity of access.

13

Synergies

A synergy exists when the whole is more than the sum of its parts. One of the benefits of increasing access is not merely the increase in the number of choices, but the increase in benefit from those same choices, or the decrease in cost. This occurs in a number of different ways. This chapter explores the common transport cost reductions associated with economies of scale and scope. These economies make transport networks more efficient (less costly) than simply moving in a straight line over space. Network economies that drive processes like hubbing and intertechnology effects illustrate the potential efficiencies of consuming multiple services.

Economies of agglomeration are a related concept in urban economies, and help explain why cities are more productive than the sum of their parts.

	Increase with	Supply (Cost Reduction)	Demand (Benefit Increase)
Intra-firm	Quantity	Scale (§13.1).	Network (§13.4).
	Variety	Scope (§13.3).	Intertechnology (§13.5).
Inter-firm	Quantity and Variety	Agglomeration (§13.6).	Amenity (§13.7).

Table 13.1: Different types of economies

13.1 Economies of Scale

Figure 13.1: The World Wildlife Zoo train benefits from scale economies. Photo by L. Levinson.



Some types of economies of scale are referred to as *economies of density* or *economies of capacity utilization*, which address the utilization of a piece of infrastructure or vehicles (like buses (§8.)), rather than the provision of infrastructure or vehicles. Economies of density are pervasive in transport. The use of a largely empty road by two cars is less than twice as expensive as a road that is designed for only one car because there is a large fixed cost of building the road and a small variable cost for each user. These economies of density or capacity utilization are not inexhaustible though. When a road is congested, adding one more car spreads the construction cost of the road across another traveler but also increases the *delay* (§5) to all other travelers, which usually is more than the benefit. The train for the privately operated Arizona Wildlife World Zoo in Figure 13.1 shows a high degree of capacity utilization.

AN ORGANIZATION IS SAID TO OPERATE WITH ‘ECONOMIES OF SCALE’ IF PRODUCING TWO OUTPUTS IS LESS THAN TWICE AS EXPENSIVE AS PRODUCING ONE OUTPUT.

Defining the output is tricky. Is it cheaper per unit to operate two roads or one? Is it cheaper per unit to operate two trucks or buses or one? At this level, there is no universal answer. On the one hand, there are fixed costs, and spreading them across more units lowers costs per unit. But on the other, there are additional administrative costs associated with larger organizations, for instance more layers of management, which add significant overhead.

13.2 Containerization



Figure 13.2: Containerization is a feature of modern ports, including Vancouver, though not all freight is containerized. Photo by D. Levinson.

PUTTING THINGS IN CONTAINERS IS AS OLD AS CIVILIZATION ITSELF.

Civilization's earliest artifacts include clay pots, which were for carrying things (appropriately, pottery is among the first technologies available in the popular computer game series *Sid Meier's Civilization*). Standardizing the size of the container, so that machines could use them directly, and so they could be packed tightly, emerged with the industrial revolution. But applying this systematically and uniformly so that transport could be made more efficient is a mid-20th century innovation. Malcom McLean saw the prospect of truck trailers being placed directly on ships, eliminating the costly and laborious loading and unloading process that tied up ships in ports for weeks at a time. Soon the trailers carried containers (no need to transport the axles and wheels), and giant cranes, like the ones in Vancouver in Figure 13.2, were used rather than driving on and off the ship, greatly increasing the flow¹ of goods through the port.²

Containerization of shipping radically restructured ports, which now needed space for cranes and different types of loading and unloading processes. Important ports of the past such as London's Docklands, San Francisco, New York City were now obsolete, and new replacements were constructed at Felixstowe, Oakland, and

¹ §6.1.

² (Levinson 2016b; Garrison and Levinson 2014).

New Jersey. Old industrial areas were now repurposed for new development.

Along with making trucking that much more efficient, containerization accompanied the freeway system and helped make ³§5.5. [just-in-time](#)³ production possible.

Maturing (and deteriorating) infrastructure is coming just as globalization, the logistics revolution, and rise of containerization place additional demands on the transport system to be [reliable](#).⁴ Industry has established a just-in-time production system that relies on infrastructure. The economy demands transport systems that do not merely have a low average time but have a low variance in that time, so that the system is predictable. The industry seeks systems that can make more material at lower cost. While railroads are effective at long-haul trips, trucks can go places trains cannot.

13.3 Economies of Scope



Figure 13.3: Train station in Tokyo. Photo by D. Levinson.

AN ORGANIZATION IS SAID TO OPERATE WITH 'ECONOMIES OF SCOPE' IF THE COST OF PRODUCING TWO OUTPUTS WITH ONE ORGANIZATION IS LESS THAN THE COST OF PRODUCING EACH OUTPUT WITH TWO DIFFERENT ORGANIZATIONS.

Like [economies of scale](#),⁵ economies of scope are also everywhere in transport. We build roads that serve multiple origins and multiple destinations, rather than have a separate road for each origin-destination pair, because it is less expensive to serve multiple markets with one road.

Buses and trains serve passengers boarding and alighting at different places, because it is less expensive than having multiple but more direct trains. Train stations, like that in the opening picture, serve trains with different destinations, which is less expensive than building a separate station for each service. We have trucks carrying goods from multiple locations to multiple locations. The post office is a perfect example. Rather than each letter sender hiring their own courier, they bundle their mail with other senders, organized as the post office, to save costs.

The boundary between economies of scale vs. economies of scope may depend on the definition of markets. Is the market just the flow on the link (in which case more traffic is an economy of scale), or between the origin and destination (in which case more traffic from

⁵§13.1.

the same origin and destination is an economy of scale, but more traffic from different origins and definitions would be considered an economy of scope)? The [mobility perspective](#)⁶ tends to look more narrowly, while the [accessibility](#)⁷ perspective considers the ultimate origin and destination.

While in some respects, these differences are semantics, they do frame thinking. But the blurriness is why we often speak of 'economies of scale and scope' in a single breath.

The ideas of economies of scale and scope relate to [network effects](#),⁸ which are about the benefits accruing to the consumer rather than the lower costs for the producer, and [intertechnology effects](#),⁹ which again are about the benefits of multiple services, not the costs.

13.4 Network Economies

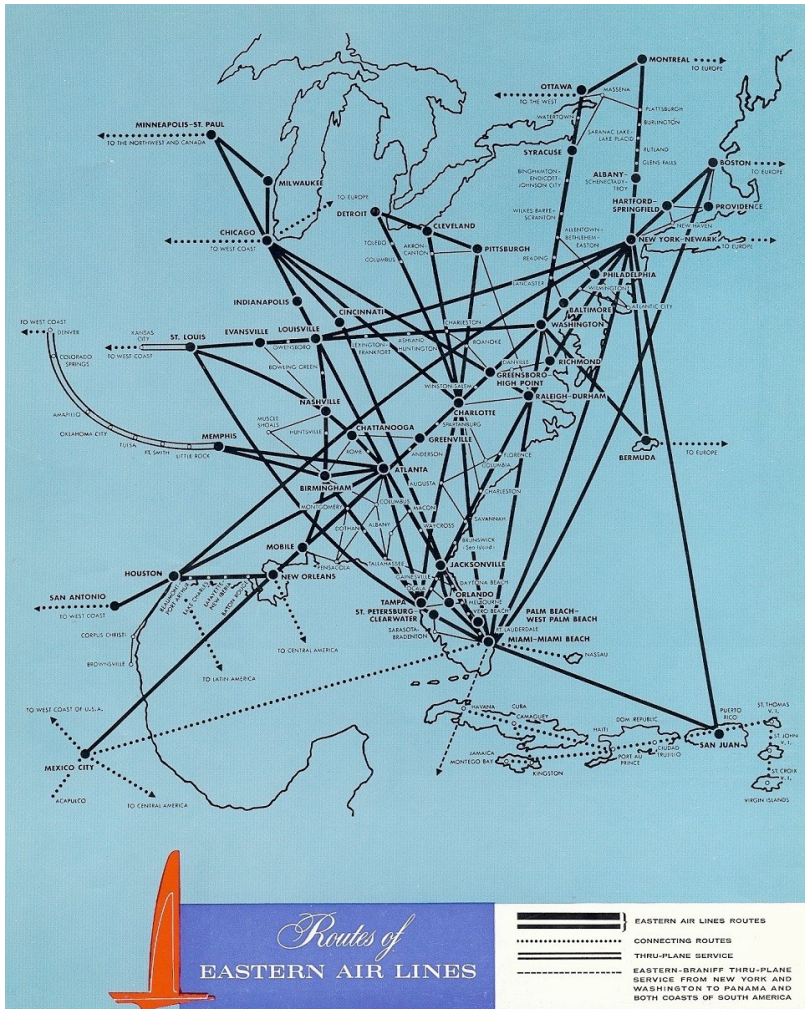


Figure 13.4: Route map, Eastern Airlines in 1961.

THE IDEA OF NETWORK EFFECTS IS THAT BENEFITS RISE WITH DEMAND.

Network effects contrast with and complement the idea of economies of scale¹⁰. That is, there are network effects when a system with more users is more valuable than one with fewer users. The other users form a network, and being part of that network produces gains.

Think about hub airports. Atlanta (ATL) was once a hub for the now defunct Eastern Airlines (and remains one for Delta). As they say in the Southeastern United States: “It doesn’t matter whether you

¹⁰ §13.1, that costs drop with demand

are going to heaven or hell, there is always a connection in Atlanta.”

Figure 13.4 shows that in 1961 there weren't enough passengers from Memphis to Miami or to Tampa to justify a direct flight. Similarly there weren't enough passengers from Nashville either. But you put the passengers from Memphis and Nashville together in Atlanta, with passengers from other origins and to other destinations, let people change planes, and everyone can get where they are going. The more passengers using ATL, the more flights there will be. The more flights, the more potential destinations and the shorter the wait between flights (there is less schedule delay).

To be clear, this differs from economies of scale or scope. Costs of operating flights do not necessarily drop because they run through a hub (they might rise with congestion), but the benefits of using the airport increase, and flights might have increased load factors, and thus more revenue (and **profit!**).¹¹ Because there are more connections, airlines will run more flights through this airport, increasing demand further.

Hubs are everywhere in transport, from public transport to seaports. The more people who want to ride public transport in my neighborhood, the more buses will serve the stop in my neighborhood. Even on a single route, this increase in frequency reduces wait time. This increases demand further. In public transport, this is called the Mohring effect, named for the famed Transport Economist Herbert Mohring.¹² The hierarchy of roads is a type of hub, with travelers moving up the hierarchy to share the less **direct**,¹³ but faster **freeway**,¹⁴ with travelers from other origins going to other destinations.

Applications in communications technology are even more obvious. How useful would a telephone or the internet be with only a few users? How useful is it with the whole world connected?

Collectively, these network effects are a form of positive feedback that is all too often ignored in transport analysis. By increasing the number of places that can be reached in less time, hubs expand access.

¹¹ §12.

¹² (Mohring 1972).

¹³ §10.9.

¹⁴ §7.1.

13.5 Intertechnology Effects



Figure 13.5: Automobile Club of Southern California (AAA) Number 1 Highway Patrol Service. Photo from (Justacarguy.blogspot.com 2011). Used with permission.

INTERTECHNOLOGY ECONOMIES EXIST IF THE BENEFITS OF CONSUMING TWO (OR MORE) TECHNOLOGIES JOINTLY EXCEED THE BENEFITS OF CONSUMING THEM SEPARATELY.¹⁵

Intertechnology effects are the demand analog to economies of scope, just as network effects are the demand analog to economies of scale.

There are many claims about various kinds of Intelligent Transportation Systems (ITS), such as highway helper or freeway service patrols (which were once operated by auto clubs and are now funded by road agencies, as in Figure 13.5), variable message signs, and ramp meters.¹⁶ While each has been shown to be valuable for travelers in terms of reducing delay, it's not clear whether working together the benefits are super-additive¹⁷ or sub-additive.¹⁸

While this is hard to measure in the field, *in silico* simulations indicate that most of the gains were obtained from the technology deployed first, successive gains were smaller, that is, benefits were sub-additive.¹⁹ For non-recurring congestion like incidents,²⁰ the benefits from freeway patrols were more than electronic signs or ramp meters. The more severe the incident, the greater the benefits from ITS.

¹⁵ Intertechnology diseconomies exist if the benefits of consuming two (or more) technologies together are lower than consuming them separately.

¹⁶ §6.8.

¹⁷ Super-additive - there are synergies and the whole is greater than the sum of the parts

¹⁸ Sub-additive - each technology saves the same travel time as the others – the sign diverts people whose time otherwise would have been saved by a faster incident response.

¹⁹ (Kanchi et al. 2002).

²⁰ §5.4.

13.6 *Economies of Agglomeration*

Figure 13.6: Skyline in Hong Kong. Photo by D. Levinson.



CITIES EXHIBIT ‘ECONOMIES OF AGGLOMERATION;’ PEOPLE AND FIRMS THAT CHOOSE TO LOCATE IN CITIES TEND TO BE MORE PRODUCTIVE THAN IF THEY DID NOT.

Economies of agglomeration are a type of inter-firm *economy of scale*²¹ or *scope*.²² There are many reasons why such economies exist, including access to ideas and a strong labor pool, as well as suppliers and customers; these are detailed in the table below showing intra-firm economies of scale and inter-firm economies of agglomeration. There are detailed in Figure 13.2. If these economies didn’t exist, there would be no economic reason for cities like Hong Kong (Figure 13.6) and no value of *accessibility*.²³ ²⁴

While we hail the benefits of agglomeration, there must be some costs (diseconomies); otherwise, since the word is out, everyone would agglomerate as quickly as possible. Diseconomies include land costs, labor costs, and *traffic congestion*.²⁵ While markets get bigger, the number of competitors increases – forcing specialization. In cities, for every increase in skilled labor available to the firm, there are other competitors who will poach its labor supply, and thus its ideas and processes – costing competitive advantage. For each increase in the number of customers, there will be other suppliers entering the market seeking to raid them.

²¹ §13.1.

²² §13.3.

²³ §1.

²⁴ Economies of agglomeration certainly vary by industry. Sectors like finance tend to locate in the largest cities; other sectors often find the costs outweigh the benefits.

²⁵ §6.5.

Type of scale economy	Example
<i>Internal</i>	
1. Pecuniary	Being able to purchase intermediate inputs at volume discounts
<i>Technological</i>	
2. Static technological	Falling average costs because of fixed costs of operating a plant
3. Dynamic technological	Learning to operate a plant more efficiently over time
<i>External or Agglomeration</i>	
<i>Localization</i>	
<i>Static</i>	
4. Shopping	Shoppers are attracted to places where there are many sellers
5. Specialization (Smith)	Outsourcing allows both the upstream input suppliers and downstream firms to profit from productivity gains because of specialization
6. Labor pooling (Marshall)	Workers with industry-specific skills are attracted to a location where there is a greater concentration.
<i>Dynamic</i>	
7. Learning-by-doing (Marshall, Arrow, Romer)	Reductions in costs that arise from repeated and continuous production activity over time and which spill over between firms in the same place
<i>Urbanization</i>	
<i>Static</i>	
8. Innovation (Jacobs)	The more that different things are done locally, the more opportunity there is for observing and adapting ideas from others
9. Labor pooling (Marshall)	Workers in an industry bring innovations to firms in other industries; similar to no. 6 above, but the benefit arises from the diversity of industries in one location.
10. Division of labor (Smith)	Similar to no. 5 above, the main difference being that the division of labor is made possible by the existence of many different buying industries in the same place
<i>Dynamic</i>	
11. Endogenous growth (Romer)	The larger the market, the higher the profit; the more attractive the location to firms, the more jobs there are; the more labor pools there, the larger the market – and so on
12. Pure agglomeration	Spreading fixed costs of infrastructure over more taxpayers; diseconomies arise from congestion and pollution

Table 13.2: Sources of agglomeration economies. Based on (Kilkenny 1998).

13.7 Economies of Amenity

Figure 13.7: Chicken restaurants in Shanghai, China. Photo by D. Levinson.



CITIES EXHIBIT ‘ECONOMIES OF AMENITY;’ PEOPLE THAT CHOOSE TO LOCATE IN CITIES TEND TO CONSUME MORE AND BETTER GOODS THAN IF THEY DID NOT.

²⁶ §13.5.

²⁷ §13.4.

Economies of amenity are a type of inter-firm *inter-technology*²⁶ or *network* economy.²⁷ There are many reasons why such economies exist, but the main is that the specialization in the market enabled by a larger number of consumers induces a greater variety of goods, and thus a better match to individual preferences. People, goods, and services achieve a better fit. There are not merely restaurants, but restaurants from a variety of countries, and not just different countries but different regions from countries. There are competing fried chicken restaurants with your choice of grey-bearded proprietors, as in Figure 13.7. The same is true for all sorts of goods and services. While a network effect is an aspect of this, it refers to consumer benefits from presence on a single network (perhaps a single firm’s network), while including access to

amenity comes from the many different people and firms providing services in ways that we don't typically think of as networks. They are like intertechnology effects, but consider services we don't think of as technologies, and for which the sub- or super-additivity is only a small aspect, the key point is the choice is additive. Entertainment, culture, parks, shops, and restaurants are some of the amenities whose value often increases with city size.

The diseconomy of amenity comes in part from what has been called the 'paradox of choice.'²⁸ More choices require more thought, consuming time and perhaps leading to increased dissatisfaction with the outcome.

²⁸ (Schwartz 2004).

Economies of amenity and agglomeration are often captured by landowners, who charge higher rents for the benefits associated with place. This in turn drives up the costs of wages.

Part VI

The Progress

The fundamentals of accessibility – in terms of the People, Places, Plexus, and Production – have entrenched themselves over generations of empirical data, but the future of accessibility is by no means static. Our last section –The Progress – delves into the dynamic coevolution of place and plexus over time. This begets fascinating discussions of accessibility as it relates to topics such as technology, biology, scaffolding, modularity, and origami. While this section on The Progress is the final section of this book, our story of accessibility is not complete. As we close the book by considering the possibilities – good, bad, and indifferent – of autonomous transport, we recognize that the future of accessibility is a story that future generations will need to continue to write. To do so, and to do so well, hopefully involves a comprehensive understanding of *The Elements of Access*.

14

Lifecycle Dynamics

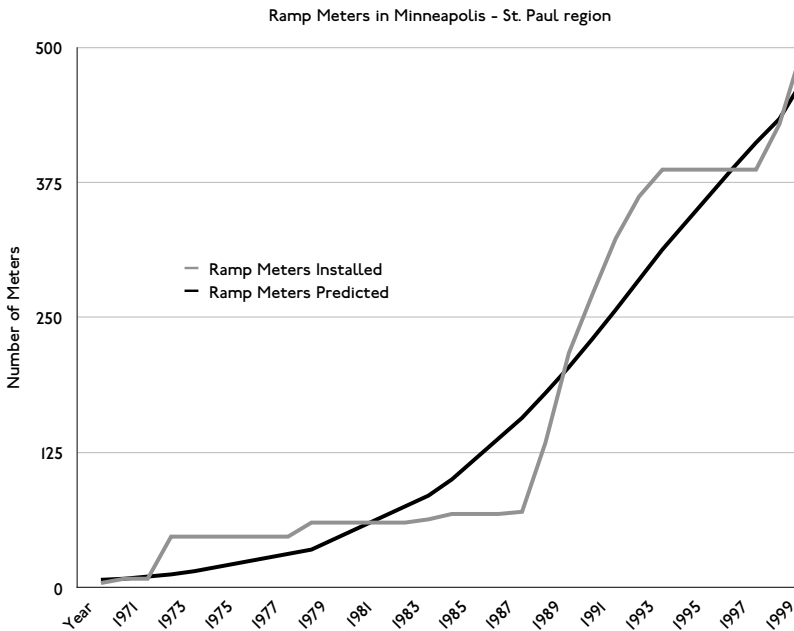


Figure 14.1: Deployment of ramp meters in the Minneapolis - St. Paul region.

LIFECYCLE THEORY TRACES OUT THE DEPLOYMENT PATH OF TECHNOLOGIES, FROM BIRTH, THROUGH GROWTH, TO MATURITY, AND THEN DECLINE.

These S-shaped curves have successfully described the deployment of many technologies. Transport networks, among the slowest technologies to deploy, may take decades to reach maturity. Railroads were first deployed in the US in 1830, they peaked in 1920, and their length has been dropping since. Freight railways continue to gain traffic, but they do so on fewer and fewer tracks every year.

Understanding this process, and where any given technology is on this curve, is important for making investment decisions. As one might imagine, the best investments are made in the early years, so there are diminishing returns setting in by maturity. Adding capacity to a mature network is often foolish. Generally forecasts in early years of an important and fast growing technology underestimate demand. In the later years of a slower growing mature technology, the forecasts overestimate demand.

Just as important as the growth stage is understanding decline and disinvestment. Many technologies see collapse eventually, public transit¹ in the US saw a massive collapse from the late 1940s through the 1980s. The post office is going through a similar trend today. As shown in Figure 14.1, ramp meters in the Twin Cities have more or less reached maturity. New technologies² may make ramp meters³ obsolete.

The reasons for decline are often related to the emergence of new technologies. While models of growth are perhaps adequate within the domain of an existing growing technology, predicting the emergence of a new technology is much harder. Old technologies have both advantages and disadvantages. Whether a new technology, however much better it would be if were deployed, can overcome the defensive moat of an existing technology's sunk costs, is very much unclear. The new technology has to not only be better if fully deployed, garnering all the network effects, but also at a much smaller scale, while it is just transforming from birth to growth stages. Investors and venture capitalists promoting the new technology network can only run losses for so long before they need to show some profits to justify continuing investments.

The difficulty remains for predicting modes that are still growing: when will they reach market saturation? For instance, how many flights will people take per year? This requires examination of fundamental factors.

One of the lessons though is what appears to be exponential growth in the early years of a technology is in fact just logistic growth. All technologies have limits. Exponential growth cannot go on forever, it is unsustainable. That which is unsustainable does not sustain.

Another lesson is that it is hard to determine the shape of the logistic curve, that is, where it levels off, from the shape of the exponential curve in the early growth period.

14.1 *Technology Substitutes for Proximity*



Figure 14.2: St. Anthony Falls, the birthplace of Minneapolis, which like many cities was founded based on a steady source of power. Photo by D. Levinson.

IN THE DARK AGE BEFORE ELECTRICITY, GREAT, IF NOT SATANIC, MILLS WERE LOCATED ADJACENT TO WATERFALLS TO PROVIDE DIRECT ENERGY.

This is the origin story of many early industrial revolution cities. The development of the electric grid – first DC then AC – untethered milling from the falls (Figure 14.2).

In the age before the streetcar, people lived within walking distance of their jobs. Downtown was very important. With the streetcar and subway, downtown remained important, as the destination of a radial commuting, shopping, or entertainment trip began farther out in suburbs. But with the automobile, not only residences, but first shops and then workplaces could become untethered from their downtown anchor at the head-end of the transit system. Downtowns in some US cities haven't added employment in many decades, and many more have lost market share to the greater metropolitan regions.

Yet the individual's daily activity pattern itself was still confined to a roughly 30-minute radius, sometimes referred to 'Marchetti's constant,' but identified by Zahavi and others earlier when studying [travel time budgets](#).⁴ This helped glue cities together.

⁴§2.5.

With forthcoming mobility technologies like [autonomous vehicles](#),⁵ this commuting budget range could theoretically expand.

⁵§15.

Thirty minutes of actively engaged traveling by driving, biking, or walking is not the same as thirty minutes as a disengaged passenger in a commuter train or autonomous vehicle. If autonomous vehicles were to actually reduce congestion, travel could be faster during the peaks (although induced demand could counter these gains). Autonomous vehicles, with shorter reaction times, could speed travel on uncongested roads as well (depending on the speed we allow them to go). All of which suggests less physical tethering between home and work and perhaps even more decentralization.

⁶ Admittedly, people have been talking about telecommunications substituting for transport for ages, but the technologies are steadily getting better (Novak 2013).

With new and better telecommunications technologies, the requirement for in-person meetings could also drop.⁶ With fewer in-person meetings, there are fewer days per week that one needs to go to work, which means a weekly commuting budget may be a more appropriate concept than a daily one. It also means off-peak travel is more likely. Non-work trips could also replace some of the time saved by reduced work trips, but they may not be as long or as peaked. Telecommunications also no longer require wires, as wireless gets more efficient. So the need to be on the wired telecom network to conduct business would not be required either.

⁷ §2.4.

Two other sources of tethering of people to their infrastructure are energy and water supply. It is worth noting that rooftop solar energy is increasingly becoming technically feasible. Without the need to attach to an electric grid (though maybe still wanting to due to [load balancing](#)⁷ – though with enough energy conversion efficiency, this doesn't matter), and with more power available on large rooftops where land is less scarce, score one more for decentralization. People have long lived with cisterns, wells, and septic systems.

To be clear, cities are getting better too. While life for the disconnected may be improving, the quality of life for highly tethered urbanites is also rising. Urban air pollution will drop as renewable energy and electric vehicles become standard, and social amenities will always be closer in terms of travel time.

Where anyone will live depends on their preferences and the opportunities available to them. The good news is advances in technology suggests more opportunities will be available. The bad news is your opportunities depend on the preferences of others. We cannot be alone in wanting to live a city of 100 million people (imagine the specialization in food, stores, and entertainment possible at that scale) and expect to be satisfied. We just need more than 99 million of our closest friends to agree.

14.2 Conurbation



Figure 14.3: Hastings, Minnesota. Photo by D. Levinson.

ENDOSYMBIOSIS IN BIOLOGY REFERS TO THE IDEA THAT ORGANELLES OF EUKARYOTIC CELLS (LIKE MITOCHONDRIA AND CHLOROPLASTS) WERE ORIGINALLY FREE-LIVING MICRO-ORGANISMS THAT COMBINED SYMBIOTICALLY (TO MUTUAL BENEFIT).

The electric self-starter is the internal combustion engine automobile example of what we might call techno-endosymbiosis.

Charles Kettering developed the electric starter, which temporarily overloaded the motor. Interestingly Kettering modeled his innovation on the self-starter with the his work on motorizing the cash register when he was an engineer at National Cash Register in Dayton, Ohio. Kettering later founded Dayton Engineering Laboratories Company (DELCO) soon acquired by General Motors. The self-starter eliminated the disease of 'Ford's Fracture' a broken arm resulting from cranking accidents.

After Kettering, the automobile become an electric system in miniature: Its generator (with the battery) was the central station, which distributed current through a network to uses like starting the car, but also for headlights, and later radios and other purposes. Surprisingly, battery makers boomed not from selling batteries to makers of EVs but from selling to makers of gasoline-powered cars containing an electric self-starter.

... The internal combustion engine adopted the battery as a self-starter, and is a technological version of this biological process [endosymbiosis]. Hybrid vehicles, which ramp up the battery so that the vehicle can travel on either electric or gasoline power, are another version of this.⁸

Techno-endosymbiosis can be seen as the gasoline powered car adopting the best feature of the electric vehicles of the early 20th century. Technologies are analogous to species in many ways. Cities

⁸ Quoted from (Garrison and Levinson 2014) (Chapter 7: Good Roads).

are not; rather, they are analogous to colonies in the insect world. The complex of technologies called cities reproduce by spawning what are appropriately called colonies, as Rome famously did this two millennia ago, with its *colonia*.

Cities, or somewhat more precisely, metropolitan areas, are not simply legal jurisdictions but have an economic definition based on economic in-flow and out-flow.⁹ Cities (commuting regions) grow spatially and incorporate formerly independent cities and towns that become subcenters through the process of conurbation. They also spawn their own *colonia*, often local, that are called suburbs.

As one builds not just intra-metropolitan but inter-metropolitan networks, one can certainly imagine multiple cities forming large *megaregions*¹⁰ with overlapping flows, particularly as networks of subcenters arise.

To illustrate the concept, consider the Twin Cities. Saint Paul and Minneapolis were once somewhat independent market areas, connected by the river and trails, but for which transport costs were expensive to go between them on a daily basis. With the advent of the horsecar, the steam railroad, then the streetcar, and finally the motorcar, interaction costs declined, and the cities were bound together as a single economic unit, even if governance remains divided to this day. But these are not the only two cities in the region ultimately forged into a single unit by urban-endosymbiosis.

There are places that existed before the Twin Cities, or before the Twin Cities became ‘The Twin Cities’ and were later incorporated, and those that were spawned by the Twin Cities.

Just among the counties in the local Metropolitan Council region, we find 7 county seats that were founded independently.

In general, the county seats (shown in Table 14.1) are on one of the major rivers (the Mississippi, Saint Croix, and Minnesota) and were founded from downstream to upstream as places continued to develop and settlers moved farther inland in search of unclaimed land and resources. Which of these County Seats (or any other early town) was to be the eventual winner (the title of which is now held by the primary city, Minneapolis) was contingent both on geography and history. Hastings (Figure 14.3), for instance, was promoted as a ‘New Chicago’ by Ignatius Donnelly until 1857. Minneapolis grew because of the power of the *Saint Anthony Falls waterfall*,¹¹ which was important only because electric grids were not yet developed. Had history been a little different, Minneapolis might be a suburb of Hastings, or farther afield, Red Wing (1853) home to 1,250 people in 1860.

At some point after the construction of intercity railroads

⁹ The U.S. Office of Management and Budget defines a set of core based statistical areas (CBSAs) throughout the country. CBSAs are delineated on the basis of a central urban area or urban cluster – in other words: a contiguous area of relatively high population density. CBSAs are composed of counties and county equivalents. The counties containing the core urban area are known as the central counties of the CBSA. Additional surrounding counties, known as outlying counties, can be included in the CBSA if these counties have strong social and economic ties to the central counties as measured by commuting and employment. Outlying counties are included in the CBSA if the employment interchange measure (total of in- and out-commuting) is 25% or more, although these numbers are estimates and exceptions are made.

¹⁰ §14.3.

Why are the Twin Cities called the Twin Cities? *InfoPlease* says:

“According to the Minnesota Historical Society, the nickname ‘Twin Cities’ originally had nothing to do with Saint Paul, the state capital of Minnesota.

The term was first applied to two settlements on either side of the Mississippi River—Saint Anthony’s Falls on the east and Minneapolis on the west—in the 1840s.

The two towns were later linked by a suspension bridge. Minneapolis was chartered as a city in 1867 and in 1872 it and Saint Anthony’s Falls were united to form one city. As a result, nearby Saint Paul assumed the nickname while it and the new Minneapolis grew during the 20th century to become Minnesota’s two biggest cities.”

¹¹ §14.1.

County	Seat	Founded	1860 Population
Anoka	Anoka	1844	1229
Carver	Chaska	1851	331
Dakota	Hastings	1833	1,642
Hennepin	Minneapolis	1838	5,809
Ramsey	Saint Paul	1838	10,401
Scott	Shakopee	1851	1,138
Washington	Stillwater	1837	2,380

Table 14.1: County seats in Minneapolis - Saint Paul region

† Author estimates.

(beginning in the 1850s), streetcars (beginning in the late 1880s), and paved state highways (from the 1920s), these semi-independent outposts, firmly attached to their location in the ground, became more and more mutually interdependent. Today, development is contiguous and people are as likely to identify with the primary city of Minneapolis, the metropolitan area, or the state as they are with their most local city or township level of government, much less their county.

There is no exact date when an independent town becomes more part of the metropolitan system and less an isolated entity. The change is a process that develops over time, not an instantaneous phase shift.

There is also no obvious threshold (25% out-commuters, 50% out-commuters, etc.) for independent vs. interdependent. Yet at some point, a town is so enmeshed in a larger urban web, it can no longer simply stand alone if its links were cut off or returned to earlier (say 1860s) levels. The Twin Cities region as a whole may not existentially need the city of Shakopee, which among other things, notably exports entertainment like racing and amusement parks to the pleasure-seekers in the rest of the region, but Shakopee does need the rest of the region. The body can cut off the hand and survive, if diminished. The hand cannot cut off the body.

to New York), usually on a less-than-daily basis, but often enough.

While transport had gotten faster over time, recently it seems to have stagnated. Some people view high-speed rail (very fast trains) as the next logical step. Others see virtual connectivity via the Internet as the next step, which leads to a more global community with worldwide interactions, rather than high speed rail (HSR). It depends very much on the context as to whether HSR is economic. [Autonomous vehicles](#)¹⁶ will emerge as well and inevitably lead to people who own such vehicles being willing to travel longer distances, as it will lower the costs of travel (since people will not need to engage in the driving task and can do other things with their time in motion).

¹⁶ §15.

The key planning problem is that land use decisions are made very locally (at the township or municipality level), while important transport decisions are made at the regional or state or national level. Yet land use decisions generate demand for streets and highways outside of the local jurisdiction that permitted them, while transport decisions affect local governments and their residents and workers. Clearly local governments are not keen to let metropolitan areas make land use decisions, or even have veto powers, and similarly cannot be responsible for regional transport decisions.

Following loosely the model of London, the United Kingdom is beginning to devolve national transport and other spending powers to 'metro mayors' who over-see so-called 'city regions.' This has been perceived as a transfer of powers away from the local to the region, rather than devolution from Westminster to the city. The devolution process has been rather ad hoc, relying on co-operation and self-selection of regions.

and the Dutch and English chose New York in large part for their capability as ports) remain important even after that technology becomes obsolete. With the logistics revolution and the new dominance of container shipping, London's shipping has moved northeast to Felixstowe as large container ships cannot easily ply the Thames, while New York City's shipping has migrated to the wide-open spaces of New Jersey.

The one-time advantages result in a set of complementary investments and inter-related decisions that take on a life of their own. Because of local trading advantages, commodities markets, banks, insurers, and other related organizations located nearby. A **critical mass**¹⁸ of those institutions felt no need to migrate just because their initial reason for being vanished. While a building is under construction, temporary framing will often be used until the more permanent structure is erected. Once the final building can stand on its own, the **falsework**¹⁹ is dismantled. In a sense, everything is falsework for what comes after.

¹⁸ §13.1.

¹⁹ §14.5.

This kind of mutual complementarity happens repeatedly in transport. Airplanes are the perfect example of mobile capital. If Amalgamated Airlines no longer wants to serve a particular city pair, the airplane can easily be redeployed elsewhere. Yet 80 years into the commercial aviation industry, airlines today serve mostly the same hubs their predecessors did on the Airmail routes of the 1930s. American Airlines is still in Dallas, United in Chicago, Delta (Northwest) in Minneapolis, and so on.²⁰ A similar example occurs with today's urban bus networks, which often are all but identical in routes to the streetcars and trams that preceded them.

²⁰ (Levinson and Xie 2011).

While very few decisions are completely irreversible, transport decisions come close. Where we place a right-of-way or an airport will explain where that facility will be decades, or possibly even centuries, from now.

A slight deviation from the efficient path to solve a short-term problem today will cost travelers time for years to come. It is important to get the design right for the longer term. Yet, getting it right from a transport efficiency perspective might be getting it wrong in terms of social costs; see the damage wrought by many urban freeways, such as the well-documented case of I-94 through the Rondo in Saint Paul.²¹

²¹ (Altshuler 1966; Cavanaugh 2006).

But a slight deviation from the path will also change what the long term is. Build a bridge 'here' rather than 'there,' and then you will adjust all of the roads feeding into the bridge to meet it 'here' (instead of 'there'). And then land will be developed along the road to 'here' to take advantage of the newly created accessibility,

properties will be platted, buildings will be built, travel and trade patterns established, and other critical dependencies will come to assume that the bridge is 'here.' At some point, say 50 years in the future, the bridge will need to be replaced. Even if 'there' was a better location than 'here' initially, after five decades of adaptation, it is quite likely that 'here' is better now. The whole may have been better were a different initial decision made, given conditions at the time.²² Given current realities, however, that path must now be foregone. Even if we didn't run freeways through cities, new cities would grow up around freeways. These edge cities are twentieth century products and built on greenfields, so there was little to no community to sever. But the freeway hardly brings both sides together.

²² In Stillwater, Minnesota, an old lift bridge was replaced by a new highway bridge outside of town. This alters the dynamic of the local economy. The gas station will see less business, but the pedestrian-oriented shops on Main Street may get more.

In transport, we say build it right the first time because there won't be a second chance. And that is true. But also remember the world will adapt to whatever we do, and we cannot let the perfect be the enemy of the good.

14.5 Urban Scaffolding



Figure 14.6: Construction at train station Arnhem, Gelderland, Netherlands. Photo by D. Levinson.

LIKE BUILDINGS UNDER CONSTRUCTION, CITIES ARE BUILT WITH SCAFFOLDING.

Remove the scaffolding and cities remain. Yet what is ‘scaffolding’ and what is ‘permanent’ is not at all clear. Yesterday’s permanent structure is today’s scaffolding.

Take for instance the deployment of streetcars in the late 1800s and early 1900s. These streetcars enabled (not coincidentally) suburbs from which their customers, resident travelers, would use on a regular basis to commute to jobs and journey to shops. Yet in the mid-20th century, these streetcars (urban scaffolding if you will) were removed (just as horsecars before them), and the city itself remained. Those streetcar suburbs still exist sans streetcars. That which enabled their construction and occupancy was eventually unnecessary and removed.

Ports were the *raison d’être* for many cities; yet in today’s era of containerization,²³ ports that failed to make the transition for whatever reason withered. The city that port enabled remains. These include such places like San Francisco, New York City, and London, which today lack significant port operations, yet have maintained or gained in status. The port scaffolding was removed, and the rest of the city was self-sufficient without. However in the absence of an initial port, those cities may never have been more

²³ §13.2.

than hamlets.

The scaffolding enables the construction of an urban web of social, economic, and technological elements that eventually becomes thick and secure enough that the initial framework can disappear without taking the system down with it. But at the time, no one envisioned the port or streetcar as temporary. They seemed quite permanent.

Past accessibility catalysts repurpose to new roles. So what today seems permanent may simply be the scaffolding for tomorrow's city.

14.6 Modularity



Figure 14.7: Just another brick in the road, Columbia, Maryland. Bricks are the archetype of modular construction. Photo by D. Levinson.

DESIGNS MIGHT BE COMPRISED OF INDIVIDUAL, DISCRETE BUILDING BLOCKS (MODULES) THAT ARE COMBINED INTO A PATTERN, OR MAY BE WHOLISTIC (UNITARY) SO THAT A SMALL PART CANNOT SIMPLY BE INTERCHANGED WITH SOMETHING SIMILAR WITHOUT BREAKING THE WHOLE DESIGN.

Most things are combinations of the two. Software has moved very much to modular architecture, and as systems become large and complex, this is a logical way of reducing complexity. On the other hand, there are advantages of integration, as evidenced by unibody construction on a car, compared with componentized body on frame.

In surface transport, we have lots of modules: vehicles and infrastructure are often separated (in elevators, they are not), bridges and roads are distinct, each link is a separate module, but you can't build half a link and expect it to function. Poured asphalt is more unitary than individual bricks (Figure 14.7).

Traffic signal engineers operate on a system designed by highway engineers and planners; consider, for instance, the traffic signal timings as a distinct element that can be optimized with everything else (lane configurations, pavement, etc.) fixed. Traffic engineers recognize that traffic signal timings affect the quality of flow upstream and downstream, and so will often [time signals](#)²⁴ as a system to optimize flow, not just at the signal, but for a corridor or a city network. This is recognition of a unitary aspect of the road network.

²⁴ §5.3.

However, this unitary nature of network logic breaks the unitary nature of a neighborhood, where we might want the signals configured for pedestrians, and we might want roads redesigned to serve local rather than regional needs.

A modular architecture, where the signal is timed independently, obviously can do no better than the systemwide performance metric of the overall system but at least enables the maximization of the quality of the neighborhood, if the appropriate local settings are chosen for the traffic signal module (at the expense of system-wide optimality on a mobility dimension). Similarly, a purely local design may wreak havoc with systemwide flow and have implications elsewhere on the network. While a module can only optimize for one master (and potentially less optimally than a unitary design), it can alternatively satisfy across multiple masters. In contradistinction, a unitary architecture must sacrifice one master for the sake of another. Modularization provides flexibility at the cost of at least one dimension of optimality. A unitary design lacks flexibility and adaptability. A new wing design will not help a unitary airframe.

The job of the designer is to understand these tradeoffs and select appropriate architectural strategies (unitary vs. modularity for particular design choices), and then design the modules as appropriate. While there is no one true path (this is not religious), there are consequences and values at play.

14.7 Network Origami

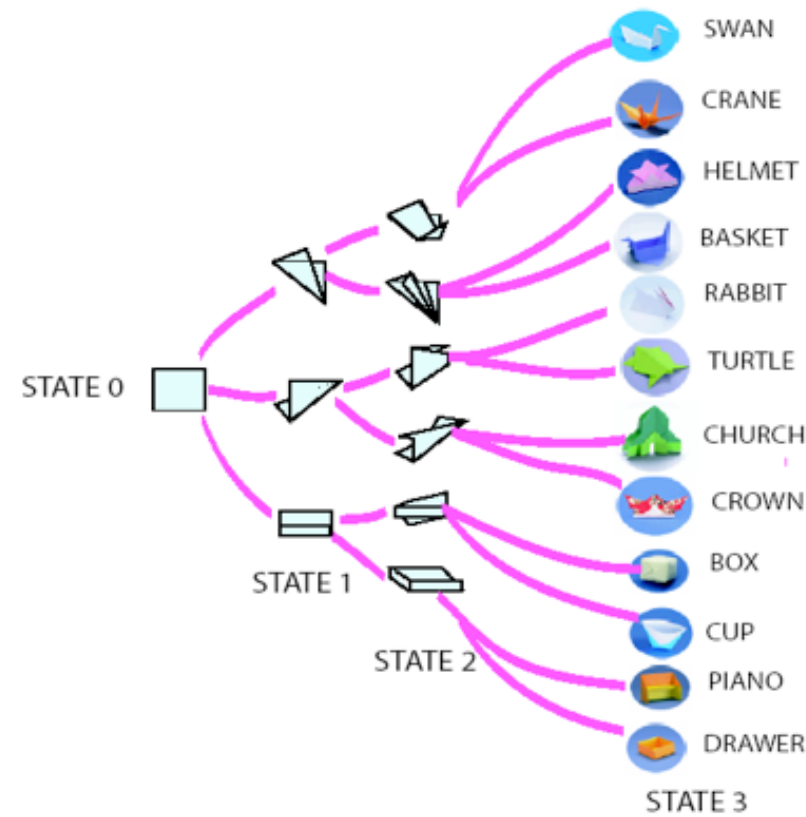


Figure 14.8: Figure: Allen's Evolution tree drawn by Wei Chen. (Allen et al. 1984).

THE DYNAMIC EVOLUTION OF URBAN SYSTEMS ANALOGIZES TO THE ART OF ORIGAMI, OR PAPER FOLDING.²⁵

Allen's paper folding story is as follows: as shown in Figure 14.8, the original state is a flat sheet of paper, then folds are made in it, and the paper gradually changes into various objects and displays different attributes. Then folds in the paper generate different traits and let the paper take various forms.

There are many choices to fold the paper at the beginning states while fewer and fewer choices remain in the ensuing states. In the 'evolution tree,' each of the objects has a past state and a future state, which both differ from its present state. In a dynamic system, if we limit the scope to a particular state and model the system in terms of the attributes present at that state, we ignore the important factors leading to system evolution, and we cannot properly capture the future changes of the system.

Allen states that the 'essence' of an object (a bird or a box) is not

²⁵ Idea for this section adapted from (Chen 2004).

only contained in the folds, but also in the order in which they are made, and the order plays a vital role in forming the object. If we describe the folds simply with their present attributes without considering the order of the formation, and then we try to predict the future style of the object based on the information we get from the current folding, we necessarily obtain the wrong conclusion. Choosing one fold (development path) requires the abandonment of other folds. Furthermore, each fold influences the emergence of the folds in the ensuing states; that is, it will increase the emergence probability of some folds while decreasing the emergence probability of some other folds. For example, folding the original flat paper along the diagonal line will increase its chance of changing into a bird but decrease the chance of changing into a box in the future. A fold is called a 'good' fold when it increases the probability that the paper evolves to the object we desire.

Now let's imagine a transport project (capacity expansion or adding new routes) as a fold of the evolution tree. Evaluating the project should not only be based on its current capability in improving traffic performance, but it also depends on how long and how well the project could help the system sustain functional operation and depends on what evolutionary direction it is leading the system to follow. Some highway projects increase flexibility and adaptability; others foreclose future opportunities. Effective highway planning that makes the system function over the long run cannot be found through myopic system optimization or ²⁶ §9.2. [equilibrium](#)²⁶ at a particular moment. Because even though we can plan the system to reach such optimization or equilibrium at some state, we cannot guarantee the system would still be optimum or in equilibrium in the next state. For an evolutionary system, system optimization or equilibrium does not necessarily exist.

14.8 *Volatility Begets Stability*



Figure 14.9: Komatsu bulldozer, Columbia, Maryland. Photo by D. Levinson.

THE HUMAN BODY EVOLVED²⁷ OVER TIME THROUGH NATURAL SELECTION.

²⁷ §14.7.

The entire body depends on various components (heart, lung, brains, etc.), and if any one of them fails, the whole fails. Thus, it made no sense to evolve a brain that would noticeably outlast the heart, or a heart that would outlast the lungs. Any effort in a longer-lived brain would be moot as the heart would fail first, and similarly, any attempt to have a heart that would beat longer than the lungs could breathe would be over-engineering. The marginal rate of return of extending the life of any critical organ would probably be equalized in such a scenario.

A similar logic has been alleged to apply to autos, with planned obsolescence. Why design a frame that outlasts the engine? The ideal, from a narrow efficiency point of view, is for all parts to fail simultaneously with no point in spending money on repairs, and no excess wasted at the outset by having parts last longer than the whole. In fast evolving technological systems, as automobiles may once again become, replacement is often more effective than repair.

Both the human body and technological artifacts like automobiles are finite systems. While the date of reckoning for bodies or cars may not be known in advance, nobody naturally lives past about 120 years of age, and intensively used cars do not economically (or

typically) last past the age of 20.

But there are other systems that are potentially infinite. These include cities and networks. While a city or network may not last forever, its potential lifespan is quite uncertain.²⁸

²⁸ We may assume that if the sun dies out and planet earth is abandoned, cities will be to, but the 5,000,000,000 year upper limit is beyond comprehension.

These seemingly infinite systems last significantly longer than their component artifacts. Just as Heraclitus said “no one ever steps into the same river twice,” one never steps into the same city twice. It is continuously evolving, as parts are abandoned, destroyed, replaced, or rebuilt. Quite often, the city while changing its buildings, maintains its networks, whose topographical and [topological](#)²⁹ structure outlasts its buildings, in part due to property ownership regulations. The example of London being rebuilt using essentially the same streets and property lines after the 1666 fire illustrates this case.

²⁹ §10.

Cities (and their networks) last longer because their components fail at different rates. If all of the components – buildings, plexus (networks, social structures) – failed at the same time (a fire plus a breakdown of the legal system ensuring property rights), then the site could be abandoned. But as long as most components last, a few failing will not destroy the city. The resources from the remaining components can help rebuild the failed ones. Similarly, [resilient](#)³⁰ networks do not fail together, and the failure of one link (given some redundancy) will not cause the network to collapse.

³⁰ §10.8.

This volatility in failure rates of components leads to a more stable whole. The price is that only piecemeal, rather than systematic overhaul of the system, is permitted.

Our Autonomous Future



Figure 15.1: Our autonomous future. ‘Uh you fellas have nothing to worry about. I’m a professional.’ Scene from *Ferris Bueller’s Day Off*.

SEVERAL TIMES IN THIS BOOK, WE HIGHLIGHT THE THEORETICAL ADVANTAGES OF AUTONOMOUS VEHICLES SUCH AS REDUCED HEADWAYS THAT DRAMATICALLY INCREASE THE CAPACITY OF EXISTING INFRASTRUCTURE AND THE ABILITY TO SAVE COUNTLESS LIVES.

We also highlight that reduced travel ‘costs’ such as the burden of driving could lead to induced demand (undermining at least some of the congestion reducing benefits of the increased [capacity](#))¹ and sparser developments (sapping advantages of [residential density](#)).²

While the fundamental technology for autonomous cars³ is close to ready, we might also want to acknowledge that we may not be quite as ready ourselves. One issue is that many of the advantages of AVs can only come with a fully autonomous fleet. The transition period may be long and perhaps even painful. One reason is that many people actually love driving (at least in some situations). Just picture the parking lot attendants (Figure 15.1) that ‘borrow’ the Ferrari in the classic 1980s coming-of-age film *Ferris Bueller’s Day Off*. Or think about the fact that the *Fast & the Furious* movie

¹ §6.9.

² §3.1.

³ Autonomous vehicles are alternatively called AVs, automated vehicles, auto autos, robo-cars, self-driving vehicles. While some authors intend there to be distinctions in the meaning of these words, most readers take them to be interchangeable.

franchise has made over \$5 billion and counting. While driving is not exactly a Constitutionally protected right in the US, the debate might parallel the gun debate in some circles. In other words, some will say: ‘you’re going to have to pull this steering wheel out of my cold, dead hands.’

There are also the important policy decisions that need to be made regarding how we use our auto autos. For one, will they be a shared resource or will we continue the personal ownership model? If the former, I can get dropped off at work while the car continues to be a workhorse for others. If the latter, would I park my autonomous car all day, or would it be able to drive around without a passenger? If so, could I send it home by itself and tell it to come pick me up later? Or could I have it circle the block a dozen times while I attend a meeting or grab dinner? If we make poor policy decisions, our autonomous future could actually lead to a marked increase in vehicle travel and congestion, despite the increased perception-reaction times and reduced headways. One hopes that poor policy choices will be quickly revised. History gives us mixed support for those hopes.

⁴ (Marshall et al. 2017).

In a recent survey of nearly 18,000 people about their scofflaw transport behavior, every single respondent admitted to breaking the law in transport in some fashion.⁴ While we aren’t all criminals, we also take it for granted that driving 5 mph (8 km/h) over the speed limit is essentially permissible even though it is illegal. Will people be able to wrap their heads around a car that actually drives the speed limit? Or will we allow our autonomous cars to bend the laws? Or will we bend the laws somehow?

When and if we do finally get to a fully autonomous fleet, there are many technical issues worth considering beyond the typical discourse about the congestion benefits of autonomous vehicles. Will they always function correctly? Given that mobile phones don’t always respond as quickly as we would like, it is difficult to imagine a device that is left out in the elements all the time being consistently instantaneous and reliable. What happens if dust or mud gets on the sensors? Will we be washing our cars daily? What about issues such as sun glare? Or even the possibility of ‘hackers’ with such connected technologies?

And how will AVs interact with pedestrians and bicyclists? It’s easy to think that we will program our vehicles not to run over pedestrians. However if we assume a fully autonomous fleet, will people be able to walk into a busy road and part traffic like the Red Sea? The simple understanding that autonomous cars will acquiesce to potential conflicts could provide pedestrians (especially

teenagers) and bicyclists – as well as the remaining drivers – free reign to ‘bully’ the autonomous cars. Should the AVs be carrying passengers, physical conflicts could arise. The AVs will undoubtedly have cameras. Will the ‘bully’ be publicly identified? Empty AVs on the other hand might just sit patiently. New laws might need to be developed, defining ‘robot harassment’ and trying to figure out what to do about it.

While this sort of thought-experiment about the cultural, political, ethical, and logistical complications of autonomous vehicles can go on indefinitely, the promise of accessibility – combined with the potential of millions of lives saved from preventable crashes – means that we should continue progressing down this path. But if we truly want to achieve the benefits of this autonomous future, we need to be mindful of the possible forks in the road.

Alphabetical List of Topics by Section

activity 2.8
agglomeration economies 13.6
amenity economies 13.7
automobile city 3.9
autonomous 15

balance 4.6
betweenness 10.4
bicycle city 3.5
bicycle network 3.6
block size 11.2
boulevards 7.2
Braess paradox 9.5
bunching 8.6

capability 2.12
choice 2.10
circuitry 10.9
clustering 10.5
complete streets 7.9
conflict points 7.7
conflicts 7.6
conservation 9.1
containerization 13.2
conurbation 14.2
coordination 2.3

cost perception 12.3

dedicated spaces 7.10
degree 10.3
deterministic queueing 5.1
directionality 7.13
distancedecay 3.2
distribution 2.6

elemental 1
equilibrium 9.2
externalities 12.4

fares 8.7
firm interactions 4.3
first mile 8.2
flow map 6.2
flow 6.1
flux 6.3
fundamental diagram 6

graph 10.1
grid 11.1

happiness 2.16
hex 11.3

hierarchy 10.2
high occupancy 6.10
highway capacity 6.9
highways 7.1

incidents 5.4
induced demand 12.1
induced supply 12.2
interaction 4.2
intertechnology effects 13.5

isochrone 1.1

junctions 7.5
just-in-time 5.5

labor markets 4.4
lanes 7.14
principle of least effort 2.11
level of service 6.5
life-cycle costing 12.5
technology lifecycle 14
linehaul 8.4

- macroscopic diagram 6.13
- markets 4
- megaregions 14.3
- meshedness 10.6
- metropolitan diagram 6.13
- mismatch 4.7
- mobility 6.5
- mode choice 8.1
- modeling people 2
- modes 8
- modularity 14.6

- neighborhood 3.4
- network economies 13.4
- networkorigami 14.7

- observation paradox 2.13

- park-and-ride 8.3
- parking capacity 7.16
- path dependence 14.4
- peaking 2.4
- pedestrian city 3.3
- platooning 5.3
- pricing 9.7

- prism 2.9
- proximity 14.1

- queues 5

- ramp metering 6.8
- rationing 9.6
- relative capacity 2.14
- reliability 9.3
- residential density 3.1
- resilience 10.8
- ring-radial 11.4
- rings 1.2
- risk compensation 2.17
- roundabouts 7.8
- routing 9

- scale economies 13.1
- schedule 2.2
- scope economies 13.3
- serendipity 4.1
- shared space 7.11
- shockwaves 6.7
- signs 7.4
- snow business 6.11
- social interactions 2.7
- space-time 2

- speed 6.6
- spontaneous priority 7.12
- stochastic queueing 5.2
- street furniture 7.3
- street types 7
- surplus 12
- synergies 13

- timetables 8.5
- topology 10
- traffic density 6.4
- transect 3
- transit capacity 8.8
- transit city 3.7
- travel time budget 2.5
- treeness 10.7
- trips 2.1

- urban scaffolding 14.5
- vertical separations 7.15
- volatility 14.8

- walkshed 3.8
- wasteful 4.5

Bibliography

AASHTO

2010. *Highway Safety Manual*, volume 529. American Association of State Highway and Transportation Officials, Washington, DC.

AASHTO

2011. *Policy on geometric design of highways and streets*. American Association of State Highway and Transportation Officials, Washington, DC.

Adams, J.

1995. *Risk*. London: UCL Press.

Aho, K.

2011. Here's how many car accidents you'll have. *Fox Business*.

Alexander, C.

1968. A city is not a tree. *Ekistics*, 139:344–348.

Allen, P., F. Boon, G. Engelen, and M. Sanglier

1984. Modeling evolving spatial choice patterns. *Applied Mathematics and Computation*, 14(1):97–129.

Altshuler, A. A.

1966. *The city planning process: A political analysis*. Cornell University Press.

Appleyard, D. and M. Lintell

1972. The environmental quality of city streets: the residents' viewpoint. *Journal of the American Institute of Planners*, 38(2):84–101.

Axelrod, R. M.

2006. *The evolution of cooperation: revised edition*. Basic books.

Axhausen, K. W. and M. Kowald

2015. *Social networks and travel behaviour*. Ashgate Publishing, Ltd.

Baco, M. E.

2009. *One-way to two-way street conversions as a preservation and*

downtown revitalization tool: the case study of Upper King Street, Charleston, South Carolina. Clemson University.

Ball, A., D. Rao, R. Haussman, and S. P. Robinson

2013. The random walking dead: Effects of urban street network topology on rates of infection spreading in zombie epidemics. Technical report, MIT.

Ballon, H.

2012. *The greatest grid: The master plan of Manhattan, 1811-2011.* Columbia University Press.

Barnes, G.

2001. Population and employment density and travel behavior in large US cities. Technical Report MN/RC – 2001-24, Minnesota Department of Transportation.

Barrington-Leigh, C. and A. Millard-Ball

2015. A century of sprawl in the United States. *Proceedings of the National Academy of Sciences*, 112(27):8244–8249.

Bavelas, A.

1948. *Some mathematical properties of psychological space.* PhD thesis, Massachusetts Institute of Technology.

BBC

2009. New bus routes under way in city. *BBC News*.

Beard, M.

2015. *SPQR: A history of ancient Rome.* Profile Books.

Beaurepaires

2017. Line haul cost breakdown.

Ben-Joseph, E. and D. Gordon

2000. Hexagonal planning in theory and practice. *Journal of Urban Design*, 5(3):237–265.

Bern, B. and W. E. Marshall

2013. Capacity analysis of pedestrian treatments at large arterial intersections and comparison with a lane-equivalent, small intersection gridded network. *Journal of Urban Planning and Development*, 139(4):241–249.

Berry, B. J. and A. Okulicz-Kozaryn

2011. An urban-rural happiness gradient. *Urban geography*, 32(6):871–883.

- Black, J.
2016. Free city bus loop service launched in canberra. *ABC News Australia*.
- Black, S.
2008. Metro Transit cuts two park and rides. *Minneapolis St. Paul Business Journal*.
- Boeing, G.
2017. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65:126–139.
- Boyce, D. E. and H. C. Williams
2015. *Forecasting Urban Travel: Past, Present and Future*. Edward Elgar Publishing.
- Braess, D.
1968. Über ein paradoxon der verkehrsplanung, unternehmensforschung 12, 258–268. *CrossRef Google Scholar*.
- Braess, D., A. Nagurney, and T. Wakolbinger
2005. On a paradox of traffic planning. *Transportation science*, 39(4):446–450.
- Bronson, R. and W. Marshall
2014. Alternative and adaptive transportation: What household factors support recovery from a drastic increase in gas price? *International Journal of Environmental Science and Technology*, 11(8):2245–2258.
- Brosnan, M. and D. M. Levinson
2015. Automobile accessibility and the allocation of time 1990-2010. *electronic International Journal of Time Use Research*, 12(1):115–132.
- Buhl, J., J. Gautrais, N. Reeves, R. Solé, S. Valverde, P. Kuntz, and G. Theraulaz
2006. Topological patterns in street networks of self-organized urban settlements. *The European Physical Journal B-Condensed Matter and Complex Systems*, 49(4):513–522.
- Carcopino, J.
1939. *Daily Life in Ancient Rome - The People and the City at the Height of the Empire*. Read Books Ltd.
- Carrion, C. and D. M. Levinson
2012a. Uncovering the influence of commuters' perception on the reliability ratio.

- Carrion, C. and D. M. Levinson
2012b. Value of travel time reliability: A review of current evidence. *Transportation research part A: policy and practice*, 46(4):720–741.
- Carroll, D.
1959. *Chicago Area Transportation Study*. Chicago Area Transportation Study.
- Carroll, L.
1917. *Through the looking glass: And what Alice found there*. Rand, McNally.
- Carvalho, R. and A. Penn
2004. Scaling and universality in the micro-structure of urban space. *Physica A: Statistical Mechanics and its Applications*, 332:539–547.
- Cavanaugh, P.
2006. Politics and freeways: Building the Twin Cities interstate system. Technical report, University of Minnesota: Center for Urban and Regional Affairs.
- Center for Neighborhood Technology
2010. H+T Affordability Index. Technical report, Center for Neighborhood Technology.
- Cervero, R.
1989. Jobs-housing balancing and regional mobility. *Journal of the American Planning Association*, 55(2):136–150.
- Cervero, R.
1996. Jobs-housing balance revisited: trends and impacts in the San Francisco Bay Area. *Journal of the American Planning Association*, 62(4):492–511.
- Cervero, R. and M. Hansen
2002. Induced travel demand and induced road investment: a simultaneous equation analysis. *Journal of Transport Economics and Policy (JTEP)*, 36(3):469–490.
- Cervero, R., J. Kang, and K. Shively
2009. From elevated freeways to surface boulevards: neighborhood and housing price impacts in San Francisco. *Journal of Urbanism*, 2(1):31–50.
- Chen, W.
2004. *Highway network evolution models*. PhD thesis, University of Minnesota.

- Chin, H.-C. and S.-T. Quek
1997. Measurement of traffic conflicts. *Safety Science*, 26(3):169–185.
- Christaller, W.
1966. *Central places in southern Germany*. Prentice-Hall.
- Clark, C.
1951. Urban population densities. *Journal of the Royal Statistical Society. Series A (General)*, 114(4):490–496.
- Coase, R. H.
1937. The nature of the firm. *economica*, 4(16):386–405.
- Cohen, J. E. and F. P. Kelly
1990. A paradox of congestion in a queuing network. *Journal of Applied Probability*, 27(3):730–734.
- Crabtree, S. and G.-H. W.-B. Index
2010. Wellbeing lower among workers with long commutes. *GallupNews Service*.
- Cropper, M. L. and P. L. Gordon
1991. Wasteful commuting: a re-examination. *Journal of Urban Economics*, 29(1):2–13.
- Dafermos, S. and A. Nagurney
1984. On some traffic equilibrium theory paradoxes. *Transportation Research Part B: Methodological*, 18(2):101–110.
- Daganzo, C. F., M. J. Cassidy, and R. L. Bertini
1999. Possible explanations of phase transitions in highway traffic. *Transportation Research Part A: Policy and Practice*, 33(5):365–379.
- Dahlgren, J.
1998. High occupancy vehicle lanes: Not always more effective than general purpose lanes. *Transportation Research Part A: Policy and Practice*, 32(2):99–114.
- Dahlgren, J.
2002. High-occupancy/toll lanes: where should they be implemented? *Transportation Research Part A: Policy and Practice*, 36(3):239–255.
- Dannenberg, A. L., R. J. Jackson, H. Frumkin, R. A. Schieber, M. Pratt, C. Kochtitzky, and H. H. Tilson
2003. The impact of community design and land-use choices on public health: a scientific research agenda. *American journal of public health*, 93(9):1500–1508.

- De Jong, P.
2012. The health impact of mandatory bicycle helmet laws. *Risk analysis*, 32(5):782–790.
- Di, X., H. X. Liu, and D. M. Levinson
2015. Multiagent route choice game for transportation engineering. *Transportation Research Record: Journal of the Transportation Research Board*, (2480):55–63.
- Dill, J. and N. McNeil
2013. Four types of cyclists? examination of typology for better understanding of bicycling behavior and potential. *Transportation Research Record: Journal of the Transportation Research Board*, (2387):129–138.
- Dobzhansky, T.
1973. The american biology teacher. *Presented at the 1972 NABT convention.*, 35:125–129.
- Domencich, T. A. and D. McFadden
1975. Urban travel demand-a behavioral analysis. Technical report.
- Downs, A.
2005. *Still stuck in traffic: coping with peak-hour traffic congestion*. Brookings Institution Press.
- Drum, C. E.
2009. Models and approaches to disability. *Disability and public health*, Pp. 27–44.
- Dumbaugh, E.
2006. Design of safe urban roadsides: An empirical analysis. *Transportation Research Record: Journal of the Transportation Research Board*, (1961):74–82.
- Dumbaugh, E. and J. Gattis
2005. Safe streets, livable streets. *Journal of the American Planning Association*, 71(3):283–300.
- Eash, R., B. N. Janson, and D. Boyce
1979. Equilibrium trip assignment: advantages and implications for practice. *Transportation Research Record*, (728).
- Ekelund Jr, R. B. and R. F. Hébert
1999. *Secret origins of modern microeconomics: Dupuit and the engineers*. University of Chicago Press.

- Euler, L.
1953. Leonhard euler and the königsberg bridges. *Scientific American*, 189(1):66–70.
- Ewing, R. and E. Dumbaugh
2009. The built environment and traffic safety: a review of empirical evidence. *Journal of Planning Literature*, 23(4):347–367.
- Fan, Y.
2012. The planners' war against spatial mismatch: Lessons learned and ways forward. *CPL bibliography*, 27(2):153–169.
- Fan, Y., R. Allen, and T. Sun
2014. Spatial mismatch in beijing, china: Implications of job accessibility for chinese low-wage workers. *Habitat International*, 44:202–210.
- Fan, Y., A. Guthrie, and D. M. Levinson
2016. Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security. *Transportation Research Part A: Policy and Practice*, 88:251–264.
- Fan, Y., A. E. Guthrie, and D. M. Levinson
2012. Impact of light rail implementation on labor market accessibility: A transportation equity perspective. *The Journal of Transport and Land Use*, 5(3):28–39.
- Ferenchak, N. and W. Marshall.
2017. Spontaneous order of pedestrian and vehicle interactions in shared spaces. (*under review*).
- Ferrer i Cancho, R. and R. V. Solé
2001. The small world of human language. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1482):2261–2265.
- Fielding, G. J. and D. B. Klein
1993. How to franchise highways. *Journal of Transport Economics and Policy*, Pp. 113–130.
- Fisk, C.
1979. More paradoxes in the equilibrium assignment problem. *Transportation Research Part B: Methodological*, 13(4):305–309.
- Fitzpatrick, K., P. J. Carlson, M. D. Wooldridge, and M. A. Brewer
2000. Design factors that affect driver speed on suburban arterials.
- Fleischer, M.
1941. *Superman: The Mechanical Monsters (cartoon)*. Paramount Pictures.

- Florip, E.
2010. Fallen tree blocks access to lake. *The Oregonian*.
- Freeman, L. C.
1978. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239.
- Garrison, W. L. and D. M. Levinson
2014. *The transportation experience: policy, planning, and deployment*. Oxford university press.
- Gayah, V. and C. Daganzo
2012. Analytical capacity comparison of one-way and two-way signalized street networks. *Transportation Research Record: Journal of the Transportation Research Board*, (2301):76–85.
- Geroliminis, N. and C. F. Daganzo
2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759–770.
- Geroliminis, N., C. F. Daganzo, et al.
2007. Macroscopic modeling of traffic in cities. In *TRB 86th annual meeting*, number 07-0413.
- Geroliminis, N., J. Haddad, and M. Ramezani
2013. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. *IEEE Transactions on Intelligent Transportation Systems*, 14(1):348–359.
- Geroliminis, N. and J. Sun
2011. Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks. *Transportation Research Part A: Policy and Practice*, 45(9):966–979.
- Giacomin, D. J. and D. M. Levinson
2015. Road network circuitry in metropolitan areas. *Environment and Planning B: Planning and Design*, 42(6):1040–1053.
- Giuliano, G. and K. A. Small
1993. Is the journey to work explained by urban structure? *Urban studies*, 30(9):1485–1500.
- Gonzales, E., C. Chavis, Y. Li, and C. F. Daganzo
2011. Multimodal transport in nairobi, kenya: Insights and recommendations with a macroscopic evidence-based model. In *Transportation Research Board 90th Annual Meeting*, number 11-3045.

- Granovetter, M. S.
1973. The strength of weak ties. *American journal of sociology*, 78(6):1360–1380.
- Gårder, P. E.
2004. The impact of speed and other variables on pedestrian safety in maine. *Accident Analysis & Prevention*, 36(4):533–542.
- Hägerstraand, T.
1970. What about people in regional science? *Papers in regional science*, 24(1):7–24.
- Hamilton, B. W.
1989. Wasteful commuting again. *Journal of political economy*, 97(6):1497–1504.
- Hamilton, B. W. and A. Röell
1982. Wasteful commuting. *Journal of political economy*, 90(5):1035–1053.
- Handy, S.
1994. Highway blues: Nothing a little accessibility can't cure. *Access Magazine*, 1(5).
- Hansen, G., N. W. Garrick, J. N. Ivan, and T. Jonsson
2007. Variation in free flow speed due to roadway type and roadside environment. In *86th annual meeting of the Transportation Research Board*. Transportation Research Board, Washington DC, USA.
- Hansen, W. G.
1959. How accessibility shapes land use. *Journal of the American Institute of planners*, 25(2):73–76.
- Hanson, S. and G. Pratt
1995. *Gender, work, and space*. Psychology Press.
- Hart, J.
1998. Converting back to two-way streets in downtown lubbock. *Institute of Transportation Engineers. ITE Journal*, 68(8):38.
- Harwood, D. W.
1990. *Effective utilization of street width on urban arterials*, number 330. Transportation Research Board.
- Hauer, E. and P. Garder
1986. Research into the validity of the traffic conflicts technique. *Accident Analysis & Prevention*, 18(6):471–481.

Hawkins, G.

1992a. Evolution of the MUTCD: Early editions of the MUTCD. *ITE Journal*, Vol. August, Pp. 17–23.

Hawkins, G.

1992b. Evolution of the MUTCD: Early standards for traffic control devices. *ITE Journal*, Pp. 23–26.

Hawkins, G.

1992c. Evolution of the MUTCD: The MUTCD since World War II. *ITE Journal*, Pp. 17–23.

Hawkins, G.

1994. New developments with the MUTCD. *ITE Journal*, Pp. 16–21.

Hirschman, A. O.

1970. *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*, volume 25. Harvard university press.

Hocherman, I., A. Hakkert, and J. Bar-Ziv

1990. Safety of one-way urban street. *Transportation Research Record*, 1270:22–27.

Huang, A. and D. M. Levinson

2010. The effects of daylight saving time on vehicle crashes in minnesota. *Journal of safety research*, 41(6):513–520.

Iacono, M. and D. M. Levinson

2016. Mutual causality in road network growth and economic development. *Transport Policy*, 45:209–217.

Institute of Transportation Engineers

1999. *Traffic Engineering Handbook*. Washington, DC: Institute of Transportation Engineers.

Institute of Transportation Engineers

2004. *Trip Generation Handbook: An ITE Recommended Practice*, 2nd ed. edition. Washington, DC: Institute of Transportation Engineers.

Isard, W.

1956. Location and space-economy.

Ivan, J. N., N. W. Garrick, and G. Hanson

2009. Designing roads that guide drivers to choose safer speeds. Technical report, Connecticut Transportation Institute, Connecticut Department of Transportation.

- Jacobs, A. B., E. Macdonald, and Y. Rofé
2002. *The boulevard book: History, evolution, design of multiway boulevards*. Mit Press.
- Janson, M. and D. M. Levinson
2014. Hot or not: Driver elasticity to price on the mnpass hot lanes. *Research in Transportation Economics*, 44:21–32.
- Jayakrishnan, R., H. S. Mahmassani, and T.-Y. Hu
1994. An evaluation tool for advanced traffic information and management systems in urban networks. *Transportation Research Part C: Emerging Technologies*, 2(3):129–147.
- Jeong, H., B. Tombor, R. Albert, Z. Oltvai, and A. Barabasi
2000. The large-scale organization of metabolic networks. *Nature*, 407(6804):651.
- Jiang, B.
2007. A topological pattern of urban street networks: universality and peculiarity. *Physica A: Statistical Mechanics and its Applications*, 384(2):647–655.
- Jiang, B.
2009. Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science*, 23(8):1033–1048.
- Jiang, B. and C. Claramunt
2004. Topological analysis of urban street networks. *Environment and Planning B: Planning and design*, 31(1):151–162.
- Justacarguy.blogspot.com
2011.
- Kain, J. F.
1968. Housing segregation, negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics*, 82(2):175–197.
- Kain, J. F.
1992. The spatial mismatch hypothesis: three decades later. *Housing policy debate*, 3(2):371–460.
- Kain, J. F.
2004. A pioneer's perspective on the spatial mismatch literature. *Urban studies*, 41(1):7–32.
- Kanchi, S., D. M. Levinson, and D. Gillen
2002. Intertechnology effects in intelligent transportation systems.

Transportation Research Record: Journal of the Transportation Research Board, (1800):1–5.

Keating, P.

2001. The danger of safer equipment. *ESPN The Magazine*.

Kerner, B. S. and H. Rehborn

1996. Experimental properties of complexity in traffic flow. *Physical Review E*, 53(5):R4275.

Khattak, A. and K. Knapp

2001. Interstate highway crash injuries during winter snow and nonsnow events. *Transportation Research Record: Journal of the Transportation Research Board*, (1746):30–36.

Kilkenny, M.

1998. Economies of scale. In *Lecture for Economics 376. Rural, Urban, and Regional Economics*. Iowa State University, Ames Iowa.

Klein, D. B.

2006. Rinkonomics: a window on spontaneous order. *Economic Affairs*, 26(4):64–67.

Knittel, C. R. and R. Sandler

2011. Cleaning the bathwater with the baby: The health co-benefits of carbon pricing in transportation. Technical report, National Bureau of Economic Research.

Knödel, W.

1969. Graphentheoretische methoden und ihre anwendung.

Knoop, V., S. P. Hoogendoorn, and H. van Zuylen

2009. Empirical differences between time mean speed and space mean speed. In *Traffic and Granular Flow'07*, Pp. 351–356. Springer.

Kolata, G.

1990. What if they closed 42d street and nobody noticed. *New York Times*, 25:38.

Kowald, M., P. van den Berg, A. Frei, J.-A. Carrasco, T. Arentze, K. Axhausen, D. Mok, H. Timmermans, and B. Wellman

2013. Distance patterns of personal networks in four countries: a comparative study. *Journal of Transport Geography*, 31:236–248.

Krackhardt, D.

1990. Assessing the political landscape: Structure, cognition, and power in organizations. *Administrative science quarterly*, Pp. 342–369.

- Kyte, M., Z. Khatib, P. Shannon, and F. Kitchener
2001. Effect of weather on free-flow speed. *Transportation research record: Journal of the transportation research board*, (1776):60–68.
- Lagune-Reutler, M., A. Guthrie, Y. Fan, and D. Levinson
2016. Transit stop environments and waiting time perception: Impacts of trees, traffic exposure, and polluted air. *Transportation Research Record: Journal of the Transportation Research Board*, (2543):82–90.
- Lari, A., D. M. Levinson, Z. J. Zhao, M. Iacono, S. Aultman, D. Vardhan, J. Junge, K. Larson, and M. Scharenbroich
2009. Value capture for transportation finance: Technical research report.
- Lenntorp, B.
1977. Paths in space-time environments: A time-geographic study of movement possibilities of individuals. *Environment and Planning A*, 9(8):961–972.
- Levinson, D.
2005. Micro-foundations of congestion and pricing: A game theory perspective. *Transportation Research Part A: Policy and Practice*, 39(7):691–704.
- Levinson, D.
2013. Access across america: Auto 2013. Technical report, Center for Transportation Studies, University of Minnesota.
- Levinson, D. and P. Rafferty
2004. Delayer pays principle examining congestion pricing with compensation. *International Journal of Transport Economics/Rivista internazionale di economia dei trasporti*, Pp. 295–311.
- Levinson, D. and F. Xie
2011. Does first last? the existence and extent of first mover advantages on spatial networks. *Journal of Transport and Land Use*, 4(2):47–69.
- Levinson, D. M.
2007. Density and dispersion: the co-development of land use and rail in London. *Journal of Economic Geography*, 8(1):55–77.
- Levinson, D. M.
2008. The orderliness hypothesis: The correlation of rail and housing development in London. *The Journal of Transport History*, 29(1):98–114.

- Levinson, D. M.
2012. Network structure and city size. *PloS one*, 7(1):e29721.
- Levinson, D. M.
2016a. *Spontaneous Access*. Network Design Lab.
- Levinson, D. M. and A. El-Geneidy
2009. The minimum circuitry frontier and the journey to work. *Regional science and urban economics*, 39(6):732–738.
- Levinson, D. M., D. Giacomini, and A. Badsey-Ellis
2015. Accessibility and the choice of network investments in the london underground. *Journal of Transport and Land Use*, 9(1).
- Levinson, D. M., K. Harder, J. Bloomfield, and K. Carlson
2006. Waiting tolerance: ramp delay vs. freeway congestion. *Transportation Research Part F: Traffic Psychology and Behaviour*, 9(1):1–13.
- Levinson, D. M. and K. J. Krizek
2007. *Planning for place and plexus: Metropolitan land use and transport*. Routledge.
- Levinson, D. M. and K. J. Krizek
2017. *Metropolitan land use and transport: Planning for place and plexus*, second edition. Routledge.
- Levinson, D. M. and A. Kumar
1994. The rational locator: why travel times have remained stable. *Journal of the american planning association*, 60(3):319–332.
- Levinson, D. M., B. Marion, A. Owen, and M. Cui
2017. The city is flatter: Changing patterns of job and labor access. *Cities*, 60:124–138.
- Levinson, D. M. and L. Zhang
2006. Ramp meters on trial: Evidence from the twin cities metering holiday. *Transportation Research Part A: Policy and Practice*, 40(10):810–828.
- Levinson, M.
2016b. *The box: how the shipping container made the world smaller and the world economy bigger*. Princeton University Press.
- Madison, L. and S. Boxer
2012-01-09. Mitt Romney: “I like being able to fire people” for bad service. *CBS News*.

- Mairs, J.
2017. Exhibition road accident shows it's 'time to review shared space' says Labour MP Emma Dent Coad. *Dezeen*.
- Marchetti, C.
1994. Anthropological invariants in travel behavior. *Technological forecasting and social change*, 47(1):75–88.
- Marshall, A.
2017. Inside the 'Tunnel' Elon Musk Is Already Digging Under Los Angeles. *Wired Magazine*.
- Marshall, J. D.
2007. Urban land area and population growth: a new scaling relationship for metropolitan expansion. *Urban Studies*, 44(10):1889–1904.
- Marshall, W. and N. Garrick
2009. The shape of sustainable street networks for neighborhoods and cities. In *Congress for the New Urbanism XVII, Denver, CO*.
- Marshall, W. and N. Garrick
2010a. Effect of street network design on walking and biking. *Transportation Research Record: Journal of the Transportation Research Board*, (2198):103–115.
- Marshall, W., N. Garrick, and G. Hansen
2008. Reassessing on-street parking. *Transportation Research Record: Journal of the Transportation Research Board*, (2046):45–52.
- Marshall, W. E. and N. W. Garrick
2010b. Street network types and road safety: A study of 24 California cities. *Urban Design International*, 15(3):133–147.
- Marshall, W. E. and N. W. Garrick
2011. Does street network design affect traffic safety? *Accident Analysis & Prevention*, 43(3):769–781.
- Marshall, W. E. and N. W. Garrick
2012. Community design and how much we drive. *Journal of Transport and Land Use*, 5(2).
- Marshall, W. E. and A. Henao
2015. The shock heard round the suburbs: Assessing the vulnerability, resilience, and transportation affordability of higher fuel price scenarios. *Transportation Research Record: Journal of the Transportation Research Board*, (2531):63–75.

- Marshall, W. E., D. Piatkowski, and A. Johnson
2017. Scofflaw bicycling: Illegal but rational. *Journal of Transport and Land Use*, 10(1):805–836.
- Martens, M., S. Compte, and N. A. Kaptein
1997. The effects of road design on speed behaviour: a literature review.
- Mekuria, M. C., P. G. Furth, and H. Nixon
2012. Low-stress bicycling and network connectivity. Technical Report Final Report CA-MTI-12-1005, Mineta Transportation Institute.
- Metz, D.
2008. The myth of travel time saving. *Transport reviews*, 28(3):321–336.
- Migletz, D., W. Glauz, and K. Bauer
1985. Relationships between traffic conflicts and accidents. us department of transportation, federal highway administration. Technical report, FHWA/RD-84/042.
- Milton, J. C. and F. L. Mannering
1998. The relationship between highway geometrics, traffic related elements, and motor vehicle accidents. *Transportation*, 25:395–413.
- Mohring, H.
1972. Optimization and scale economies in urban bus transportation. *The American Economic Review*, 62(4):591–604.
- Mokhtarian, P. L. and C. Chen
2004. Ttb or not ttb, that is the question: a review and analysis of the empirical literature on travel time (and money) budgets. *Transportation Research Part A: Policy and Practice*, 38(9):643–675.
- Montoya, J. M. and R. V. Solé
2002. Small world patterns in food webs. *Journal of theoretical biology*, 214(3):405–412.
- Mooney, J. D.
1969. Housing segregation, negro employment and metropolitan decentralization: An alternative perspective. *The Quarterly Journal of Economics*, 83(2):299–311.
- Murchland, J. D.
1970. Braess's paradox of traffic flow. *Transportation Research*, 4(4):391–394.

- Murphy, B., D. M. Levinson, and A. Owen
2017. Evaluating the safety in numbers effect for pedestrians at urban intersections. *Accident Analysis & Prevention*, 106:181–190.
- Mustafa, Z. and M. Birdsall
2014. The great streets movement: Identifying how to make our streets great. *ITE Journal*, 84(3).
- Nader, R.
1972. *Unsafe at Any Speed: The Designed-in Dangers of the American Automobile*. [Expanded Ed.]. Grossman.
- Nash, J.
1951. Non-cooperative games. *Annals of mathematics*, Pp. 286–295.
- Nash, J. F. et al.
1950. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49.
- National Association of City Transportation Officials
2013. *Urban Street Design Guide*. Island Press.
- National Safety Council
2016. Injury facts chart.
- National Travel Survey 2014
2015. Statistical release. Technical report, United Kingdom.
- Neal, Z. P.
2012. *The connected city: How networks are shaping the modern metropolis*. Routledge.
- Nelson, G. D. and A. Rae
2016. An economic geography of the United States: from commutes to megaregions. *PloS one*, 11(11):e0166083.
- Nelson-Nygaard Consulting Associates, Inc.
2015. Salt lake city transit master plan (p.24). Technical report.
- Newman, M. E.
2003. The structure and function of complex networks. *SIAM review*, 45(2):167–256.
- Noland, R. B.
2003. Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. *Accident Analysis & Prevention*, 35(4):599–611.

- Nottingham City Transport
2017. Catching your bus in the city centre (city centre bus map). Technical report, Nottingham City Transport.
- Novak, M.
2013. A brief history of the videophone that almost was. *Gizmodo*.
- O'Hare, S. J., R. D. Connors, and D. P. Watling
2016. Mechanisms that govern how the price of anarchy varies with travel demand. *Transportation Research Part B: Methodological*, 84:55–80.
- O'Regan, K. M. and J. M. Quigley
1998. Cars for the poor. *Access magazine*, 1(12).
- Owen, A. and D. M. Levinson
2015. Modeling the commute mode share of transit using continuous accessibility to jobs. *Transportation Research Part A: Policy and Practice*, 74:110–122.
- Parthasarathi, P., H. Hochmair, and D. M. Levinson
2015. Street network structure and household activity spaces. *Urban Studies*, 52(6):1090–1112.
- Parthasarathi, P., D. Levinson, and H. Hochmair
2013. Network structure and travel time perception. *PloS one*, 8(10):e77718.
- Parthasarathi, P. and D. M. Levinson
2012. Network structure and the journey to work: An intra-metropolitan analysis. *Working paper*.
- Pas, E. I. and S. L. Principio
1997. Braess' paradox: Some new insights. *Transportation Research Part B: Methodological*, 31(3):265–276.
- Perkins, S. R. and J. L. Harris
1968. Traffic conflict characteristics-accident potential at intersections. *Highway Research Record*, (225).
- Porta, S., P. Crucitti, and V. Latora
2006. The network analysis of urban streets: a primal approach. *Environment and Planning B: planning and design*, 33(5):705–725.
- Porter, M. E.
2011. *Competitive advantage of nations: creating and sustaining superior performance*. Simon and Schuster.

- Potter, J. M.
2011. Estimating the offsetting effects of driver behavior in response to safety regulation: The case of formula one racing. *Journal of Quantitative Analysis in Sports*, 7(3).
- Potts, I., D. Harwood, and K. Richard
2007. Relationship of lane width to safety on urban and suburban arterials. *Transportation Research Record: Journal of the Transportation Research Board*, (2023):63–82.
- Pushkarev, B. and J. M. Zupan
1977. *Public transportation and land use policy*. Indiana Univ Pr.
- Raphael, S.
1998. The spatial mismatch hypothesis and black youth joblessness: evidence from the San Francisco Bay Area. *Journal of Urban Economics*, 43(1):79–111.
- Rapoport, A., T. Kugler, S. Dugar, and E. J. Gisches
2009. Choice of routes in congested traffic networks: Experimental tests of the braess paradox. *Games and Economic Behavior*, 65(2):538–571.
- 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.
- Regional Plan Association
1929. *Regional Plan of New York and Its Environs*, volume 2. Regional Plan Association.
- Renault, M.
2014. Inventor says synchronized lights could boost Green Line travel time. *Minnesota Public Radio*.
- Riggs, W. and J. Gilderbloom
2016. Two-way street conversion: evidence of increased livability in louisville. *Journal of Planning Education and Research*, 36(1):105–118.
- Roth, C., S. M. Kang, M. Batty, and M. Barthelemy
2012. A long-time limit for world subway networks. *Journal of The Royal Society Interface*, P. rsif20120259.
- Roughgarden, T.
2005. *Selfish routing and the price of anarchy*, volume 174. MIT press Cambridge.

- San Francisco Planning Department
2016. No great "LOS". Technical report.
- Schafer, A. and D. G. Victor
2000. The future mobility of the world population. *Transportation Research Part A: Policy and Practice*, 34(3):171–205.
- Schoner, J. E. and D. M. Levinson
2014. The missing link: Bicycle infrastructure networks and ridership in 74 us cities. *Transportation*, 41(6):1187–1204.
- Schwartz, B.
2004. The paradox of choice: Why less is more. *New York: Ecco*.
- Scott, A. J.
1988. Metropolis. *From the division of labour to urban form*. Berkeley.
- Sears, B.
2014. Incorporating complete streets into transportation master plans. *ITE Journal*, 84(4):32.
- Small, K. A. and S. Song
1992. "wasteful" commuting: A resolution. *Journal of political economy*, 100(4):888–898.
- Steinberg, R. and W. I. Zangwill
1983. The prevalence of braess' paradox. *Transportation Science*, 17(3):301–318.
- Storey, G.
1997. The population of ancient Rome. *Antiquity*, 71(274):966–978.
- Storper, M.
1992. The limits to globalization: technology districts and international trade. *Economic geography*, 68(1):60–93.
- Sudakov, B. and V. H. Vu
2008. Local resilience of graphs. *Random Structures & Algorithms*, 33(4):409–433.
- Sugiyama, Y., M. Fukui, M. Kikuchi, K. Hasebe, A. Nakayama, K. Nishinari, S.-i. Tadaki, and S. Yukawa
2008. Traffic jams without bottlenecks—experimental evidence for the physical mechanism of the formation of a jam. *New journal of physics*, 10(3):033001.
- Svensson, r.
1992. Vidareutveckling och validering av den svenska konflikttekniken. *Further Development and Validation of the Swedish*

Traffic Conflicts Technique., Department of Traffic Planning and Engineering, Lund University.

Swaminathan, N.

2007. Strange but true: Helmets attract cars to cyclists. *Scientific American*.

Taylor, B. D.

1991. Unjust equity: An examination of california's transportation development act. *University of California Transportation Center*.

Tilahun, N. and Y. Fan

2014. Transit and job accessibility: an empirical study of access to competitive clusters and regional growth strategies for enhancing transit accessibility. *Transport Policy*, 33:17–25.

Tilahun, N. and D. M. Levinson

2011. Work and home location: Possible role of social networks. *Transportation Research Part A: Policy and Practice*, 45(4):323–331.

Tilahun, N. and D. M. Levinson

2013. An agent-based model of origin destination estimation (abode). *The Journal of Transport and Land Use*, 6(1):73–88.

Tindale, S., D. Noyce, R. Seyfried, et al.

2004. Toolbox on intersection safety and design.

Tomko, M., S. Winter, and C. Claramunt

2008. Experiential hierarchies of streets. *Computers, Environment and Urban Systems*, 32(1):41–52.

Transportation Research Board

2010. Highway Capacity Manual 2010. Technical report, National Research Council, Washington, DC.

Treiber, M., A. Kesting, and D. Helbing

2010. Three-phase traffic theory and two-phase models with a fundamental diagram in the light of empirical stylized facts. *Transportation Research Part B: Methodological*, 44(8):983–1000.

Truong, L. and W. Marshall

2014. Are park-and-rides saving the environment or just saving parking costs? case study of denver, colorado, light rail system. *Transportation Research Record: Journal of the Transportation Research Board*, (2419):109–117.

US Department of Transportation

2017. Reducing Non-Recurring Congestion. Technical report, Federal Highway Administration: Office of Operations.

- Vanderbilt, T.
2008. *Traffic*. Vintage.
- Velasco, D.
2002. Boulevards and parkways. Technical report, Open Space Seattle 2100, City of Seattle, Washington.
- Vitaliano, D. F. and J. Held
1991. Road accident external effects: an empirical assessment. *Applied Economics*, 23(2):373–378.
- Walker, G. W., W. M. Kulash, and B. T. McHugh
2000. Downtown streets: Are we strangling ourselves in one-way networks? *Transportation Research Circular*, 501.
- Walker, J.
2011. Basics: walking distance to transit. *Human Transit Blog*.
- Wang, K. and D. M. Levinson
2016. Towards a metropolitan fundamental diagram using travel survey data. *PloS one*, 11(2):e0148660.
- Wardrop, J. G.
1952. Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(3):325–362.
- Wazana, A., V. L. Rynard, P. Raina, P. Krueger, and L. W. Chambers
2000. Are child pedestrians at increased risk of injury on one-way compared to two-way streets? *Can J Public Health*, 91(3):201–6.
- Weingroff, R. F.
2003. President dwight d. eisenhower and the federal role in highway safety. *Retrieved January*, 11:2005.
- Weinstein Agrawal, A., M. Schlossberg, and K. Irvin
2008. How far, by which route and why? a spatial analysis of pedestrian preference. *Journal of urban design*, 13(1):81–98.
- Weis, C. and K. W. Axhausen
2009. Induced travel demand: Evidence from a pseudo panel data based structural equations model. *Research in Transportation Economics*, 25(1):8–18.
- White, M. J.
1988. Urban commuting journeys are not" wasteful". *Journal of Political Economy*, 96(5):1097–1110.

- Williamson, O. E.
1985. *The economic institutions of capitalism: firms, markets, relational contracting*. Free Press.
- World Health Organization
2012. Health effects of black carbon. Technical report, World Health Organization.
- Wu, X., D. M. Levinson, and H. Liu
2009. Perception of waiting time at signalized intersections. *Transportation Research Record: Journal of the Transportation Research Board*, (2135):52–59.
- Xie, F. and D. M. Levinson
2007. Measuring the structure of road networks. *Geographical analysis*, 39(3):336–356.
- Xie, F. and D. M. Levinson
2008. The weakest link: The decline of the surface transportation network. *Transportation Research Part E: Logistics and Transportation Review*, 44(1):100–113.
- Youn, H., M. T. Gastner, and H. Jeong
2008. Price of anarchy in transportation networks: efficiency and optimality control. *Physical review letters*, 101(12):128701.
- Zahavi, Y.
1974. Travel time budgets and mobility in urban areas. Technical Report FHWA PL 8183, Federal Highway Administration, US Department of Transportation, Washington, DC.
- Zhang, L. and D. M. Levinson
2010. Ramp metering and freeway bottleneck capacity. *Transportation Research Part A: Policy and Practice*, 44(4):218–235.
- Zhu, S. and D. M. Levinson
2015. Do people use the shortest path? an empirical test of wardrop’s first principle. *PloS one*, 10(8):e0134322.
- Zhu, S., D. M. Levinson, and H. X. Liu
2017. Measuring winners and losers from the new I-35W Mississippi River Bridge. *Transportation*, 44:905–918.
- Zhu, S., D. M. Levinson, H. X. Liu, and K. Harder
2010. The traffic and behavioral effects of the I-35W Mississippi River Bridge collapse. *Transportation research part A: policy and practice*, 44(10):771–784.